Search for strongly lensed counterpart images of binary black hole mergers in the first two LIGO observing runs

Connor McIsaac[1,*] David Keitel[1,2,†] Thomas Collett[1] Ian Harry[1] Simone Mozzon[1]
Oliver Edy[1] and David Bacon[1]

[1]*University of Portsmouth, Institute of Cosmology and Gravitation,*
Portsmouth PO1 3FX, United Kingdom
[2]*Departament de Física, Universitat de les Illes Balears, IAC3–IEEC,*
Carretera Valldemossa km 7.5, E-07122 Palma, Spain

Strong gravitational lensing can produce multiple images of the same gravitational-wave signal, each arriving at different times and with different magnification. Previous work has explored if lensed pairs exist among the known high-significance events from the LIGO and Virgo Collaboration's GWTC-1 catalog and found no evidence of this. However, the possibility remains that weaker counterparts of these events are present in the data, unrecovered by previous searches. We conduct a targeted search specifically looking for subthreshold lensed images of known binary black hole (BBH) observations from GWTC-1. We recover candidates matching three of the additional events first reported by Venumadhav *et al.* [Phys. Rev. D **101**, 083030 (2020)] but find no evidence for additional BBH events. We also find no evidence that any of the Venumadhav *et al.* observations are lensed counterparts. We demonstrate how this type of counterpart search can constrain hypotheses about the overall source and lens populations and we rule out at very high confidence the extreme hypothesis that all heavy BBH detections are in fact lensed systems at high redshift with intrinsic masses $< 15 \, M_\odot$.

I. INTRODUCTION

Dense accumulations of matter, such as galaxies or galaxy clusters, can bend the path of light from sources behind them, an effect known as gravitational lensing [1]. Similarly, gravitational waves (GWs) can be lensed by masses between the source and observer (see e.g., [2,3]). In the strong lensing regime, multiple images are produced with a delay between arrival times.

Since 2015, Advanced LIGO and Virgo [4,5] are regularly detecting GWs [6,7] from coalescing binary black holes (BBHs) and binary neutron stars (BNSs). Such a signal is described by a set of parameters including the masses, spins, location and orientation of the source. The observed (detector-frame) masses are increased by cosmological redshift [8]. For lensing of GWs with wavelengths in the Advanced LIGO band by galaxy or cluster lenses, geometric optics apply [9,10]. This means that for multiple images of the same event, the lensed waveforms will be identical up to different arrival times, amplitudes and phases [11], with apparent positions on the sky that are indistinguishable to GW observatories.

The official LIGO-Virgo event catalog GWTC-1 [6] from the O1 and O2 observing runs contains coalescences of ten BBHs and one BNS. The possibility that some of these are lensed images of a single event has previously been suggested [12–15]. A systematic study [16] found no evidence for multiple images, or any other lensing effects, among the ten BBHs.

However, one observational signature of strong lensing has not yet been systematically tested. Large relative magnification between images of the same event can lead to "sub-threshold" counterparts [17] of the known events, which the broad searches conducted for GWTC-1 were not able to confidently extract from the data. In this paper, we perform ten separate targeted reanalyses of Advanced LIGO O1 and O2 data [18–20] searching for faint lensed counterparts.

We describe the general setup of our subthreshold targeted searches and subsequent candidate validation in Sec. II. This method can also readily form the basis for robust searches for lensed counterparts in future observing runs. We then present our results in Sec. III, finding no evidence for previously unknown BBH candidates above background but recovering some of the non-GWTC-1 events previously reported in [21,22]. Additional Bayesian inference demonstrates that these are more likely to be independent mergers with parameters coincidentally similar to GWTC-1 events than actual lensed counterparts.

While the expected rate of strongly lensed events at current sensitivity is very low [14,23], even our null results

---
[*]connor.mcisaac@port.ac.uk
[†]david.keitel@ligo.org

can provide relevant constraints on the astrophysical population of lensed BBH sources. As a demonstration of this, in Sec. IV we rule out the hypothesis of [13] who proposed that most, if not all, of the heavy BBH observations in GWTC-1 could be strongly lensed images of BBH mergers with component masses below 15 solar masses.

II. SEARCH SETUP

The problem we wish to solve is the following. We have a confident GW observation that has already been identified in the data. We want to observe lensed counterpart images to that observation. Such counterpart images, if they exist, would be identical to the original image except that they would be shifted in time, have a different amplitude and potentially have a shifted coalescence phase [11]. In addition to potential counterpart images the data will also contain additional GW signals from other sources, including weak signals not identified with previous searches. The problem, therefore, is not only to identify potential lensed counterpart images in the data, but also to distinguish between lensed counterpart images and unrelated GW signals from other mergers.

The optimal method to identify lensed counterpart images of a GW observation would be to perform Bayesian inference on the full dataset [24]. This would need to include prior knowledge of the rate and properties of counterpart images given the information known about the GW signal. It would also need to include prior information about the population properties of any other, unrelated GW signals. One would then attempt to determine the posterior probability of a lensed counterpart image, or images, being present in the data. However, performing such a search over significant stretches of an observation run would have a prohibitively high computational cost, and the choice of population priors would require significant modeling work.

We therefore pursue a more pragmatic approach, based on current standard search methods to identify isolated GW signals from compact binary mergers. Most of these methods use matched filtering over a grid of template waveforms, using a set of assumptions and analytical maximization to restrict the search parameter space to that of the component masses and component spins [25–28]. Building on this standard approach, we perform our search for lensed counterpart images in two steps. In our initial step we use the PyCBC search method [27,29] to search for compact binary mergers using only a single search template waveform. This allows us to extract weak candidates from the data that match well with that single template, such as the lensed counterparts that we target. However, given that LIGO and Virgo have observed numerous BBHs in a relatively small region of parameter space [6,30], it is also possible that our search will observe additional GW signals from separate sources. We therefore perform a second step

where, assuming a particular candidate observed in the first step is astrophysical, we use fully coherent Bayesian inference [31] to compute the posterior probability between the hypothesis that it is a lensed image of the original event and the hypothesis that it came from a separate source. These steps are described in more detail in the rest of this section.

A. Data selection

For each BBH event from GWTC-1, we search the entire observing run in which it was found. Here we assume that the break between O1 and O2 (316 days) is longer than any time delays expected from typical astrophysical lenses—typically less than ∼1 month [32]. The total two-detector coincident time of publicly available data [18–20] from the two LIGO detectors [5] is 48 days for O1 and 117 days for O2. These durations already exclude any stretches of problematic data quality using the data quality categories discussed in [18] along the same criteria as laid out in [27] and used in the GWTC-1 PyCBC search. Similar to the standard PyCBC search used for GWTC-1 [6,33], we do not use Virgo [4] data for the initial search stage, though we use it for the Bayesian inference step on some events in the last month of O2.

We cannot include the short data stretch around GW170608 when the LIGO Hanford detector was nominally out of observing mode [34]: For robust statements on any candidates from this stretch, we would have needed to include and characterize all data from other periods when one or both of the detectors were in similar states to produce a consistent background, and data from such periods is not public. Instead, we accept this short stretch as just another blind period, similar to any other when there were no data in nominal observing mode from both detectors at the same time.

B. Template selection

We perform a separate search for lensed images of each BBH from GWTC-1. We do not include the BNS event GW170817 in this analysis, since its close distance and extensive electromagnetic observational coverage [35] already rule out strong lensing. Each search uses a single aligned-spin template waveform, generated using the SEOBNRv4_ROM model [36–38].

To select the template parameters we have started from the public posterior samples [39] produced with the precessing IMRPhenomPv2 waveform [40–42]. Since there is no evidence for precession in any of these events [6] and the PyCBC search pipeline in its standard configuration currently relies on aligned-spin waveforms, we select the aligned-spin subset of parameters from the peak of each posterior. To do so, we obtain a four-dimensional Gaussian kernel density estimator (KDE) [43] in $\{m_1, m_2, a_{1z}, a_{2z}\}$ on each set of samples and determine the maximum-posterior (MaP) set of parameters from it.

TABLE I.   Configurations and sensitivity metrics of the ten searches for counterparts of GWTC-1 BBH events. The first four columns after the event name give the parameters of the aligned-spin template used in each search (masses in the detector frame). d_{post} is the sensitive distance of the search, determined by using injections drawn from the full posterior. d_{single} is the sensitive distance when using only the search template parameters for injections. Sensitive distances here are not reinterpreted to account for lensing magnification.

Event	m_1 [$M_\odot$]	m_2 [$M_\odot$]	a_{1z}	a_{2z}	d_{post} [Mpc]	d_{single} [Mpc]
GW150914	37.46	34.47	−0.02	−0.02	1634^{+4}_{-4}	1752^{+12}_{-13}
GW151012	25.13	17.89	0.00	0.00	923^{+4}_{-4}	1048^{+14}_{-14}
GW151226	14.72	8.57	0.26	0.04	562^{+2}_{-2}	604^{+7}_{-7}
GW170104	37.45	23.79	−0.03	−0.03	1342^{+5}_{-5}	1433^{+16}_{-16}
GW170608	10.55	9.01	0.01	0.01	643^{+2}_{-2}	675^{+7}_{-7}
GW170729	79.45	48.50	0.60	0.05	2782^{+10}_{-10}	2965^{+29}_{-29}
GW170809	41.27	28.01	0.03	0.03	1548^{+8}_{-9}	1602^{+15}_{-15}
GW170814	32.48	29.42	0.06	0.03	1536^{+4}_{-4}	1580^{+11}_{-12}
GW170818	40.62	34.79	−0.06	−0.05	1604^{+5}_{-5}	1757^{+16}_{-16}
GW170823	49.06	40.75	0.03	0.03	2013^{+8}_{-8}	2076^{+25}_{-25}

Here, m_1 and m_2 are the BBH's component masses in the detector frame while a_{1z} and a_{2z} are the spin magnitudes along the binary's orbital angular momentum axis. We then generate a SEOBNRv4_ROM waveform at each set of MaP parameters to serve as our search template. The parameters of these ten templates are listed in Table I.

Our choice to only use a single template to search for lensed counterpart images for each GWTC-1 observation is very different from the approach presented by [17] who reduced the original GWTC-1 template bank by a more modest factor. To validate our choice, we compute the matches

$$m(h_1, h_2) = \max_{t_0, \phi_0}\left(\frac{h_1}{\sqrt{(h_1|h_1)}} \middle| \frac{h_2}{\sqrt{(h_2|h_2)}} \right), \qquad (1)$$

defined using the usual inner product $(h_1|h_2)$ [25] which depends on the detector noise curve, between our (aligned) MaP waveforms and any draws from the whole set of (precessing) posterior samples. This corresponds to the fraction of the signal-to-noise ratio (SNR) that we expect to recover under ideal conditions when searching for signals drawn from the full posterior using the single template. We compute these matches separately for each detector, estimating its power spectral density (PSD) using the Welch method [44] with public data around each event.

The resulting matches are shown in Fig. 1. They are consistently high for the higher-SNR events (e.g., a worst match of 95% for GW150914) and still quite high for the bulk of the posteriors for all other events. For lower-SNR events, there is a tail of a small number of the posterior samples for which the matches are lower, in particular for GW151012 where the lowest match is 0.5. However, even for GW151012 we find 90% of samples with matches >0.89. This indicates that a single aligned-spin template matches well to the majority of each event's posterior

samples. Therefore our targeted single-template search provides the ability to match well to most potential lensed counterpart signals while greatly reducing the rate of background compared to a full template bank search, as we demonstrate below.

C. Candidate identification and significance

We use the PyCBC search pipeline [27,29] to identify potential lensed counterpart images in the data. Overall, our setup is quite similar to that of the original "offline" PyCBC analysis as reported in [6,33]. We briefly describe the
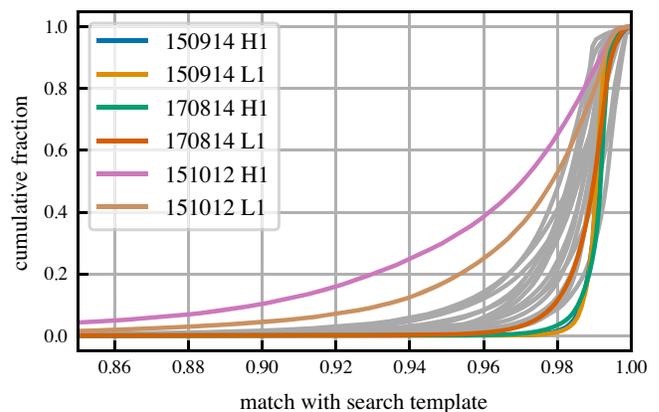

FIG. 1.   The matches between the aligned-spin single template that we use in each of our searches for lensed counterparts of GWTC-1 events and the (precessing) public posterior samples [39] for the corresponding event. This is plotted individually for each event and for the PSD of each detector at the time of the event. We highlight GW151012, the observation with the lowest SNR for which the distributions of matches are broadest, and GW150914 and GW190814, the events with the highest SNR for which the distributions are narrowest. All other events are plotted using gray lines.

search method here, highlighting differences between our targeted search and the original PyCBC analyses.

First, matched filtering is performed separately for each LIGO detector, and we record all single-detector candidates ("triggers") above a minimum SNR of 4. This is a lower threshold than in the original GWTC-1 PyCBC search, which used a threshold of 5.5. This is particularly important for identifying candidates in times when the sensitivities of the two detectors differ from each other. For each single detector candidate, signal consistency tests are applied [45,46] and the single-detector statistic described in [46] is calculated. For identifying these initial triggers we do not use any information about the GWTC-1 events other than their apparent intrinsic parameters—the observed component masses and spins. We do however include the time delay between the observed candidate events and the original GWTC-1 observation in an alternative approach to assessing the significance of such candidates, as described below. Other properties of the source, in particular the sky location, only enter our method when performing our second step of Bayesian hypothesis testing, as described in Sec. II E.

After using the search template to find a list of single-detector triggers, we identify coincident candidates by testing for consistency in arrival time between detectors. Each of these is assigned a ranking statistic (from [47]) and we then compute their significance by comparing these values against a background distribution. This background is measured by performing a large number of unphysical time shifts of the datasets [25,26] and applying the same method as described above to find a new set of coincident triggers. The same ranking statistic can then be calculated for this new set of background triggers. Using this background we can estimate, for each candidate, the rate at which we expect to see events with ranking statistic as loud or louder than that candidate to occur in the background; we refer to this as the candidate's false-alarm rate. With only a single search template, our targeted searches have much less freedom to match background noise than the GWTC-1 searches, which used O(100 000) search templates [6]. This means that a signal that does match our single search template well will have a lower false-alarm rate than in the original GWTC-1 search. This is illustrated in Fig. 2 where we compare the mapping from detection statistic to false-alarm rate when using only a single template against using the full set of template waveforms as in the PyCBC GWTC-1 search [33]. For a background rate of one per year the required detection statistic is decreased by ≈1, corresponding to ∼13% smaller detectable amplitudes.

With the settings used in the GWTC-1 search, PyCBC reports no more than one candidate signal within a ±10 s window [27]. This was motivated by the fact that the event rate is small enough that it is highly unlikely to observe two astrophysical signals from separate sources in a 10 s
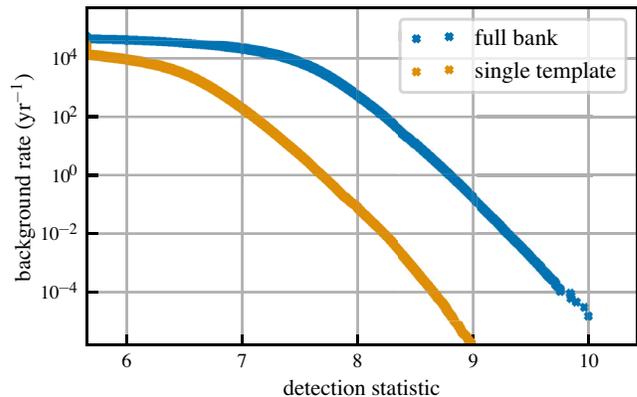

FIG. 2. Background distribution, as a function of the detection statistic, for the single-template GW150914 counterpart search compared to a full template bank as used in the PyCBC search in [6]. The detection statistic is roughly proportional to the signal strain amplitude.

window. Using the same setting, we would not have been able to recover any lensed counterpart signals with time delays shorter than 10 s around each GWTC-1 event, regardless of their strength. To ensure that there are no lensed counterparts within a 10 s window, we perform additional reanalyses with this clustering criterion disabled over these narrow time windows, for all nine events besides GW170608.

In addition to the standard significance estimate of a candidate in terms of a false-alarm rate from the time-shifted background, it is also informative to include the candidate's time delay from the associated GWTC-1 event when ranking possible lensed counterparts. Astrophysical models for strong lensing favor shorter time delays, but the specific distribution varies greatly between models (see e.g., [48]). As a simple first step, we define a new ranking statistic as the inverse false-alarm rate divided by the absolute value of the time delay. In the case that the time delay is less than 10 s, we divide by exactly 10:

$$\varrho = \begin{cases} \text{IFAR}/|\Delta t|, & \text{if } |\Delta t| > 10, \\ \text{IFAR}/10, & \text{otherwise,} \end{cases} \quad (2)$$

where ϱ is the new detection statistic, IFAR is the inverse false-alarm rate and Δt is the time delay between the candidate and the original GWTC-1 event.

This is equivalent to imposing a log-uniform prior on the time delay between the original event and the lensed counterpart between 10 s and the length of data searched and a uniform-in-time prior between 0 and 10 s time delays. Such a prior does not directly follow from specific astrophysical expectations but is a simple choice to reflect our belief that an event with a time delay of minutes or hours is more interesting than one with a time delay of months, while not being too restrictive and setting the prior at any time such that it would be effectively 0.

To produce a background for this new statistic we draw values of the inverse false-alarm rate from our time-shifted background and combine them with a time delay randomly drawn from the analyzed time. Using this background we can calculate a new delay-weighted false-alarm rate, which is then converted into a p value:

$$p = 1 - \exp(-T \cdot \text{FAR}), \qquad (3)$$

where p is the delay-weighted p value; T is the coincident time of the observation run used in the search; and FAR is the false-alarm rate of the candidate.

This reranking could be repeated with any specific prior on time delays, for example to test a particular lens and source population model with an associated predicted time-delay distribution. To facilitate such analyses we release our full lists of coincident triggers [49], allowing others to perform such reranking accordingly.

It may also be informative to include information about the relative magnification compared to the original GWTC-1 event. This could be achieved using a joint prior between time delay and relative magnification. However, the time delay and magnification are not strongly correlated [50] so we do not include such a prior in this first analysis.

It may be possible to reduce the background further by targeting the search toward signals that would not be observable by the original search. However, we choose to leave this for future work.

In our results, we therefore quote both an inverse false-alarm rate obtained from the original ranking statistic and the p value obtained after applying the time-delay weighting. Both the unweighted false-alarm rate and delay-weighted p value are computed independently for each search and do not include an overall trials factor. Hence, under the null hypothesis of no additional GW signals in the data we typically expect one candidate with a p value of $\lesssim 0.1$ from our combined set of ten searches.

D. Search sensitivity

To assess the sensitivity of our search to lensed counterpart images we use simulated signal waveforms added to the detector data. For each of our ten searches we add a set of simulated signals whose parameters are drawn from the full GWTC-1 posterior samples [39] and generated using the precessing IMRPhenomPv2 waveform model. For each of the ten injection sets the coalescence times of the simulated signals are varied uniformly covering a full month of data either side of the original GWTC-1 event. The amplitude of each signal is also multiplied by a scale factor, equal to the square root of the lensing magnification (relative to the original GWTC-1 event). The scale factor is drawn from a uniform distribution between 0.1 and 10 in order to cover the range of signals recoverable by the search, taking into account changes in detector sensitivity over time.

The sensitivity of GW transient searches is often evaluated in terms of sensitive distance, sensitive volume or sensitive volume multiplied by accumulated time. However, it is not straightforward to discuss these measures when dealing with lensed signals, since there will be a large uncertainty on the true luminosity distance to the source. We can quote such a standard sensitivity estimate, however, if we consider our simulated signals with scaled amplitudes not as magnified through lensing but as having different luminosity distances as for unlensed signals. We can then quote the "sensitive distances" of each search, as listed in Table I. These sensitive distances are calculated by applying a detection threshold on the false-alarm rate of one per year and finding the detection efficiency in a number of distance bins. The volume of each distance bin is multiplied by the search efficiency before being summed and converted to a sensitive distance. We do not use the delay-weighted p values when computing these efficiencies.

To compare the sensitivity of our searches, where we use only a single template waveform, to that of the idealized case in which the template bank has 100% match with the full posterior, we create a second set of injections with parameters identical to the MaP values for each search, still searching for them with the same fixed single template. In this case the search template will always match well with the simulated signal. As an example, we consider GW151012, the lowest-SNR event. Figure 3 shows the difference in recovery for the two injection sets, corresponding to a loss of sensitive distance of $\sim$12% compared to the idealized case. This is the largest difference between the two injection sets for all ten events, with losses in sensitive distance of $\sim$3%–9% for the others. This again demonstrates that searching with a single template provides good coverage of the events' posteriors, with the majority
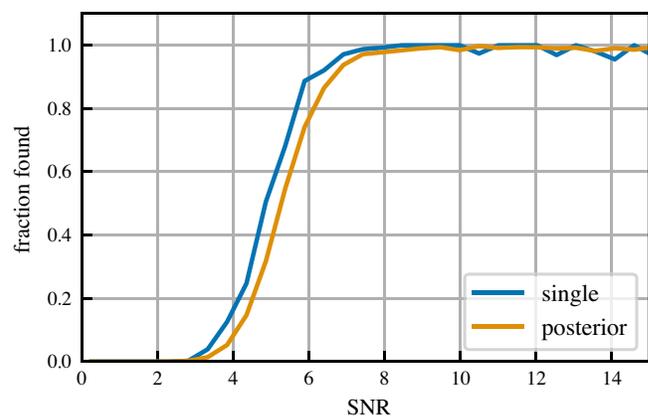

FIG. 3. Fraction of recovered signal injections for GW151012 for two injection sets: one ("single") with all parameters fixed to those of the search template and one ("posterior") with parameters drawn from the full GWTC-1 posterior. An injection is counted as recovered if found with an inverse false-alarm rate greater than one year. The horizontal axis gives the injection SNR defined as the smallest of the two optimal SNRs in each detector.

of events having a loss of $\lesssim 7\%$ compared to the idealised case. The sensitive distances for both injection sets for all ten searches are listed in Table I.

This measure provides an estimate of the loss of sensitivity due to the use of a single template to cover the posterior of each event. However, in practice a search using more templates to cover the full posterior will also increase the rate of background triggers, reducing the sensitivity. Additionally, in this test the injected signals are a perfect match to the search template, whereas a practical template bank will still have some mismatch due to the discrete placement of templates in any bank, which must be balanced with the number of templates required to cover the posterior. There will also be some mismatch between the true signal and the employed waveform models due to the lack of precession in the current search, as well as a small effect due to the limited accuracy of numerical-relativity calibration. An optimal template bank construction along these lines will be an interesting avenue for future research, but, as demonstrated, our extreme single-template choice can already produce very sensitive results.

E. Lensed events or separate gravitational-wave signals?

If any of the single-template searches recover significant candidates, we will have to ask the question whether these are actually the lensed images we are looking for. In other words, if we assume that any given candidate event is astrophysical, what is the probability that it is a lensed counterpart of the original event, compared to the probability that it is an independent GW signal?

In [16], the same question for pairs of GWTC-1 events was answered through KDE-based overlap integrals of posterior distributions from single-event Bayesian inference, following the method introduced by [51]. This approach is somewhat limited by the practical problems in robustly constructing high-dimensional KDEs and hence not enforcing the full constraints on the consistency of the two events as expected under the lensing hypothesis. Instead, here we perform this test through joint Bayesian inference on the pairs of events, fully coherently combining the data from all available detectors for the two relevant times. To compare the lensed and unlensed hypotheses we calculate the evidences $\mathcal{Z}$ for each and use these to find the Bayes' factor (evidence ratio) $\mathcal{B}_{L/U}$ between the two.

A similar approach of "joint parameter estimation" on candidate pairs of lensed events was previously suggested in the context of space-based GW observations by [52] and for the case of LIGO-Virgo observations similar methods have been developed independently [53,54], based on the LALInference and bilby parameter estimation packages [55,56].

In the unlensed hypothesis the two candidates $i = 1, 2$ can be described with independent sets of parameters $\theta_i = \{m_1, m_2, a_{1z}, a_{2z}, d, \iota, \alpha, \delta, \psi, \phi_c, t_c\}$, where α and δ

are the right ascension and declination of the source, respectively, ι is the inclination, ψ is the polarization angle, ϕ_c is the coalescence phase, and t_c is the time of coalescence. The evidence can therefore be calculated as the product of the evidences for each event $\mathcal{Z}_U = \mathcal{Z}_1 \cdot \mathcal{Z}_2$, where each

$$\mathcal{Z}_i = \int P(d_i|\theta_i, \mathcal{I}) P(\theta_i|\mathcal{I}) d\theta_i \qquad (4)$$

and d_i is the dataset for event i; $\mathcal{I}$ represents any implicit assumptions made within the model, e.g., that the data are Gaussian and locally stationary; $P(d_i|\theta_i, \mathcal{I})$ is the likelihood of producing the data given the parameters; and $P(\theta_i|\mathcal{I})$ is the prior on the parameter space.

To calculate the evidences we perform nested sampling using PyCBC Inference [31,57] with the dynesty [58] sampler and the aligned-spin IMRPhenomD waveform model [41,42], again not considering misaligned spins as there is no evidence of precession for any of the GWTC-1 events. We use a fixed prior for all candidates: t_c uniform within 0.2 s around the time reported by the search, component masses uniform within $[5, 120]$ $M_\odot$ (detector frame), spin magnitudes uniformly distributed in $[0, 0.99]$, with an equal probability of being aligned or antialigned with the orbital angular momentum, distance uniform in $[10, 5000]$ Mpc, and ψ uniform in $[0, 2\pi]$. The pairs of (ι, ϕ_c) and (α, δ) are both chosen so that they are uniformly distributed on a sphere. The PSD is calculated using 1024 s of data centred on the time reported by the search.

Under the lensed model the two signals share the same astrophysical origin and therefore share the same intrinsic parameters. The typical change in sky position between lensed signals is far smaller than the resolution of GW detectors and (α, δ) can therefore also be treated as shared parameters. The two candidates therefore have a set of common parameters $\theta' = \{m_1, m_2, a_{1z}, a_{2z}, d, \iota, \alpha, \delta, \psi\}$ with only the coalescence phase, time of coalescence and relative magnification changing between events: $\theta_i = \{\phi_c, t_c, \mu_{rel}\}$, where $\mu_{rel} = \mu_1/\mu_i$ is the magnification of the additional candidate relative to the original GWTC-1 event. The evidence is therefore given by

$$\mathcal{Z}_L = \int P(d_1|\theta', \theta_1, \mathcal{I}) P(d_2|\theta', \theta_2, \mathcal{I})$$
$$\cdot P(\theta', \theta_1, \theta_2|\mathcal{I}) d\theta' d\theta_1 d\theta_2. \qquad (5)$$

This lensed evidence is also calculated using nested sampling. For each sample, we draw a set of common parameters θ' and a set of independent parameters θ_i for each candidate. The shared parameters use the same priors as in the unlensed case. In addition, the first event has a uniform prior on ϕ_c in $[0, 2\pi]$, while the coalescence phase of the second signal, in the case of an aligned-spin system undergoing circular inspiral, can be shifted relative to the first [11] by values of $\{0, \pi/4, \pi/2, -\pi/4\}$; these values are

each given an equal prior weight. The prior on t_c for the first event is uniform in a 0.2 s window around the original time reported in GWTC-1, while the prior on t_c for the second candidate is uniform in a 0.2 s around the trigger time reported by our search. Finally, the relative magnification is 1, by definition, for the original GWTC-1 event and uses a uniform prior on $\sqrt{\mu_{\mathrm{rel}}}$ between [0.1, 10] for the second candidate.

We then combine the shared parameters and independent parameters for each signal and use these to produce a separate waveform for each. These waveforms are then used to evaluate the two likelihoods in Eq. (5), which we then combine to calculate the total likelihood for the lensed model. This method allows us to calculate the lensed evidence directly using nested sampling. Finally we combine the two evidences to calculate the Bayes' factor $\mathcal{B}_{\mathrm{L/U}} = \mathcal{Z}_{\mathrm{L}}/\mathcal{Z}_{\mathrm{U}}$.

These Bayes' factors are only meaningful if both events are of an astrophysical origin. In order to produce a Bayes' factor comparing the lensed hypothesis to the unlensed hypothesis without assuming that the candidates are astrophysical, a more detailed model for terrestrial noise sources would be required.

The Bayes' factors produced by this analysis not only test if two candidates are consistent with a single set of parameters, but also implicitly test if two independent astrophysical events are likely to occur with these parameters given the priors. A pair of events with parameters consistent with one another will produce a higher Bayes' factor when occurring in a region of low prior support than in a region of high prior support, matching the expectation that multiple independent astrophysical events should be observed more frequently in regions of high prior support. For this reason the Bayes' factors produced by this analysis will be highly prior dependent, with a wider prior favoring the lensed hypothesis more.

In the future it will be possible to use the large number of BBHs observed in O3 to produce an astrophysical prior for this analysis, but for the current analysis we will see in the next section that the standard wide uniform priors are already sufficient to reject the possible candidate pairs that we find in this case.

III. RECOVERED CANDIDATES

In this section we will show the results of our ten single-template searches, along with the results of Bayesian inference carried out for each of the resulting candidates.

Several of our searches recover other GWTC-1 events with high significance; details of these cross-matches are provided in Table IV in the Appendix. This is an expected result of the well-known clustering of high-mass detections in a small part of the overall parameter space [21,22,30]. As already demonstrated by [16], none of these pairs of GWTC-1 events are viable candidates for strongly lensed double images after considering their relative time delays

TABLE II. The most significant recovered candidates from our ten single-template searches, sorted by their delay-weighted p value which prefers short time delays (first column), cut at ≤0.5. Candidates with an inverse false-alarm rate (second column, derived from the original ranking statistic) ≥1 yr are also included, regardless of their time delay. The third column gives the candidate end time (UTC) and the fourth column gives the event name for those candidates already published by [21,22] (listed separately in the lower half of the table). The GWTC-1 event whose counterpart search found the candidate is listed in the fifth column, and the absolute value of the time delay is given in the sixth column. Candidates that are themselves listed in GWTC-1 are excluded; see Table IV in the Appendix for those. The final column is the log Bayes' factor (evidence ratio) comparing the lensed vs unlensed hypothesis, assuming both events are astrophysical and using nested sampling to compute the evidence when the two events have a shared set of parameters and when their parameters are independent (see Sec. II E). Times and events that have Virgo data available have been marked with an asterisk (*); these data have been used only when calculating the log Bayes' factors.

| Delay-weighted p value	False-alarm rate^{-1} [yr]	UTC time	Known event?	Found by	$\lvert \Delta t \rvert$ [d]	$\log_e(\mathcal{B}_{\mathrm{L/U}})$
0.16	0.166	2017-07-30 08:05:26.8		GW170729	0.548	-0.77 ± 0.40
0.29	0.0000913	2017-01-04 10:12:57.9		GW170104	0.000687	4.43 ± 0.33
0.37	0.497	2017-08-04 14:57:29.3*		GW170809*	4.73	-2.25 ± 0.32
0.4	0.000550	2017-07-29 19:05:05.9		GW170729	0.00598	-1.43 ± 0.26
0.46	0.000465	2015-09-14 10:04:34.7		GW150914	0.00960	4.61 ± 0.37
0.48	0.000241	2017-07-29 18:51:31.4		GW170729	0.00345	-2.74 ± 0.18
0.48	0.000131	2017-01-04 10:14:56.3		GW170104	0.00206	3.16 ± 0.25
0.86	2.53	2017-04-01 08:19:53.0		GW170809*	130	3.41 ± 0.59
0.0067	2246	2017-01-21 21:25:36.6	GW170121	GW170818*	208	-2.55 ± 0.92
0.064	191	2017-01-21 21:25:36.6	GW170121	GW170823	214	-8.05 ± 0.52
0.097	1.57	2017-07-27 01:04:30.0	GW170727	GW170729	2.74	-1.98 ± 0.48
0.37	15.5	2017-03-04 16:37:53.4	GW170304	GW170729	147	-2.41 ± 0.40
0.48	12.4	2017-03-04 16:37:53.4	GW170304	GW170823	172	-7.30 ± 0.55
0.99	1.14	2017-03-04 16:37:53.4	GW170304	GW170818*	166	-5.54 ± 0.76
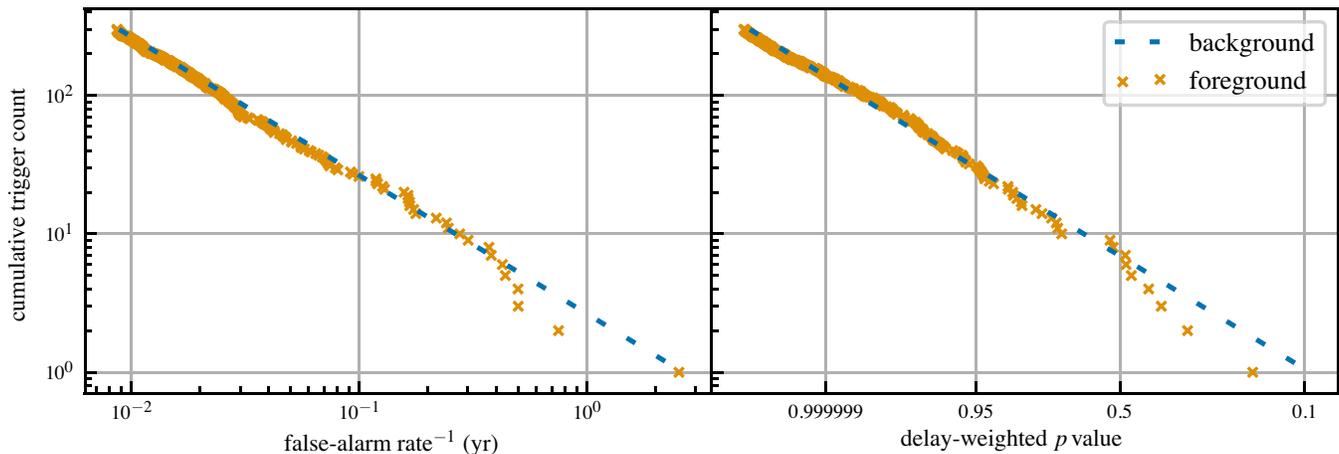
FIG. 4. Combined results from our ten single-template searches. Left panel: The cumulative count of coincident triggers with inverse false-alarm rates less than or equal to a given value. Right panel: The cumulative count of coincident triggers with delay-weighted p values greater than or equal to a given value. The dashed line is the distribution of background triggers produced from time-shifted detector data with randomly drawn time delays. The crosses are foreground triggers, excluding those from GWTC-1 and those reported in [21]. This set of foreground triggers is fully consistent with the null hypothesis.

and the degree of posterior overlap over all relevant parameters.

Table II lists the most significant candidate events from our search after removing pairs of GWTC-1 events. We separate this table into new candidates and those already reported by [21] and later by [22].

The most significant new candidate by inverse false-alarm rate has a value of 2.53 yr; this is consistent with the expectation that we would observe one noise event with a similar significance, as the ten searches covered a total of ~2.6 yr of data. Additionally, the most significant new candidate when ordered by p value has a value of 0.16, due to no trial factor being included in the statistic one event with a p value of $\lesssim 0.1$ is expected from the combined ten searches. The set of new candidates is therefore consistent with the null hypothesis when ranked either by inverse false-alarm rate or p value. This can also be seen in Fig. 4, showing that the rate of new foreground events is consistent with the empirically measured background.

However, we do recover three of the events from [21] (GW170121, GW170304, and GW170727) with false-alarm rates of less than one per year, some of them from multiple searches. These were not considered in the test for lensed pairs in [16]. For GW170121 and GW170304, the large time delays from their matching GWTC-1 events (5–7 months) already suggest that these pairs are unlikely due to lensing (at least by the more common galaxy lenses) but more probably come from two unrelated sources with similar characteristics. On the other hand, the time delay between GW170727 and GW170729 is relatively short, warranting further investigation.

As discussed in Sec. II E we perform Bayesian inference on all candidate pairs in Table II in order to calculate Bayes' factors $\mathcal{B}_{L/U}$ comparing the hypothesis that they are lensed

counterparts against the hypothesis that the two events come from different sources. Positive values indicate support for the lensed hypothesis. The analysis is performed multiple times in order to estimate the error on the Bayes' factor and the results of this analysis are included in Table II.

The $\log_e \mathcal{B}_{L/U}$ values for four of the new low-significance candidates in the first half of the table are nominally supportive of the lensed hypothesis. However, these must be considered in the context of several mitigating factors. First, the expected lensing rate at O1 and O2 sensitivities is very small under standard astrophysical assumptions [14,23], meaning that very large Bayes' factors, $\mathcal{O}(1000)$, would be needed to obtain high posterior odds after factoring this rate in as an explicit prior between the two hypotheses. Second, these results are not independent of the initial search stage. By first testing for consistency in the intrinsic parameters, selection effects are introduced that will favor larger values of $\log_e \mathcal{B}_{L/U}$. As discussed in Sec. II E the wide priors on masses used in this analysis will also favor the lensed hypothesis if the astrophysical distribution on masses [30] is much narrower than this. Most importantly, we have already shown that the set of candidates in the first half of the table are fully consistent with the background noise model, while the Bayes' factors calculated are only meaningful if both candidates are real astrophysical events. Our conclusion, therefore, is that these are most likely noise events, and the positive $\log_e \mathcal{B}_{L/U}$ values are spurious.

For the events in the second half of the table, which have been previously discovered [21,22] as independent BBH candidates, all $\mathcal{B}_{L/U}$ values favor the unlensed hypothesis. We therefore find no evidence for strong lensing present between the BBHs reported in GWTC-1 and [21,22]. This analysis therefore extends the results from [16] to
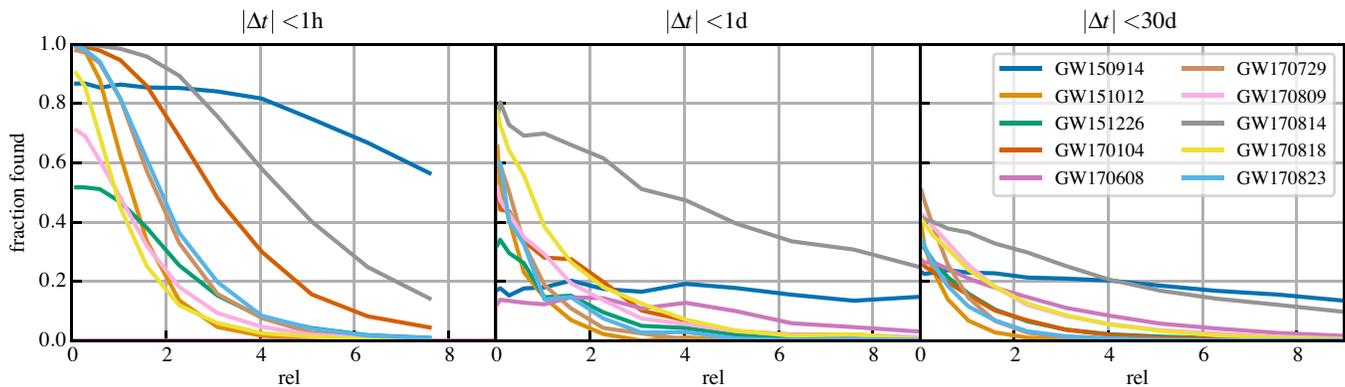
FIG. 5.   The probability of finding lensed images of the GWTC-1 events in each of the ten corresponding searches as a function of relative magnification $\mu_{\rm rel} = \mu_0/\mu_{\rm inj}$, for several ranges of time delays Δt. This is calculated using simulated signals drawn from each event's posterior with thresholds of an inverse false-alarm rate ≥ 1 yr or delay-weighted p value <0.16.

demonstrate that no pairs of candidates from O1 and O2 have evidence supporting the presence of strongly lensed double images.

While negative, these Bayesian inference results are still useful in order to investigate the types of candidates that subthreshold searches for lensed counterparts return. In future searches of this type, it will be a continuing challenge to confidently identify a lensed counterpart observation as opposed to observing a pair of similar but separate GW events. However, the $\mathcal{B}_{\rm L/U}$ can at least be used to rule out pairs that strongly disfavor the lensed hypothesis. More detailed analyses taking into account detailed lens modeling [59] or even targeted follow-up campaigns [60] may then become feasible on any strong remaining candidates. Also, as our knowledge of the astrophysical distribution of GW source systems improves, we will be able to better constrain the priors that go into this analysis and hence better distinguish promising candidates for lensed pairs from chance coincidences in parameter space.

IV. IMPLICATIONS OF THE ABSENCE OF CLEAR COUNTERPART CANDIDATES

We now explore the astrophysical implications of the null result of our searches, based on their estimated sensitivity. To find the expected recovery rate for lensed images in our searches we use simulated signals added to the GW strain data before matched filtering ("injections"). Of the injection sets described in Sec. II B, in the following we always use the injections drawn from the full precessing posterior of the original GWTC-1 events [39]. Their strain amplitudes are multiplied by a scale factor $\sqrt{\mu_{\rm rel}} = \sqrt{\mu_0/\mu_{\rm inj}}$, where μ_0 and $\mu_{\rm inj}$ are the absolute magnifications of the primary and injected signals, respectively.

We combine two injection sets, one uniformly covering a full month of data on either side of the original GWTC-1 event with $\sqrt{\mu_{\rm rel}} \in [0.1, 10]$ and one focusing on a smaller window 2 h either side of the event with $\sqrt{\mu_{\rm rel}} \in [1/3, 3]$ in order to more accurately calculate recovery rates for short

time delays. Any injection made outside of the analyzed times for either of the two detectors is considered to be missed to account for the detector duty cycle.

The recovery rate of these injections yields an estimate for the probability of finding a lensed image as a function of magnification ratio $\mu_{\rm rel}$ and time delay Δt. These results are summarized in Fig. 5, where we have now used a threshold based on the actual search results: To be counted as recovered, an injection must be more significant than anything in the top half of Table II, i.e., have an inverse false-alarm rate ≥ 1 yr or a delay-weighted p value <0.16. Along the $\mu_{\rm rel}$ dimension, we find that recovery mostly depends on the strength of the original GW event. Along the time dimension, recovery is mostly limited by the $\lesssim 50\%$ coincident livetime of the LIGO detectors during O1 and O2.

Any astrophysical interpretation of the absence of convincing counterpart candidates depends on the choice of priors for the true (unlensed) high-redshift BBH source population and the properties of lenses in the Universe. We provide supplemental data [49] for the sensitivity of all ten searches so that other authors may explore the implications of our search results under different choices for these priors.

A. Extreme lensing hypothesis

Here, we demonstrate one application of these results, testing a strict interpretation of the idea from [13]: What if *all* of the high-mass GWTC-1 BBH events really came from lighter objects at higher redshifts? To be more specific, we phrase the test hypothesis as: *The intrinsic component masses of any BBH in the Universe cannot be larger than* 15 $M_\odot$. *All apparently heavier GW events are due to lensing.*

Let us start from the following quantities under the standard no-lensing hypothesis (U) from GWTC-1 [6]: primary source-frame masses $m_{1,{\rm U}}$, redshifts $z_{\rm U}$ and luminosity distances $d_{\rm U}$. We then define the extreme lensing hypothesis (L) as there being no merging BBHs in the

TABLE III. Parameters of the eight heaviest GWTC-1 events, reinterpreted under the extreme lensing hypothesis (L) that they all should have intrinsic primary masses of $m_{1,U} = 15\ M_\odot$. Unlensed parameters (U) correspond to the median values from [6]. Luminosity distances are obtained under a standard Planck cosmology [62]. P_{single} is the probability of an event of this magnification not having a comparably bright counterimage (with $1/3 \leq \mu_{rel} \leq 3$). P_{missed} is the probability that our search fails to recover the counterimage in the LIGO data (with delay-weighted p value below 0.16). P_{found} is the probability of our search recovering a counterimage under the extreme lensing hypothesis. GW151012 is excluded from the analysis as we only consider systems with magnification >100.

Event	$m_{1,U}\ [M_\odot]$	z_U	z_L	d_U [Gpc]	d_L [Gpc]	μ	P_{single}	P_{missed}	P_{found}
GW150914	36	0.09	1.6	0.4	12	800	0.020	0.031	0.950
GW151012	23	0.21	0.9	1.1	5.8	30	$\cdots$	$\cdots$	$\cdots$
GW170104	31	0.20	1.5	1.0	11	130	0.046	0.127	0.832
GW170729	50	0.49	4.0	2.8	37	170	0.040	0.116	0.848
GW170809	35	0.20	1.8	1.0	14	200	0.038	0.072	0.893
GW170814	31	0.12	1.3	0.6	9.3	260	0.034	0.043	0.925
GW170818	35	0.21	1.8	1.1	19	200	0.038	0.051	0.913
GW170823	40	0.35	2.5	1.9	21	130	0.046	0.127	0.833

Universe with component masses over $15\ M_\odot$; we implement this by fixing the intrinsic (source-frame) primary mass of each event to $m_{1,L} = 15\ M_\odot$.[1] Having all primary masses equal to exactly $15\ M_\odot$ is an unlikely source distribution, but including a distribution of lower masses would only increase the magnifications required to produce the same detector frame masses. As discussed below, this would decrease the probability of no additional lensed images being present, therefore increasing the chance of observing these counterparts and strengthening the arguments below.

We can then easily obtain the "corrected" redshifts under the lensing hypothesis as

$$z_L = \frac{m_{1,U}}{15\ M_\odot}(1 + z_U) - 1. \qquad (6)$$

These are then converted to luminosity distances under the standard Planck cosmology [62], and the corresponding lensing magnification factors are $\mu = (d_L/d_U)^2$. Table III lists the results for the eight heaviest BBH sources (with median $m_1 > 15\ M_\odot$), using median results from [6] for simplicity.

GW151226 and GW170608 have median $m_1 < 15\ M_\odot$ and hence are already consistent with the hypothesis that all black holes have masses below $15\ M_\odot$, so we do not include these in our test. GW151012 at a median $m_1 \approx 23\ M_\odot$ is included in the table for completeness but also not used in the test, because its required magnification of 30 does not fall into the sufficiently extreme regime for the approximations discussed below, which require magnifications greater than ~ 100. As we will see, the result of the test is also already sufficiently strong from just considering the remaining seven higher-mass signals.

For the seven remaining signals the ratio of intrinsic and observed masses yields corrected redshifts and luminosity distances, from which the required magnifications are between 100 and 800 with GW150914 having the largest magnification. (Again, see Table III.)

B. Astrophysical priors on magnification and time delay

Lensing events with magnifications greater than 100 are rare in the Universe, with the distribution of magnifications following $P(\mu) \sim \mu^{-3}$, due to the small area in the lens plane where high μ can be produced. Still, such values are possible, particularly for point sources. For example a star at redshift 1.5 has been observed with a magnification of ~ 2000 [63]. The catastrophe theory of strong gravitational lensing [64] shows that images with very high magnification are formed when the source lies either: (i) just inside a fold catastrophe, forming a pair of images with the same brightness; (ii) just inside a cusp catastrophe, forming a triplet with one image twice as bright as the other two; or (iii) just outside the cusp catastrophe, forming a single highly magnified image [1]. Higher-order catastrophes can produce more complicated configurations [65,66] but are extremely rare [67,68]. The time delays between the multiple highly magnified images are extremely short, typically seconds to at most hours, with $\Delta t \sim \mu^{-3}$.

The lensing mass of galaxies is well approximated by singular isothermal ellipsoids [69]. In this case the fraction of highly magnified images without a comparably bright counterimage is given by [70]

$$P_{single} = \frac{1}{1 + 4\pi \frac{15}{16\sqrt{6}} \frac{\sqrt{1-q^2}}{1+q} \mu^{1/2}}, \qquad (7)$$

where q is the axis ratio of the lens. Thus, unless the lens is very close to spherical ($q = 1$), highly magnified images

---

[1]In [13], the $15\ M_\odot$ limit was originally phrased in terms of chirp mass. But since it is motivated by the observed masses for galactic black holes with a reference to [61], which in turn refers to the *individual* black hole masses in Cyg X-1 and GRS 1915, we reinterpret the hypothesis in terms of component masses.

are unlikely to occur without a bright counterimage. In the case of a lensed BBH, we do not know the specific lens and cannot measure q directly. Instead we must marginalize over the population of all potential lenses in the Universe. We use the lens population model of [71] to realize the population of gravitational lenses in the Universe. This model assumes lens galaxies are singular isothermal ellipsoids. The lenses are randomly distributed with uniform comoving number density out to redshift 2. Lens masses are derived from the Sloan Digital Sky Survey (SDSS) velocity dispersion function [72]. Lens ellipticities are drawn from the velocity dispersion-dependent probability density function fit to SDSS [71]. To perform the marginalization, we draw 10^8 lenses from the [71] population and weight them by their strong lensing cross section (proportional to the Einstein radius squared) for the source plane of each BBH. For each lens, we calculate the probability of seeing a single bright image using Eq. (7). The final probability of seeing a single image is thus

$$P_{\text{single}} = \sum_i \Theta_{E,i}^2 P_{\text{single},i} / \sum_i \Theta_{E,i}^2, \qquad (8)$$

where $\Theta_{E,i}$ is the Einstein radius of each lens.

Using Eq. (8) to marginalize over the lens population we find that only 2% of images with $\mu > 800$, or 4% with $\mu > 100$, have no counterimage with comparable magnification ($1/3 \leq \mu_{\text{rel}} \leq 3$). We define the probability of not having such a bright counterimage as P_{single} in Table III.

The same lens population model allows us to numerically infer the time delay and magnification ratio between counterimages. For each putative lensed BBH event, we realize 10^5 lens systems weighted by their lensing cross section given the true source redshift.

For each lens we draw a random position from a uniform distribution on the image plane. For each position we infer the magnification of a strongly lensed image forming at that location. We do this repeatedly until we find 1000 image positions for each lens that have a magnification within 10% of the putative magnification. To find the unlensed source positions that produce these highly magnified images, we trace the highly magnified images back onto the source plane using the lens equation.[2] This gives a set of source positions for each lens that produce an image of the correct magnification. For each source position, we use the lens equation to find the locations on the image plane where the counterimages form. For each counterimage we calculate the magnification and the time delays relative to the image with the putative magnification. This procedure gives a list of 10^8 lensed BBH events with time delays

---

[2]The lens equation relates the observed and true angular positions of a lensed source. The difference between these two positions is the reduced deflection angle, which is sensitive to the observed image position, the mass distribution of the lens and the angular diameter distances between observer, lens and source.

and magnification ratios relative to the highly magnified image. Assuming the lensing hypothesis, each of the events in this list could have produced the observed highly magnified event.

This model shows that for the highly magnified images as required by the extreme lensing hypothesis, 90% (99%) of the comparably bright counterimages occur within e.g., 5 (45) min for GW170823 and 2 (15) s for GW150914.

C. Results of the hypothesis test

For each simulated highly magnified lensing event, the probability of missing the counterimage is given by one minus the recovery fraction (calculated using the same thresholds as for Fig. 5, already including the duty factors of the detectors during O1 and O2) for injections with the correct time delay and strain ratio relative to the observed BBH event. Marginalizing over all of the simulated image pairs gives the overall probability of missing a comparably bright counterimage, P_{missed}. The probability of seeing a counterimage is thus

$$P_{\text{found}} = 1 - P_{\text{single}} - (1 - P_{\text{single}})P_{\text{missed}}$$
$$\approx 1 - P_{\text{single}} - P_{\text{missed}}. \qquad (9)$$

This yields recovery fractions ranging from 95% for GW150914 to 83% for GW170104. These values are shown for each event in Table III.

The product of $1 - P_{\text{found}}$ for the seven heaviest events is 1×10^{-7}. This is the probability of missing the counterimages for all of the seven events assuming they are all lensed by the magnification corresponding to a maximum component mass of 15 $M_\odot$. If the unlensed masses were lower, the required magnifications increase, decreasing both P_{single} and P_{missed}, thus in turn driving P_{found} further toward 1.

The model presented here does not include lensing by clusters. The mean Einstein radius in the model is 0.7 arcsec, and the time delays are proportional to the Einstein radius, so even if clusters dominate the lensing cross section, the expected time delays would only increase by a factor of a few as the Einstein radius of typical clusters is ~ 5 arcsec [73].

The model also does not account for deviations from isothermality. This introduces a small change in the constant of proportionality in the time delays between highly magnified image pairs. To assess the potential size of these systematics, we rerun our pipeline, but assuming all time delays are 10 times longer than in the fiducial model: In this scenario the probability of missing all of the counterimages increases to 5×10^{-7}. Therefore, possible systematics introduced by cluster lenses and deviations from isothermality will not change the result significantly. Thus, the observed lack of detecting any lensed counterimages is still clearly incompatible with the hypothesis that

all of the high-mass GWTC-1 events are lensed events with intrinsic masses below 15 $M_\odot$.

V. CONCLUSION AND OUTLOOK

We have performed the first focused search for strongly lensed counterpart images to all binary black hole mergers from the GWTC-1 catalog [6]. We recovered several candidates previously found by [21,22]. Performing Bayesian inference on these candidate events we found no evidence that these are lensed counterparts. All other new candidates are consistent with a noise-only background.

The absence of clear candidates from our search already provides useful observational constraints on astrophysical lensing scenarios. For example, if all observed BBHs with component masses greater than 15 $M_\odot$ had originated from lower-mass, highly magnified, high-redshift sources [13], then we should have observed at least one counterpart. We therefore rule out this hypothesis.

Another method to search for subthreshold lensed events has been proposed in [17]. Our single-template searches provide less freedom to match detector noise fluctuations than the template bank employed in [17]. For future applications, the optimal template bank size per event might lie between our single-template method and the larger banks of [17]. For example, one could construct a template bank to obtain a certain minimal match across each event's posterior distribution. We will explore this in more detail in future work.

As discussed in Sec. II, our initial matched-filter search stage currently does not check for consistency in the extrinsic parameters (especially the sky location) between any new candidates and the original event used to provide the search template. This is a key area where the search could be improved in the future. In order to include the extrinsic parameters of the signal a multidetector coherent search [74] could be used. This would improve the sensitivity of the search by removing candidates that match well in the intrinsic parameters but have poor overlap in sky position.

Looking ahead, the LIGO-Virgo O3 run has already yielded a rich crop of additional GW candidates [7], and future observing runs promise many more [75]. It has also been suggested that the first detection of a lensed source is expected within the next 5 yr [23,32]. The framework of targeted subthreshold searches for lensed counterparts as presented in this paper can be readily applied to new detections in O3 and beyond. Observing strongly lensed BBHs before the detector network reaches design sensitivity [75] would imply that the merger rate increases much more steeply with redshift than expected [14,30,76] or challenge the established understanding of lensing statistics.

More generally, once strongly lensed pairs of events can be identified, joint parameter inference on the combined images (as introduced in this paper and independently by [53,54]) can significantly improve estimates of the source properties and location [59]. Hence, our approach is a key contribution toward the discovery of lensed GWs, which will advance our understanding of both cosmology and high-redshift black hole populations.

Supplemental data for this paper are available at [49].

APPENDIX

Table II in the main part of the paper provides only those candidates (found with a delay-weighted p value < 0.5 or an inverse false-alarm rate of more than one year) which do not themselves correspond to another GWTC-1 event. An extended version of the full search results is provided here as Table IV, which includes all candidates above either of

TABLE IV.   List of all candidates with a delay-weighted p value (first column) <0.5 *or* an inverse false-alarm rate (second column) of more than one year. In contrast to Table II, all new candidates as well as any known events from both GWTC-1 and [21] are combined together, all sorted by the delay-weighted p value. The third column gives the candidate end time (UTC) and the fourth column notes if an event at this time has already been published. The GWTC-1 event whose search found the reported candidate is listed in the fifth column, and the absolute value of the time delay between the two is given in the final column.

Delay-weighted p value	False-alarm rate^{-1} [yr]	UTC time	Known event?	Found by	$|\Delta t|$ [d]
$\ll 0.001$	3.26×10^7	2017-08-14 10:30:43.5	GW170814	GW170104	222
$\ll 0.001$	3.26×10^7	2017-01-04 10:11:58.6	GW170104	GW170814	222
$\ll 0.001$	8.15×10^6	2017-08-23 13:13:58.5	GW170823	GW170818	5.5
$\ll 0.001$	3.62×10^6	2017-08-23 13:13:58.5	GW170823	GW170729	25
$\ll 0.001$	1.09×10^6	2017-01-04 10:11:58.6	GW170104	GW170809	217
$\ll 0.001$	1.78×10^3	2017-08-09 08:28:21.8	GW170809	GW170814	5.1
$\ll 0.001$	4.89×10^3	2017-08-09 08:28:21.8	GW170809	GW170818	8.7
0.002	151	2017-08-14 10:30:43.5	GW170814	GW170809	5.1
0.0035	3.94×10^3	2017-01-04 10:11:58.6	GW170104	GW170818	226
0.0067	2.25×10^3	2017-01-21 21:25:36.6	GW170121	GW170818	208
0.0075	1.93×10^3	2017-08-09 08:28:21.8	GW170809	GW170104	217
0.0082	30	2017-08-14 10:30:43.5	GW170814	GW170818	3.7
0.011	74	2017-08-23 13:13:58.5	GW170823	GW170809	14
0.031	43	2017-07-29 18:56:29.3	GW170729	GW170823	25
0.064	191	2017-01-21 21:25:36.6	GW170121	GW170823	214
0.097	1.6	2017-07-27 01:04:30.0	GW170727	GW170729	2.7
0.14	5.3	2017-08-09 08:28:21.8	GW170809	GW170823	14
0.14	11	2015-09-14 09:50:45.4	GW150914	GW151012	28
0.16	0.17	2017-07-30 08:05:26.8		GW170729	0.55
0.17	30	2017-08-23 13:13:58.5	GW170823	GW170814	9.1
0.29	9.13×10^{-5}	2017-01-04 10:12:57.9		GW170104	6.87×10^{-4}
0.37	0.50	2017-08-04 14:57:29.3		GW170809	4.7
0.37	16	2017-03-04 16:37:53.4	GW170304	GW170729	147
0.4	5.50×10^{-4}	2017-07-29 19:05:05.9		GW170729	5.98×10^{-3}
0.44	19	2017-01-04 10:11:58.6	GW170104	GW170823	231
0.46	4.65×10^{-4}	2015-09-14 10:04:34.7		GW150914	9.60×10^{-3}
0.48	12	2017-03-04 16:37:53.4	GW170304	GW170823	172
0.48	2.41×10^{-4}	2017-07-29 18:51:31.4		GW170729	3.45×10^{-3}
0.48	1.31×10^{-4}	2017-01-04 10:14:56.3		GW170104	2.06×10^{-3}
0.86	4.1	2017-08-23 13:13:58.5	GW170823	GW170104	231
0.86	2.5	2017-04-01 08:19:53.0		GW170809	13
0.91	3.3	2017-01-04 10:11:58.6	GW170104	GW170729	206
0.99	1.1	2017-03-04 16:37:53.4	GW170304	GW170818	166

those two thresholds from any of the ten searches, and both sets of known events from GWTC-1 and from [21] are mixed in with the new candidates.

We also provide the full set of search results, without thresholds on p value or false-alarm rate, as machine-readable supplemental data files [49].

[1] P. Schneider, J. Ehlers, and E. E. Falco, *Gravitational Lenses*, Astronomy and Astrophysics Library (Springer, Berlin, 1992), https://doi.org/10.1007/978-3-662-03758-4.

[2] Y. Wang, A. Stebbins, and E. L. Turner, Gravitational Lensing of Gravitational Waves from Merging Neutron Star Binaries, Phys. Rev. Lett. **77**, 2875 (1996).

[3] M. Oguri, Strong gravitational lensing of explosive transients, Rep. Prog. Phys. **82**, 126901 (2019).

[4] F. Acernese *et al.* (Virgo Collaboration), Advanced Virgo: A second-generation interferometric gravitational wave detector, Classical Quantum Gravity **32**, 024001 (2015).

[5] J. Aasi *et al.* (LSC Collaboration), Advanced LIGO, Classical Quantum Gravity **32,** 074001 (2015).

[6] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs, Phys. Rev. X **9,** 031040 (2019).

[7] LIGO Scientific and Virgo Collaborations, GraceDB— Gravitational-wave candidate event database—O3 LIGO/ Virgo Public Alerts, https://gracedb.ligo.org/superevents/ public/O3/.

[8] A. Krolak and B. F. Schutz, Coalescing binaries—Probe of the universe, Gen. Relativ. Gravit. **19,** 1163 (1987).

[9] R. J. Bontz and M. P. Haugan, A diffraction limit on the gravitational lens effect, Astrophys. Space Sci. **78,** 199 (1981).

[10] R. Takahashi and T. Nakamura, Wave effects in gravitational lensing of gravitational waves from chirping binaries, Astrophys. J. **595,** 1039 (2003).

[11] L. Dai and T. Venumadhav, On the waveforms of gravitationally lensed gravitational waves, arXiv:1702.04724.

[12] G. P. Smith, M. Jauzac, J. Veitch, W. M. Farr, R. Massey, and J. Richard, What if LIGOs gravitational wave detections are strongly lensed by massive galaxy clusters?, Mon. Not. R. Astron. Soc. **475,** 3823 (2018).

[13] T. Broadhurst, J. M. Diego, and G. Smoot, Reinterpreting low frequency LIGO/Virgo events as magnified stellar-mass black holes at cosmological distances, arXiv: 1802.05273.

[14] M. Oguri, Effect of gravitational lensing on the distribution of gravitational waves from distant binary black hole mergers, Mon. Not. R. Astron. Soc. **480,** 3842 (2018).

[15] T. Broadhurst, J. M. Diego, and G. F. Smoot, Twin LIGO/ Virgo detections of a viable gravitationally-lensed black hole merger, arXiv:1901.03190.

[16] O. A. Hannuksela, K. Haris, K. K. Y. Ng, S. Kumar, A. K. Mehta, D. Keitel, T. G. F. Li, and P. Ajith, Search for gravitational lensing signatures in LIGO-Virgo binary black hole events, Astrophys. J. Lett. **874,** L2 (2019).

[17] A. K. Y. Li, R. K. L. Lo, S. Sachdev, T. G. F. Li, and A. J. Weinstein, Targeted sub-threshold search for strongly-lensed gravitational-wave events, arXiv:1904.06020.

[18] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Open data from the first and second observing runs of Advanced LIGO and Advanced Virgo, arXiv:1912.11716.

[19] GWOSC (LIGO Scientific and Virgo Collaborations), Advanced LIGO O1 Data Release, 2017, https://doi.org/ 10.7935/K57P8W9D.

[20] GWOSC (LIGO Scientific and Virgo Collaborations), Advanced LIGO O2 Data Release, 2019, https://doi.org/ 10.7935/CA75-FM95.

[21] T. Venumadhav, B. Zackay, J. Roulet, L. Dai, and M. Zaldarriaga, New binary black hole mergers in the second observing run of Advanced LIGO and Advanced Virgo, Phys. Rev. D **101,** 083030 (2020).

[22] A. H. Nitz, T. Dent, G. S. Davies, S. Kumar, C. D. Capano, I. Harry, S. Mozzon, L. Nuttall, A. Lundgren, and M. Tpai, 2-OGC: Open gravitational-wave catalog of binary mergers from analysis of public Advanced LIGO and Virgo data, Astrophys. J. **891,** 123 (2019).

[23] K. K. Y. Ng, K. W. K. Wong, T. Broadhurst, and T. G. F. Li, Precise LIGO lensing rate predictions for binary black holes, Phys. Rev. D **97,** 023012 (2018).

[24] G. Ashton, E. Thrane, and R. J. E. Smith, Gravitational wave detection without boot straps: A Bayesian approach, Phys. Rev. D **100,** 123018 (2019).

[25] B. Allen, W. G. Anderson, P. R. Brady, D. A. Brown, and J. D. E. Creighton, FINDCHIRP: An algorithm for detection of gravitational waves from inspiraling compact binaries, Phys. Rev. D **85,** 122006 (2012).

[26] S. Babak *et al.*, Searching for gravitational waves from binary coalescence, Phys. Rev. D **87,** 024033 (2013).

[27] S. A. Usman *et al.*, The PyCBC search for gravitational waves from compact binary coalescence, Classical Quantum Gravity **33,** 215004 (2016).

[28] C. Messick *et al.*, Analysis framework for the prompt discovery of compact binary mergers in gravitational-wave data, Phys. Rev. D **95,** 042001 (2017).

[29] A. N. Nitz *et al.*, gwastro/pycbc: Post-O2 release 12I (v1.11.14), 2018, https://doi.org/10.5281/zenodo.1407251.

[30] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Binary black hole population properties inferred from the first and second observing runs of Advanced LIGO and Advanced Virgo, Astrophys. J. Lett. **882,** L24 (2019).

[31] C. M. Biwer, C. D. Capano, S. De, M. Cabero, D. A. Brown, A. H. Nitz, and V. Raymond, PyCBC Inference: A PYTHON-based parameter estimation toolkit for compact binary coalescence signals, Publ. Astron. Soc. Pac. **131,** 024503 (2019).

[32] S.-S. Li, S. Mao, Y. Zhao, and Y. Lu, Gravitational lensing of gravitational waves: A statistical perspective, Mon. Not. R. Astron. Soc. **476,** 2220 (2018).

[33] T. Dal Canton and I. W. Harry, Designing a template bank to observe compact binary coalescences in Advanced LIGO's second observing run, arXiv:1705.01845.

[34] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), GW170608: Observation of a 19-solar-mass binary black hole coalescence, Astrophys. J. **851,** L35 (2017).

[35] B. P. Abbott *et al.*, Multi-messenger observations of a binary neutron star merger, Astrophys. J. Lett. **848,** L12 (2017).

[36] M. Pürrer, Frequency domain reduced order models for gravitational waves from aligned-spin compact binaries, Classical Quantum Gravity **31,** 195010 (2014).

[37] M. Pürrer, Frequency domain reduced order model of aligned-spin effective-one-body waveforms with generic mass-ratios and spins, Phys. Rev. D **93,** 064041 (2016).

[38] A. Bohé, L. Shao, A. Taracchini, A. Buonanno, S. Babak, I. W. Harry, I. Hinder, S. Ossokine, M. Pürrer, V. Raymond, T. Chu, H. Fong, P. Kumar, H. P. Pfeiffer, M. Boyle, D. A. Hemberger, L. E. Kidder, G. Lovelace, M. A. Scheel, and B. Szilágyi, Improved effective-one-body model of spinning, nonprecessing binary black holes for the era of gravitational-wave astrophysics with advanced detectors, Phys. Rev. D **95,** 044028 (2017).

[39] GWOSC (LIGO Scientific and Virgo Collaborations), A gravitational-wave transient catalog of compact binary mergers observed by LIGO and Virgo during the first and second observing runs, 2015–2017 (2018), https://doi .org/10.7935/82H3-HH23.

[40] M. Hannam, P. Schmidt, A. Bohé, L. Haegel, S. Husa, F. Ohme, G. Pratten, and M. Pürrer, Simple Model of Complete Precessing Black-Hole-Binary Gravitational Waveforms, Phys. Rev. Lett. **113,** 151101 (2014).

[41] S. Husa, S. Khan, M. Hannam, M. Pürrer, F. Ohme, X. J. Forteza, and A. Bohé, Frequency-domain gravitational waves from nonprecessing black-hole binaries. I. New numerical waveforms and anatomy of the signal, Phys. Rev. D **93,** 044006 (2016).

[42] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. J. Forteza, and A. Bohé, Frequency-domain gravitational waves from nonprecessing black-hole binaries. II. A phenomenological model for the advanced detector era, Phys. Rev. D **93,** 044007 (2016).

[43] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, Scikit-learn: Machine learning in PYTHON, J. Mach. Learn. Res. **12,** 2825 (2011), https://www.jmlr.org/papers/v12/pedregosa11a.html.

[44] P. Welch, The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms, IEEE Trans. Audio Electroacoust. **15,** 70 (1967).

[45] B. Allen, χ^2 time-frequency discriminator for gravitational wave detection, Phys. Rev. D **71,** 062001 (2005).

[46] A. H. Nitz, Distinguishing short duration noise transients in LIGO data to improve the PyCBC search for gravitational waves from high mass binary black hole mergers, Classical Quantum Gravity **35,** 035016 (2018).

[47] A. H. Nitz, T. Dent, T. D. Canton, S. Fairhurst, and D. A. Brown, Detecting binary compact-object mergers with gravitational waves: Understanding and improving the sensitivity of the PyCBC search, Astrophys. J. **849,** 118 (2017).

[48] S. Mao, Gravitational lensing, time delay, and gamma-ray bursts, Astrophys. J. Lett. **389,** L41 (1992).

[49] C. McIsaac, D. Keitel, T. Collett, I. Harry, S. Mozzon, O. Edy, and D. Bacon, Data release for the paper 'Search for strongly lensed counterpart images of binary black hole mergers in the first two LIGO observing runs', https://github.com/icg-gravwaves/lensed-o1-o2-data-release.

[50] M. Oguri and P. J. Marshall, Gravitationally lensed quasars and supernovae in future wide-field optical imaging surveys, Mon. Not. R. Astron. Soc. **405,** 2579 (2010).

[51] K. Haris, A. K. Mehta, S. Kumar, T. Venumadhav, and P. Ajith, Identifying strongly lensed gravitational wave signals from binary black hole mergers, arXiv:1807.07062.

[52] N. Seto, Strong gravitational lensing and localization of merging massive black hole binaries with LISA, Phys. Rev. D **69,** 022002 (2004).

[53] X. Liu, I. M. Hernandez, and J. Creighton, Identifying strong gravitational lensing from the second observing Run of Advanced LIGO and Advanced Virgo (to be published).

[54] R. K. L. Lo and I. Hernandez Magana, A Bayesian statistical framework for identifying strongly-lensed gravitational-wave signals (to be published).

[55] J. Veitch *et al.*, Parameter estimation for compact binaries with ground-based gravitational-wave observations using the LALInference software library, Phys. Rev. D **91,** 042003 (2015).

[56] G. Ashton *et al.*, BILBY: A user-friendly Bayesian inference library for gravitational-wave astronomy, Astrophys. J. Suppl. **241,** 27 (2019).

[57] A. N. Nitz *et al.*, gwastro/pycbc: Release 1.13.8 of PyCBC (2019), https://doi.org/10.5281/zenodo.2801307.

[58] J. S. Speagle, DYNESTY: A dynamic nested sampling package for estimating Bayesian posteriors and evidences, Mon. Not. R. Astron. Soc. **493,** 3132 (2020).

[59] O. A. Hannuksela, T. E. Collett, M. Alkan, and T. G. F. Li, Localizing merging black holes with sub-arcsecond precision using gravitational-wave lensing, Mon. Not. R. Astron. Soc., staa2577 (2020).

[60] G. Smith, M. Bianconi, M. Jauzac, J. Richard, A. Robertson, C. L. Berry, R. Massey, K. Sharon, W. Farr, and J. Veitch, Deep and rapid observations of strong-lensing galaxy clusters within the sky localization of GW170814, Mon. Not. R. Astron. Soc. **485,** 5180 (2019).

[61] M. Dominik, K. Belczynski, C. Fryer, D. E. Holz, E. Berti, T. Bulik, I. Mandel, and R. O'Shaughnessy, Double compact objects II: Cosmological merger rates, Astrophys. J. **779,** 72 (2013).

[62] P. A. R. Ade *et al.* (Planck Collaboration), Planck 2015 results. XIII. Cosmological parameters, Astron. Astrophys. **594,** A13 (2016).

[63] P. L. Kelly *et al.*, Extreme magnification of an individual star at redshift 1.5 by a galaxy-cluster lens, Nat. Astron. **2,** 334 (2018).

[64] A. O. Petters, H. Levine, and J. Wambsganss, *Singularity Theory and Gravitational Lensing*, Progress in Mathematical Physics Vol. 21 (Birkhäuser, Basel, 2001), https://doi.org/10.1007/978-1-4612-0145-8.

[65] C. R. Keeton, S. Mao, and H. J. Witt, Gravitational lenses with more than four images. I. Classification of caustics, Astrophys. J. **537,** 697 (2000).

[66] N. W. Evans and H. J. Witt, Are there sextuplet and octuplet image systems? Mon. Not. R. Astron. Soc. **327,** 1260 (2001).

[67] G. Orban de Xivry and P. Marshall, An atlas of predicted exotic gravitational lenses, Mon. Not. R. Astron. Soc. **399,** 2 (2009).

[68] T. E. Collett and D. J. Bacon, Compound lensing: Einstein zig-zags and high-multiplicity lensed images, Mon. Not. R. Astron. Soc. **456,** 2210 (2016).

[69] M. W. Auger, T. Treu, A. S. Bolton, R. Gavazzi, L. V. E. Koopmans, P. J. Marshall, L. A. Moustakas, and S. Burles, The sloan lens ACS survey. X. Stellar, dynamical, and total mass correlations of massive early-type galaxies, Astrophys. J. **724,** 511 (2010).

[70] R. Kormann, P. Schneider, and M. Bartelmann, Isothermal elliptical gravitational lens models, Astron. Astrophys. **284,** 285 (1994), http://articles.adsabs.harvard.edu/full/1994A%26A...284..285K.

[71] T. E. Collett, The population of galaxy-galaxy strong lenses in forthcoming optical imaging surveys, Astrophys. J. **811,** 20 (2015).

[72] Y.-Y. Choi, C. Park, and M. S. Vogeley, Internal and collective properties of galaxies in the Sloan digital sky survey, Astrophys. J. **658,** 884 (2007).

[73] A. Zitrin, T. Broadhurst, M. Bartelmann, Y. Rephaeli, M. Oguri, N. Benítez, J. Hao, and K. Umetsu, The universal Einstein radius distribution from 10 000 SDSS clusters, Mon. Not. R. Astron. Soc. **423,** 2308 (2012).

[74] I. W. Harry and S. Fairhurst, A targeted coherent search for gravitational waves from compact binary coalescences, Phys. Rev. D **83,** 084002 (2011).

[75] B. Abbott *et al.* (KAGRA, LIGO Scientific, and VIRGO Collaborations), Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA, Living Rev. Relativity **21,** 3 (2018).

[76] M. Fishbach, D. E. Holz, and W. M. Farr, Does the black hole merger rate evolve with redshift? Astrophys. J. Lett. **863,** L41 (2018); Erratum, Astrophys. J. Lett. **863,** L41 (2018).