Detecting transient gravitational waves in non-Gaussian noise with partially redundant analysis methods

Rahul Biswas,[1] Patrick R. Brady,[2] Jordi Burguet-Castell,[3] Kipp Cannon,[4] Jessica Clayton,[2] Alexander Dietz,[5] Nickolas Fotopoulos,[6] Lisa M. Goggin,[7] Drew Keppel,[8,9] Chris Pankow,[2] Larry R. Price,[6] and Ruslan Vaulin[10]

[1]*University of Texas-Brownsville, Brownsville, Texas 78520, USA*
[2]*University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201, USA*
[3]*Universitat de les Illes Balears, E-07122 Palma de Mallorca, Spain*
[4]*Canadian Institute for Theoretical Astrophysics, University of Toronto, Toronto, Ontario, M5S 3H8, Canada*
[5]*The University of Mississippi, University, Mississippi 38677, USA*
[6]*LIGO - California Institute of Technology, Pasadena, California 91125, USA*
[7]*University of California San Francisco, San Francisco, California 94143 USA*
[8]*Albert-Einstein-Institut, Max-Planck-Institut für Gravitationsphysik, D-30167 Hannover, Germany*
[9]*Leibniz Universität Hannover, D-30167 Hannover, Germany*
[10]*LIGO - Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*

There is a broad class of astrophysical sources that produce detectable, transient, gravitational waves. Some searches for transient gravitational waves are tailored to known features of these sources. Other searches make few assumptions about the sources. Typically events are observable with multiple search techniques. This work describes how to combine the results of searches that are not independent, treating each search as a classifier for a given event. This will be shown to improve the overall sensitivity to gravitational-wave events while directly addressing the problem of consistent interpretation of multiple trials.

I. INTRODUCTION

A variety of astrophysical sources are capable of producing transient gravitational-wave signals with sufficient strength to be detectable by ground-based gravitational-wave detectors such as Laser Interferometer Gravitational-wave Observatory (LIGO) [1]. Such systems include coalescing compact binaries (CBC) consisting of neutron stars and/or black holes [2]. Currently, the same experimental data sets are searched by several analysis methods. These methods use various signal models and data processing algorithms and may have different responses to signals and non-Gaussian noise artifacts [3]. Since these searches are not independent, a single, more powerful result can be obtained by combining the results of multiple search methods.

The analysis methods can be divided into two classes by their assumptions about the signal properties. The first class assumes that the waveforms are well modeled and typically employs matched-filtering. For this reason, these methods is referred to as template-based searches. The second class assumes only basic time frequency properties about the signals. These methods are referred to as unmodeled searches. In the template-based searches, there are often several ways to construct signal models. This means that if a detectable signal exists in the data, it may not perfectly match the signal model chosen for the analysis. For example, the templates may be constructed using different approximation techniques, or they may correspond to different parts of the gravitational-wave signal

(e.g. inspiral or ringdown stages of the compact binary coalescence). The unmodeled searches make minimal assumptions about the shape of the signal and are designed to detect any short outburst of gravitational radiation in a given frequency band. Both search classes employ algorithms for identifying and discarding the non-Gaussian noise artifacts. To their advantage, unmodeled searches are able to detect a wide class of signals. However, the template-based searches generally achieve higher detection efficiency for signals matching the templates.

A gravitational-wave search produces a list of candidate events. Reanalyzing the data with multiple methods may increase the odds of detecting a gravitational wave. At the same time it has the negative effect of generating redundant lists of gravitational-wave candidates and increasing the number of trials, which makes it more difficult to assess the significance of an event. Sensitivity domains of many searches overlap, meaning that multiple searches may detect the same gravitational-wave signal. The detection efficiency of a given search depends on a variety of factors and it can be difficult to interpret results of multiple searches.

It is apparent that gravitational-wave searches would benefit from a procedure to consistently combine results into a joint detection or model exclusion statement for a given population of gravitational-wave sources. We apply the general framework for detection of gravitational waves in the presence of non-Gaussian noise developed in our earlier paper [4] to this problem. Treating the output of

each search method as a *classifier* for gravitational-wave candidate events, we construct a unified ranking for all candidate events that is easy to implement and interpret. We test it by combining candidate events from four different search methods that analyze simulated gravitational-wave signals from compact binary coalescence embedded in the data taken during LIGO's S4 science run. We find that this procedure is robust and can be used in ongoing and future searches. Interpretation of the combined results is straightforward. In particular, the calculation of the posterior probability distribution or upper limit for the rate of coalescing binaries can be carried out as it is normally performed for a single search. The combined upper limit calculation was addressed in [5], assuming multiple search methods were performed. We briefly discuss the relationship between the method suggested in that paper and ours.

The paper is organized as follows. In Sec. II, we formulate the problem of combining results from multiple searches and construct a statistic for the joint analysis. We conclude this section with a discussion of a rate upper limit calculation for the combined search and its relation to the method suggested in [5]. In Sec. III, we test our procedure by combining results from four different search methods. We briefly describe each method, the data, and the model signals. This is followed by details of the simulations and a discussion of the results. In Appendices A and B we derive a formal expression for the multivariate statistic, which accounts for correlations between the searches, and analyze the limits of applicability of our procedure.

II. METHOD FOR COMBINING SEARCHES

In this section, we establish a method for combining results from different gravitational-wave searches performed over the same data. The method builds on the general approach described in [4]. We construct a unified statistic for searches by treating each as a separate, possibly redundant, classifier for a given event.

Each search method aims to classify observational data into a list of candidate events, ranked by their likelihood to be a gravitational-wave signal. In the data analysis process, the data are analyzed and assigned a rank, r, a real number reflecting the odds that the data contain a gravitational-wave signal. Ordering time series data by amplitude is one simple method for ranking candidate events. The rank (or amplitude) is compared to a preestablished threshold, a boundary that separates signal-like data with sufficient confidence. In this way, the procedure classifies data on a scale from not signal-like to signal-like. A search method may classify events by complicated consistency tests and noise rejection schemes, but conceptually any search can be thought of as a mapping from the space of data to the space of real numbers that indicate their rank. We will assume that such a ranking procedure exists for any search method, i, and that the result, r_i, indicates the likelihood that a signal is present in a given search.

Different search methods employ a variety of techniques, data processing algorithms and waveform models. As there are a number of potential gravitational-wave sources, the search targets may vary as well. Separate searches may provide different information about a particular population of sources. Hence, it is important to extract as much information as possible by combining the results of various searches. When multiple searches analyze the same data, the output of each search, r_i, can be further processed to make the most informative detection or rate limit statement for a population of gravitational-wave sources. In doing so, it is important to ensure that there is no loss of detection efficiency when one or more of the methods has a high false alarm rate or is uninformative or irrelevant for the targeted source population.

For a given event with rank, r_i, one can compute the posterior probability that it is a gravitational-wave signal, $p(1|r_i)$. Following the steps outlined in [4], this probability can be expressed as

$$p(1|r_i) = \frac{\Lambda(r_i)}{\Lambda(r_i) + p(0)/p(1)}, \qquad (1)$$

where the likelihood ratio, $\Lambda(r_i)$, is defined by

$$\Lambda(r_i) = \frac{\int p(r_i|\mathbf{h}, 1)p(\mathbf{h}|1)d\mathbf{h}}{p(r_i|0)}, \qquad (2)$$

and $p(r_i|\mathbf{h}, 1)$ is the probability of observing r_i in the presence of the signal $\mathbf{h}$, $p(\mathbf{h}|1)$ is the prior probability to receive that signal, and $p(r_i|0)$ is the probability of observing r_i in the absence of any signal. The targeted astrophysical population of sources is completely described by $p(\mathbf{h}|1)$, where $\mathbf{h}$ denotes all possible intrinsic (e.g. masses of compact objects in the binary) and extrinsic (e.g. distance to the source, sky location) source parameters.

If an event is identified as a plausible candidate by several search methods, $p(1|r_i)$ can be calculated for each search based on the ranking, r_i, the event received. Thus, information from each search can be directly compared. The most relevant search results in the highest posterior probability for a signal to be present in the data. According to Eq. (1), this probability is a monotonically increasing function of the likelihood ratio, $\Lambda(r_i)$. As such, comparing likelihood ratios is equivalent to comparing the posterior probabilities, $p(1|r_i)$. Therefore, the likelihood ratio can be used as a unified ranking statistic to combine the output of all searches.

Strictly speaking, the denominator in Eq. (2) should contain contributions from all gravitational-wave sources not included in the targeted population, $p(\mathbf{h}|1)$. We neglect these terms because typically their contribution is very small. If a population of sources induces a response very similar to that of the targeted sources, then these classifiers will not distinguish signals from the two different populations. This can lead to an overestimation of event rates for

the targeted population, however no signal would be missed. Further refinement of the data analysis techniques or detectors themselves would be required to distinguish between the signals from these sources.

We define the ranking statistic for the joint search to be

$$r_{\text{joint}} = \max\{\Lambda(r_1), \Lambda(r_2), \ldots, \Lambda(r_n)\}, \quad (3)$$

where maximization is carried out over simultaneous events. Though this choice does not make use of all available information (we neglect correlations between the classifier's ranks; see Appendix A for multivariate treatment of the problem), it does offer some advantages. It is straightforward to compute $\Lambda(r_i)$ for each search method and simply take the largest. This has a simple interpretation: events from each search are compared based on the ratio of sensitivity of the method to the targeted sources and the search's false alarm rate. The event that is most likely to be a gravitational wave is kept. As a result, the searches are combined according to the best classifier for each event. Events classified by noisy, insensitive searches receive a low likelihood-ratio ranking and therefore do not contaminate the overall sensitivity of the analysis. We further discuss the limits of applicability for this ranking statistic in Appendix B.

As in the case of a single search method, the result of combining searches using the maximum likelihood-ratio statistic, Eq. (3), is a list of events. They can be treated as the output of a single search, with their significance evaluated by estimating the background. In the next section, we discuss how to do this. Having a single list of events allows for straightforward interpretation of results. The most significant events can be further studied and possibly promoted to the list of detected gravitational-wave signals. The posterior probability distribution or the upper limit on the rate of coalescence can be calculated following any of the methods developed for a single search [6–8].

The upper limit calculation for multiple searches described in [5] differs from our method. In [5], searches are treated as counting experiments and the upper limit on the rate of events is calculated using the total number of events above some fixed threshold and Poisson statistics. To apply this method when multiple searches are performed, a prescription is needed for determining how many events each search should contribute to the total count. In [5] events are classified by combinations of searches that generated them. The problem is reduced to the choice of foliation by a family of exclusion surfaces, $S(\zeta)$, of the space of the number of events in each category, N_q, where q runs through all possible combinations of m ($m \le n$) out of n searches. In the paper, the authors suggest and discuss several plausible choices of linear surfaces, $S(\zeta)$, that lead to different upper limits. By construction, the maximum likelihood ratio, Eq. (3), ensures that the total number of events each search

contributes on average to the joint search is proportional to the ratio of its efficiency to detect the targeted signals to its background. This construction is closely related to the "single combination" option of [5], in which only the most sensitive search contributes to the upper limit. Note though, that in our method the most sensitive search is determined during the analysis on event by event basis. This relieves an analyst from determining before hand which of the searching methods is the most sensitive. Often sensitivity is a very complicated function of signal's parameters and there might not be a single most sensitive search method that covers all signals. In this case, one would have to split the signal parameter space into subdomains, within which a single most sensitive search method exist, and carry out the upper limit calculation for each of the domains independently. In practice, this may prove to be a formidable task. The maximum-likelihood ratio procedure is universal and is almost trivial to implement, as we show it the next section. Its other important advantage is accounting for background noise present in each of the search methods. The choice between the methods is based not only on their sensitivity but also their susceptibility to the noise artifacts. In [5], the authors also mention the necessity to include information about the background to achieve more optimal upper limits.

The maximum likelihood construction ignores correlation between the searches. The optimal way to account for it is to define the multivariate likelihood-ratio ranking described in Appendix A. Unfortunately, implementing this ranking for more that two search methods is not feasible. Also, we argue that in most practical situations the net positive effect of correlations is small. The multivariate likelihood-ratio, Eq. (A1), defines the optimal exclusion surfaces, $S(\zeta)$, for the upper limit calculation method of [5]. These surfaces generally are nonlinear and, therefore, do not directly correspond to any of the choices considered in [5]. The closest in spirit is the "efficiency-weighted combination" suggested by the authors of [5], where contributions from each combination of the search methods are weighted proportionally to their sensitivity to signals. Using notation of [5] and accounting for noise contribution, the corresponding exclusion planes are defined by the normal vector $\mathbf{k} \equiv (\epsilon_q/b_q)$, where ϵ_q is the probability of a signal to be detected by the qth combination of the searches and b_q is the number of background events in this combination.

III. TESTING THE MAXIMUM LIKELIHOOD-RATIO STATISTIC WITH NON-GAUSSIAN DATA

The maximum likelihood-ratio statistic, Eq. (3), provides a natural way to combine results of several search methods into a single joint search. It possesses several attractive qualities and is expected to result in no loss of efficiency in the most practical situations (see Appendix B

for discussion of this). Still, it is important to verify this in conditions that mimic the strong non-Gaussian noise that is encountered in the search for gravitational waves in real data. To simulate a real life application of our procedure, we employ four search methods that are currently used to detect gravitational waves from coalescing binaries with the LIGO and Virgo observatories. We analyze simulated signals inserted in the data from LIGO's fourth scientific run (S4) and combine results of these analyses using the maximum likelihood-ratio statistic. We estimate the efficiency of the combined search in detecting these signals in the typical LIGO noise and compare it to the efficiencies of the individual searches.

A. Data and signals

We insert simulated signals into data collected between February 24 and March 24, 2005, during LIGO's S4 run. The data was taken by three detectors: the H1 and H2 colocated detectors in Hanford, WA and L1 in Livingston, LA. Several searches for gravitational waves were performed in these data, but no gravitational-wave candidates were identified [9–11]. For this work, 15 days of triple coincidence data were used, which is the sum of all times during S4 when all three detectors were simultaneously operating in science mode.

Our signals include three kinds of binaries: neutron star-neutron star (BNS), neutron star-black hole (NSBH) and black hole-black hole (BBH) binaries. We use nonspinning waveforms to model signals from these binaries. For BNS, these are post-Newtonian waveforms [12–22], Newtonian order in amplitude and second order in phase, calculated using the stationary phase approximation [13,20,21] with the upper cut-off frequency set by the Schwarzschild innermost stable circular orbit (ISCO). Signals from all other binaries are approximated by the effective one body numerical relativity (EOBNR) waveforms [15,23–33]. The former waveforms describe only the inspiral phase of the coalescence, whereas the latter also include merger and ringdown phases.

The simulated signals are injected into nonoverlapping 2048-second blocks of data. To improve the statistic, multiple signal populations are inserted in the data and independently analyzed. Signals are split into three categories by total mass of the binary: 2–$6M_\odot$, 6–$100M_\odot$, and 100–$350M_\odot$. The lowest mass range includes only BNS systems. Within each mass range, signals are distributed uniformly in distance or the inverse of distance. The BNS range covers 1–20 Mpc while other systems reach from 1–200 Mpc. In order to represent realistic astrophysical population with probability density function scaling as distance squared, the simulated signals are appropriately reweighted and are counted according to their weights. All other parameters of the signals have uniform distribution. In total, there are 943, 2245, and 2237 signals injected within each mass category, respectively.

B. Search methods

Four search methods, each representing one of the standard searches for transient gravitational-wave signals in LIGO and Virgo data, are used to perform this joint analysis. Brief descriptions of the search methods are given below. The first three are template-based searches, while the last one does not rely on any specific signal model.

The low-mass CBC pipeline targets binaries with total mass below $35M_\odot$. The data from each interferometer are match-filtered with a bank of nonspinning post-Newtonian waveforms [12–22] covering binary mass combinations with total mass in the range 2–$35M_\odot$. The template waveforms are calculated in the frequency domain using the stationary phase approximation [13,20,21] to Newtonian order in amplitude and second PN order in phase. The waveforms are extended up to the Schwarzschild ISCO. When the signal-to-noise ratio (SNR) time series for a particular template crosses the threshold of 5.5, a single-interferometer trigger is recorded. These triggers are required to pass waveform consistency and coincidence tests with triggers from other interferometers. The surviving triggers are ordered by a ranking statistic and form a ranked set of candidate events. For detailed descriptions of this pipeline and recent search results, see [11,34–37].

The high-mass CBC pipeline is similar to its low-mass counterpart, however it is designed to target binaries with total mass between 25–$100M_\odot$. The EOBNR family of templates used in a high-mass search has waveforms covering the evolution of a coalescing binary from late inspiral to ringdown. Other than the choice of templates, the analysis is quite similar to the low-mass search. The high-mass CBC pipeline was used to search for gravitational waves from binary black holes in the S5 LIGO data [38].

The ringdown pipeline was developed to search for gravitational-wave signals corresponding to the post-merger phase of the binary coalescence. After two compact objects merge, a single, highly perturbed black hole forms and radiates energy while it settles down to a stable Kerr solution. The pipeline constructs its template bank from the dominant, $l = 2$ and $m = 2$, black hole quasinormal modes characterized by a single frequency and quality factor. The template bank used in the ringdown search spans the most sensitive part of the LIGO frequency band, 50 Hz–2 kHz. Quality factors between 2–20 are used, corresponding to a range for the final black hole spin between nonspinning and $\hat{a} = 0.994$. As with the other search methods previously described, candidate events are ranked after being detected by multiple interferometers and passing several consistency tests. This pipeline was used to search for gravitational waves in the S4 data [9].

Coherent WaveBurst is a gravitational-wave burst analysis pipeline designed to detect signals from transient gravitational-wave sources. It uses minimal information

about the signal model, instead using the cross-correlated excess power from the gravitational-wave signal across a network of interferometers [39]. The pipeline enforces the signal hypothesis by maximizing a likelihood functional that describes the expected signal response of an impinging gravitational wave given its source location in the sky combined over the network of interferometers [40]. Triggers are generated from the interferometer network by combining time-frequency maps using wavelet transformations of the interferometer time-series data. From these maps, the likelihood of a trigger is calculated by the pipeline from the correlation of the whitened data streams weighted by the network's antenna patterns. This pipeline was used in the LIGO S4 [41] and LIGO/Virgo S5/VSR1 [42,43] searches for unmodeled short duration transients, as well as searches for black hole binaries [44].

We note that our analysis does not include the most recent innovations developed to improve the efficiency of each of these pipelines. In particular, we do not categorize the candidate events by the template mass and coincidence type—a novelty introduced in the low- and high-mass CBC pipelines during the analyses of S5 LIGO data and the S5/VSR1 data from LIGO and Virgo. Also, we use the default settings for the S5 analysis (S4 for the ringdown search) for numerous pipelines' parameters without attempting to re-tune them. We choose to perform simulations without the most up-to-date and fully tuned versions of the pipelines to save time. This is justified because our algorithm for combining searches is ignorant of the inner-workings of each pipeline. This makes the combined search robust against small changes in the individual analysis algorithms. For our purpose, it is sufficient to use somewhat simplified versions of the pipelines, as we do not expect these results to change dramatically when incremental changes occur as the pipelines evolve.

C. Algorithm for combining searches

The procedure for combining candidate events identified by different classifiers is straightforward and based on Eq. (3). The first step is the calculation of the likelihood ratio, $\Lambda(r_i)$, defined by Eq. (2), for every event. Notice that since the numerator depends on the population of signals through $p(\mathbf{h}|1)$, $\Lambda(r_i)$ is not just a trivial rescaling of a rank assigned by a classifier to an event. The likelihood ratio estimates the significance of each event in the context of a gravitational-wave detection from the targeted population of sources specified via $p(\mathbf{h}|1)$ and $p(r_i|\mathbf{h}, 1)$. As a result, events are ranked by the odds of being produced by the classifier in response to the targeted signals, rather than noise. Depending on the population of sources, some classifiers may not provide any useful information. In that case, events provided by such classifiers receive a very low likelihood ratio rank and are effectively removed from the search. This is a desirable feature that makes the procedure robust against nuisance classifiers.

In order to compute the one-dimensional likelihood ratio given in Eq. (2), we need to measure the classifiers' response to the gravitational-wave signals interposed over noise and to background noise only. For the latter, we use a common background estimation technique for gravitational-wave searches—shifting recorded data from the non-colocated interferometers in time with respect to each other [11,34–37]. If the shift is much longer than the gravitational-wave travel time between the detector sites (≈ 10 ms for a Hanford-Livingston detector pair), then the resulting time-shifted data are guaranteed to contain no coherent gravitational-wave signals. These data, when analyzed by the classifier, represent the background of the search. In the low-mass CBC, high-mass CBC, and ringdown pipelines, we perform 2000 forward-in-time shifts of the L1 data with time steps of 7 seconds relative to H1 and H2. The Coherent WaveBurst search uses 100 forward-in-time shifts of the L1 data with 5-second time steps. Each time-shifted data set produces an independent sample of the background. The time-shifted data are analyzed and all background events are recorded.

For background events, it is significantly easier to estimate the cumulative probability density function (cdf), $P(r_i|0) = \int_{r_i}^{\infty} p(r_i'|0)dr_i'$, than the probability density function (pdf), $p(r_i|0)$. Each background data set represents an independent trial observation of duration, T. Therefore, for a trigger, r_i, the ratio of the number of the trial observations that produced a trigger with the rank $r_i' \geq r_i$ to the total number of trial observations provides an estimate of the probability, $P_T(r_i|0)$, of the classifier producing an equally or higher ranked event in the analysis of noise alone. This probability is a monotonic function of $P(r_i|0)$ and experiment duration, T,

$$P_T(r_i|0) = 1 - (1 - P(r_i|0))^{T/T_0}, \qquad (4)$$

where T_0 is the duration of a unit experiment, which can be classified by a single rank, r_i. The scale for T_0 is set by the duration of the gravitational-wave signal, the time scale on which data samples can be considered uncorrelated. In practice, given that all methods analyze the same amount of data, the two probabilities $P_T(r_i|0)$ and $P(r_i|0)$ are equivalent for the purpose of ranking the candidate events since one is a monotonic function of the other. Computation of the background cdf curve, $P_T(r_i|0)$, is a trivial task. First, the single event with the highest rank from every background data set is chosen. Then, for any value of r_i, one simply counts the number of these events with rank $r_i' \geq r_i$, divided by the total number of background data sets. In this way, the background cdf curves are calculated for each search method.

In order to measure the response of each search method to the targeted gravitational-wave signals and calculate the numerator in Eq. (2), several populations of simulated signals are injected into the data and processed by the

pipelines. For each search method, events identified with the injected signals are recorded along with the parameters of the signals. As with background events, it is much easier to compute the cdf, $P(r_i|1)$, of the signals. For an event with rank r_i, this probability is approximated by the ratio of the number of injected signals with rank $r_i' \geq r_i$ to the total number of injected signals. Using this algorithm we compute cdf curves, $P(r_i|S_j, 1)$, for each search method, i, and mass category of the simulated signals, S_j: 2–6$M_\odot$, 6–100$M_\odot$, and 100–350$M_\odot$, which represent the intended targets of the low-mass CBC, high-mass CBC, and burst/ringdown pipelines, respectively.

Having precomputed the background and signal cdf curves, the algorithm for combining the analysis pipelines can be summarized in the following two steps. First, every event, r_i, produced by the ith classifier is assigned the log-likelihood-ratio ranking given by

$$L(r_i|S_j) = \ln\left[\frac{P(r_i|S_j, 1)}{P_T(r_i|0)}\right], \qquad (5)$$

in which the ratio of pdfs in Eq. (2) is approximated by the ratio of cdfs. For the sake of brevity in what follows we omit the "log" from the ranking name and refer to $L(r_i|S_j)$ as the likelihood-ratio ranking. Second, all events from all classifiers are mixed together and clustered, retaining events with the highest likelihood-ratio ranking, $L(r_i|S_j)$, within the specified time window. These events form the final list of gravitational-wave candidates. The time window is approximately equal to the autocorrelation time for an average signal injected in the data. In our simulations, we set it to 10 seconds. Events separated in time by more than 10 seconds are uncorrelated and therefore may correspond to different signals [11,34–37]. This last step effectively implements maximization in Eq. (3) over the likelihood ratio for coincident events identified by multiple search methods.

We expect the cdf approximation used in Eq. (5) to be fair in the context of our simulations. Injection and background distributions are one-dimensional, monotonic functions of rank. They generally fall off as some negative power of rank. Detectable signals lie on the tail of the background distribution. Under these conditions, the difference between using the pdf or cdf in the likelihood ratio is insignificant. Nevertheless, we should stress that this may not be the case in general and proper estimation of signal and background probability distributions may be required.

Before proceeding to the discussion of simulation results, we note that $L(r_i|S_j)$, defined by Eq. (5), depends on the population of injected signals, S_j. Therefore, events identified by the classifiers in each search must be reprocessed according to the algorithm described above for each population of sources, S_j.

D. Simulation results

After multiple search methods are used on the data injected with gravitational-wave signals, the events selected by each search are processed with the algorithm sketched in the previous subsection. To estimate the background for the combined search, we again perform time shifts of the L1 data with respect to data from H1 and H2. Although the time shifts performed in III C are independent for each classifier, the time shifts must be synchronized for all classifiers when estimating the background for the combined search. For this purpose, 100 5-second time shifts of the L1 data are performed with respect to the H1 and H2 data. The background sample is processed with the same algorithm as the main data.

As previously mentioned, we consider three target populations of compact binaries, categorized by their total mass: the binary neutron stars with total mass 2–6$M_\odot$, the compact binaries with total mass 6–100$M_\odot$, and the binaries with total mass 100–350$M_\odot$. These define three independent searches. The data with injected signals from each category are analyzed independently. The resulting events are ranked by the likelihood ranking, Eq. (5), with S_j being one of the considered target populations. The background events are ranked and combined in the same way, providing an estimate of the background for the combined searches.

To compare combined searches with the individual search methods, we compute their sensitivities to the signals in the presence of typical background. We summarize this in Figs. 1–3, which show curves of visible volume versus false alarm rate for each of the search methods and for the combined search. For each point on these curves, the calculation proceeds as follows. First, using
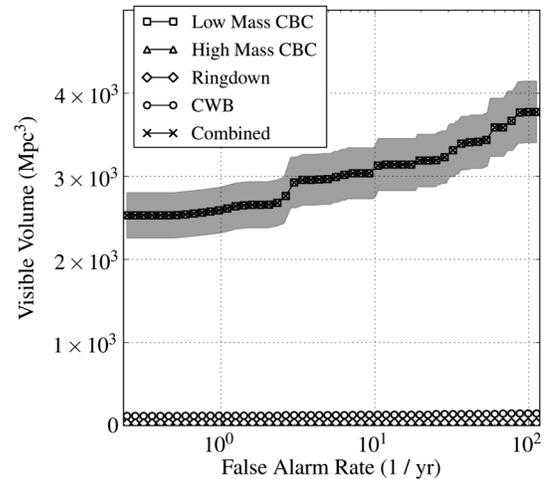

FIG. 1. Visible volume versus false alarm rate for binary neutron stars. The shaded area around a curve represents its 1σ Poisson error. The "Combined" and the "Low Mass CBC" curves coincide, whereas all other curves drop to near zero in visible volume.
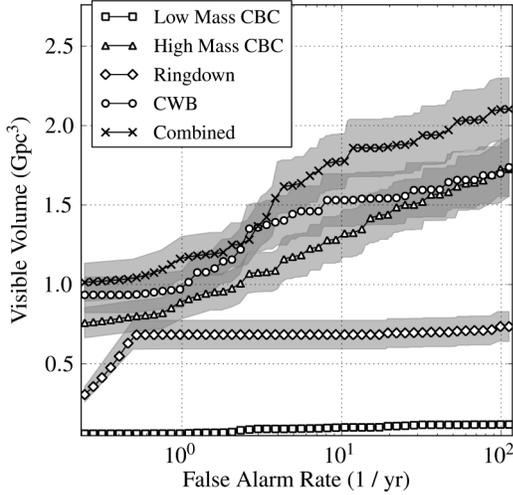
FIG. 2. Visible volume versus false alarm rate for CBC with total mass 6–$100M_\odot$. The shaded area around a curve represents its 1σ Poisson error.

background events, we determine the value of the rank corresponding to a given false alarm rate. Next, the efficiency as a function of distance to the source, $\epsilon(D)$, is estimated by the fraction of the signals at distance D ranked above that value. This efficiency is then converted to the visible volume.

The binary neutron star search is a case study in which only one of the classifiers, namely, the low-mass CBC pipeline, is effective in detecting the particular type of gravitational-wave signal. All other classifiers are designed to detect either black hole binaries or short duration bursts and are inefficient in detecting the long inspiral signal sweeping through the whole LIGO frequency band. This is properly accounted for in the likelihood-ratio ranking,
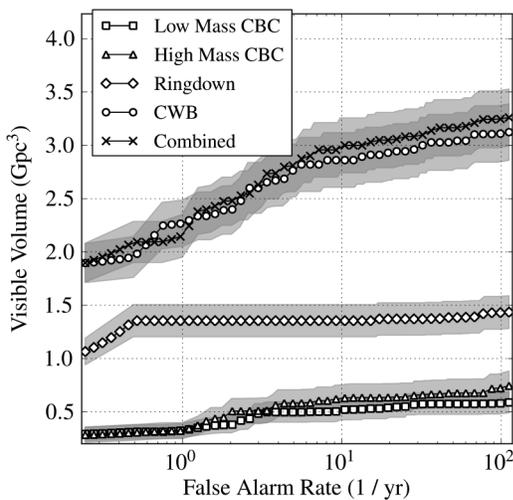

FIG. 3. Visible volume versus false alarm rate for CBC with total mass 100–$350M_\odot$. The shaded area around a curve represents its 1σ Poisson error.

which is very low for all events from these other classifiers. The only events not deweighted are those from the low-mass CBC pipeline. As a result, the combined search is equivalent to the low-mass CBC search in this case. The sensitivity curves for both searches, shown on Fig. 1, coincide. This shows that our algorithm is robust against uninformative, nuisance classifiers.

The picture changes dramatically for compact binaries in the medium mass range, shown in Fig. 2. In this case, the efficiency of the low-mass CBC pipeline is negligible in comparison to the other classifiers. In this category, the Coherent WaveBurst pipeline has the best overall sensitivity. Further inspection reveals that the high-mass CBC pipeline is the most sensitive of the three in the 6–$50M_\odot$ mass region, whereas the ringdown pipeline, despite being subdominant, tends to detect signals with high mass ratio that are either missed or not ranked high enough by the other pipelines. Thus, in this case, all but one classifier contribute detected signals to the combined search (the "$\times$-$\times$" curve on Fig. 2). This is a desired effect of incorporating the detection sensitivities of different pipelines, which results in a more sensitive and robust combined search.

We observe similar effects for the high-mass binaries, although this is not obvious from Fig. 3. The figure shows the Coherent WaveBurst pipeline dominating over the ringdown or the high-mass CBC pipelines. The sensitivity curve of the combined search tends to be just above the Coherent WaveBurst curve and occasionally drops below it. However, these drops are well within the error bars. The detailed investigation shows that the high-mass CBC and the ringdown pipelines still contribute detections of extra signals missed by the Coherent WaveBurst pipeline. In particular, the ringdown pipeline has the highest sensitivity of all searches in the 270–$350M_\odot$ region. Most of these extra signals are in the near or mid range zone (less then 100 Mpc) and therefore do not contribute as much to the total visible volume as those at far distances. As a result, overall gain for the combined search is not that significant when compared to the Coherent WaveBurst search alone. Moreover, occasionally, due to background fluctuations, the threshold for the combined search fluctuates upward, which results in a loss of a few distant signals detected by the Coherent WaveBurst pipeline. The loss of visible volume associated with these signals is not compensated by the gain of efficiency in the mid range. We should note that the measurement of detection efficiency for signals beyond 150 Mpc has large uncertainties due to low counts for detected signals. Therefore, the actual loss of efficiency in this case may be overestimated.

For demonstration of these effects and further insight, we plot cumulative 50% efficiency contours on the distance/total mass plane for signals to be detected above the threshold, Fig. 4. The threshold is set by the lowest measured false-alarm rate of 0.28 events per year
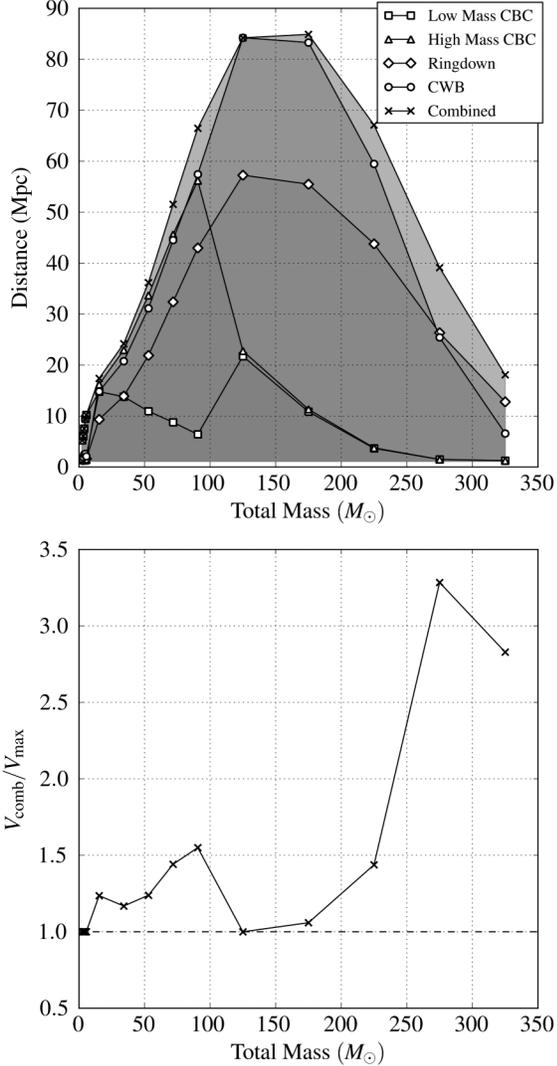
FIG. 4. The upper pane shows the cumulative 50% efficiency contours at false alarm rate of 0.28 events per year. The combined search contour envelops the single classifier contours from above. The lower pane shows the plot of the ratio of the visible volume of the combined search to the visible volume of the most sensitive classifier in each mass bin. Having a ratio greater than one indicates that the combined search gains sensitivity across the entire mass range.

(corresponding to the left most point on the sensitivity curves in Figs. 1–3). In Fig. 4, the contour for the combined search envelops contours of the other pipelines. Furthermore, we calculate the corresponding visible volume for the joined search and plot its ratio to the visible volume of the most sensitive pipeline in each mass bin, shown in the lower pane of Fig. 4.

IV. CONCLUSIONS

We consider the problem of combining outputs of partially redundant search methods analyzing the same data in the context of gravitational wave searches. We suggest that

the likelihood ratio, Eq. (2), provides a natural unified ranking for the candidate events identified by the search methods. It has a straightforward interpretation—events from each method are ranked according to the ratio of the method's sensitivity to its background. After forming the joined list of candidate events, calculation of the posterior probability distribution or an upper limit on the rate of gravitational-wave emissions can proceed exactly as it would for a single search method. Therefore, there is never a problem consistently accounting for multiple trials in the analysis. If the combined search is interpreted as a counting experiment, the procedure for calculating the upper limit from multiple searches is similar to that suggested in [5]. In that case, classifiers contribute in proportion to their sensitivities. Additionally, our method accounts for information about the classifier's background, which is important when dealing with experimental data containing non-Gaussian noise artifacts.

We test our procedure by simulating a search for gravitational waves from compact binary coalescence in the data from LIGO's S4 science run. We combine outputs from four search methods—the low-mass CBC, high-mass CBC, ringdown, and Coherent WaveBurst analysis pipelines—analyzing data with injected gravitational-wave signals from compact coalescing binaries in a wide range of masses. We find that our algorithm is robust against nuisance pipelines—those that are not sensitive to the targeted gravitational-wave sources. Moreover, the combined search proves to have greater or comparable sensitivity to any individual pipeline. In the simulations, we observe that the pipelines we use contribute different events to the total count of detected signals, thus increasing robustness and the overall probability of detecting gravitational waves from coalescing binaries. This effect is especially pronounced for sources in near to mid range distances. Overall, our simulations show that searches for gravitational waves from coalescing binaries can benefit from combining results of multiple analysis methods by means of the likelihood-ratio statistic.

APPENDIX A: MULTIVARIATE LIKELIHOOD-RATIO STATISTIC

The maximum likelihood-ratio procedure described in Sec. II does not account for potentially useful information contained in the correlations between classifiers. In this section, we derive the formal expression for the multivariate statistic that is optimal by construction and includes all available information. We discuss some of its properties and its relation to the maximum likelihood-ratio statistic, Eq. (3).

In the absence of internal thresholds, each classifier assigns a rank, r_i, to every data sample. In this case, the vector of ranks, $\vec{r} \equiv (r_1, r_2, \dots r_n)$, can be interpreted as the reduced experimental data, and the problem of combining searches is analogous to a detection problem. The general problem of detection in the presence of arbitrary noise was discussed in detail in [4]. Here, we state the final result and refer the interested reader to [4] for derivation and further discussion. The optimal solution, assuming the Neyman-Pearson criteria that requires the maximization of the signal detection probability at a fixed rate of false alarms, ranks data samples by the likelihood-ratio detection statistic. For n classifiers, this takes the form

$$\Lambda(r_1, r_2, \dots, r_n) = \frac{\int p(r_1, r_2, \dots, r_n | \mathbf{h}, 1) p(\mathbf{h}|1) d\mathbf{h}}{p(r_1, r_2, \dots, r_n | 0)},$$

(A1)

where $\mathbf{h}$ stands for a gravitational-wave signal, $p(r_1, r_2, \dots, r_n | \mathbf{h}, 1)$ is the probability distribution for the vector of detection statistics $(r_1, r_2, \dots, r_n)$ in the case when the gravitational-wave signal $\mathbf{h}$ is present in the data, $p(r_1, r_2, \dots, r_n | 0)$ is the analogous distribution for the noise, and $p(\mathbf{h}|1)$ is the distribution of signal parameters for the targeted population of gravitational-wave sources.

The joint likelihood ratio, Eq. (A1), includes the output from all classifiers and by construction provides the optimal ranking. We can simplify the expression in Eq. (A1) by noting that for $n = 2$,

$$\begin{aligned}
\Lambda(r_1, r_2) &= \frac{\int p(r_1, r_2 | \mathbf{h}, 1) p(\mathbf{h}|1) d\mathbf{h}}{p(r_1, r_2 | 0)} \\
&= \frac{1}{2} \bigg(\frac{\int p(r_1 | r_2, \mathbf{h}, 1) p(\mathbf{h}|1) \Lambda(r_2, \mathbf{h}) d\mathbf{h}}{p(r_1 | r_2, 0)} \\
&\quad + \frac{\int p(r_2 | r_1, \mathbf{h}, 1) p(\mathbf{h}|1) \Lambda(r_1, \mathbf{h}) d\mathbf{h}}{p(r_2 | r_1, 0)} \bigg),
\end{aligned}$$

(A2)

and extending linearly for $n > 2$.

In practice, computing the conditional probabilities in Eq. (A2) can be a nontrivial task. Moreover, as the number of classifiers increases, computing the necessary conditional probabilities becomes a less and less viable option and it is necessary to develop an approximation. The maximum likelihood-ratio procedure can be regarded as

such even though it does not follow from the multivariate expression (A1) in a straightforward way. In order to find a relation between the two, it is useful to consider the limiting cases of Eq. (A2). First, assume there is no correlation between the measurements made by each classifier. In that case, $p(r_i | r_j, \mathbf{h}, 1) = p(r_i | \mathbf{h}, 1)$ (and similarly for the denominators), so that

$$\Lambda(r_1, r_2) = \int \Lambda(r_1, \mathbf{h}) \Lambda(r_2, \mathbf{h}) p(\mathbf{h}|1) d\mathbf{h} \approx \Lambda(r_1) \Lambda(r_2).$$

(A3)

Factorization in the last step is justified because only one of the classifiers exhibits a nontrivial response in the presence of a signal in the data. As a consequence, the unmarginalized likelihood ratio, $\Lambda(r, \mathbf{h})$, for that classifier is a function of the signal, $\mathbf{h}$, sharply peaked around the true parameters of the signal, whereas the other likelihood ratio is almost constant.

At the other extreme, consider the case of two strongly correlated classifiers. Then

$$\begin{aligned}
\frac{\int p(r_1, r_2 | \mathbf{h}) p(\mathbf{h}|1) d\mathbf{h}}{p(r_1, r_2 | 0)} &= \frac{\delta(r_1 - r_2) \int p(r_2 | \mathbf{h}) p(\mathbf{h}|1) d\mathbf{h}}{\delta(r_1 - r_2) p(r_2 | 0)} \\
&= \Lambda(r_1).
\end{aligned}$$

(A4)

Both cases can be easily generalized for $n > 2$ classifiers.

When classifiers are strongly correlated, the maximum likelihood ratio is trivially equivalent to (A4). In the opposite case, the absence of correlation between the classifiers implies their complementarity. If one classifier identifies a significant event, the others do not. This means that typically only one of the likelihood ratios in the product in Eq. (A3) will be significantly different from unity. Therefore, in this case, picking the maximum of the single classifier likelihood ratios or calculating their product has similar effect. Based only on these extreme situations, it is difficult to determine how good of an approximation the maximum likelihood ratio is in the intermediate case. Nevertheless, we conjecture that the truly useful information can only be in correlations between the classifiers using incomplete, but complementary, information about the signal (e.g. template-based searches using inspiral and merger or ringdown waveforms). Even in this situation, the inclusion of correlations should be a next-order effect.

APPENDIX B: MAXIMUM LIKELIHOOD-RATIO STATISTIC

One can gain further insight into the statistic defined by Eq. (3) by mapping the ranks, r_i, to their likelihood ratios, $\Lambda_i(r_i)$. The mapping is defined by Eq. (2). The data space of the combined search is $\vec{\Lambda} \equiv (\Lambda_1, \Lambda_2, \dots, \Lambda_N)$. For the ith classifier, the probabilities of detection and of false alarm are given by

$$P_1^i = \int \Theta(\Lambda_i - \Lambda_i^*) p(\vec{\Lambda}|1) p(1) d\vec{\Lambda} \qquad \text{(B1)}$$

$$P_0^i = \int \Theta(\Lambda_i - \Lambda_i^*) p(\vec{\Lambda}|0) p(0) d\vec{\Lambda}. \qquad \text{(B2)}$$

and for the combined search by

$$\tilde{P}_1 = \int \Theta(\max(\vec{\Lambda}) - \Lambda_c^*) p(\vec{\Lambda}|1) p(1) d\vec{\Lambda} \qquad \text{(B3)}$$

$$\tilde{P}_0 = \int \Theta(\max(\vec{\Lambda}) - \Lambda_c^*) p(\vec{\Lambda}|0) p(0) d\vec{\Lambda}, \qquad \text{(B4)}$$

where Λ_i^* and Λ_c^* are detection thresholds determined from the threshold value for the false alarm probability, P_0^*, which is the same for all classifiers.

The efficiency of the combined search is expected to be, at the very least, no less than the efficiency of any of the classifiers being used ($\tilde{P}_1 \geq P_1^i$). This is a necessary condition for the maximum likelihood-ratio procedure to be applicable. To get a better understanding of what this condition implies and when it is expected to hold, consider the simple case of combining a pair of classifiers. This can be generalized in a straightforward way to arbitrary number. The data space, in this case, is a positive quarter in the (Λ_1, Λ_2) space. The lines of constant likelihood ratio, Λ_i, are horizontal or vertical lines. The lines of constant joint likelihood ratio, given by Eq. (A1), can be complicated curves even in the (Λ_1, Λ_2) plane and define the optimal detection surfaces. The corresponding surfaces for the maximum likelihood-ratio statistic form a square, centered at the origin, with sides parallel to the Λ_1 and Λ_2 axes. This configuration is visualized on Fig. 5, where Λ_1^* (vertical dashed line), Λ_2^* (horizontal dashed line), and Λ_c^* (dotted line) are the thresholds corresponding to a particular value of the probability of false alarm, for single and combined searches, respectively. Detection regions for each classifier consist of all points for which the argument of the theta function in the expressions for detection and false alarm probabilities, Eqs. (B1) and (B2), is positive. The detection region for the *i*th classifier is defined by the condition $\Lambda_i > \Lambda_i^*$. All data points in the plane satisfying this condition are counted as detection of a signal. The detection region for the combined search defined by Eqs. (B3) and (B4) consists of the points satisfying two conditions: $\Lambda_1 > \Lambda_c^*$ and $\Lambda_2 > \Lambda_c^*$. Recall that the false alarm probability for all searches is the same, which implies that

$$\int_{V_i} p(\Lambda_1, \Lambda_2|0) d\vec{\Lambda} = \int_{V_c} p(\Lambda_1, \Lambda_2|0) d\vec{\Lambda}, \qquad \text{(B5)}$$

where V_i and V_c denote the detection regions for either of the individual searches and for the combined search, respectively. This implies that $\Lambda_c^* > \Lambda_i$—the threshold for the combined search is higher than the threshold for any of the individual pipelines. Indeed, if it was not true, then Eq. (B5) could not be satisfied since $V_i \subset V_c$. This would

correspond to moving the vertical dashed line to the right of the solid square in Fig. 5, as an example for classifier Λ_1. Thus, the diagram shown in Fig. 5 represents the only allowed configuration. Continuing with the classifier Λ_1, one can identify the set of points gained by the combined search, V_+, which are points not included in V_1, and the set of points lost, V_-, those belonging to V_1 but not to V_c. Both sets are shown in Fig. 5 as shaded regions. It is clear that the efficiency of the combined search will be greater or equal to the efficiency of the classifier Λ_1 if and only if

$$\int_{V_+} p(\Lambda_1, \Lambda_2|1) d\vec{\Lambda} \geq \int_{V_-} p(\Lambda_1, \Lambda_2|1) d\vec{\Lambda}. \qquad \text{(B6)}$$

Note that at the same time

$$\int_{V_+} p(\Lambda_1, \Lambda_2|0) d\vec{\Lambda} = \int_{V_-} p(\Lambda_1, \Lambda_2|0) d\vec{\Lambda}, \qquad \text{(B7)}$$

by virtue of Eq. (B5). Thus, Eq. (B6) states that the joint likelihood for the points in V_+ must be (on average) greater than or equal to the joint likelihood for the points in V_-. For this case, exchanging V_+ for V_- results in a positive gain. It is not unjustified to expect that Eq. (B6) would be satisfied in most practical situations. After all, according to the classifier Λ_2, points in V_+ have a better chance of being a signal than those in V_-, because Λ_2 for any point in V_+ is greater than Λ_1 for any point in V_-. In effect, when combining searches using the maximum likelihood-ratio approximation, one exchanges points from V_- with decent Λ_1 and low Λ_2 in favor of points in V_+ with low Λ_1 but high Λ_2. Consider an extreme case when the classifier Λ_2 is not informative. The probability of getting a high
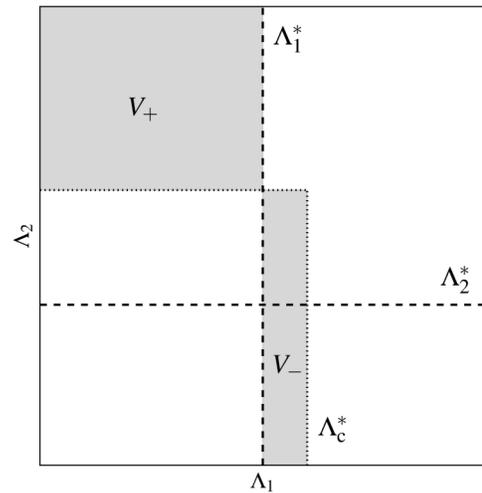

FIG. 5. The detection thresholds can be visualized for individual and combined searches in the (Λ_1, Λ_2) space. For individual searches, the detection threshold appears as the vertical dashed line (Λ_1^*) and horizontal dashed line (Λ_2^*). The Λ_c^* threshold is the dotted line. The shaded regions represent the data points gained, V_+, and lost, V_-, by the combined search when referencing a search performed with Λ_1.

likelihood ratio in the absence of the signal is very low. Then, the total probability $\int_{V_+} p(\Lambda_1, \Lambda_2|0)$ is negligible, effectively making V_- a null set. This implies robustness of the approximation against nuisance, noninformative classifiers. The above steps can be mirrored for the classifier Λ_2, resulting in the same conclusions.

In conclusion, we should stress that, although condition (B6) does not hold in general, it is expected to be satisfied when combining well designed classifiers that are sufficiently different to be able to complement each other's detection efficiencies. These are the typical cases that arise in practice.

———————————————

[1] B. Abbott *et al.* (LIGO Scientific Collaboration), Rep. Prog. Phys. **72,** 076901 (2009).

[2] J. Abadie *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), Classical Quantum Gravity **27,** 173001 (2010).

[3] J. Slutsky, L. Blackburn, D. A. Brown, L. Cadonati, J. Cain, M. Cavagli, S. Chatterji, N. Christensen, M. Coughlin, S. Desai, G. Gonzlez, T. Isogai, E. Katsavounidis, B. Rankins, T. Reed, K. Riles, P. Shawhan, J. R. Smith, N. Zotov, and J. Zweizig, *Methods for Reducing False Alarms in Searches for Compact Binary Coalescences in LIGO Data* (2010).

[4] R. Biswas *et al.*, this issue, Phys. Rev. D **85,** 122008 (2012).

[5] P. J. Sutton, Classical Quantum Gravity **26,** 245007 (2009).

[6] G. J. Feldman and R. D. Cousins, Phys. Rev. D **57,** 3873 (1998).

[7] P. R. Brady, J. D. E. Creighton, and A. G. Wiseman, Classical Quantum Gravity **21,** S1775 (2004).

[8] R. Biswas, P. R. Brady, J. D. E. Creighton, and S. Fairhurst, Classical Quantum Gravity **26,** 175009 (2009).

[9] B. Abbott *et al.* (LIGO Scientific Collaboration), Phys. Rev. D **80,** 062001 (2009).

[10] B. Abbott *et al.* (LIGO Scientific Collaboration), Classical Quantum Gravity **24,** 5343 (2007).

[11] B. Abbott *et al.* (LIGO Scientific Collaboration), Phys. Rev. D **77,** 062002 (2008).

[12] L. Blanchet, B. R. Iyer, C. M. Will, and A. G. Wiseman, Classical Quantum Gravity **13,** 575 (1996).

[13] S. Droz, D. J. Knapp, E. Poisson, and B. J. Owen, Phys. Rev. D **59,** 124016 (1999).

[14] L. Blanchet, Living Rev. Relativity **5,** 3 (2002).

[15] A. Buonanno, G. B. Cook, and F. Pretorius, Phys. Rev. D **75,** 124018 (2007).

[16] M. Boyle *et al.*, Phys. Rev. D **76,** 124038 (2007).

[17] M. Hannam *et al.*, Phys. Rev. D **77,** 044020 (2008).

[18] Y. Pan *et al.*, Phys. Rev. D **77,** 024014 (2008).

[19] M. Boyle, D. A. Brown, and L. Pekowsky, Classical Quantum Gravity **26,** 114006 (2009).

[20] K. S. Thorne, in *Three Hundred Years of Gravitation*, edited by S. W. Hawking and W. Israel (Cambridge University Press, Cambridge, 1987), Chap. 9, pp. 330–458.

[21] B. S. Sathyaprakash and S. V. Dhurandhar, Phys. Rev. D **44,** 3819 (1991).

[22] B. J. Owen and B. S. Sathyaprakash, Phys. Rev. D **60,** 022002 (1999).

[23] A. Buonanno and T. Damour, Phys. Rev. D **59,** 084006 (1999).

[24] A. Buonanno and T. Damour, Phys. Rev. D **62,** 064015 (2000).

[25] T. Damour, P. Jaranowski, and G. Schäfer, Phys. Rev. D **62,** 084011 (2000).

[26] T. Damour, B. R. Iyer, P. Jaranowski, and B. S. Sathyaprakash, Phys. Rev. D **67,** 064028 (2003).

[27] Y. Pan, A. Buonanno, J. G. Baker, J. Centrella, B. J. Kelly, S. T. McWilliams, F. Pretorius, and J. R. van Meter, Phys. Rev. D **77,** 024014 (2008).

[28] A. Buonanno *et al.*, Phys. Rev. D **76,** 104049 (2007).

[29] T. Damour, A. Nagar, E. N. Dorband, D. Pollney, and L. Rezzolla, Phys. Rev. D **77,** 084017 (2008).

[30] T. Damour, A. Nagar, M. Hannam, S. Husa, and B. Brügmann, Phys. Rev. D **78,** 044039 (2008).

[31] M. Boyle, A. Buonanno, L. E. Kidder, A. H. Mroué, Y. Pan, H. P. Pfeiffer, and M. A. Scheel, Phys. Rev. D**78,** 104020 (2008).

[32] T. Damour and A. Nagar, Phys. Rev. D **79,** 081503 (2009).

[33] A. Buonanno, Y. Pan, H. P. Pfeiffer, M. A. Scheel, L. T. Buchman, and L. E. Kidder, Phys. Rev. D **79,** 124028 (2009).

[34] B. Abbott *et al.* (LIGO Scientific Collaboration), Tech. Report. No LIGO-T070109-01, 2007.

[35] B. Abbott *et al.* (LIGO Scientific Collaboration), Phys. Rev. D **79,** 122001 (2009).

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

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

[38] B. Abbott *et al.* (LIGO Scientific Collaboration) (unpublished).

[39] S. Klimenko, I. Yakushin, A. Mercer, and G. Mitselmakher, Classical Quantum Gravity **25,** 114029 (2008).

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

[41] B. Abbott *et al.*, Classical Quantum Gravity **25,** 245008 (2008).

[42] B. P. Abbott *et al.* (LIGO Scientific Collaboration), Phys. Rev. D **80,** 102002 (2009).

[43] J. Abadie *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. D **81,** 102001 (2010).

[44] C. Pankow, S. Klimenko, G. Mitselmakher, I. Yakushin, G. Vedovato, M. Drago, R. A. Mercer, and P. Ajith, Classical Quantum Gravity **26,** 204004 (2009).