Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **526,** 3832–3860 (2023) https://doi.org/10.1093/mnras/stad2909
Advance Access publication 2023 September 26

Follow-up analyses to the O3 LIGO–Virgo–KAGRA lensing searches

J. Janquart,[1,2]★† M. Wright,[3]★† S. Goyal,[4]† J. C. L. Chan,[5]‡ A. Ganguly,[4,6]‡ Á. Garrón,[7]‡ D. Keitel,[7,8]‡ A. K. Y. Li,[9]‡ A. Liu,[10]‡ R. K. L. Lo,[9]‡ A. Mishra,[6]‡ A. More,[6,11]‡ H. Phurailatpam,[10]‡ P. Prasia,[6]‡ P. Ajith,[4,12]§ S. Biscoveanu,[13,14]§ P. Cremonese,[7]§ J. R. Cudell,[15]§ J. M. Ezquiaga,[5]§ J. Garcia-Bellido,[16]§ O. A. Hannuksela,[10]§ K. Haris,[1,2]§ I. Harry,[8]§ M. Hendry,[3]§ S. Husa,[7]§ S. Kapadia,[6]§ T. G. F. Li,[10,17,18]§ I. Magaña Hernandez,[19,20]§ S. Mukherjee,[21]§ E. Seo,[3]§ C. Van Den Broeck[1,2]§ and J. Veitch[3]§

Affiliations are listed at the end of the paper

Accepted 2023 September 21. Received 2023 September 21; in original form 2023 June 15

ABSTRACT

Along their path from source to observer, gravitational waves may be gravitationally lensed by massive objects leading to distortion in the signals. Searches for these distortions amongst the observed signals from the current detector network have already been carried out, though there have as yet been no confident detections. However, predictions of the observation rate of lensing suggest detection in the future is a realistic possibility. Therefore, preparations need to be made to thoroughly investigate the candidate lensed signals. In this work, we present some follow-up analyses that could be applied to assess the significance of such events and ascertain what information may be extracted about the lens-source system by applying these analyses to a number of O3 candidate events, even if these signals did not yield a high significance for any of the lensing hypotheses. These analyses cover the strong lensing, millilensing, and microlensing regimes. Applying these additional analyses does not lead to any additional evidence for lensing in the candidates that have been examined. However, it does provide important insight into potential avenues to deal with high-significance candidates in future observations.

Key words: gravitational lensing: micro – gravitational lensing: strong – gravitational waves.

1 INTRODUCTION

Gravitational lensing of gravitational waves (GWs) happens when they pass nearby a massive object and the deformation of space–time caused by that object modifies their propagation. The observed modifications depend on the exact properties of the lens and include repeated events, phase shifts, changes in amplitude, beating patterns, and distortions (Ohanian 1974; Thorne 1983; Deguchi & Watson 1986; Wang, Stebbins & Turner 1996; Nakamura 1998; Takahashi & Nakamura 2003).

When the lens is an extended high-mass object (e.g. a galaxy or galaxy cluster), the GW frequency evolution is unaffected as the geometric optics limit applies and results in multiple signals – called images – with different magnification, phase shift, and time delay (Wang et al. 1996; Dai & Venumadhav 2017; Ezquiaga et al. 2021). The relative time delays between the images range from minutes to years depending on the scale of the lens with shorter time delays for galaxies (Ng et al. 2018; Oguri 2018; Li et al. 2018), and longer ones for galaxy clusters (Smith et al. 2017, 2018a, b, 2019; Robertson et al. 2020; Ryczanowski et al. 2020). This lensing regime is referred to as strong lensing and the predicted rates imply non-negligible chances of detection for current ground-based detectors, with around one event in a thousand being strongly lensed (Ng et al. 2018; Oguri 2018; Li et al. 2018; Wierda et al. 2021; Xu, Ezquiaga & Holz 2022; Abbott et al. 2023).

For lighter compact lensing objects (e.g. individual stars) significant frequency-dependent wave optics effects occur in the waveform (Deguchi & Watson 1986; Nakamura 1998; Takahashi & Nakamura 2003; Cao, Li & Wang 2014; Christian, Vitale & Loeb 2018; Dai et al. 2018; Lai et al. 2018; Diego et al. 2019; Jung & Shin 2019; Çalışkan et al. 2023b). Fields of objects such as stars can also lead to these effects, usually with more complex characteristics than for a single object (Diego 2020; Pagano, Hannuksela & Li 2020; Cheung et al. 2021; Mishra et al. 2021; Meena et al. 2022). This regime of lensing is known as microlensing to refer to the small size of the lens compared to the GW wavelength.

In between these two scales of lenses is a region in which the multiple images produced by the lens have a time separation of the order of milliseconds, making the different images overlap which produces similar beating patterns to microlensing. Objects in this scale include, for instance, galactic sub-halos. With appropriate analysis tools (Liu et al. 2023) and the ability to consider this scale in the geometric optics regime, the different images can be resolved and analyzed, leading to information about the lensing object at the root of the phenomenon.

★ E-mail: j.janquart@uu.nl (JJ); m.wright.3@research.gla.ac.uk (MW)
† Editorial Team.
‡ Analysts.
§ Contributors.

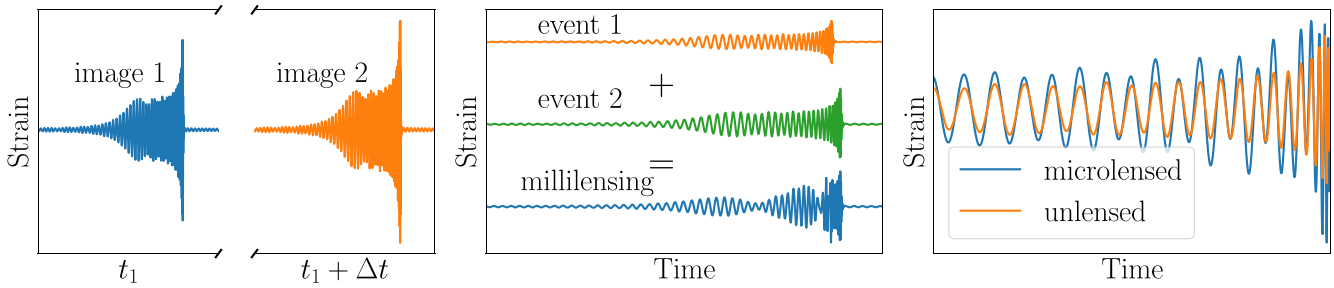

Figure 1. A representation of the effects of the different types of lensing considered in this work on GW strain. From the left to right panels, we have strong lensing – where one has multiple distinct images – millilensing – where one has multiple images with a time separation such that they overlap, giving a modulated signal in the detector – and microlensing – where one has frequency-dependent beating patterns.

The search for GW lensing is motivated by the diverse scientific opportunities its observation would offer. Examples include, but are not limited to, precise localization of the source (Hannuksela et al. 2020; Yu, Zhang & Wang 2020), characterization of the lens (Lai et al. 2018; Diego et al. 2019; Oguri & Takahashi 2020), precision cosmology (Sereno et al. 2011; Liao et al. 2017; Cao et al. 2019; Li, Fan & Gou 2019b; Hannuksela et al. 2020), statistical cosmology (Xu et al. 2022), and tests of general relativity (Baker & Trodden 2017; Collett & Bacon 2017; Fan et al. 2017; Ezquiaga et al. 2021; Goyal et al. 2021b, 2023a).

The LVK collaboration has searched for strong lensing and for microlensing signatures in the following LVK observing runs: O1–O2 (Hannuksela et al. 2019), O3a (Abbott et al. 2021f), and the full O3 run (Abbott et al. 2023), yielding no confident signatures. In parallel, other searches have been performed, confirming that no lensing features have been confidently detected so far (Li et al. 2023; Dai et al. 2020; McIsaac et al. 2020; Liu, Hernandez & Creighton 2021).

Nevertheless, in these searches, interesting candidates have been found. This is the case, for example, for GW190412, that shows some support for being a type II image, the GW191103–GW191105 pair for strong lensing – discarded only after the inclusion of both the population priors and selection effects – and GW200208_130117 which displays some features which are compatible, at low significance, with microlensing (Abbott et al. 2023). Although, ultimately, not confirmed as lensed, such events contain features representative of signatures one could find in genuinely lensed events. It is therefore important to see what sort of follow-up analyses one could do on such events to have a better grasp on their significance, and to extract a maximum of information about the systems.

Possible avenues to achieve this in the future are presented in this work by applying them to these interesting O3 candidates. We follow up on strongly lensed candidates by making a background distribution of simulated unlensed events in order to compute each candidate's false-alarm probability (FAP). We also compare the candidates to the most recent simulations as to see if we can identify the lens that could be at the root of such a lensed event. Additionally, we look for lensed electromagnetic (EM) counterparts by cross-matching with galaxy-lens catalogues. Moreover, since some fainter counterparts are likely to be present in a strongly lensed multiplet, we also follow up on an additional strongly lensed candidate containing a supra-threshold event GW191230_180458 and a weaker 'sub-threshold' event LGW200104_184028 identified for investigation by a new method (Goyal et al. 2023b). We analyse this pair in more details in this work, showing that it is an intriguing pair but is unlikely to be lensed. We also analyse the most significant candidate microlensing events using different lens models, inferring the parameters of the lens models at the same time. We compare these models to investigate which is most likely. Moreover, we analyse the most significant of these microlensing candidates with a millilensing framework to see if the signatures could come from a source in this lensing regime. We do not report any additional evidence for lensing but outline some important next steps to further deal with a possibly lensed event.

We stress that whilst the events discussed in this paper may be treated as though they were lensed, they do not display significant evidence for lensing (Abbott et al. 2023). The goal of this work is to demonstrate the methodologies that can be used to dig deeper in the case of genuinely lensed events and to better assess the importance of candidates. To represent this, we refer to the events as 'lensed candidates' in what follows. Additionally, since the events and event pairs analyzed in this work have been selected because they present interesting features, it is often the case that they lead to higher Bayes factors. However, this is generally not enough to claim lensing, and we would also require to have posteriors converging to a given value of the lensing parameters or a high significance compared to a background before considering an event as lensed.

The rest of the paper is structured as follow. In Section 2, we introduce GW lensing and its different regimes. Then, in Section 3, we show the results for different new analyses performed on GW190412, an event flagged with some support for a type II image. Next, in Section 4, we analyze the GW191103–GW191105 event pair, found to have some characteristics resembling the ones expected for strongly lensed event pairs. We continue in Section 5 by analyzing another event pair, GW191230–LGW200104, a pair made of a supra and a sub-threshold event. In Section 6, we analyze GW200208, an event found to have a mild support for microlensing in past searches. We then give our conclusions and prospects in Section 7.

2 LENSING REGIMES AND ANALYSIS METHODS

Depending upon the characteristics of the lens and the configuration of the lens-source system, the effects of the space–time distortions on the GW will be different. Categories of lensing are usually grouped depending upon whether *geometric optics* is valid, or *wave optics* must be taken into account. The former grouping contains *strong lensing* and *millilensing*, whereas the latter grouping contains *microlensing*. A representation of the effect of each of these types of lensing on GWs is given in Fig. 1. Typically, in the wave optics regime, we see one single distorted waveform, while in the geometric

optics regime, we get multiple images, possibly sufficiently separated in time to be distinguished from one another.

2.1 Strong lensing

When the GW travel path is close enough to a massive lens, gravitational lensing leads to several – possibly detectable – images having the same frequency evolution. This is called the strong lensing regime. The number of images and their specific characteristics depend on both the parameters of the lens and the lens-source configuration. Each of the images experiences a magnification, time delay, and phase shift compared to the unlensed waveform (Schneider, Ehlers & Falco 1992) and their strain, $h_{\rm L}^{j}(f)$, may be described by modifying the unlensed strain, $h_{\rm U}(f)$, such that

$$h_{\rm L}^{j}(f;\theta,\mu_j,\Delta t_j,n_j) = \sqrt{|\mu_j|}\,\mathrm{e}^{2i\pi f\Delta t_j - i\pi\,\mathrm{sign}(f)n_j}h_{\rm U}(f;\theta)\,, \qquad (1)$$

where μ_j is the magnification factor related to the inverse of the determinant of the lensing Jacobian matrix, Δt_j is the time delay related to the different geometrical path taken by the wave and the Shapiro delay (Shapiro 1964), and n_j is the Morse factor which may take one of three distinct values, 0, 1/2, and 1 corresponding to the so-called type of the image which is, respectively, I, II, and III. θ represents the usual compact binary coalescence (CBC) source parameters. In this paper, we only consider the lensing of binary black holes (BBHs). Whilst the magnification and the time delay do not cause any changes to the waveform morphology, the three values of Morse phase may contribute distinct features. Most notable are type II images, where the overall de-phasing between the modes can lead to distortions in the waveform when there is a significant contribution from sub-dominant modes (Dai & Venumadhav 2017; Ezquiaga et al. 2021; Wang et al. 2021; Janquart et al. 2021b; Vijaykumar, Mehta & Ganguly 2023). By contrast, type I images have no extra shift at all, and type III images lead to a sign flip in the GW phase – completely degenerate with a $\pi/2$ shift in polarization angle.

This characterization of strong lensing led to two distinct ways of searching for such events in the O3 data (Abbott et al. 2023): looking only for the type II image distortion in single images and looking for pairs of events that are compatible with the strong lensing hypothesis. For the multiple event search case, two main scenarios can be identified; when the images are sufficiently magnified so as to be detected by the usual CBC search pipelines, termed 'supra-threshold', and when one or more of the images is not sufficiently magnified or is demagnified so as the resultant signal is termed 'sub-threshold'. In the latter case, one can look for the possible lensed counterpart of a supra-threshold event by building an event-specific template bank and searching the data (Li et al. 2023; McIsaac et al. 2020). So far, these searches have yielded no compelling evidence for strong lensing image pairs. A new method to identify lensed sub-threshold candidates is proposed in Goyal et al. (2023b), and used in this work. To avoid the problem of performing computationally expensive parameter estimation (PE) for many merely potential candidates, it relies on the matched filter-estimated chirp masses and the rapidly produced BAYESTAR skymaps (Singer & Price 2016). The matched filter estimations also provide millisecond-level precision on the event's time delay relative to the supra-threshold counterpart allowing the use of the lens model to further assess the probability of the pair being lensed (Haris et al. 2018). Here, we follow up on the most promising candidate from this analysis using the PE-based pipelines.

In this work, several methods are used to analyze the interesting lensing candidates and compute Bayes factors for lensed versus unlensed hypotheses: $\mathcal{B}^{L}_{U} = \mathcal{Z}_L/\mathcal{Z}_U$, where $\mathcal{Z}_H = P(d_1,d_2|H)$ is the evidence under the hypothesis H ($H = L$ for lensed or $H = U$ for unlensed), and $d_i = n_i + h_i^H$ is the data stream for the $i_{\rm th}$ image, made of a noise (n_i) and a GW (h_i^H) component. In this work, we adopt the same conventions as those used in Lo & Magaña Hernandez (2023) and Janquart et al. (2021a), referring to the evidence ratio as the Bayes factor when it includes both the population and selection effects, and the coherence ratio when these are not included.

The first analysis method is called posterior overlap (PO, Haris et al. 2018). Since the frequency evolution of lensed images is unchanged, the detector frame parameters should match (except for those modified by lensing), there should be a significant overlap between the posteriors obtained for these images under the unlensed hypothesis. Therefore, one can compute a detection statistic comparing the evidence for the lensed and unlensed hypotheses

$$\mathcal{B}^{\rm overlap} = \int d\theta\,\frac{P(\theta|d_1)P(\theta|d_2)}{P(\theta)}, \qquad (2)$$

where $P(\theta|d_i)$ is the posterior obtained from data i, and $P(\theta)$ is the prior.

Typically, kernel-density estimators (KDEs) are performed on the posteriors in equation (2) for a subset of parameters (component masses, spins, inclination angle and sky location), and $\mathcal{B}^{\rm overlap}$ is computed using those KDEs. The posteriors used often come from the usual unlensed PE (Veitch et al. 2015; Ashton et al. 2019a).

Another method used is joint parameter estimation (JPE), where one performs the joint inference of the lensed images, linking them through the lensing parameters (Dai et al. 2020; McIsaac et al. 2020; Liu et al. 2021; Lo & Magaña Hernandez 2023; Janquart et al. 2021a, 2023). These pipelines have two different ways to tackle the problem. Some compute the full joint evidence $p(d_1, d_2|H_L)$ at once, such as those outlined in Liu et al. (2021) and Lo & Magaña Hernandez (2023). Here, we use the HANABI pipeline from Lo & Magaña Hernandez (2023). The alternative approach is to instead consider the evidence for the second image as conditional on that of the first (Janquart et al. 2021a, 2023). This makes the computation faster and is equivalent to JPE under the lensed hypothesis. However, some level of approximation is added by doing sub-sampling to improve the speed. The pipeline undertaking this method used within this work is called GOLUM (Janquart et al. 2021a, 2023). The analysis done using the pipeline for the first image also directly offers the possibility to search for type II images when higher order modes are present (Ezquiaga et al. 2021; Lo & Magaña Hernandez 2023; Wang et al. 2021; Janquart et al. 2021b; Vijaykumar et al. 2023; Abbott et al. 2023).

In addition to matching purely on the observational parameters of the system, one can also use models of the lens to inform the strong-lensing detection process (Haris et al. 2018; Lo & Magaña Hernandez 2023; Wierda et al. 2021; Janquart, More & Van Den Broeck 2022; More & More 2022; Lo & Oguri, in preparation). Lensed events do not only have matching frequency-domain evolution but they are also linked via the lensing parameters. Their measured values can be used to assess how likely it is for the observed events to be lensed for a given model. To do this, one compares the probability of having the apparent lensing parameters under the lensed and unlensed hypotheses. This may be done for all of the lensing parameters, or a subset of them. To obtain the probabilities, an unlensed population of BBH mergers is constructed using given population models (mass, redshift, spin,...,distributions) and the phase differences, relative magnifications, and time delays are computed between these events. In parallel, the same process is performed on a lensed population, produced from a BBH population and a lens population following a specific model such as a singular isothermal sphere (SIS; Witt 1990;

Haris et al. 2018) or a singular isothermal ellipsoid (SIE; Koopmans et al. 2009; Wierda et al. 2021; More & More 2022) for the galaxy lenses. From the two computed distributions, a probability density function can be obtained via, for example, KDE reconstruction. It is then possible to evaluate the ratio

$$S^{\rm gal} = \frac{P\left(\Phi|{\rm H_L}\right)}{P\left(\Phi|{\rm H_U}\right)}, \tag{3}$$

where $\rm H_L$ and $\rm H_U$ designate the lensed and unlensed hypotheses, respectively, and Φ is the set of lensing parameters under consideration.

Specific examples of the statistics computed with this method for this work are $\mathcal{R}^{\rm gal}$ (Haris et al. 2018) using specifically the time delay, and $\mathcal{M}^{\rm gal}$ (More & More 2022) which may use all or a subset of the lensing parameters. Both models can be used either with an SIS or an SIE lens model. These statistics are used to select candidates to be followed up by the more extensive analyses or are multiplied with the detection statistics evaluating the match between the parameters. Though model dependent, this in general decreases the risk of false alarm detections (Haris et al. 2018; Wierda et al. 2021; Janquart et al. 2022).

One can also formally combine the effect of these lensing statistics with the coherence ratio. To do this, one reweights the evidence under the lensed and unlensed hypothesis with a weight computed based on the lens model (Janquart et al. 2022). This model-reweighted coherence ratio thus evaluates the ratio of lensed and unlensed evidence for a given lens model directly relating the observed binary parameters and the lensing parameters for a given model. Additionally, the above lensing statistics can also be used when computing the selection effects to obtain the final Bayes factor.

Alternatively, one can consistently incorporate the information from a lens and a source population model (Lo & Magaña Hernandez 2023), where the lens and the source population model affect both the probability of observing a given set of data, in this case (d_1, d_2), under the lensed and the unlensed hypothesis. Specifically, the lens population model informs the joint probability distribution on the magnification, the image type, and the time delay between images, as well as the optical depth for strong lensing, while the source population model informs the distribution of the (true) redshift and the source parameters of a lensed source. This was already done in Abbott et al. (2021f; 2023) using the simple SIS lens model.

In practice, it is difficult to write down an analytical form for the above-mentioned joint probability distribution from a lens model except for some simple lens models (e.g. the SIS model), and instead one usually resorts to constructing a surrogate that approximates the probability density function, such as the aforementioned KDE technique. However, it can be computationally expensive to use KDE-based schemes to construct an estimate for the probability density from a catalogue of simulated lensed images that contains many (e.g. millions of) samples, which, in turn, is evaluated over a set of (roughly tens of thousands of) posterior samples.

In this work, we use the probability density surrogate described in Lo & Oguri (in preparation) that fits the joint probability density on the magnification and the image type conditioned on the time delay between images from a catalogue of mock lens images used in Oguri (2018) using a normalization-flow-based method (Lo 2023). The underlying strong lensing model adopted in the simulation is a population of galaxy-scale SIE lenses with external shear. The lens-redshift-dependent velocity dispersion function is constructed from hybridizing the velocity dispersion measurement for the local Universe derived from the Sloan Digital Sky Survey Data Release 6 (Bernardi et al. 2010) with the Illustris simulation result for the velocity dispersion function at higher lens redshifts (Torrey et al. 2015). The ellipticity and the external shear follow a Gaussian distribution and a log-normal distribution, respectively with additional detail found in Oguri (2018).

2.2 Millilensing

When the mass of the lens and its extent are reduced, the different images produced by lensing can start overlapping. In such a case, they interfere, and one gets only one image exhibiting beating patterns. This is called millilensing (Liu et al. 2023), which is expected to take place for lens masses in the range $M^{z}_{\rm Lens} \in [10^2, 10^6]{\rm M}_\odot$ for which the geometric optics approximation continues to hold. Therefore, the observed GW signal ($h_{\rm Milli}$) that results from the sum of the different images produced can be expressed as

$$h_{\rm Milli}\left(f;\theta,\left\{d_j^{\rm eff}, t_j, n_j\right\}_{j=\{1,\ldots,K\}}\right) = \left(\sum_{j=1}^{K}\frac{d_L}{d_j^{\rm eff}}{\rm e}^{2i\pi f t_j - i{\rm sign}(f)\pi n_j}\right) h_{\rm U}(f;\theta)\,, \tag{4}$$

where θ represents the set of usual BBH parameters, K is the total number of signals produced by lensing, and $\{d_j^{\rm eff}, t_j, n_j\}$ is the set of lensing parameters for image j: the effective luminosity distance,[1] relative time-delay and Morse factor, respectively. Note that the signal for each image has the same frequency evolution as the unlensed signal. However, the interaction between them makes for a non-trivial total lensed signal (see middle panel of Fig. 1 for an illustration).

To search for such signals, one needs to analyze the GW signal assuming that several lensed images are interfering with each other. Usually, the number of signals is not known beforehand. Therefore, it can either be fixed in the search or it can be a variable one tries to infer (Liu et al. 2023).

2.3 Microlensing

For lens-source systems such that the wavelength of the GW is comparable to the Schwarzschild radius of the lens, frequency-dependent modulation of the amplitude occurs. Observing such a phenomenon can offer insights into the nature of the lens itself (Takahashi & Nakamura 2003; Cao et al. 2014; Christian et al. 2018; Dai et al. 2018; Lai et al. 2018; Diego et al. 2019; Jung & Shin 2019; Diego 2020; Pagano et al. 2020; Cheung et al. 2021; Mishra et al. 2021; Meena et al. 2022). Expected lenses in which this regime is applicable and could be detected by current ground-based detectors – objects with masses up to $\sim 10^5{\rm M}_\odot$ – include stellar-mass objects and intermediate-mass black holes. However, it is unlikely that a GW travelling to Earth would cross paths with only one such object as they are most often found in larger structures. As a consequence, a more realistic microlensing scenario would be the case of one or more microlenses embedded within a larger macrolens, such as a galaxy or galaxy cluster. In this case, the effect on the unlensed waveform is much more complicated (Diego 2020; Speagle 2020; Cheung et al. 2021; Mishra et al. 2021; Yeung et al. 2021). This scenario is often unaccounted for because its modelling requires very computationally expensive modifications to the standard waveform

[1]Note here that the magnification μ_j and the effective luminosity distance $d_j^{\rm eff}$ are linked through the source unbiased luminosity distance d_L as $\mu_j = (d_L/d_j^{\rm eff})^2$.

models. In addition, this can also lead to a joint effect with strong lensing leading to multiple microlensed images (Seo, Hannuksela & Li 2022).

To maintain computational tractability, in Hannuksela et al. (2019) and Abbott et al. (2021f, 2023), the microlensing search has been performed using an isolated point mass model. Regardless of the exact model of the lens, the lensed ($h_{\rm Micro}$) and unlensed ($h_{\rm U}$) waveforms are linked as

$$h_{\rm Micro}(f;\theta, M^z_{\rm Lens}, y) = h_{\rm U}(f;\theta) \times F(f; M^z_{\rm Lens}, y), \quad (5)$$

where, as before, θ are the parameters defining the unlensed GW signal, y is the dimensionless source position, $M^z_{\rm Lens}$ is the redshifted lens mass, and $F(f; M^z_{\rm Lens}, y)$ is the amplification factor which is dependent upon both the frequency and the mass–density profile of the lensing object (e.g. more details can be found in Takahashi & Nakamura 2003), hence on the lens model.

Whilst the isolated point mass model has been used for its simplicity and consequent speed, there are other density profiles that may describe microlenses. These include, but are not limited to, the SIS (Witt 1990), the SIE (Koopmans et al. 2009), or the Navarro–Frenk–White (Navarro, Frenk & White 1997) profiles. Efforts have been made to incorporate some of these models into microlensing GW analyses (Wright & Hendry 2022) to determine more information about the lensing object in the event of microlensing detection. This work will use these models to analyse the data and recover information about potential lenses that could be at the origin of a lensed event system.

To explore these multiple models, microlensing candidates are analysed using GRAVELAMPS (Wright & Hendry 2022). This algorithm performs PE analyses of the GW data by jointly inferring the source and lens parameters, assuming a particular lens model. Therefore, it can not only extract information on the lens, but also perform lens-model selection by comparing the evidence obtained for different models and see which one is favoured based on the data. In addition, to cross-check the results obtained with this pipeline, the data is also analysed with the GWMAT pipeline (Mishra, in preparation), a similar but independent PYTHON/CYTHON analysis package implementing the point lens model.

3 GW190412

GW190412 (Abbott et al. 2020a) is a well-known event as it is, alongside GW190814 (Abbott et al. 2020b), one of the events with significant higher order modes (HOMs) detected during O3. It is identified as the coalescence of an $\sim 30\,{\rm M}_\odot$ black hole, with an $\sim 8\,{\rm M}_\odot$ one. During the O3 lensing search (Abbott et al. 2023), this event was found to be the most likely candidate for being a type II image. However, the evidence ($\log_{10}(\mathcal{B}^{\rm II}_{\rm I}) = 0.61$ and $\log_{10}(\mathcal{B}^{\rm II}_{\rm III}) = 0.30$) is inconclusive and could be related to other effects as well, such as noise or waveform effects. In this section, we investigate possible origins of this feature. In particular, we re-analyse the data searching for type II images using other waveform models, and see if the observed feature could be related to microlensing effects.

3.1 Check for waveform systematics

For any astrophysical inference about CBCs from GW data, it is crucial to understand the possible systematic errors due to approximations in the waveform models used or whether observed features could not originate from the model used. Since full numerical relativity simulations are only available as a reference for a few points in parameter space, the most convenient way to study the impact of waveform systematics is to compare results from different models. PE runs for GWTC data releases have always used at least two waveforms from independent modelling approaches and additional studies on events of special interest regularly compare larger numbers of models (see e.g. Abbott et al. 2017, 2020c; Colleoni et al. 2021; Hannam et al. 2022; Mateu-Lucena et al. 2022; Estellés et al. 2022b).

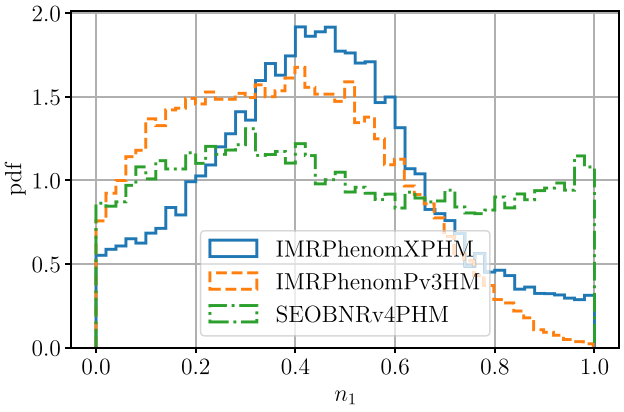

Figure 2. Recovery of the Morse factor for the GW190412 event with different waveforms: `IMRPhenomXPHM`, `IMRPhenomPv3HM`, and `SEOBNRv4PHM`. The support for a type II image is present for the two waveforms from the `Phenom` family. However, the feature is less prominent for the `IMRPhenomPv3HM` waveform, and only marginally present for the `SEOBNRv4PHM` waveform. Therefore, the observed feature is probably spurious and the event is not a type II image.

In Abbott et al. (2023), the type II image (Morse factor of 0.5) search was performed with the `IMRPhenomXPHM` waveform (Pratten et al. 2021), including HOM and precession effects. A feature similar to the observed one – meaning a peak around a value of 0.5 in the Morse factor posterior – was recovered in various scenarios. For example, by injecting a signal with the maximum likelihood parameters in simulated noise with a given waveform family and recovering with `IMRPhenomXPHM`. This seemed to indicate that the feature was consistent with a real signal, at least given the used waveform.

Here, we re-analyse the data using two other commonly used waveforms: `IMRPhenomPv3HM` (Khan et al. 2020) and `SEOBNRv4PHM` (Ossokine et al. 2020). These two waveforms encapsulate the same type of physics as `IMRPhenomXPHM` with HOMs and precession included. The analyses are performed using the same setup as the one used for the Type II image search performed in Abbott et al. (2023).

The posterior recovery of the Morse factor (n_1), allowing it to take any value from 0 to 1 instead of discrete, for the different waveforms is shown in Fig. 2. The posterior peaks at 0.5 using `IMRPhenomXPHM`. When analyzing the data with the `IMRPhenomPv3HM` waveform we still observe a peak but it is less prominent. On the other hand, with the `SEOBNRv4PHM` waveform the posterior distribution has an earlier maximal value, is much broader, and is lacking a pronounced peak. This is different from the results seen with the `Phenom`-family waveforms. Therefore, the observed feature seems to come from a combination of noise and waveform effects, and the event is probably not a type II image.

3.2 Microlensing analysis

The possible signs of de-phasing that generated initial interest in GW190412 as a type II lensing candidate may also be a mistaken

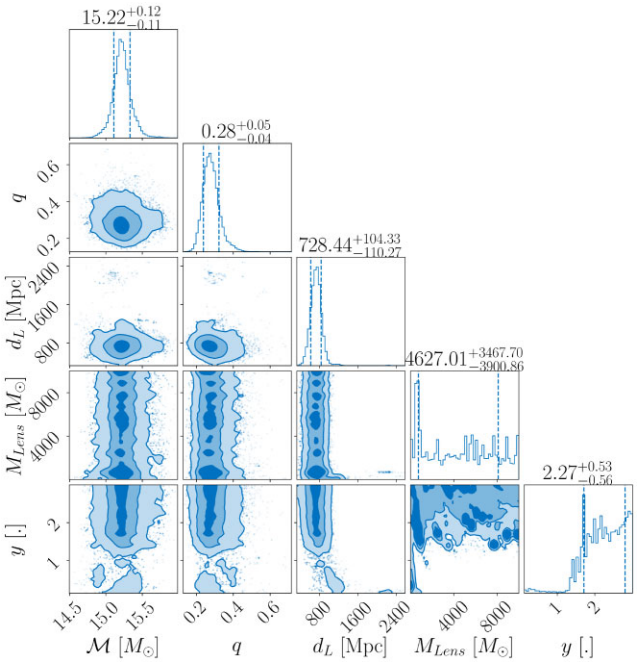

Figure 3. Posteriors for a subset of the parameters including the detector frame chirp mass and mass ratio, the luminosity distance, the lens frame lens mass, and the dimensionless displacement of the source from the optical axis (i.e. the source position). These posteriors were produced during the point mass microlensing analysis done for GW190412 using GRAVELAMPS. As can be seen, similarly to GW191103 (shown in Fig. 8), the lensing parameter posteriors are extremely broad and uninformative. This is consistent with the expectations for a non-microlensing event.

interpretation of frequency-dependent microlensing effects. This motivates an analysis of the event using the GRAVELAMPS pipeline to determine if there is any potential favouring of this interpretation of the signal's features.

Looking at the results of the analysis for GW190412 shown in Fig. 3[2] marks the event as unique amongst those analysed for this paper in that it favours the point mass lensing model over the SIS model with $\log_{10}(\mathcal{B}^{\rm L}_{\rm U})$ values of 0.6 and 0.4, respectively. This remains quite marginal preference for the microlensing hypothesis in either case, although it is worth noting that this support is higher than for GW191103 or GW191105, the events analysed in a strong-lensing context in Section 4. This is consistent with the apparent signs of dephasing being present only in GW190412. Whether the correlation holds would need to be confirmed with additional examinations of simulations or additional type II lensing candidates.

However, whilst support for this event is higher in terms of the raw Bayes factors, the posteriors for the lensing parameters need to be examined. Fig. 3 presents these posteriors for the marginally more preferred point mass lensing model. The source posteriors are tightly constrained but the lensing parameters are extremely broad, leading to the conclusion that this event does not indicate any particular signs of microlensing either, and again the apparent features could be related to noise or waveform issues.

[2]The redshifted lens mass $M^z_{\rm Lens}$ $(1 + z_{\rm Lens})M_{\rm Lens}$ where $M_{\rm Lens}$ is the lens frame lens mass. In the GRAVELAMPS analyses, the redshift of the lens is calculated based on the lens being halfway between the source and the observer.

4 GW191103–GW191105

GW191103 and GW191105 were BBHs detected during the third observing run (Abbott et al. 2021a). In the main LVK analyses, the standard treatment of the signals revealed nothing out of the ordinary for these events. However, when treating the events as potential lensing candidates, the pair display some intriguing characteristics. There is a notable amount of overlap between some of the reported source parameters, such as the sky location and masses; see Fig. 4 for a representation of the posteriors. Moreover, the two events have about 2 d delay between their merger times which is consistent with galaxy-scale lenses (Wierda et al. 2021; More & More 2022). However, in the LVK lensing search, these events were ultimately discarded once the JPE Bayes factor had been computed (Abbott et al. 2023), meaning that the observed overlap is unlikely to be coming from a lensed BBH and is more likely to be coincidental. Nevertheless, as was stated in the introduction, in the following analyses, we have disregarded this and treated the event as though it were a lensed pair.

4.1 PO candidate identification

During the O3 strong lensing search, the PO analysis and a machine learning pipeline, LENSID (Goyal et al. 2021a), were used to identify potential lensing candidate pairs from the detected events. The top 1 per cent of these – approximately a hundred pairs – were then passed on to GOLUM (Janquart et al. 2021a, 2023) for filtering, and then to HANABI (Lo & Magaña Hernandez 2023) for final analysis. The GW191103–GW191105 pair was identified as one of the most likely candidates by the PO analysis using the posteriors of some of the parameters obtained with the `IMRPhenomXPHM` waveform (Pratten et al. 2021) released publicly on the Gravitational Wave Open Science Centre (GWOSC) (Abbott et al. 2021b, e), whereas LENSID – which uses Q-transform images and BAYESTAR (Singer & Price 2016) skymaps – had not classified it as a candidate. Appendix A discusses possible reasons for the non-identification of the pair by the machine-learning-based pipeline. For PO, the pair ended up having the highest overall statistic: $\log_{10}\mathcal{B}^{\rm overlap} = 3.03$ and $\log_{10}\mathcal{R}^{\rm gal} = 1.14$ for the SIS model giving a total of 4.17. Fig. 4 shows the posteriors of both events where one may see the varying degrees of overlap between the events.

Fig. 5 shows the candidate event pairs identified by PO analysis on the $\log\mathcal{B}^{\rm overlap}$–$\log\mathcal{R}^{\rm gal}$ plane considering both the SIS and SIE galaxy models. The choice of model affects only the $\log\mathcal{R}^{\rm gal}$ value. The PO analysis is marginalized over phase and is, therefore, insensitive to the relative Morse phase ($\Delta\phi$) between the two events. As a result of this insensitivity, the SIE cases $\Delta\phi = 0$ and $\pi/2$ are considered separate models, hence we compute the $\mathcal{R}^{\rm gal}$ expected distributions for each case.

Posteriors of events detected by the LVK detectors can overlap by random coincidence meaning that unlensed pairs could also give high Bayes factors. For this reason, a background injection study with ∼1000 unlensed events (the combinations of which yield about half a million pairs) is done to calculate the FAP (Çalışkan et al. 2023a) of the observed log Bayes factor for the candidate pair. The FAP per-pair ($\rm FAP_{PP}$) for the candidate, hence the number of unlensed events with a Bayes factor higher than the one observed for the pair of interest, is found to be 1 in 10 000. Taking into consideration that a total of ∼68 BBH events were detected in O3 the total FAP (given by $\rm FAP = 1 - (1 - FAP_{PP})^{N_{pairs}}$) is found to be 0.3, i.e. a significance of slightly above 1σ. As seen in the figure, the time delay for this event pair is more compatible with an SIE with $\Delta\phi =$

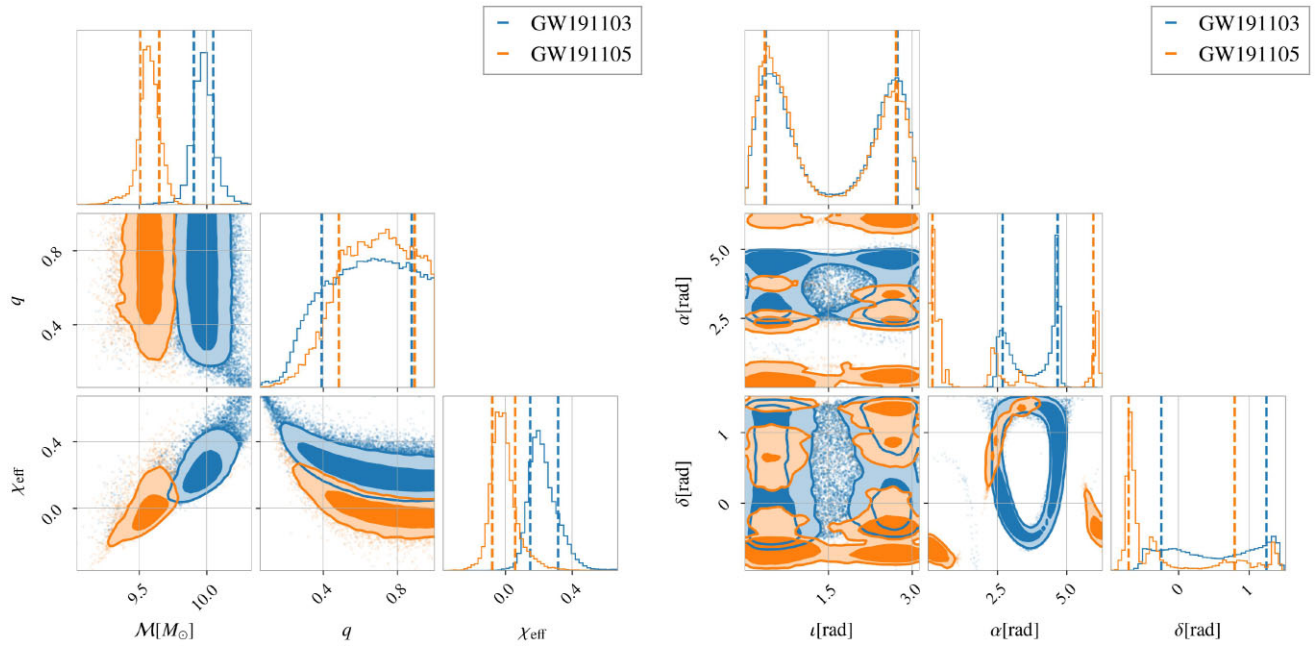

Figure 4. Posteriors for some of the parameters obtained using the `IMRPhenomXPHM` waveform for GW191103 (blue) and GW191105 (orange). The overlap in the extrinsic parameters (e.g. sky location) is larger than that for the intrinsic parameters (e.g. detector-frame chirp mass and spins).

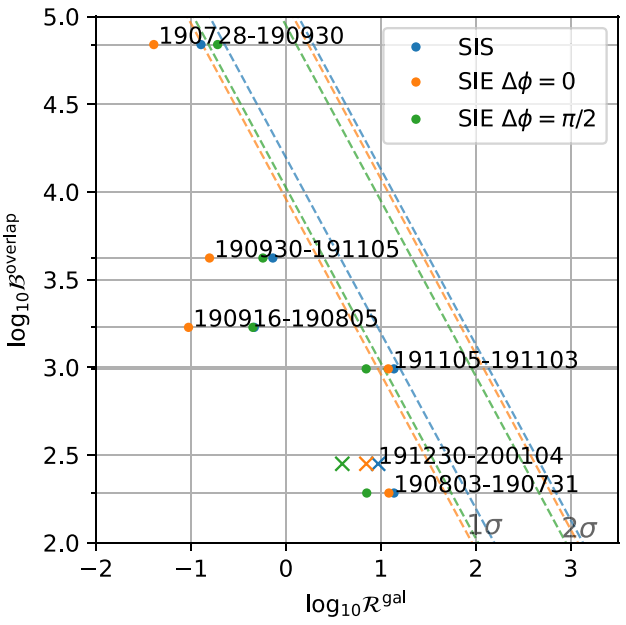

Figure 5. The highest ranked candidate strong lensing pairs from the PO analysis considering all the event pairs found based on the O3 data (dots) (Abbott et al. 2023) and the supra-sub pair analyzed in this work. The dashed lines correspond to the 1σ and 2σ confidence levels for the combined PO statistic ($\mathcal{B}^{\rm overlap} \times \mathcal{R}^{\rm gal}$) with different lensing models computed from the background simulations. We note that beside GW191103–GW191105, the pair analyzed in this work, GW190728–GW190930 is also close to 1σ for PO. However, the pair has been discarded in previous searches with a lower overall significance than GW191103–GW191105 (Abbott et al. 2021f). Therefore, it is not considered for further analyses in this work.

0 as compared to an SIE with $\Delta\phi = \pi/2$ and SIS. After this step, to extract more information about the event pair, it is passed to more extensive pipelines for further investigation.

4.2 Waveform systematics study on PO candidate identification

At the time of writing, no dedicated studies of waveform systematics have been conducted for gravitational lensing analyses. As an initial check, we report on a comparison of PO calculations on single-event PE performed with different waveforms. This is a practical first step as the single-event PE is significantly cheaper computationally than JPE, and detailed studies of waveform systematics on the latter are left for future work. The results presented here for the GW191103–GW191105 pair are an excerpt from a larger pioneer study on waveform model systematics in GW lensing that will be published separately (Garrón & Keitel 2023).

The PO statistics reported in Abbott et al. (2023) and used to initially qualify the GW191103–GW191105 pair as sufficiently interesting for further follow-up were based on the `IMRPhenomXPHM` waveform (Pratten et al. 2021). Besides the posterior samples for this waveform, the data release (Abbott et al. 2021b) for GWTC-3 (Abbott et al. 2021a) contains samples obtained with the `SEOBNRv4PHM` waveform (Ossokine et al. 2020). For this study, we performed additional PE runs on GW191103 and GW191105 for several other models, using `PARALLEL_BILBY` (Ashton et al. 2019a; Smith et al. 2020) with the `DYNESTY` sampler (Speagle 2020), using settings and priors consistent with the GWTC-3 `IMRPhenomXPHM` runs (Abbott et al. 2021a).

The additional waveforms covered are three variants from the same family of frequency-domain phenomenological waveforms as `IMRPhenomXPHM`, as well as one time-domain phenomenological waveform:

(i) `IMRPhenomXAS` is an aligned-spin frequency-domain waveform for dominant-mode-only GW emission (Pratten et al. 2020);

Table 1. PO statistic values for the GW191103–GW191105 pair using different waveform models in the single-event PE.

Waveform	$\log_{10}(\mathcal{B}^{\rm overlap})$
IMRPhenomXAS	3.37
IMRPhenomXHM	3.48
IMRPhenomXP	3.08
IMRPhenomXPHM	3.03
IMRPhenomTPHM	2.70
SEOBNRv4PHM	2.65

(ii) IMRPhenomXHM is an aligned-spin frequency-domain waveform including HOMs (García-Quirós et al. 2020);

(iii) IMRPhenomXP is a frequency-domain waveform allowing for spin precession but for dominant-mode-only GW emission (Pratten et al. 2021);

(iv) IMRPhenomTPHM is a time-domain waveform allowing for spin precession and including HOMs (Estellés et al. 2022a).

The three reduced-physics IMRPhenomX waveforms allow us, in comparison with the most complete family member IMRPhenomXPHM, to check if neglecting any of these features has a significant impact on the PEs for each event, and hence on their overlap. In addition, the IMRPhenomTPHM waveform shares its time-domain nature with SEOBNRv4PHM but much of its modelling approach with IMRPhenomXPHM, making it an ideal tool to further cross-check for consistency between different modelling strategies.

We have followed the original KDE-based calculation from Haris et al. (2018) to compute the PO statistic $\mathcal{B}^{\rm overlap}$, with the modification of computing sky overlaps and intrinsic-parameter overlaps separately and then multiplying them, as done in Abbott et al. (2023).

Table 1 lists the $\mathcal{B}^{\rm overlap}$ resulting from comparing the posteriors from both events with each waveform. There are changes up to factors of $\sim$6 in overlap statistics, with IMRPhenomXHM producing the highest $\mathcal{B}^{\rm overlap}$ and SEOBNRv4PHM resulting in the lowest. We first notice that the aligned-spin waveforms produce the highest $\mathcal{B}^{\rm overlap}$. These have fewer free parameters than the precessing models and hence also different prior volumes. By inspecting the posteriors, we find that the aligned-spin waveforms prefer a_2 closely peaked to zero for both events which gives a high contribution to the overlap, while the precessing waveforms have broader distributions in this parameter, compensating with the additional freedom in the tilt angles. In addition, the two time-domain waveforms produce lower $\mathcal{B}^{\rm overlap}$ than the frequency-domain waveforms. However, these changes are not significantly larger than expected from other sources such as prior choice and KDE implementation details, and all results are consistently in favour of the lensing hypothesis. Hence, we conclude that waveform choice does influence the PO method to some degree, but that for this specific event pair, it is sufficiently robust under waveform choice, in the sense that all results agree qualitatively on identifying the pair as possibly lensed and interesting for follow-up.

One caveat on this type of study is that a full interpretation of $\mathcal{B}^{\rm overlap}$ requires extensive simulation studies on both lensed and unlensed pairs, and the respective distributions could easily be different for different waveforms. However, in Abbott et al. (2023), this factor was used purely as a ranking statistic. So, as long as the number does not drop strongly for any of the considered waveforms, we can conclude that the identification of the candidate pair is robust under waveform choice.

4.3 Compatibility with lensing models

Once an event has been identified as a potential candidate through the aforementioned PO or machine learning searches, it can be followed up by other pipelines. However, an additional check can be made by comparing the observed lensing parameters with the ones predicted by specific lensing models (Haris et al. 2018; Lo & Magaña Hernandez 2023; Wierda et al. 2021; Janquart et al. 2022; More & More 2022). In this work, we focus on galaxy lenses. A comparison of the lensing parameters observed in the O3 search with the catalogue presented in More & More (2022) and Wierda et al. (2021) is given in Fig. 6. Most of the events are compatible with the values expected for unlensed events. Noticeably, two supra-threshold event pairs have lensing parameters that are more consistent with a lensed hypothesis: GW191103–GW191105, which is the one the most on the left and therefore the most favoured, and GW190706–GW190719. One can also compute the $\mathcal{M}^{\rm gal}$ (More & More 2022) statistics for the pairs. This number is the ratio of the probabilities for observing the lensing parameters under the lensed and unlensed hypotheses given by the lensing catalogue though it does not account for the rate of lensing.[3] For GW191103–GW191105, we find $\log_{10}(\mathcal{M}^{\rm gal}) = 1.3$, while for GW190706–GW190719, $\log_{10}(\mathcal{M}^{\rm gal}) = 0.8$. This shows that in the two cases, based on the catalogue, the observed lensing parameters agree more with the expected values for the lensed hypothesis than for the unlensed one.

Whilst such a comparison is valuable to gain information on the chances of lensing from specific models, it does not account for the compatibility of the binary parameters. Since the lensed hypothesis is favoured over the unlensed one both based on PO and lensing statistics, we need to further ascertain the lensed nature by turning to JPE methods.

4.4 JPE-based investigations

In the case of events being genuine lensed images, in addition to the lensing parameters being compatible with at least one lensing model, parameters whose estimation are unaffected by the lensing – e.g. sky location, component masses – should be the same between the events. For the GW190706–GW190719 pair, the GOLUM analysis gives $\log_{10}(\mathcal{C}_{\rm U}^{\rm L}) = 0.6$, which is a value slightly favouring lensing, but still well within the values one can expect for unlensed events (Janquart et al. 2022). On the other hand, for GW191103–GW191105, we find $\log_{10}(\mathcal{C}_{\rm U}^{\rm L}) = 2.6$. However, despite this higher value, this does not guarantee that the signals are not merely coincidentally similar (Janquart et al. 2022). Here, we follow up only on the GW191103–GW191105 pair, which has the highest coherence ratio observed in the O3 pairs (Abbott et al. 2023), with the methodologies described in Janquart et al. (2022) and Lo & Magaña Hernandez (2023) to include information from a lens model into the analyses.

We use the method from Janquart et al. (2022) to include a lens model in the detection statistic coming from the GOLUM pipeline (Janquart et al. 2021a, 2023), modifying the coherence ratio so that it also accounts for the observed lensing parameters and a given lens model. This reduces the risks of false alarms in lensing searches and makes the detection of genuinely lensed pairs more robust. This also enables one to compare the compatibility of the observation with different lens models, constraining the nature of the lens.

[3]Such a catalogue is obtained through extensive lensed and unlensed populations simulations (Haris et al. 2018; Wierda et al. 2021; More & More 2022).

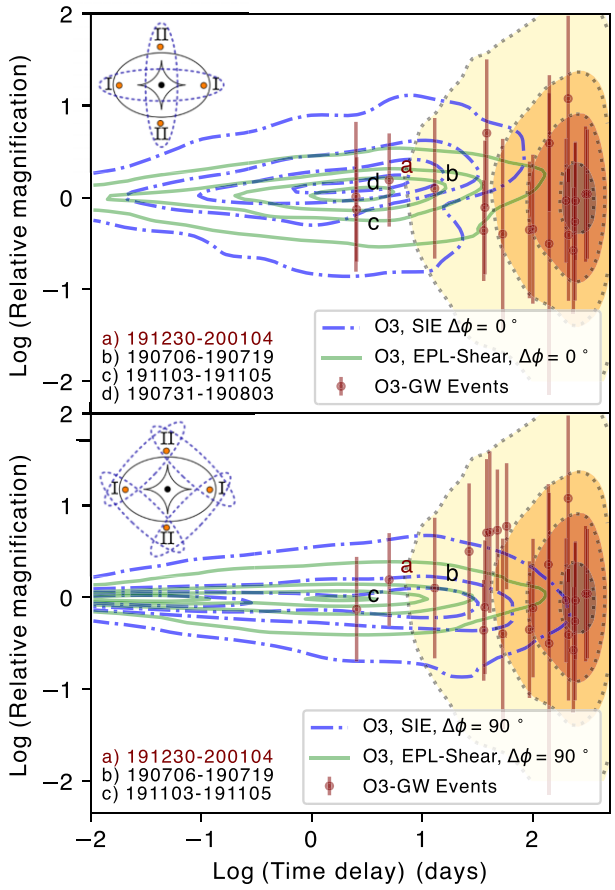

Figure 6. Comparison of the posterior for the observed relative magnification and time delays for the O3 event pairs with the expected distributions for the lensed population of mergers from More & More (2022) (dashed blue, using an SIE model) and from Wierda et al. (2021) [solid green, using an elliptical model with a power-law density profile and external shear (EPL-Shear)] and unlensed population (yellow–orange–red), both assuming galaxy lenses. Overlayed in brown are the observed values for selected O3 event pairs, and the letters mark the event pairs more compatible with the lensing hypothesis. Written in brown, and denoted with letter a, is a pair made of a supra-threshold and a sub-threshold event, and further analysed in this work. The top panel corresponds to the expected distribution when the two images are of the same type, i.e. there is no phase difference between the two images (see the top left illustration), while the bottom panel corresponds to a configuration where the two image types differ, i.e. there is a $\pi/2$ shift between the images. Most of the observed event pairs are well outside the lensed distribution. The GW191103–GW191105, GW190706–GW190719, and GW101230–LGW200104 pairs are more compatible with the lensed hypothesis than with the unlensed one. In particular, the GW191103–GW191105 pair lies in a higher probability density region than the other pairs. One also sees that the GW191230–LGW200104 pair – made of a supra and a sub-threshold event – lies in a higher density region, even if it is less important than the GW191103–GW191105 pair. This pair is discussed further in Section 5.

To explore the event's significance, we make an injection study by generating an unlensed background. We simulate 250 unlensed BBHs, sampling their masses from the POWERLAW + PEAK distributions; the spins and redshift are also sampled from the latest LVK observations, using the maximum likelihood parameters to generate the distributions (Abbott et al. 2021c). The sky location is sampled from a uniform distribution over the sphere of the sky. The inclination is uniform in cosine, and the phase and polarizations are sampled uniformly on their domain. The merger times are set uniformly throughout the third observing run. For each set of parameters, we randomly associate a real event from the GWTC-3 catalogue and assume the same observation conditions (the same detectors are online, and the noise is generated from the event's PSD). The 250 events selected are such that their network SNR is higher than 8.[4]

Table 2. Values of the detection statistic obtained using the GOLUM framework for the GW191103–GW191105 lensed candidate pair without lens model ($\mathcal{C}_U^L$), and with an SIE lens, with ($\mathcal{C}_{\mathcal{M}_{\mu,t}}$) and without ($\mathcal{C}_{\mathcal{M}_t}$) relative magnification accounted for.

Statistic	$\log_{10}$ value	FAP_{PP}
$\mathcal{C}_U^L$	2.5	2.0×10^{-3}
$\mathcal{C}_{\mathcal{M}_{\mu,t}}$	2.4	1.6×10^{-3}
$\mathcal{C}_{\mathcal{M}_t}$	2.9	9.8×10^{-4}

Notes. The FAP_{PP} is decreased when using an SIE model.

Based on this population, we can compute the FAP_{PP} for the coherence ratio for a given lens model. Table 2 summarizes the detection statistic and the FAP_{PP} for the analysis with and without model.

As a first step, we can verify the FAP_{PP} for the event pair when considering the coherence ratio. In this case, $\text{FAP}_{\text{PP}} = 2.007 \times 10^{-3}$, meaning that for our 250 unlensed BBHs, the chance is 1 in 500 that these events are not a genuine lensed pair. Thus, based only on the match between the parameters, the probability for these events to originate from two unrelated unlensed events is relatively high. Statistically, this means that the combination of only 33 randomly selected unlensed BBH mergers is capable to make at least one pair with at least the same coherence ratio as the GW191103–GW191105 pair.

Including a lens model helps study which astrophysical object could be at the origin of the lensed event. Here, we consider an SIE model (Koopmans et al. 2009; More & More 2022), which is a good model for a galaxy lens. We consider the case where we account for the time delay and the relative magnification in the model, and the case where we only consider the time delay. The combinations of the coherence ratio and the lensing statistics are written $\mathcal{C}_{\mathcal{M}_{\mu,t}}$ and $\mathcal{C}_{\mathcal{M}_t}$ for the case with and without relative magnification, respectively. The values for detection statistics and the FAP for the two cases are given in Table 2. The inclusion of the SIE model, with and without the relative magnification, decreases the FAP, meaning that the candidate pair becomes more significant. It is the case even if the SIE model with the time delay and the relative magnification slightly decreases the coherence ratio. The reduction in FAP is slightly larger when only the time delay is considered. This is because the observed relative magnification is slightly outside the highest density regions for the two models under the lensed hypothesis. Moreover, the overlap between lensed and unlensed populations for this parameter is high, making it less helpful to discriminate between the two situations. The results have also been cross checked using an SIE-based catalogue from Wierda et al. (2021), giving the same conclusions.

We note here that whilst the FAP_{PP} seem relatively small for the SIE model, it is insufficient to claim the pair to be lensed. The smallest value found is 9.8×10^{-4}. Whilst this is an improvement over the original FAP_{PP}, it represents only an increase from the

[4]Whether an event can be considered to be of astrophysical origin is not dependent only on its SNR, and recent GWTC papers also put a threshold on the probability p_{astro} (Dent 2021; Abbott et al. 2021a). Here, we consider the SNR threshold sufficient to assess detectability.

Table 3. $\log_{10} \mathcal{B}^{\rm L}_{\rm U}$ for the GW191103–GW191105 pair from HANABI assuming three different merger rate density models and two different lens models. The values computed using the SIS model are reproduced from Abbott et al. (2023) for the sake of comparison. We see that the values with the SIE + external shear model are consistently higher than that with the SIS model, indicating a higher compatibility of the pair with a more realistic strong lensing model. However, since the values remain negative, the event is still most likely to be unlensed considering a more realistic lensing population with the most recent population models.

Lens model \ Merger rate density	Madau–Dickinson	$\mathcal{R}_{\rm min}(z)$	$\mathcal{R}_{\rm max}(z)$
SIS (Abbott et al. 2023)	−3.27	−3.21	−2.33
SIE + external shear (Lo & Oguri, in preparation)	−2.60	−2.46	−1.28

combinations of 33 randomly selected unlensed BBH mergers to the combinations of 47 such mergers to, statistically, have a pair display a detection statistic higher or equal to the one reported for the observed pair. Consequently, whilst GW191103–GW191105 displays some interesting behaviours, these are insufficiently significant to claim a first strong lensing detection.

Additionally, we repeat the Bayes factor calculation comparing the probability ratio of the lensed versus the unlensed hypothesis as described in Abbott et al. (2023) using the more realistic lens population model described in Oguri (2018) (see Lo & Oguri, in preparation and also Section 2) using HANABI (Lo & Magaña Hernandez 2023). We use the same set of source population models as in Abbott et al. (2023), for example, the POWERLAW + PEAK model for the source masses from the GWTC-3 observations (Abbott et al. 2021c) and three models for the merger rate density: Madau-Dickinson (Madau & Dickinson 2014), $\mathcal{R}_{\rm min}(z)$, and $\mathcal{R}_{\rm max}(z)$. Table 3 shows the log-10 Bayes factors computed using the three merger rate density models with the simple SIS lens model reported in Abbott et al. (2023) and the SIE + external shear model reported in Lo & Oguri (in preparation). We see that the values calculated using the SIE + external shear model are consistently higher than those using the SIS model, indicating that the pair is more consistent with a more realistic strong lensing model. Still, the $\log_{10} \mathcal{B}^{L}_{U}$ values are negative, and therefore the event pair is most likely unlensed.

4.5 EM follow-up

In the case of a genuinely lensed GW signal, the light emitted by the host galaxy should also be lensed (Hannuksela et al. 2020; Wempe et al. 2022). As a result of this, in the case of a high-significance lensing candidate, one practical avenue would be to initiate an EM follow-up search. Identifying the host galaxy would be a way to verify the lensed nature of the signal independently.

The EM counterpart of a signal may be searched for in two ways: cross-matching of the joint GW sky localization with EM catalogues containing known lens-source systems or a dedicated EM search on a per-event basis. Both of these make use of the improved GW sky localiztation from the observation of multiple images which provide a virtually extended detector network (Dai et al. 2020; Liu et al. 2021; Lo & Magaña Hernandez 2023; Janquart et al. 2021a, 2023). A dedicated, per-event, EM search could be peformed by looking for lenses within the sky localization area and performing lens reconstruction to try to identify the specific lens at the origin of the observation (Hannuksela et al. 2020; Wempe et al. 2022).

We cross-matched the GW191103–GW191105 pair with a few lens catalogues from optical surveys such as SuGOHI (Sonnenfeld et al. 2017; Wong et al. 2018; Sonnenfeld et al. 2019, 2020; Chan et al. 2020; Jaelani et al. 2020; Jaelani et al. 2021; Wong et al. 2022), AGEL (Tran et al. 2022) and the Master Lens Database (MLD, Moustakas et al. 2012). Whilst no matches were found from the SuGOHI and AGEL catalogues, the grade A and grade B lenses[5] selected from MLD at galaxy scales showed five matches – four doubly lensed systems and 1 quadruply lensed system (see Fig. 7). Two of the doubles are predicted to have time delays >50 d based on the best-fitting lens mass models (Johnston et al. 2003; Paraficz, Hjorth & Elíasdóttir 2009). For the remaining double, we infer a time delay of ∼20 d given the redshifts of the lens galaxy and the source as well as the velocity dispersion (Brewer et al. 2012) under the assumption of an SIS lens mass distribution. All of these time delays are too long to be consistent with the 2-d time gap of the GW191103–GW191105 pair.

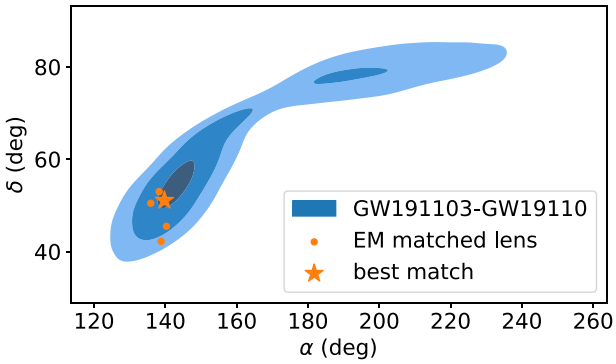

Figure 7. From dark to lighter, the 10, 50, and 90 per cent confidence sky localization for the GW191103–GW191105 pair. Overlaid are the cross-matched five candidates from the MLD.

Lastly, we determine the time delays expected for the quadruplet SDSS J0918 + 5104 using the best-fit mass model and results from Ritondale et al. (2019). The expected time delay for the closest pair in SDSS J0918 + 5104 is around ∼0.5−1 d. Given the uncertainties in the lens model, this ballpark estimate of time delay could be consistent with that of the GW191103–GW191105 pair. However, a more detailed mass modelling analysis and exploring different physical scenarios for the offset between the host galaxy and the GW source can lead to slightly different degrees of compatibility. Still, the predicted relative magnification is both qualitatively and

[5]In this context, grade A lenses have a higher observational quality and accuracy than grade B ones.

quantitatively consistent since the latter GW event is found to be weaker than the former GW event according to the HANABI analysis. Furthermore, the closest pair of SDSS J0918 + 5104 images corresponds to a minimum (Type I) and a saddle point (Type II) suggesting a phase shift of $\Delta\phi = \pi/2$. This is somewhat less favoured but yet plausible for the GW191103–GW191105 pair based on the PO analysis (see Fig. 5).

We note that the cross-matching analysis is limited by both the incompleteness of the databases of known EM lenses and the algorithms used to find matching lenses. A more detailed investigation (see Wempe et al. 2022 for an example of how to investigate the link between EM and GW lensed systems) is warranted to assess the probability of a lens like SDSS J0918 + 5104 to be a genuine EM counterpart of the GW191103–GW191105 pair, if the latter is considered to be lensed. In the future, having dedicated EM follow-up of the lensed event candidates using optical telescopes could also help to gather more information about the lens and localize the source to the host galaxy.

4.6 Microlensing analysis

Whilst interest in the GW191103–GW191105 pair was triggered from the strong lensing perspective, it is worth paying additional attention to whether either of the single events in the pair displays any signs of frequency-dependent effects associated with microlensing. As has been noted above, the most likely microlensing scenario is a microlens embedded within a macrolens. In such a scenario, one or both of the individual signals could display the signatures of the microlens (Mishra et al. 2021; Seo et al. 2022). To determine if that scenario is plausible both GW191103 and GW191105 have been analysed using the GRAVELAMPS pipeline to determine the evidence for an isolated point mass or SIS microlens.

The main result of this analysis is that neither event shows particular favouring for either the point mass or the SIS microlensing models over the unlensed model. The preferred case is that of an SIS microlens in GW191103 which produces a $\log_{10}(\mathcal{B}^{\rm L}_{\rm U}) = 0.38$. In the point lens case, however, support drops to a $\log_{10}(\mathcal{B}^{\rm L}_{\rm U})$ of 0.09 meaning that neither case posits strong evidence. To further compound this, the posteriors for the SIS case are given in Fig. 8. The source parameters' posteriors are well constrained, but those of the lensing parameters are extremely broad and uninformative. This is consistent with the expectation of an unlensed event and, combined with the marginal Bayes factor, leads to the conclusion that there are no observable microlensing signatures within this event.

The case of GW191105 is similar, albeit with even less favouring for the lensing models. Here, $\log_{10}(\mathcal{B}^{\rm L}_{\rm U}) = 0.21$ for the SIS case, and it is -0.35 for the point mass lens model. With even the more optimistic of the two models having a lower favouring, as well as a repetition of the posterior behaviour for the lensing parameters, one again concludes that there are no observable microlensing signatures within this event either.

4.7 Targeted sub-threshold search

Whilst lensing may produce multiple images, it is not guaranteed that all of the images will be detected. However, if it can be ascertained that a detected signal (or signal pair) is lensed, this allows deeper investigation for events below the detection threshold used for standard searches. Reciprocally, finding a sub-threshold counterpart to images with a low probability of being lensed could increase the support for the lensing hypothesis. As such, we conducted searches for sub-threshold lensed counterparts to GW191105 and GW191103

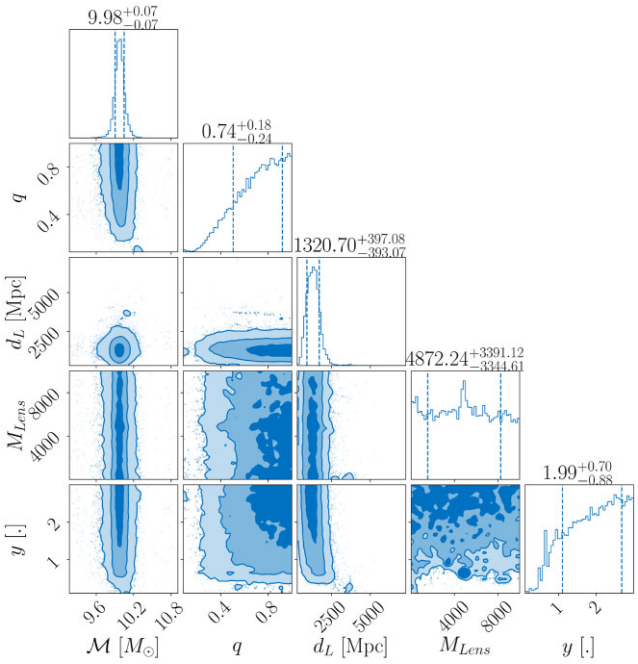

Figure 8. Posteriors for a subset of detector-frame source parameters and the lensing parameters produced during the GRAVELAMPS microlensing analysis of GW191103. As can be seen, whilst the source posteriors are well constrained, the lensing parameter posteriors are extremely broad and uninformative. This shows there are no observable microlensing features in this signal.

individually in the O3 data with the event-specific template banks constructed out of the intrinsic parameter posterior samples (Li et al. 2023). These searches yielded 7 triggers for GW191105 and 15 triggers for GW191103 above the false-alarm rate (FAR) threshold of 1 in 69 yr as defined in Abbott et al. (2023). None of these were reported as a potential lensed counterpart to any of the GW191103 and GW191105 events. One of the interesting triggers found was LGW191106_200820 which arrived just about a day after GW191105, agreeing with galactic scale lens models. However, this trigger was ruled out as a lensed counterpart to a GW191103–GW191105 pair as the overlap in the sky location is poor and the evidence for the event being a real event is very small. It was thus concluded that no promising candidates for an additional sub-threshold counterpart image for GW191103–GW191105 was found within the O3 data.

5 GW191230_180458–LGW200104_180425

During the O3 sub-threshold lensing counterpart search, the TESLA pipeline (Li et al. 2023) based on the GSTLAL software (Messick et al. 2017; Cannon et al. 2021) found roughly 470 triggers which could be potential strong-lensing counterparts to the supra-threshold events. Of these, two had a FAR lower than 1 in 69 yr (Abbott et al. 2023) though none were found to have support for the lensing hypothesis and all were ultimately discarded. An alternative method for identifying the sub-threshold triggers as possible lensed counterparts to supra-threshold events, developed in Goyal et al. (2023b), uses the BAYESTAR localization skymaps, matched-filter chirp mass estimates and the time delay priors to rank all the supra-sub pairs. It identifies

Table 4. PyCBC targeted sub-threshold results for counterpart candidates to GW191230 ranked by IFAR.

Rank	Name	Event	ΔT (d)	IFAR (yr)	SNR	90 per cent CR overlap
0	LGW191222_033537	GW191230	8.60	125822.11	10.99	0.00
1	LGW191230	GW191230	0.00	312.15	10.11	0.75
2	LGW191212_220841	GW191230	17.83	0.57	16.38	0.00
3	LGW191214_055524	GW191230	16.51	0.10	7.16	0.02
4	LGW200104	GW191230	5.02	0.09	8.02	0.62

Notes. From left to right, the columns represent the event, the time delay compared to the supra-threshold event used to make the template, the IFAR, the signal-to-noise ratio (SNR), and the 90 per cent confidence region (CR) overlap for the sky posteriors.

the sub-threshold event termed LGW200104_180425[6] as a possible lensed counterpart to the supra-threshold GW191230_180458 event. It is the most promising supra-sub pair according to this method as it has significant sky and mass overlap, coupled with apparent lensing parameters matching the expected values for a galaxy-lensed models (see Goyal et al. 2023b for more detail). In the rest of this section, we denote the supra- and the sub-threshold events GW191230 and LGW200104, respectively.

LGW200104 was detected with both LIGO detectors with an SNR of 6.31 in Hanford and 4.94 in Livingston. The GstLAL matched-filter estimates on its chirp mass place it at 67.39 $M_\odot$ with the individual component masses being 82.48 and 72.71 $M_\odot$. These high component masses combined with the faintness of the signal contribute to a very low $p_{\rm astro}$ of 0.01 during usual unlensed super-threshold searches, where the event was found with the SPIIR (Luan et al. 2012; Chu et al. 2022) and cWB (Klimenko et al. 2016) pipelines, signifying a significant lack of probability of the event being a genuine detection. Likewise, the FAR found for this event during the superthreshold searches is 4824/yr, also favouring a terrestrial origin for the signal (Kapadia et al. 2020). Since the sub-threshold searches have a more focused template bank, they also reduce the FAR for the events when they are in the correct region of the parameter space (Li et al. 2023; McIsaac et al. 2020). Therefore, the FAR for the event decreases to 6.59/yr when it is found with the TESLA pipeline (Li et al. 2023), still higher than the threshold used for following-up on sub-threshold events in O3 (Abbott et al. 2023). In keeping with the analyses done within this work, whilst we do not claim that the event is both genuine and genuinely lensed, we treat it as though it were. Consequently, we investigate the pair using the lensing identification tools used for supra-threshold pairs.

5.1 PyCBC sub-threshold search

To further verify this candidate and look for sub-threshold counterparts, an independent search pipeline, based on PyCBC (Usman et al. 2016; Davies et al. 2020), has been applied (McIsaac et al. 2020). In contrast to the GstLAL-based TESLA pipeline (Li et al. 2023), the PyCBC-based approach uses a single template based on the posterior distribution of each target event. This search method was previously applied to O1–O2 data (McIsaac et al. 2020) and O3a data (Abbott et al. 2021f). Whilst this search was not deployed across the totality of the O3b data, we have applied it as a cross-check on the chunk of data containing LGW200104 starting on 2019 December 3 at 15:47:10, and ending on 2020 January 13 at 10:28:01, looking for counterparts to GW191230.

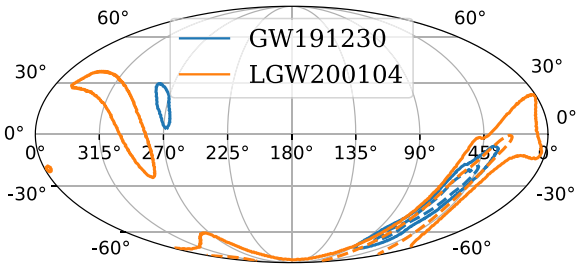

Figure 9. Overlaid LGW200104 and GW191230 skymaps with 90 and 50 per cent CRs.

For the template, we selected the maximum-posterior point from the `IMRPhenomXPHM` samples for GW191230 released with GWTC-3 (Abbott et al. 2021b), from a KDE after removing transverse spins, as to obtain the parameters for a single `IMRPhenomXAS` template with the following values: $m_1 = 82.48 M_\odot$, $m_2 = 72.71 M_\odot$, $a_1 = -0.0037$, and $a_2 = 0.026$. After running PyCBC over the chunk using the same clustering steps as in McIsaac et al. (2020), the results are collected and events are ranked by the inverse of their false alarm rate (IFAR).

For the examined chunk, we found five candidates above an IFAR threshold of 1 month, with 2 previously known GW events topping the list, one being GW191230 itself. To check the correlation of the remaining three events with the target supra-threshold event, we performed a sky overlap estimation of each pair, following the idea described in Wong et al. (2021). The results are shown in Table 4. The sky overlap is computed as the fractional overlap between the sky map obtained using PE for GW191230 and the sky map produced using Bayestar (Singer & Price 2016) for each sub-threshold candidate. Since these two methods do not match exactly, it leads to a 75 per cent overlap for the supra-threshold event with itself.

In interpreting sub-threshold search results, one has to take into account that there is a good chance that, in addition to the potential counterpart images, there will be candidates originating from instrumental glitches or also from different, weaker, GW events that were not identified in previous searches. Here, the candidate corresponding to LGW200104 is ranked fifth (including GWTC events) with an IFAR of 0.09 yr. Its network SNR is recovered as 8.02 (with an SNR of 6.31 and 4.94 in H1 and L1, respectively) and its sky localization overlap with GW191230 is 62 per cent. The sky overlap map is given in Fig. 9.

The third-ranked event has a higher SNR, but no sky overlap with GW191230 and can be clearly identified as a glitch since there is

[6]Here, we follow the usual naming convention, adding an L at the start of the event name to specify it is a sub-threshold candidate. Therefore the name of the sub-threshold trigger is LGWYYMMDD_hhmmss, where YY is the year, MM the month, DD the day, hh the hour, mm the minutes, and ss the second in UTC time.

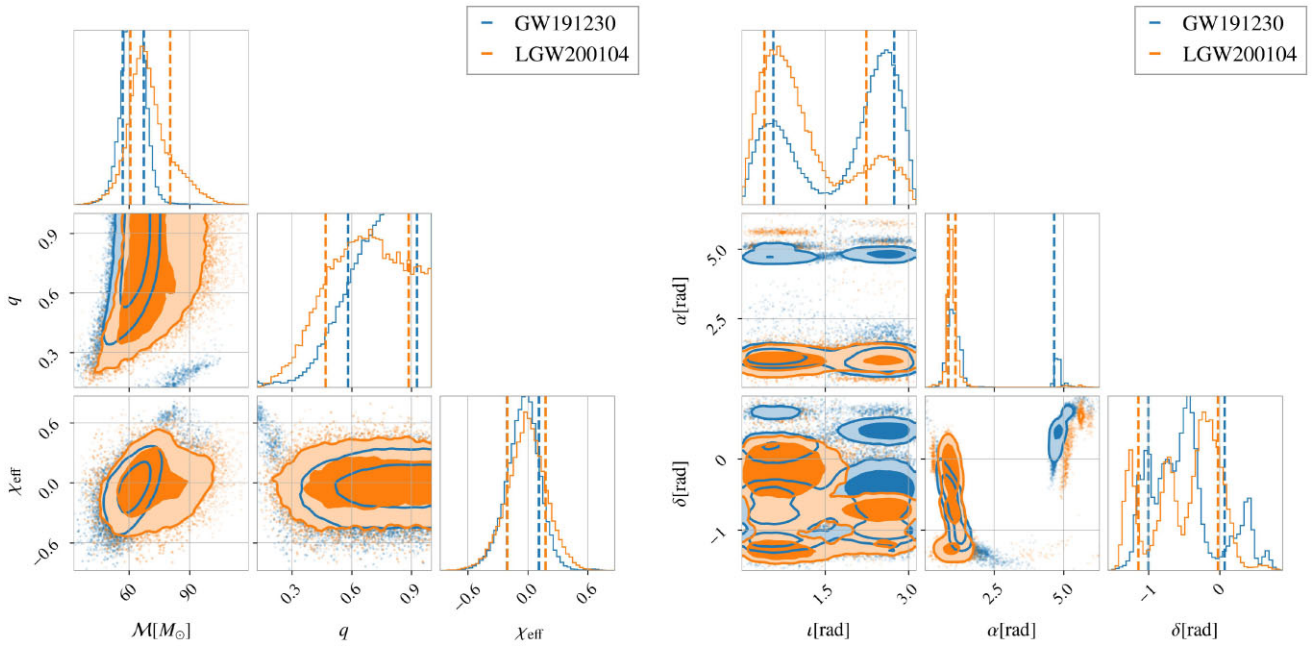

Figure 10. Posteriors for GW191230 (blue) and LGW200104 (orange). The posteriors, though broad, have significant overlap for both the intrinsic (left panel) and extrinsic (right panel) parameters.

simply an excess in power for all frequencies at a given time and no time–frequency evolution similar to that expected for a CBC signal. The fourth-ranked event is clearly a case of a scattered-light glitch (Soni et al. 2021a, b; Tolley et al. 2023). Appendix B further details these two events. In the end, the PyCBC sub-threshold search also finds LGW200104 as the most plausible lensed GW sub-threshold counterpart to GW191230 consistent with the GstLAL pipeline (Li et al. 2023) and the ranking method proposed in Goyal et al. (2023b).

5.2 PO analyses

From the PO analysis this pair has $\log_{10}\mathcal{B}^{\rm overlap} = 2.45$. Since the combined SNR of the sub-threshold trigger is close to 8, it is reasonable to treat the event pair the same way we did for other candidates. Using the same time delay priors as for the supra-threshold events we find $\log_{10}\mathcal{R}^{\rm gal} = 0.97$ which makes the log of the overall PO statistic 3.43. Fig. 10 shows the posteriors for LGW200104 and GW191230. Visually, the degree of overlap in both extrinsic and intrinsic parameters is high. However, the intrinsic parameters posteriors are broader as compared to GW191103–GW191105. For events having high masses in the detector frame, such as these, the number of cycles in the waveform within the LIGO–Virgo frequency band is smaller. This leads to broader posteriors which in turn reduce the overlap statistics, while increasing the rate of coincidental overlaps (Finn & Chernoff 1993; Cutler & Flanagan 1994). In addition, lensed events are more likely to have higher detector frame masses than unlensed events due to the their magnification. Hence, it is a challenge to identify high-mass lensed candidates. Including the population priors and selection effects might help (Haris et al. 2018; Lo & Magaña Hernandez 2023; Janquart et al. 2022).

We also compute the significance of the pair using the supra-threshold background introduced in Section 4.1 and find it to be $\lessapprox 1\sigma$, as shown in Fig. 5. This implies that this pair, though not conclusively lensed, is one of the most significant candidates amongst all the O3 event pairs.

Table 5. Posterior-overlap factors for the GW191230–LGW200104 pair using different waveform models in the single-event PE.

Waveform	$\log_{10}(\mathcal{B}^{\rm overlap})$
`IMRPhenomXAS`	3.20
`IMRPhenomXHM`	3.13
`IMRPhenomXP`	2.52
`IMRPhenomXPHM`	2.45
`IMRPhenomTPHM`	2.55

To look for potential waveforms systematics, we perform the same analysis as in Section 4.2 using results from PARALLEL_BILBY runs with the same five waveform models in the PO calculation. The results are shown in Table 5 and we find relatively consistent results. However, we again notice that the aligned-spin waveforms produce higher $\mathcal{B}^{\rm overlap}$, by a factor of $\sim$6. In this case, both a_1 and a_2 peak towards zero for the aligned-spin models for the two events, leading to a better overlap. However, all waveforms agree on identifying this pair as consistent with lensing.

The PO analysis can quickly identify the lensed candidates but it does not take into account the full correlation between the data streams, the selection effects, and the lensing parameters. Hence, the candidates are passed on to JPE pipelines for further investigations.

5.3 JPE-based investigation

After discovering the candidate with the sub-threshold searches and confirming interest with PO, it was analysed in more detail using GOLUM (Janquart et al. 2021a, 2023) with the samples of the supra-threshold event as the prior for the sub-threshold one. The evidence of this run can be compared to the results of a standard unlensed BILBY (Ashton et al. 2019a) investigation to yield the coherence ratio. In this case, the run yielded $\log_{10}(\mathcal{C}_U^L) = 1.1$. This is lower than that

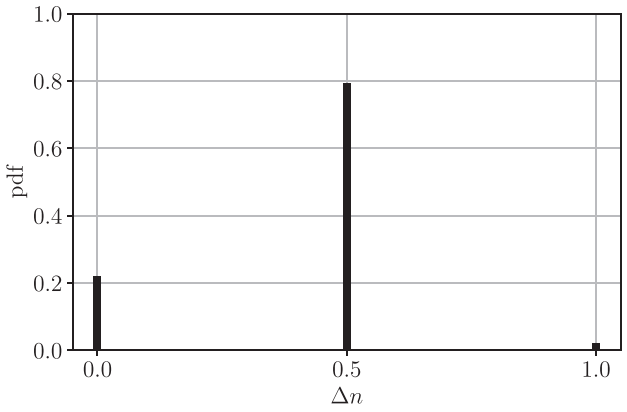

Figure 11. The difference in Morse factor for the GW191230–LGW200104 event pair according to the GOLUM pipeline. The preferred value is 0.5 but there is also some support for 0.

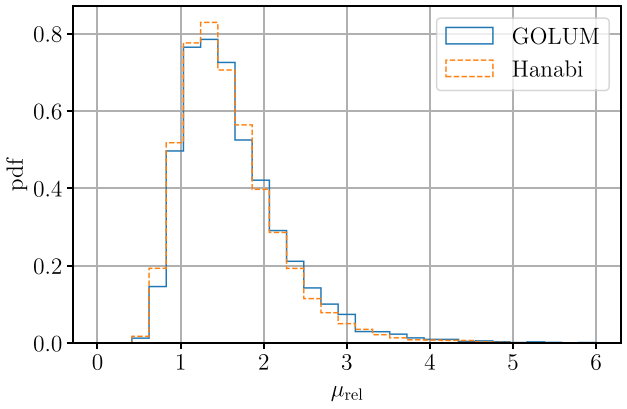

Figure 12. Posterior distribution for the relative magnification for the GW191230–LGW200104 event pair measured with the GOLUM (solid blue) and HANABI (dashed orange) pipelines. One sees that the measured values are consistent between the two pipelines.

calculated for GW191103–GW191105. However, in this case, one of the two images is very close to the limit of a detectable event and this may impact the coherence ratio. By itself, the coherence ratio also is still high enough to favour the lensing hypothesis. To initially investigate the pair's significance, it was compared with the same background as outlined in Section 4. This results in a $\mathrm{FAP_{pp}}$ of 1.4 per cent and thus a FAP of 0.6 which indicates the event is consistent with a coincidental unlensed background event. However, the background resulting in this FAP consisted entirely of supra-threshold events and the exact effects of sub-threshold events in such studies have not been deeply explored.

The GOLUM analysis also offers the possibility to gain insight into the lensing parameters. In particular, it gives information about the difference in Morse factor and relative magnification.[7] Their posterior distributions are given in Figs 11 and 12, respectively, in which it can be seen that the relative magnification peak is at ∼1.5, meaning that its value is close to the highest probability region expected for an SIE lens model (see for example Fig. 6). On the other hand, the difference in Morse factor is less well constrained, with the main support being for $\Delta n = 0.5$, but also some support for the $\Delta n = 0$ case. We note that, generally, for well-detectable lensing events, the difference in Morse factor is well recovered (Janquart et al. 2021a, 2023). This observation may indicate that the event is unlensed but also simply that the lower SNR of the signal makes the identification harder. These lensing parameters and the time delay, however, are consistent with expected values for a galaxy-scale lens.

[7]It also gives the possibility to constrain the time delay, but since the arrival times are very well measured already in GW data analysis, this does not provide much additional information.

Table 6. Values of the detection statistic for the GW191230–LGW200104 lensed candidate pair without lens model ($\mathcal{C}_U^L$), and with an SIE lens, with ($\mathcal{C}_{\mathcal{M}_{\mu,t}}$) and without ($\mathcal{C}_{\mathcal{M}_t}$) relative magnification accounted for. The $\mathrm{FAP_{PP}}$ is decreased when using an SIE model, showing that the observed characteristics are in line with the expected behaviour for the given model and population.

Statistic	$\log_{10}$ value	$\mathrm{FAP_{PP}}$
$\mathcal{C}_U^L$	1.105	1.401×10^{-2}
$\mathcal{C}_{\mathcal{M}_{\mu,t}}$	3.427	1.167×10^{-3}
$\mathcal{C}_{\mathcal{M}_t}$	1.915	2.017×10^{-3}

Based on the GOLUM results, we may also investigate how the coherence ratio and the FAP evolve with the inclusion of expected parameter values from a lens model, as was done in Section 4. Using the same background, and the same models as within that section, we compute the population-reweighted coherence ratios. These values are reported in Table 6. Notably, the coherence ratio found for the SIE model including both the relative magnification and the time delay is now higher than that for the GW191103–GW191105 event pair, meaning that the observed characteristics are more in agreement with the expected value for a galaxy-lens model and the currently observed population than for that pair. This is a demonstration of the fact that the candidate pair – even though the sub-threshold event is not confirmed to be of astrophysical origin – is interesting.

The GW191230-LGW200104 pair was also analyzed by the full JPE package HANABI (Lo & Magaña Hernandez 2023) where the joint parameter space of the two events was simultaneous explored by the stochastic sampler DYNESTY (Speagle 2020) with settings identical to those used in Abbott et al. (2023). The parameters found are consistent with the ones found using the GOLUM framework – see Fig. 12 for a comparison of the relative magnifications. In particular, Fig. 13 shows the posterior probability mass function for the possible image types of the GW191230-LGW200104 pair. We see that the image type configurations for the two events that have non-zero support have the difference in the Morse phase factor Δn either 0 (i.e. the I-I, II-II, and III-III configuration) or 0.5 (i.e. the II-I and III-II configuration). Again, this is consistent with the values shown in Fig. 11 obtained using GOLUM.

We also performed the Bayes factor calculation comparing the probability ratio of the lensed versus the unlensed hypothesis for this pair in the same fashion that we did for the GW191103–GW191105 pair as in Section 4. Again, we use the same set of source population models as in Abbott et al. (2023), for example, the POWERLAW + PEAK model for the source masses from the GWTC-3 observations (Abbott et al. 2021c) and three models for the merger rate density: Madau–Dickinson (Madau & Dickinson 2014), $\mathcal{R}_{\rm min}(z)$, and $\mathcal{R}_{\rm max}(z)$. Table 7 shows the $\log_{10}$ Bayes factors computed using the three merger rate density models with the simple SIS lens model (Abbott et al. 2023) and the SIE + external shear model (Lo & Oguri, in preparation). We see that the values calculated using the SIE + external shear model are positive but only mildly (<1), and they are also consistently higher than the values computed using

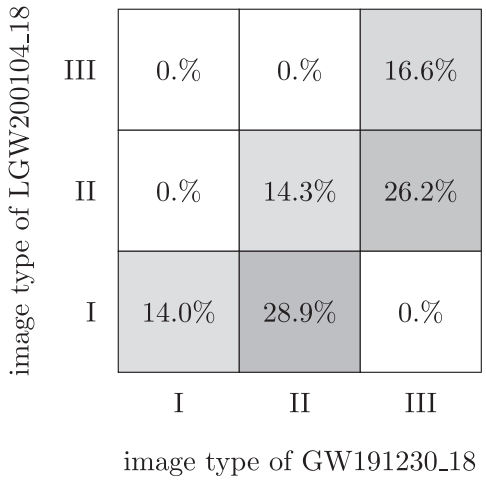

Figure 13. Posterior probability mass function for the image type of GW191230 and the image type of LGW200104 from HANABI. It is consistent with the GOLUM result that it is more likely for the difference in Morse factor to be $\Delta n = 0.5$ (i.e. the II-I and III-II configuration) than to be $\Delta n = 0$ (i.e. the I-I, II-II, and III-III configuration).

the SIS model (which are all negative), indicating that the pair is more consistent with a more realistic strong lensing model. It should be noted that the calculations assumed that both GW events are astrophysical of origin and the second is treated as a supra-threshold event.

Despite some of the evidence for this event aligning relatively well with the expectations for a lensed event, there remain several key arguments against a claim of lensing for this pair. The first is that whilst it is the case that the event has the highest currently observed Bayes factor, it is insufficient to yield a positive log posterior odds considering that the $\log_{10}$ prior odds is between -2 and -4 (Ng et al. 2018; Oguri 2018; Li et al. 2018; Buscicchio et al. 2020; Mukherjee et al. 2021; Wierda et al. 2021). The second argument is the nature of the trigger itself. The sub-threshold event is not convincing – consider for instance the extremely low $p_{\rm astro}$ and FAR – and there is no clear evidence to claim that the event is a genuine GW detection.

In the end, although the event pair is unlikely to be lensed, the analyses performed on this event pair serve as a powerful demonstration of the necessity for searching for such sub-threshold counterparts and the kinds of information that they may yield.

5.4 EM follow-up

Even though the lensing hypothesis is disfavoured, we investigate if there are any EM lens systems with consistent lensing properties from the literature for this event pair. As in Section 4.5, we cross-matched with the MLD. The grade A and grade B lenses selected from the catalogue at galaxy scales showed 21 matches (see Fig. 14). There are two major lens samples that fall within the 90 per cent CR of the sky localization in addition to a handful of systems from heterogeneous studies (e.g. Treu et al. 2006; Shu et al. 2016). The Strong Lensing Legacy Survey lens systems are those with RA $< 40\,\mathrm{deg}^2$ (Gavazzi et al. 2012; More et al. 2012; Sonnenfeld et al. 2013) and South Pole Telescope (SPT)–ALMA lens systems are those with RA $> 50\,\mathrm{deg}^2$ (e.g. Weiß et al. 2013; Spilker et al. 2016). Some of these lens systems do not have sufficient information (for instance, are lacking source redshift) and the others do not have best-fitting mass model parameters or sufficiently high-resolution imaging to identify the multiple lensed image positions. In the absence of these constraints, the time delay of a few days (i.e. $\sim$5 d for the GW191230–LGW200104 pair) can easily be consistent with many of the lens systems. As a result, a rudimentary analysis to determine the likelihood of these lens systems and to select the most likely EM counterparts is not possible, and a more detailed mass modelling for all of the 21 systems would be necessary to find the most promising EM counterpart. Whilst the observation of a third or a fourth GW image would also help further in constraining the characteristics, the lack of high-resolution imaging to clearly and accurately measure the image positions are still anticipated to be the limiting factor. Lastly, we again caution the reader that any incompleteness of the survey (both telescope imaging and subsequent lens searches) may mean that additional potential EM counterparts may have been missed from our initial list of 21 candidates. We found more candidates as compared to the EM follow up of GW191103–GW191105 (Section 4.5) merely because there are more optical surveys that have looked towards the sky region of interest here with respect to the sky overlap region of GW191103–GW191105 which is nearer to the poles. Hence in order to have a robust association one needs to incorporate the selection effects for both the EM and GW observations (see Wempe et al. 2022 for possible avenues).

6 GW200208_130117

GW200208_130117, denoted GW200208 from here on, was selected for follow-up in this paper for two reasons. The first was because it was the event with the highest $\log_{10} \mathcal{B}^L_U$ in the O3 microlensing analysis (Abbott et al. 2023), with a value of 0.8 which, whilst positive, is still inconclusive. The secondary reason was that the event had a relatively narrow posterior on the redshifted lens mass which is atypical of unlensed events. In the O3 lensing paper, it was considered that the cause of the apparent favouring of microlensing for the event could be due to short-duration noise fluctuations causing an apparent dip in the signal, mimicking the beating pattern of microlensing (Abbott et al. 2023).

6.1 Microlensing model investigation

As has been done with a selection of the other events within this paper, GW200208 was re-examined using the GRAVELAMPS pipeline (Wright & Hendry 2022) to investigate the potentiality of model selection in the case of a microlensing candidate. Whilst the isolated point mass model used by GRAVELAMPS is similar to that used by the O3 microlensing search pipeline, there are sufficient implementation differences to warrant re-examination with GRAVELAMPS for this model rather than simply comparing the results of the SIS investigation with those of the O3 microlensing analysis pipeline.

For all of the models examined, GRAVELAMPS had increased favouring for the microlensing hypothesis with this event as compared to the O3 microlensing anlysis pipeline. In the point lens case, the $\log_{10} \mathcal{B}^L_U$ increases to 1.20. This confirms the result from the O3 analysis and shows the event warrants additional investigations. In the SIS case, the preference increased further with the $\log_{10} \mathcal{B}^L_U = 1.77$. This again is sufficiently high as to warrant additional scrutiny, but not high enough to make a statement by itself.

One stage of preparatory work that would shed additional light on the potential significance of the $\log_{10} \mathcal{B}^L_U$ figures would be a detailed background study to determine the range over which unlensed events may appear as microlensing candidates. Such a study had been done for the microlensing search in Abbott et al. (2021f, 2023) and allowed

Table 7. $\log_{10} \mathcal{B}^{\rm L}_{\rm U}$ for the GW191230 and LGW200104 pair from HANABI assuming three different (source) merger rate density models and two different lens models.

Lens model \ Merger rate density	Madau–Dickinson	$\mathcal{R}_{\rm min}(z)$	$\mathcal{R}_{\rm max}(z)$
SIS	− 0.76	− 0.35	− 0.57
SIE + external shear	0.14	0.57	0.30

Notes. We see that the values with the SIE + external shear model are all positive (but only mildly) and consistently higher than that with the SIS model which are all negative, indicating a higher compatibility of the pair with a more realistic strong lensing model. Note that the calculations assumed that both GW events are astrophysical of origin. These values are not sufficient to claim the event pair to be lensed as we would require a positive $\log_{10}$ posterior odds, and the observed Bayes factors are not high enough to balance the low prior odds for strong lensing.

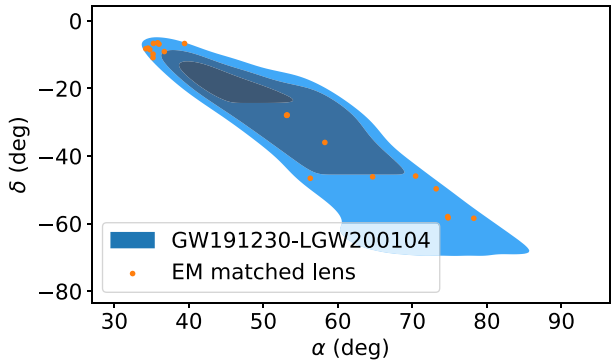

Figure 14. Sky localization 10, 50, and 90 per cent CR (from dark to light) for the GW191230-LGW200104 pair. Overlaid are the cross-matched 21 candidates from the MLD.

the statement above that for the case of that pipeline, the $\mathcal{B}^{L}_{U}$ recorded for GW200208 was within the expected range for an unlensed event. Due to computational constraints, this has not yet been performed for the GRAVELAMPS models, although it is one of the steps that should be taken during O4 so that the significance of candidates may be evaluated quickly. What we note is that the increase between the two pipelines would not necessarily have rendered GW200208 outside of expectations for the O3 microlensing search pipeline, and the general trend of events analysed would appear to indicate that, in general, SIS is preferred over the point mass model. This is likely due to a lens with similar parameters producing lower peak amplifications in the SIS model as compared with the point mass model which would yield smaller deformations from the unlensed template.

For the other events that have been examined, the posteriors for lensing parameters have been a factor in determining that the microlensing hypothesis is unlikely. However, in the case of GW200208 – the posteriors of which in the SIS case are shown in Fig. 15 – this same test yields results more consistent with the microlensing hypothesis. As can be seen in the figure, the lensing parameter posteriors are relatively narrowly constrained around a 2000 $M_\odot$ lensing object with a source position value of 0.60. Fig. 16 shows that in the more pessimistic point lens case, the lensing parameters are constrained to similar values which further cements the need for additional scrutiny of this event. For the two models, we see that the 3σ confidence intervals for the lensing parameters are a bit noisy. However, the peaks in the density distributions remain clearly visible.

We further investigate GW200208 using various sampler and prior settings, as well as testing different waveform models, as listed in

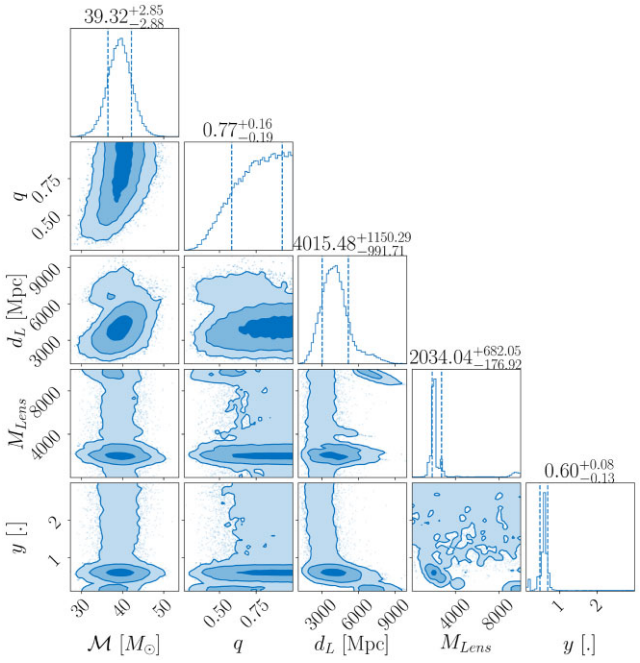

Figure 15. Posteriors of a subset of source parameters as well as lensing parameters for GW200208 in the SIS microlensing case. Unlike the other events that have been examined within this work, the lensing parameters for the model are well constrained, even if the 3σ confidence interval is a bit noisy. This means that this event, unlike the others, cannot be immediately ruled out as a lensing candidate from this test.

Table 8. These tests are designed to verify whether noise artefacts could be at the root of the observed support for microlensing, and whether the results are robust for different sampler settings. We assume the microlensing model of an isolated point-lens and do PE using the GWMAT framework (Mishra, in preparation).

This pipeline utilizes a CYTHON implementation of the amplification factor calculation for the isolated point mass model serves as an independent cross-check for the study. Additionally, the pipeline incorporates a dynamic cutoff based on the source position y to transition to a geometric optics approximation. The resulting probability density functions representing the recovered microlensing parameters are illustrated in Fig. 17.

First, we observe that the posteriors for both parameters, y and $\log_{10} M^{z}_{\rm Lens}$, are consistent across different runs, with median values and 1σ errors of $1.07^{+0.61}_{-0.32}$ and $3.15^{+0.18}_{-0.21}$, respectively. However, the posteriors for y show signs of railings at the upper end, particularly

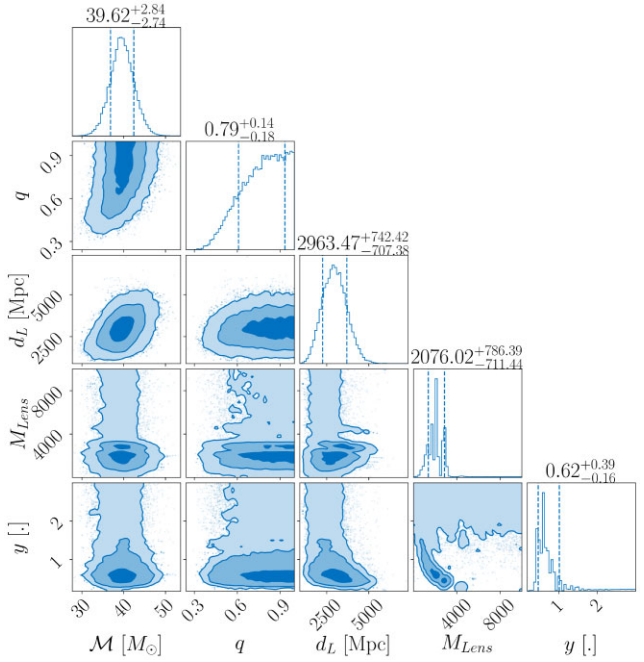

Figure 16. Posteriors of a subset of source parameters as well as lensing parameters for GW200208 in the point mass microlensing case. For this case, the posteriors are similarly constrained, notably arriving at similar lensing parameters as the SIS case even if the 3σ confidence interval is a bit noisy.

in runs with an upper limit of 5 in the prior. Since the SNR is around 11, we limit the prior on y to a maximum of 5, which would already require an SNR $\gtrsim 40$ to make the microlensing signatures detectable.[8] It is worth noting that the posteriors for $\log_{10} M_{\rm Lz}$ are relatively well-converged, with a sharp peak, except in the case where we used a uniform prior in $M^z_{\rm Lens}$, which shows a tendency towards bimodality with another peak at $\log_{10} M^z_{\rm Lens} \gtrsim 5$.

As shown in Table 8, we primarily use $f_{\rm high} = 448$ Hz[9] and a duration of 4 s due to the heavy mass nature of the GW200208 event, which has a total binary mass of approximately 90 $M_\odot$ and negligible spin content. Comparing the first entry in the table with the second-to-last entry, we find that `NRSur7dq4` (Varma et al. 2019) yields a slightly higher Bayes factor value than `IMRPhenomXPHM` for similar settings, except that the `NRSur7dq4` case imposes a total mass constraint of greater than 66 $M_\odot$, considering its region of validity for $f_{\rm low} = 20$ Hz. However, since the event has a total mass $\geq$74 $M_\odot$ with 3σ certainty, we also analyzed the event with $f_{\rm low} = 18$ Hz (last row), resulting in a slight decrease in $\log_{10} \mathcal{B}^{\rm L}_{\rm U}$. On the other hand, when we lowered the value of $f_{\rm low}$ to 15 Hz for the `IMRPhenomXPHM` case, the $\log_{10} \mathcal{B}^{\rm L}_{\rm U}$ increased compared to a similar run with $f_{\rm low} = 20$ Hz (see first and fourth entry).

When we choose $f_{\rm high} = 896$ Hz and a duration of 8 s (third row), both $\mathcal{B}^{\rm U}_{\rm noise}$ and $\mathcal{B}^{\rm L}_{\rm noise}$ decrease, as does $\log_{10} \mathcal{B}^{\rm L}_{\rm U}$, resulting in the lowest value among all the different settings used in the table. Additionally, $\log_{10} \mathcal{B}^{\rm L}_{\rm U}$ also decreases when we broaden the prior on y or $\log_{10} M_{\rm Lz}$ (compare, for example, the fourth and fifth row), which could be additionally lowered due to railing and bimodalities at the higher values of y and $\log_{10} M^z_{\rm Lens}$, as shown in Fig. 17.

The apparent railing and the bimodality can be attributed to the fact that if the likelihood fails to exhibit strong unimodality, the posterior densities may vary depending upon the prior beliefs. A higher upper limit in the prior of y with a power-law profile $p(y) \propto y$ will assign more weight to higher values of y. Similarly, a uniform prior in $M^z_{\rm Lens}$ places a higher weight on heavier microlenses than a log-uniform or log–log-uniform prior, thereby increasing the posterior density in that region. However, if the SNR is high, or if the event is truly microlensed, the likelihood values are better constrained and the posterior densities would not be expected to change much with the priors.

We also note that the Bayes factors presented in Table 8 show more variability. These results indicate that we cannot make a firm conclusion on whether the event is microlensed or not based on the Bayes factor, and the event can only be deemed interesting probabilistically depending on the prior beliefs we choose.

6.2 Maximum-likelihood injection

One avenue of investigation to determine whether an event with the parameters that are suggested by the lensing models within GRAVELAMPS would be detected, and if it was detected, how significant a detection would we expect is to examine a simulated waveform with the maximum likelihood parameters injected into simulated detector noise.

Whilst as stated above, a full-scale injection campaign was not undertaken for the GRAVELAMPS analysis due to temporal and computational constraints, we can investigate if the $\mathcal{B}^L_U$ figures would be plausible for a genuine microlensing event of the suggested parameters by injecting a signal with the parameters of the maximum likelihood sample of the GRAVELAMPS analysis into a realization of Gaussian noise assuming a representative PSD for the noise around the time of detection and analysing this injection with the GRAVELAMPS models in the same fashion as the real event.

Performing this analysis yields value for the $\log_{10} \mathcal{B}^L_U$ of 0.37 and 0.79 for the isolated point mass and SIS profiles, respectively which are lower than those given for the event. This suggests that it would be difficult to confidently confirm an event with these parameters, and therefore this test does not rule out either the possibility of a genuine microlensing event or a noise fluctuation in the data. This again highlights the need for additional investigations such as the aforemnetioned full-scale injection campaign to given greater context to the significance of the calculated Bayes factors.

6.3 Residual power examination

An additional means of scrutinizing a microlensing candidate event is to examine the residual power that is left within the data when the maximum-likelihood waveform fit from the non-lensed PE carried out on the event is removed from the data. In the case of a genuine microlensing event, one would expect to see remaining oscillating amounts of power in each of the detectors due to the unaccounted for oscillating behaviour of the amplification factor. In the case of a non-lensed event, absent any other systematic errors, one would expect to see the fluctuations associated with the noise of the detector. This type of analysis is also performed when looking for deviations from general relativity, where one looks at the residual power in the data after the maximum-likelihood general-relativity template –

[8]The minimum SNR required to distinguish two waveforms with a match value of m is roughly $\sqrt{2/(1-m^2)}$, where we used an estimate of the Bayes factor and set a threshold of unity for distinguishability (Cornish et al. 2011; Vallisneri 2012; Del Pozzo et al. 2014). For $M^z_{\rm Lens} = \mathcal{O}(10^3\ M_\odot)$, the match value between the unlensed and lensed waveform with $y = 5$ comes out to be $\sim$0.9993, which implies a minimum SNR of $\sim$40.

[9]We choose $f_{\rm high} \leq 0.875*(f_s/2)$, where f_s is the sampling frequency (see appendix E of Abbott et al. 2021a).

Table 8. This table presents the results of a Bayesian model comparison study between the unlensed and the microlensed hypotheses for GW200208, with microlensing model corresponding to an isolated point-lens mass. The study was conducted for different configurations and sampler settings, as indicated by the different columns, to verify for possible noise artefacts and check the influence of the sampler settings on the results. The table includes the waveform approximant used, the lower and higher frequency cutoffs used for likelihood evaluation ($f_{\rm low}$, $f_{\rm high}$), duration of the data segment used, and the priors on the redshifted microlens mass ($M^z_{\rm Lens}$) and the impact parameter (y), represented by $p(M_{\rm Lz})$ and $p(y)$, respectively. The Bayes factor for the support of microlensing over the unlensed waveform model is given by $\log_{10}\mathcal{B}^{\rm L}_{\rm U}$. The range of the priors is also indicated. The terms 'L.U' and 'L.L.U' refer to lognormal and log–lognormal distributions, respectively, while 'P.L' refers to a power-law profile with the index given by α.

Waveform	$f_{\rm low}$	$f_{\rm high}$	Duration	$p(M^z_{\rm Lens})$	$p(y)$	$\log_{10}\mathcal{B}^{\rm L}_{\rm U}$
IMRPhenomXPHM	20	448	4	L.U (min = 1, max = 10^5)	P.L (α = 1, min = 0.1, max = 3.0)	0.89
IMRPhenomXPHM	20	1024	4	L.U (min = 10, max = 10^5)	P.L (α = 1, min = 0.01, max = 5.00)	0.63
IMRPhenomXPHM	20	896	8	L.U (min = 10, max = 10^5)	P.L (α = 1, min = 0.01, max = 5.00)	0.46
IMRPhenomXPHM	15	448	4	L.U (min = 10, max = 10^5)	P.L (α = 1, min = 0.1, max = 3.0)	1.02
IMRPhenomXPHM	15	448	4	L.U (min = 10, max = 10^5)	P.L (α = 1, min = 0.01, max = 5.00)	0.53
IMRPhenomXPHM	15	448	4	L.U (min = 10, max = 10^5)	Uniform (min = 0.1, max = 3.0)	1.04
IMRPhenomXPHM	15	448	4	L.L.U (min = 10, max = 10^5)	P.L (α = 1, min = 0.1, max = 3.0)	0.70
IMRPhenomXPHM	15	448	4	L.L.U (min = 10, max = 10^5)	Uniform (min = 0.1, max = 3.0)	0.95
IMRPhenomXPHM	15	448	4	Uniform (min = 10, max = 10^5)	Uniform (min = 0.1, max = 3.0)	0.50
NRSur7dq4	20	448	4	L.U (min = 1, max = 10^5)	P.L (α = 1, min = 0.1, max = 3.0)	0.96
NRSur7dq4	18	448	4	L.U (min = 1, max = 10^5)	P.L (α = 1, min = 0.1, max = 3.0)	0.90

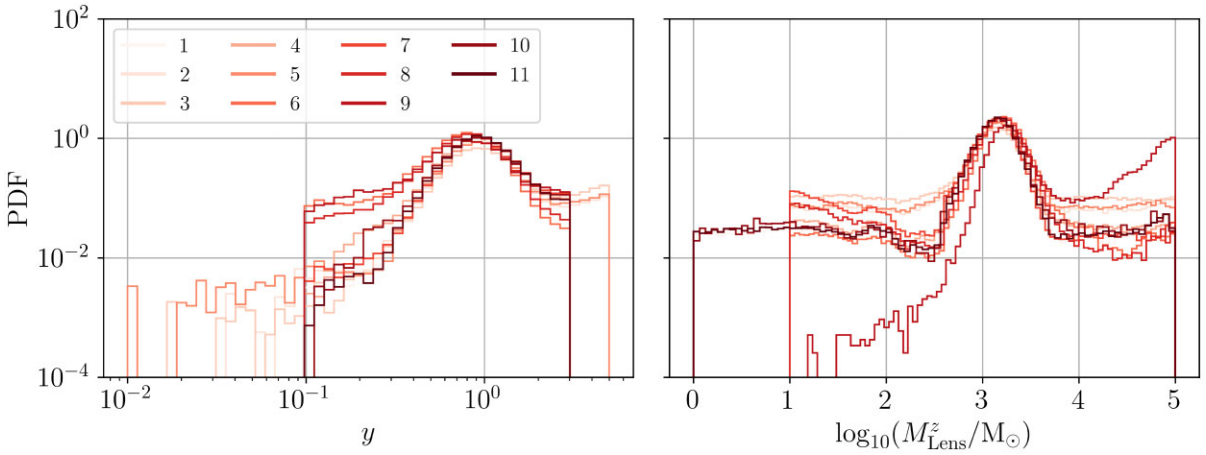

Figure 17. The posterior densities of the recovered microlensing parameters for different PE runs, as presented in Table 8. The results are visualized with varying colours from light to dark (numbered from 1 to 11, indicating different runs as we move down the table.

equivalent to the unlensed one – has been subtracted from it (Abbott et al. 2021d). The residual power investigation carried out in Abbott et al. (2021d) for this event yielded a p-value, corresponding to the probability of obtaining a background event with a residual SNR higher than the event, of 0.97. This suggests the remaining power is within expectations for residual noise.

The residuals from performing this subtraction are shown for each of the detectors, for a subset of the total frequency range, in Fig. 18. As can be seen, none of the detectors display an obvious coherent oscillation in the residual power that would be expected in the microlensing hypothesis. These residuals are more typical of the noise which may indicate that the event is unlikely to be a microlensing event. Hence, despite the increased favouring of the microlensing hypothesis under the SIS case in terms of raw PE analysis, this work draws the same conclusion as that of the lensing searched conducted by the LVK: GW200208, whilst interesting, is not a genuine microlensing event – though it does highlight the need for more systematic studies on the imapct of the noise on microlensing searches in the future.

6.4 Millilensing analysis

The range of masses favoured by the microlensing analysis both in Abbott et al. (2023) and within this work would also be within the millilensing regime as described in Section 2. In the analysis performed here, four millilensing waveform models were used – three with fixed numbers of millilensing signals (two, three, and four signals, respectively), and the fourth being a variable multisignal waveform allowing any number of signals from 1 to 6.

With each of the millilensing waveform models, we performed PE of the source and lensing parameters using the Bayesian inference library BILBY (Ashton et al. 2019b; Romero-Shaw et al. 2020) with the DYNESTY (Speagle 2020) sampler and the IMRPhenomXPHM waveform approximant (Pratten et al. 2021), following the method developed in Liu et al. (2023).

The plots resulting from these PE runs are presented in this section and Appendix C. Before commenting on each of the results individually, we note the terminology used commonly for each of the figures. The millilensing parameters are described by a series of effective luminosity distances, $d_j^{\rm eff}$, time delays with respect to the

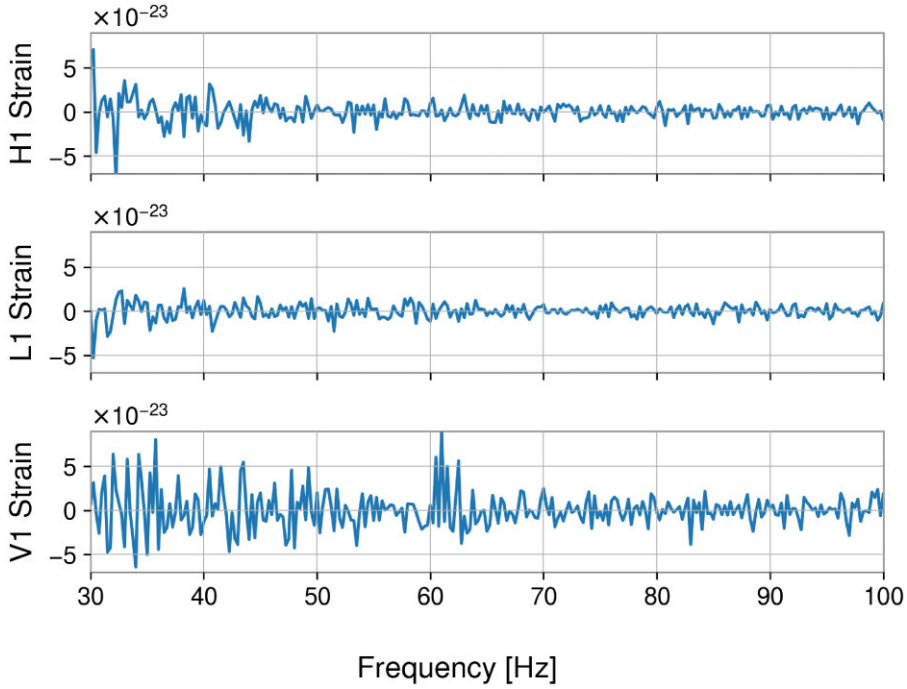

Figure 18. The residual power remaining, from top-to-bottom panels, in LIGO Hanford, LIGO Livingston, and Virgo when subtracting the best-fitting non-lensed waveform template for GW200208 as determined by the unlensed PE from the detector strain, over a subset of the total frequency range. As can be seen, there is no obviously coherent oscillatory behaviour in the residual power which would be expected in the case of the microlensing hypothesis. This absence would suggest that what remains is noise related rather than signal related.

first image t_{j+1}, and Morse phase n_j for the *jth* image. The convention for this work is that the images are referred to in time ordering.

Turning to the specific results, we begin with Fig. 19 detailing the two-signal case. The posterior distribution of the effective luminosity distance of the first signal, $d_1^{\rm eff}$, displays a clear peak as would be expected of a real signal. The posteriors for the second signal parameters – i.e. the effective luminosity distance and the time delay – both display peak-like features but also have an extended underlying posterior. Without additional evidence this would be insufficient to claim millilensing of the signal and could be explained by the presence of noise, as discussed in the microlensing analysis of the event.

The three-signal analysis results (detailed in Appendix C) are in agreement with the two-signal case, where the effective luminosity distance, $d_2^{\rm eff}$, and time delay of the second signal, t_2, display a peak-like feature in the posterior distribution. The corresponding parameters of the third signal, however, do not show significant peaks in their distributions which disfavours the presence of a third millilensing component signal. The four-signal analysis (detailed in Appendix C) also lacks any peaking features in the parameters of either the third or fourth signal – returning their uniform priors and giving additional evidence for the disfavouring of any more than two signals.

Lastly, a multisignal analysis, making the number of millilensing components signals itself a free parameter, was performed. In this analysis, the number of signals was allowed to range from 1 to 6. The posterior distributions for the millilensing parameters are shown in Appendix C. These posteriors are again consistent with the assertion that there is no favouring for any number of signals above a possible second one. The additional results of attempting to infer the number of signals are shown in Fig. 20. The discrete posterior here is notably ambiguous disallowing confident constraints on the number of signals here – despite only the posteriors of the second

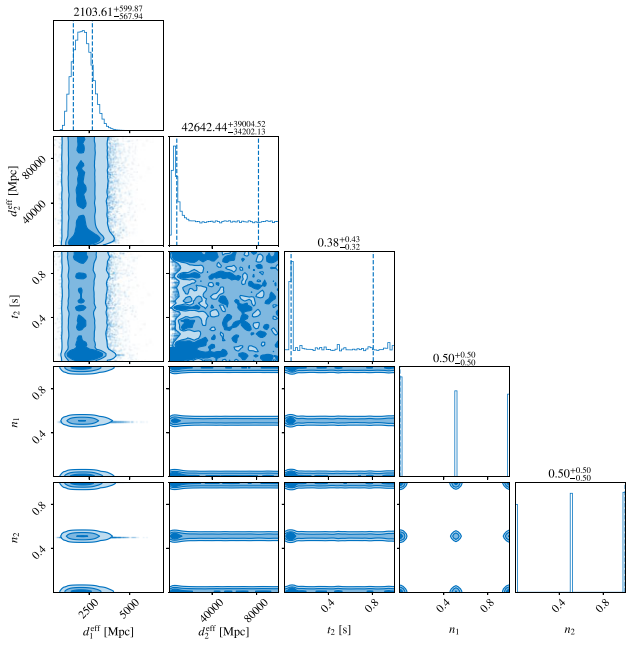

Figure 19. Corner plot of the millilensing parameters obtained from a two-signal analysis of GW200208. Notable is that there is a peak within the effective luminosity distance and time delay parameters for a potential second millilensing signal. However, this peak could be explained by the presence of detector noise.

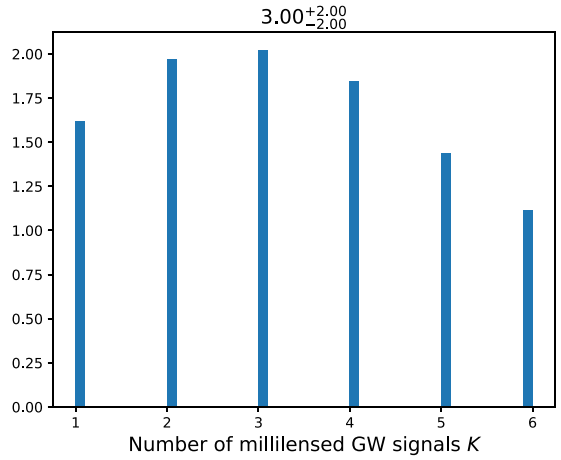

Figure 20. Discrete posterior of the number of signals in the multisignal analysis for GW200208. This posterior is insufficient to confidently assert a number of signals present.

Table 9. Comparison of Bayes factors for the evidence against the unlensed hypothesis from the millilensing runs for GW200208.

Model	$\log_{10}(\mathcal{B}^{\text{Milli}}_{\text{U}})$
Two signals	0.86
Three signals	0.92
Four signals	0.96
Multisignal	1.10

image having any notable features. This serves to underline the fact that the features within the second image posteriors are insufficient to claim a millilensing detection.

In addition to the posterior plots, we also compute the Bayes factors for the millilensing hypothesis against the unlensed one. The values are given in Table 9. They slightly favour the millilensing case, not significantly enough to truly favour this hypothesis when accounting for the astrophysical information, the prior odds, and the observed posteriors.

It is thus the conclusion of the millilensing analysis that there is insufficient evidence to support millilensing within GW2000208 despite a favouring of the millilensing hypothesis when comparing the Bayes factors of signal versus noise for each of the models.

7 CONCLUSIONS AND PROSPECTS

In this work, we have analyzed candidates found to be interesting by the LIGO–Virgo–KAGRA lensing searches in the full O3 data (Abbott et al. 2023) as though they were genuinely lensed. We considered three main types of lensing: strong lensing, millilensing, and microlensing with the types being defined by the effects they have on observed GWs. Though the events investigated do not display strong evidence of being lensed, the analyses done here demonstrate possible follow-up strategies for future observing runs in order to assess the significance of any lensing candidate event.

First for the GW190412 event – which displayed the greatest support for being a type II image – we analysed the data with two other waveform models, showing that these do not show as strong a feature. Therefore, the observed support is most likely due to combined noise and waveform systematics. Additionally, we study the event with a microlensing pipeline to see if such lensing could lead to the apparent deviations in the overall phasing, finding no evidence for this hypothesis.

Next considered was the GW191103–GW191105 pair, which was flagged as interesting because of its relatively high coherence ratio and the consistency of the relative amplitudes and time separation with the expectations for the relative magnification and time delay of galaxy lenses. Testing the effect of waveform systematics on the PO analysis showed that the lensing hypothesis is favoured regardless of waveform choice. We then went on to demonstrate that whilst the event is compatible with galaxy-lens models, inclusion of models in the coherence ratio ultimately does not yield a significant increase in support, as seen by the low coherence ratio. The disfavouring of the lensed hypothesis is further shown by including a more realistic SIE model in our analysis pipeline, still finding a negative log Bayes factor. Furthermore, we demonstrated how an EM counterpart to the host galaxy could be searched for and showed that no confident counterpart could be found. We also demonstrated that neither of the individual events has any indications of microlensing effects. Finally, we looked for a sub-threshold lensed counterpart but found no promising candidate.

A new ranking scheme for the sub-threshold counterparts of detected supra-threshold events found a new interesting candidate pair: the GW191230_180458 supra-threshold and the LGW200104_180425 sub-threshold events. We performed investigations with additional dedicated sub-threshold searches which confirmed interest in the event pair. As was done for the other event pairs, we then analysed it using the standard and follow-up tools. First with the PO analysis, we saw the event pair is an interesting candidate. Again, a waveform systematic study yielded consistent results for the various waveforms considered in this work. In this case, analysis with the joint PE tools showed that upon the inclusion of a galaxy-lens model, the coherence ratio was higher than for the previous pair though only to the extent that 40 unlensed events can produce a pairing with similar results by coincidence. Additionally, performing the computation of the proper Bayes factor with an SIS model leads to negative log Bayes factor, disfavouring the lensing hypothesis. On the other hand, the inclusion of an SIE lens model leads to a marginally positive log Bayes factor. However, it is not high enough to compensate for the prior odds, and therefore the posterior odds is disfavouring the lensing hypothesis. Besides, the low p_{astro} and FAR cast doubt on the astrophysical origin of the sub-threshold event. Finally, as with the previous pair, a search for possible EM counterparts yielded no confident matches which is in line with the expectation for the events not being lensed. Let us also re-iterate the absence of clear evidence for the sub-threshold event to be genuine in the first place.

The last event analysed was GW200208_130117 which was flagged as the event closest to the expectations for a microlensing event. First, we re-examined the event using different lens models. We found the Bayes factors to be slightly higher than those computed in Abbott et al. (2023), but still compatible with values found for background unlensed events. We also studied the variation of the Bayes factor for a point-mass lens model depending on the priors used. We found values ranging from slightly disfavouring lensing to favouring this hypothesis, in line with other analyses performed on this event. To deepen the investigation, we perform a maximum-likelihood injection, recovering the injection with a Bayes factor comparable to the one found for the event, showing the difficulty to confidently identify microlensing at this sensitivity. We then looked at the residual power remaining when subtracting the best-fitting unlensed waveform. This test did not yield any particular evidence

for the remaining power being consistent between the detectors, which is in line with the event being most likely not microlensed. Finally, since millilensing may also lead to beating patterns in the waveform, we searched for millilensing features in this event. These searches demonstrated that there was no additional evidence for any more than two lensed waveforms comprising the event, and the combination of posteriors and Bayes factors were not sufficient to conclusively favour the millilensing hypothesis in general.

In the end, the conclusions from the additional tests are in line with those given by the LIGO–Virgo–KAGRA Scientific Collaboration (Abbott et al. 2023), showing that none of the events or event pairs is likely to be genuinely lensed, regardless of their initially intriguing characteristics. By doing these additional studies, we have shown some important points for future lensing searches, such as the possibility of having waveform systematics, the impact of the lens model in the analysis, the difficulties one may have to distinguish between events resembling each other by chance and genuinely lensed ones, the interplay between microlensing and millilensing, and other additional avenues to further investigate lensing candidates in the future. These follow-up methods should be valuable in the future when more intriguing lensed candidates are found.

ACKNOWLEDGEMENTS

The authors thank Aditya Vijaykumar for the useful discussion, providing some scripts to generate plots, and carefully re-reading the manuscript. The authors thank Christopher Berry for useful discussions on $p_{\rm astro}$.

The authors are grateful for computational resources provided by the LIGO laboratory and Cardiff University and supported by the National Science Foundation Grants PHY-0757058 and PHY-0823459, and the STFC grant ST/I006285/1, respectively. The authors are also grateful to the Inter-University Center for Astronomy & Astrophysics (IUCAA), Pune, India for additional computational resources.

This material is based upon work supported by NSF's LIGO Laboratory which is a major facility fully funded by the National Science Foundation. The authors also gratefully acknowledge the support of the Science and Technology Facilities Council (STFC) of the United Kingdom, the Max-Planck-Society (MPS), and the State of Niedersachsen/Germany for support of the construction of Advanced LIGO and construction and operation of the GEO 600 detector. Additional support for Advanced LIGO was provided by the Australian Research Council. The authors gratefully acknowledge the Italian Istituto Nazionale di Fisica Nucleare (INFN), the French Centre National de la Recherche Scientifique (CNRS) and the Netherlands Organization for Scientific Research (NWO), for the construction and operation of the Virgo detector and the creation and support of the EGO consortium. The authors also gratefully acknowledge research support from these agencies as well as by the Council of Scientific and Industrial Research of India, the Department of Science and Technology, India, the Science & Engineering Research Board (SERB), India, the Ministry of Human Resource Development, India, the Spanish Agencia Estatal de Investigación (AEI), the Spanish Ministerio de Ciencia e Innovación and Ministerio de Universidades, the Conselleria de Fons Europeus, Universitat i Cultura and the Direcció General de Política Universitaria i Recerca del Govern de les Illes Balears, the Conselleria d'Innovació, Universitats, Ciéncia i Societat Digital de la Generalitat Valenciana and the CERCA Programme Generalitat de Catalunya, Spain, the National Science Centre of Poland and the European Union – European Regional Development Fund; Foundation for Polish Science (FNP), the Swiss National Science Foundation (SNSF), the Russian Foundation for Basic Research, the Russian Science Foundation, the European Commission, the European Social Funds (ESF), the European Regional Development Funds (ERDF), the Royal Society, the Scottish Funding Council, the Scottish Universities Physics Alliance, the Hungarian Scientific Research Fund (OTKA), the French Lyon Institute of Origins (LIO), the Belgian Fonds de la Recherche Scientifique (FRS-FNRS), Actions de Recherche Concerées (ARC) and Fonds Wetenschappelijk Onderzoek – Vlaanderen (FWO), Belgium, the Paris Ile-de-France Region, the National Research, Development and Innovation Office Hungary (NKFIH), the National Research Foundation of Korea, the Natural Science and Engineering Research Council Canada, Canadian Foundation for Innovation (CFI), the Brazilian Ministry of Science, Technology, and Innovations, the International Center for Theoretical Physics South American Institute for Fundamental Research (ICTP-SAIFR), the Research Grants Council of Hong Kong, the National Natural Science Foundation of China (NSFC), the Leverhulme Trust, the Research Corporation, the National Science and Technology Council (NSTC), Taiwan, the United States Department of Energy, and the Kavli Foundation. The authors gratefully acknowledge the support of the NSF, STFC, INFN and CNRS for provision of computational resources. This work was supported by MEXT, JSPS Leading-edge Research Infrastructure Program, JSPS Grant-in-Aid for Specially Promoted Research 26000005, JSPS Grant-inAid for Scientific Research on Innovative Areas 2905: JP17H06358, JP17H06361 and JP17H06364, JSPS Core-to-Core Program A. Advanced Research Networks, JSPS Grantin-Aid for Scientific Research (S) 17H06133 and 20H05639, JSPS Grant-in-Aid for Transformative Research Areas (A) 20A203: JP20H05854, the joint research program of the Institute for Cosmic Ray Research, University of Tokyo, National Research Foundation (NRF), Computing Infrastructure Project of Global Science experimental Data hub Center (GSDC) at KISTI, Korea Astronomy and Space Science Institute (KASI), and Ministry of Science and ICT (MSIT) in Korea, Academia Sinica (AS), AS Grid Center (ASGC) and the National Science and Technology Council (NSTC) in Taiwan under grants including the Rising Star Program and Science Vanguard Research Program, Advanced Technology Center (ATC) of NAOJ, and Mechanical Engineering Center of KEK.

JJ and CVDB are supported by the research programme of the Netherlands Organisation for Scientific Research (NWO). SG is supported by the Department of Atomic Energy, Government of India. JME is supported by the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No. 847523 INTERACTIONS, and by VILLUM FONDEN (grant no. 53101 and 37766). ÁG, DK, PC and SH are supported by the Universitat de les Illes Balears (UIB); the Spanish Agencia Estatal de Investigación (AEI) grants PID2022-138626NB-I00, PID2019-106416GB-I00, RED2022-134204-E, RED2022-134411-T, CNS2022-135440, funded by MCIN/AEI/10.13039/501100011033; the MCIN with funding from the European Union NextGenerationEU/PRTR (PRTR-C17.I1); Comunitat Autónoma de les Illes Balears through the Direcció General de Recerca, Innovación y transformación Digital with funds from the Tourist Stay Tax Law (PDR2020/11 - ITS2017-006), the Conselleria d'Economia, Hisenda i Innovación grant numbers SINCO2022/18146 and SINCO2022/6719, co-financed by the European Union and FEDER Operational Program 2021-2027 of the Balearic Islands; the "ERDF A way of making Europe"; and EU COST Action CA18108. ÁG is supported through SOIB, the Conselleria de Fons Europeus, Universitat i Cultura and the Conselleria de

Model Econòmic, Turisme i Treball with funds from the Mecanisme de Recuperación y Resiliencia (PRTR, NextGenerationEU). DK is supported by the Spanish Ministerio de Ciencia, Innovación y Universidades (ref. BEAGAL 18/00148) and cofinanced by UIB. The authors thank the Supercomputing and Bioinnovation Center (SCBI) of the University of Malaga for their provision of computational resources and technical support (www.scbi.uma.es/site) and thankfully acknowledge the computer resources at Picasso and the technical support provided by Barcelona Supercomputing Center (BSC) through grants No. AECT-2022-1-0024, AECT-2022-2-0028, AECT-2022-3-0024, and AECT-2023-1-0023 from the Red Española de Supercomputación (RES). JG-B acknowledges support from the Spanish Research Project PID2021-123012NB-C43 (MICI Española de SupercomputaciónNN-FEDER), and the Centro de Excelencia Severo Ochoa Program CEX2020-001007-S at IFT. PP would like to thank Prof. Sukanta Bose for his support and IUCAA, Pune for providing computational facilities. AKYL and RKLL are supported by the National Science Foundation grants PHY-1912594 and PHY-2207758. AM would like to thank the University Grants Commission (UGC), India, for financial support as a research fellow.

DATA AVAILABILITY

The data underlying this article will be shared in reasonable request to the corresponding authors.

REFERENCES

Abbott B. P. et al., 2017, Class. Quant. Grav., 34, 104002
Abbott R. et al., 2020a, Phys. Rev. D, 102, 043015
Abbott R. et al., 2020b, Astrophys. J. Lett., 896, L44
Abbott R. et al., 2020c, Astrophys. J. Lett., 900, L13
Abbott R. et al., 2021a, ApJ, 923, 14
Abbott R. et al., 2021b, preprint (arXiv:2111.03606)
Abbott R. et al., 2021c, Phys. Rev. R, 13, 011048
Abbott R. et al., 2021d, GWTC-3 Data Release Documentation. Available at: https://gwosc.org/GWTC-3/
Abbott R. et al., 2021e, SoftwareX, 13, 100658
Abbott R. et al., 2021f, Astrophys. J., 923, 14
Abbott R. et al., 2023, preprint (arXiv:2304.08393)
Ashton G. et al., 2019a, Astrophys. J. Suppl., 241, 27
Ashton G. et al., 2019b, ApJS, 241, 27
Baker T., Trodden M., 2017, Phys. Rev. D, 95, 063512
Bernardi M., Shankar F., Hyde J. B., Mei S., Marulli F., Sheth R. K., 2010, MNRAS, 404, 2087
Brewer B. J. et al., 2012, MNRAS, 422, 3574
Buscicchio R., Moore C. J., Pratten G., Schmidt P., Bianconi M., Vecchio A., 2020, Phys. Rev. Lett., 125, 141102
Çalışkan M., Ezquiaga J. M., Hannuksela O. A., Holz D. E., 2023a, Phys. Rev. D, 107, 063023
Çalışkan M., Ji L., Cotesta R., Berti E., Kamionkowski M., Marsat S., 2023b, Phys. Rev. D, 107, 043029
Cannon K. et al., 2021, SoftwareX, 14, 100680
Cao Z., Li L.-F., Wang Y., 2014, Phys. Rev. D, 90, 062003
Cao S., Qi J., Cao Z., Biesiada M., Li J., Pan Y., Zhu Z.-H., 2019, Sci. Rep., 9, 11608
Chan J. H. H. et al., 2020, A&A, 636, A87
Cheung M. H. Y., Gais J., Hannuksela O. A., Li T. G. F., 2021, MNRAS, 503, 3326
Christian P., Vitale S., Loeb A., 2018, Phys. Rev. D, 98, 103022
Chu Q. et al., 2022, Phys. Rev. D, 105, 024023
Colleoni M., Mateu-Lucena M., Estellés H., García-Quirós C., Keitel D., Pratten G., Ramos-Buades A., Husa S., 2021, Phys. Rev. D, 103, 024029
Collett T. E., Bacon D., 2017, Phys. Rev. Lett., 118, 091101
Cornish N., Sampson L., Yunes N., Pretorius F., 2011, Phys. Rev. D, 84, 062003
Cutler C., Flanagan E. E., 1994, Phys. Rev. D, 49, 2658
Dai L., Venumadhav T., 2017, preprint (arXiv:1702.04724)
Dai L., Li S.-S., Zackay B., Mao S., Lu Y., 2018, Phys. Rev. D, 98, 104029
Dai L., Zackay B., Venumadhav T., Roulet J., Zaldarriaga M., 2020, preprint (arXiv:2007.12709)
Davies G. S., Dent T., Tápai M., Harry I., McIsaac C., Nitz A. H., 2020, Phys. Rev. D, 102, 022004
Deguchi S., Watson W. D., 1986, Phys. Rev. D, 34, 1708
Del Pozzo W., Grover K., Mandel I., Vecchio A., 2014, Class. Quantum Gravity, 31, 205006
Dent T., 2021, Extending the PyCBC Pastro Calculation to a Global Network. Available at: https://dcc.ligo.org/public/0173/T2100060/001/pycbc_multi_p_astro.pdf
Diego J. M., 2020, Phys. Rev. D, 101, 123512
Diego J., Hannuksela O., Kelly P., Broadhurst T., Kim K., Li T., Smoot G., Pagano G., 2019, Astron. Astrophys., 627, A130
Estellés H., Colleoni M., García-Quirós C., Husa S., Keitel D., Mateu-Lucena M., Planas M. d. L., Ramos-Buades A., 2022a, Phys. Rev. D, 105, 084040
Estellés H. et al., 2022b, Astrophys. J., 924, 79
Ezquiaga J. M., Holz D. E., Hu W., Lagos M., Wald R. M., 2021, Phys. Rev. D, 103, 6
Fan X.-L., Liao K., Biesiada M., Piorkowska-Kurpas A., Zhu Z.-H., 2017, Phys. Rev. Lett., 118, 091102
Finn L. S., Chernoff D. F., 1993, Phys. Rev. D, 47, 2198
García-Quirós C., Colleoni M., Husa S., Estellés H., Pratten G., Ramos-Buades A., Mateu-Lucena M., Jaume R., 2020, Phys. Rev. D, 102, 064002
Garrón A., Keitel D., 2023, preprint (arXiv:2306.12908)
Gavazzi R., Treu T., Marshall P. J., Brault F., Ruff A., 2012, ApJ, 761, 170
Goyal S., D. H., Kapadia S. J., Ajith P., 2021a, Phys. Rev. D, 104, 124057
Goyal S., Haris K., Mehta A. K., Ajith P., 2021b, Phys. Rev. D, 103, 024038
Goyal S., Vijaykumar A., Ezquiaga J. M., Zumalacarregui M., 2023a, Phys. Rev. D, 108, 024052
Goyal S., Kapadia S., Cudell J.-R., Li A. K. Y., Chan J. C. L., 2023b, preprint (arXiv:2306.04397)
Hannam M. et al., 2022, Nature, 610, 652
Hannuksela O., Haris K., Ng K., Kumar S., Mehta A., Keitel D., Li T., Ajith P., 2019, Astrophys. J. Lett., 874, L2
Hannuksela O. A., Collett T. E., Çalışkan M., Li T. G. F., 2020, MNRAS, 498, 3395
Haris K., Mehta A. K., Kumar S., Venumadhav T., Ajith P., 2018, preprint (arXiv:1807.07062)
Jaelani A. T. et al., 2020, MNRAS, 495, 1291
Jaelani A. T. et al., 2021, MNRAS, 502, 1487
Janquart J., Hannuksela O. A., Haris K., Van Den Broeck C., 2021a, MNRAS, 506, 5430
Janquart J., Seo E., Hannuksela O. A., Li T. G. F., Broeck C. V. D., 2021b, Astrophys. J. Lett., 923, L1
Janquart J., More A., Van Den Broeck C., 2022, MNRAS, 519, 2046
Janquart J., Haris K., Hannuksela O. A., Van Den Broeck C., 2023, MNRAS. Available at: https://doi.org/10.1093/mnras/stad2838
Johnston D. E. et al., 2003, AJ, 126, 2281
Jung S., Shin C. S., 2019, Phys. Rev. Lett., 122, 041103
Kapadia S. J. et al., 2020, Class. Quant. Grav., 37, 045007
Khan S., Ohme F., Chatziioannou K., Hannam M., 2020, Phys. Rev. D, 101, 024056
Klimenko S. et al., 2016, Phys. Rev., D93, 042004
Koopmans L. V. E. et al., 2009, ApJ, 703, L51
Lai K.-H., Hannuksela O. A., Herrera-Martín A., Diego J. M., Broadhurst T., Li T. G., 2018, Phys. Rev. D, 98, 083005
Li S.-S., Mao S., Zhao Y., Lu Y., 2018, MNRAS, 476, 2220
Li Y., Fan X., Gou L., 2019, Astrophys. J., 873, 37
Li A. K., Lo R. K., Sachdev S., Chan C., Lin E., Li T. G., Weinstein A. J., 2023, Phys. Rev. D, 107, 123014
Liao K., Fan X.-L., Ding X.-H., Biesiada M., Zhu Z.-H., 2017, Nat. Commun., 8, 1148

Liu X., Hernandez I. M., Creighton J., 2021, Astrophys. J., 908, 97
Liu A., Wong I. C. F., Leong S. H. W., More A., Hannuksela O. A., Li T. G. F., 2023, 525, 4149
Lo R. K. L., 2023, preprint (arXiv:2305.14379)
Lo R. K. L., Magaña Hernandez I., 2023, Phys. Rev. D., 107, 123015
Luan J., Hooper S., Wen L., Chen Y., 2012, Phys. Rev. D, 85, 102002
McIsaac C., Keitel D., Collett T., Harry I., Mozzon S., Edy O., Bacon D., 2020, Phys. Rev. D, 102, 084031
Madau P., Dickinson M., 2014, ARA&A, 52, 415
Mateu-Lucena M., Husa S., Colleoni M., Estellés H., García-Quirós C., Keitel D., Planas M. d. L., Ramos-Buades A., 2022, MNRAS, 517, 2403
Meena A. K., Mishra A., More A., Bose S., Bagla J. S., 2022, MNRAS, 517, 872
Messick C. et al., 2017, Phys. Rev. D, 95, 042001
Mishra A., Meena A. K., More A., Bose S., Bagla J. S., 2021, MNRAS, 508, 4869
More A., More S., 2022, MNRAS, 515, 1044
More A., Cabanac R., More S., Alard C., Limousin M., Kneib J. P., Gavazzi R., Motta V., 2012, ApJ, 749, 38
Moustakas L. A. et al., 2012, American Astronomical Society Meeting Abstracts, #219
Mukherjee S., Broadhurst T., Diego J. M., Silk J., Smoot G. F., 2021, MNRAS, 501, 2451
Nakamura T. T., 1998, Phys. Rev. Lett., 80, 1138
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
Ng K. K., Wong K. W., Broadhurst T., Li T. G., 2018, Phys. Rev. D, 97, 023012
Oguri M., 2018, MNRAS, 480, 3842
Oguri M., Takahashi R., 2020, Astrophys. J., 901, 58
Ohanian H., 1974, Int. J. Theor. Phys., 9, 425
Ossokine S. et al., 2020, Phys. Rev. D, 102, 044055
Pagano G., Hannuksela O. A., Li T. G. F., 2020, Astron. Astrophys., 643, A167
Paraficz D., Hjorth J., Elíasdóttir Á., 2009, A&A, 499, 395
Pratten G., Husa S., Garcia-Quiros C., Colleoni M., Ramos-Buades A., Estelles H., Jaume R., 2020, Phys. Rev. D, 102, 064001
Pratten G. et al., 2021, Phys. Rev. D, 103, 104056
Ritondale E., Auger M. W., Vegetti S., McKean J. P., 2019, MNRAS, 482, 4744
Robertson A., Smith G. P., Massey R., Eke V., Jauzac M., Bianconi M., Ryczanowski D., 2020, MNRAS, 495, 3727
Romero-Shaw I. M. et al., 2020, MNRAS, 499, 3295
Ryczanowski D., Smith G. P., Bianconi M., Massey R., Robertson A., Jauzac M., 2020, MNRAS, 495, 1666
Schneider P., Ehlers J., Falco E. E., 1992, Gravitational Lenses. Springer, Berlin
Seo E., Hannuksela O. A., Li T. G. F., 2022, Astrophys. J., 932, 50
Sereno M., Jetzer P., Sesana A., Volonteri M., 2011, MNRAS, 415, 2773
Shapiro I. I., 1964, Phys. Rev. Lett., 13, 789
Shu Y. et al., 2016, ApJ, 833, 264
Singer L. P., Price L. R., 2016, Phys. Rev. D, 93, 024013
Smith G. et al., 2017, IAU Symp., 13, 98
Smith G. P. et al., 2018a, IAU Symp., 338, 98
Smith G. P., Jauzac M., Veitch J., Farr W. M., Massey R., Richard J., 2018b, MNRAS, 475, 3823
Smith G. P., Robertson A., Bianconi M., Jauzac M., 2019, preprint (arXiv:1902.05140)
Smith R. J. E., Ashton G., Vajpeyi A., Talbot C., 2020, MNRAS, 498, 4492
Soni S. et al., 2021a, Class. Quantum Gravity, 38, 025016
Soni S. et al., 2021b, Class. Quantum Gravity, 38, 195016
Sonnenfeld A. et al. 2019, A&A, 630, A71
SonnenfeldA. et al., 2020, A&A, 642, A148
Sonnenfeld A., Gavazzi R., Suyu S. H., Treu T., Marshall P. J., 2013, ApJ, 777, 97
Sonnenfeld A. et al., 2017, Publ. Astron. Soc. Japan, 70, S29
Speagle J. S., 2020, MNRAS, 493, 3132
Spilker J. et al., 2016, ApJ, 826, 112
Takahashi R., Nakamura T., 2003, Astrophys. J., 595, 1039
Thorne K., 1983, Proceedings of the Advanced Study Institute, IN: Gravitational Radiation. North-Holland Publishing Co., Amsterdam, p. 1
Tolley A. E., Davies G. S. C., Harry I. W., Lundgren A. P., 2023, Class. Quantum Gravity, 40, 29
Torrey P. et al., 2015, MNRAS, 454, 2770
Tran K.-V. H. et al., 2022, AJ, 164, 148
Treu T., Koopmans L. V., Bolton A. S., Burles S., Moustakas L. A., 2006, ApJ, 640, 662
Usman S. A. et al., 2016, Class. Quant. Grav., 33, 215004
Vallisneri M., 2012, Phys. Rev. D, 86, 082001
Varma V., Field S. E., Scheel M. A., Blackman J., Gerosa D., Stein L. C., Kidder L. E., Pfeiffer H. P., 2019, Phys. Rev. Res., 1, 033015
Veitch J. et al., 2015, Phys. Rev., D91, 042003
Vijaykumar A., Mehta A. K., Ganguly A., 2023, Phys. Rev. D, 108, 043036
Wang Y., Stebbins A., Turner E. L., 1996, Phys. Rev. Lett., 77, 2875
Wang Y., Lo R. K. L., Li A. K. Y., Chen Y., 2021, Phys. Rev. D, 103, 104055
Weiß A. et al., 2013, ApJ, 767, 88
Wempe E., Koopmans L. V. E., Wierda A. R. A. C., Hannuksela O. A., Broeck C. v. d., 2022, MNRAS, preprint (arXiv:2204.08732)
Wierda A. R. A. C., Wempe E., Hannuksela O. A., Koopmans L. V. E., Van Den Broeck C., 2021, ApJ, 921, 154
Witt H. J., 1990, A&A, 236, 311
Wong K. C. et al., 2018, ApJ, 867, 107
Wong H. W. Y., Chan L. W. L., Wong I. C. F., Lo R. K. L., Li T. G. F., 2021, preprint (arXiv:2112.05932)
Wong K. C., Chan J. H. H., Chao D. C.-Y., Jaelani A. T., Kayo I., Lee C.-H., More A., Oguri M., 2022, Publ. Astron. Soc. Japan, 74, 1209
Wright M., Hendry M., 2022, ApJ, 935, 68
Xu F., Ezquiaga J. M., Holz D. E., 2022, ApJ, 929, 9
Yeung S. M. C., Cheung M. H. Y., Gais J. A. J., Hannuksela O. A., Li T. G. F., 2021, preprint (arXiv:2112.07635)
Yu H., Zhang P., Wang F.-Y., 2020, MNRAS, 497, 204

APPENDIX A: LENSID GW191103–GW191105 INVESTIGATIONS

In Abbott et al. (2023), the pairs which had false positive probability (FPP) less than 0.01 either with PO or LENSID were passed on for the follow-up analysis. According to PO the GW191103–GW191105 pair is found to be one of the most significant candidates. However, the LENSID FPP is found to be 0.16. We cannot determine for certain why LENSID did not find the pair significant for follow-up analysis, however, we can identify some possible contributing factors. Before detailing these, we briefly summarize how LENSID works. LENSID is made up of two ML models, one which takes Q-transforms input, and another which takes skymaps as input. On the basis of the Q-transforms, the network outputs the probability for the event pair to be lensed for each detector. Additionally, there is one output lensing probability based on the sky map. The entire probability for lensing is then computed by taking the four individual probabilities mentioned above. For more details we refer the reader to Goyal et al. (2021a).

GW191103 was observed only in two detectors, LIGO Hanford (H1) and LIGO Livingston (L1), whereas, GW191105 was observed in all three detectors but was contaminated by a glitch in the Virgo detector. As seen in Fig. A1, the final PE skymap of the event (right panel in Fig. A1), which is made after deglitching the data (Abbott et al. 2021a), is different from the initial skymap (left panel in Fig. A1), reducing the sky map FPP from 0.08 to 0.02 after using the PE sky map, still not crossing the threshold.

For the Q-transforms, only the H1 and L1 detectors data are used by the framework. We notice that the Q-transforms for the events,

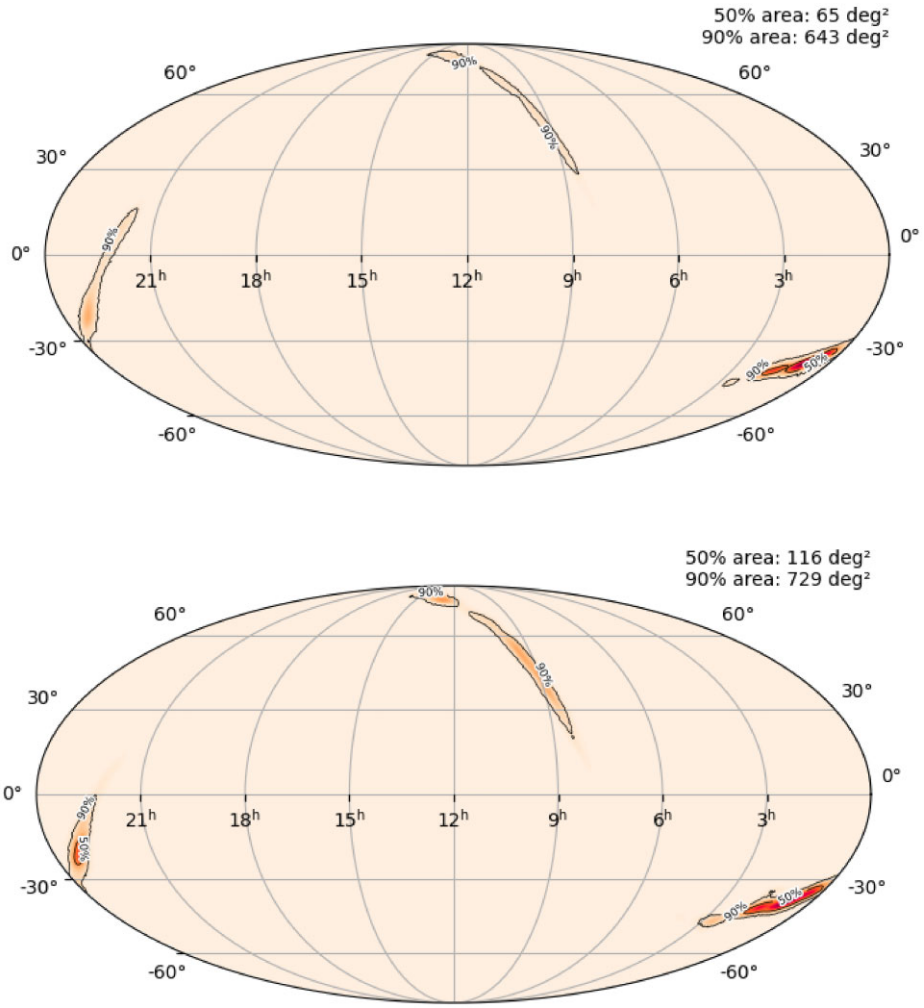

Figure A1. Comparison of sky maps for GW191105, as downloaded from GraceDB which is generated using LALINFERENCE (left) and from GWOSC which is generated using BILBY after de-glitching the Virgo data (right). The LALINFERENCE sky map is narrower as compared to BILBY, likely because of the glitch present in the data. Note that the localization patch near to the north pole is the one that overlaps with GW191103 and therefore matters more than the rest of the patches.

especially for GW191105, are visually poor. They seem to be broken in the middle, as shown in the Fig. A2. Notice that the Q-transform of GW191105 in the L1 detector has a gap in the middle of the signal with peaks of power on both sides of the gap. This is not expected from a GW chirp signal. We checked that even though the SNRs are similar for both the events in the H1 and L1 detectors, the estimated probability for lensing varies a lot between the two detectors, 0.86 for H1 and 0.12 for L1. This indicates that the ML algorithm is not robust to real noise fluctuations, which is expected as it is trained using simulated Gaussian noise signals. Additionally, from an injection study, we found that LENSID is more prone to misclassifying lensed signals with low chirp masses ($<20\,M_{\odot}$) which is the case here. In the future, to mitigate these problems, the ML models will be trained and tested on data containing real noise and lower chirp masses.

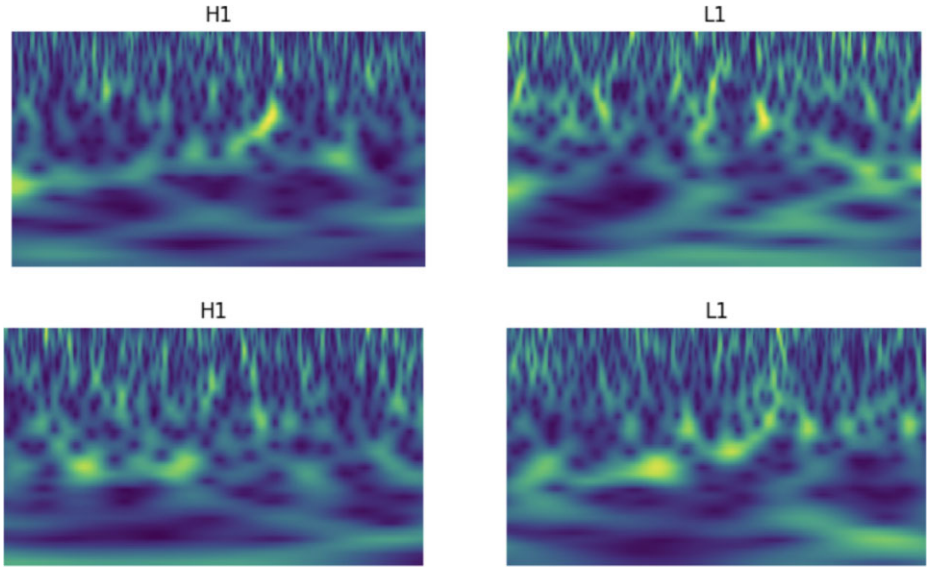

Figure A2. Q-transforms images input to the LENSID pipeline for GW191103 (top panel) and GW191105 (bottom panel). The chirping feature for GW191105 is broken in both the LIGO detectors, whereas for GW191103 the chirp signal is fairly visible in Hanford, and not so visible in Livingston. This could be one of the reasons why LENSID did not identify this pair as significant.

APPENDIX B: DISCARDED TARGETED PYCBC SUB-THRESHOLD SEARCH TRIGGERS

In this section, we show the time–frequency maps for the two discarded (third and fourth ranked) candidate triggers found as possible sub-threshold counterparts for the GW191230 event by the PyCBC-based pipeline. The two can clearly be identified as glitches, with the third-ranked clearly having a power excess across a broad frequency band at the same time without presenting a time–frequency evolution similar to the one expected for a genuine GW signal (see Fig. B1), and the fourth in ranking (see Fig. B2) clearly matching a scattered-light glitch (Soni et al. 2021a, b; Tolley et al. 2023).

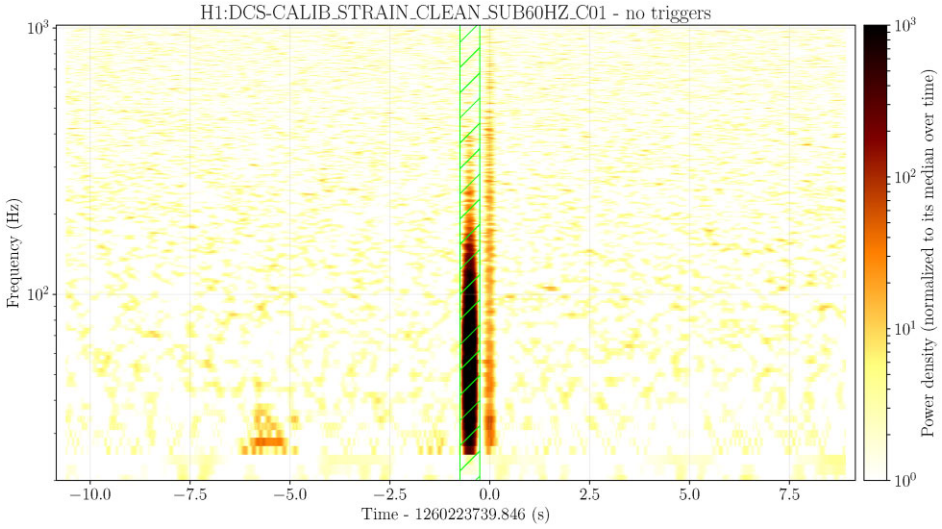

Figure B1. Time–frequency map of the third-ranked PyCBC candidate. This shows the glitch present near to GPS time 1260223739. This represents the kind of quieter glitches that gets skipped in the normal autogating procedures.

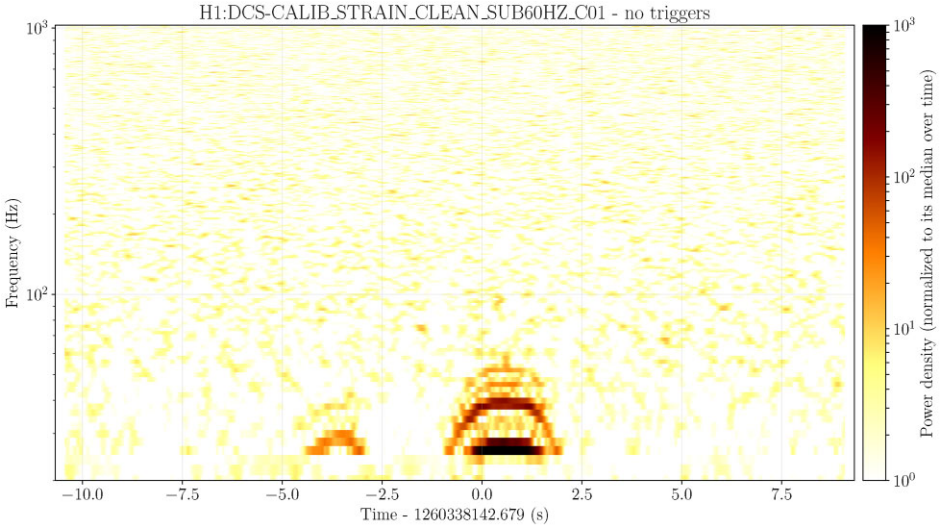

Figure B2. Time–frequency map of the fourth-ranked PyCBC candidate, consistent with a scattered-light glitch (Soni et al. 2021a, b; Tolley et al. 2023).

APPENDIX C: DETAILS OF THE MILLILENSING INVESTIGATION

This Appendix presents additional details of the millilensing investigation for GW200208.

Fig. C1 represents the result for a millilensing run done with three possible superposed images. It shows that in comparison with Fig. 19, the addition of a third image is not leading to the recovery of an extra possible image since the posterior for its lensing parameters are flat and uninformative. Similarly, the posteriors for the four-image analysis (Fig. C2) show flat posteriors for the lensing parameters of the third and fourth possible images, leading to the conclusion that no more than two images can be identified in the data.

In addition to a posterior on the possible number of images, the run where the number of images is left free also returns posteriors for the lensing parameters of the different images. These are shown in Fig. C3. Only the posteriors for a possible second image are not completely uninformative. The others are flat, meaning that the analysis does not favour anything with more than two signals.

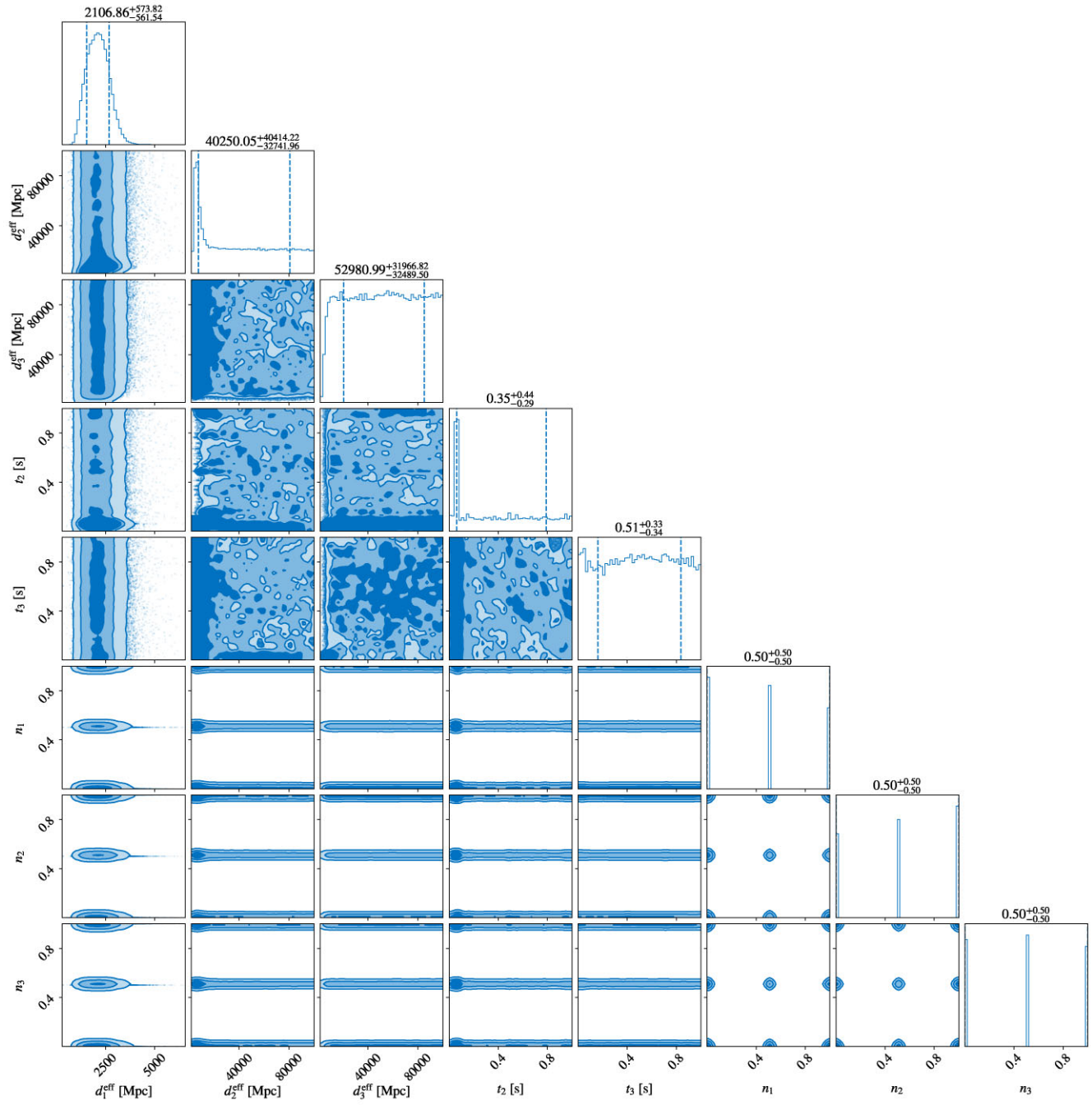

Figure C1. Corner plot of the millilensing parameters obtained from a three-signal analysis. A similar peak appears in the second image parameters as was present within the two-signal analysis shown in Fig. 19. However, no such features are present within the posteriors of the third millilensing image.

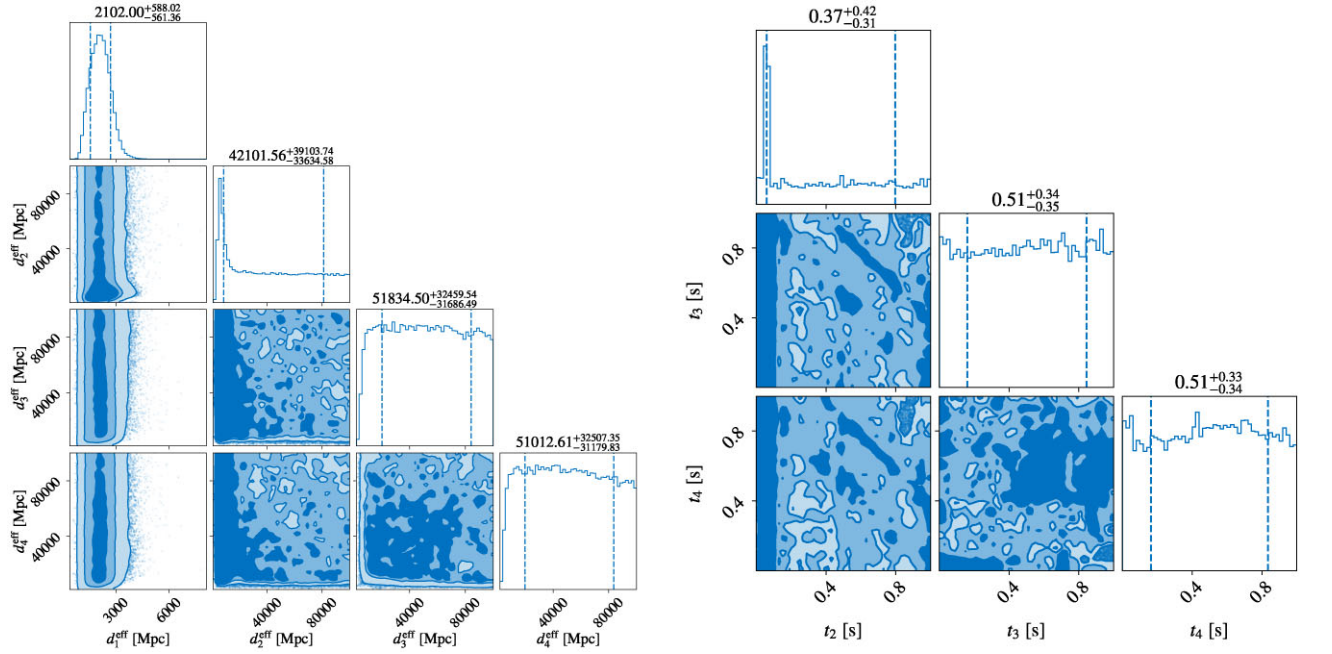

Figure C2. Corner plots of the effective luminosity distance (left panel) and time delay (right panel) parameters obtained from the four-signal analysis. Consistent with the previous analyses, there are no peaking features in the third or fourth signal posteriors.

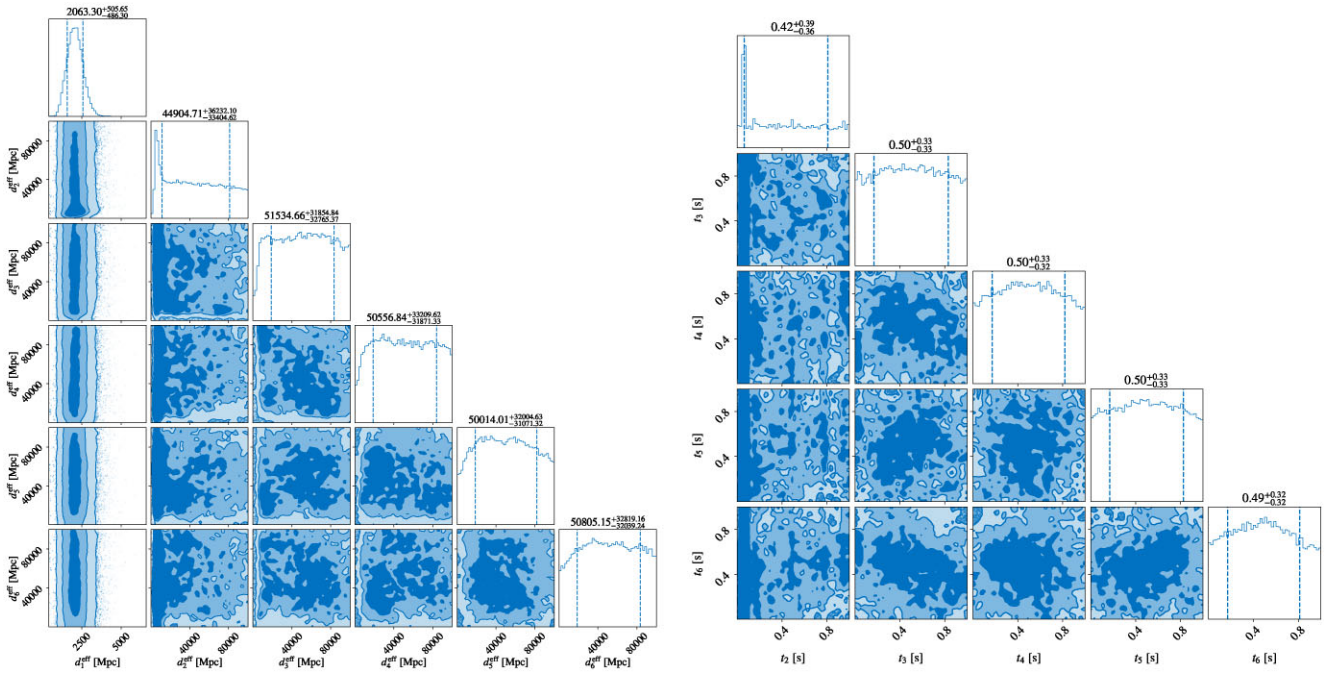

Figure C3. Corner plots of the effective luminosity distance (left panel) and relative time delays (right panel) obtained from a multisignal analysis. Again, these are consistent with what has been seen in the previous analyses with no favouring for any number of signals above two.

[1]*Department of Physics, Institute for Gravitational and Subatomic Physics (GRASP), Utrecht University, Princetonplein 1, NL-3584 CC Utrecht, The Netherlands*
[2]*Nikhef – National Institute for Subatomic Physics, Science Park, NL-1098 XG Amsterdam, The Netherlands*
[3]*SUPA, School of Physics and Astronomy, University of Glasgow, UK*
[4]*International Centre for Theoretical Science, Tata Institute of Fundamental Research, Bangalore 560089, India*
[5]*Niels Bohr International Academy, Niels Bohr Institute, Blegdamsvej 17, DK-2100 Copenhagen, Denmark*
[6]*The Inter-University Centre for Astronomy and Astrophysics, Post Bag 4, Ganeshkhind, Pune 411007, India*
[7]*Departament de Física, Universitat de les Illes Balears, IAC3–IEEC, E-07122 Palma, Spain*
[8]*University of Portsmouth, Institute of Cosmology and Gravitation, Portsmouth PO1 3FX, UK*
[9]*LIGO, California Institute of Technology, Pasadena, CA 91125, USA*
[10]*Department of Physics, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong*
[11]*Kavli Institute for the Physics and Mathematics of the Universe, 5-1-5 Kashiwanoha, Kashiwa-shi, Chiba 277-8583, Japan*
[12]*Canadian Institute for Advanced Research, CIFAR Azrieli Global Scholar, MaRS Centre, West Tower, 661 University Ave, Toronto, ON M5G 1M1, Canada*
[13]*LIGO Laboratory, Massachusetts Institute of Technology, 185 Albany St, Cambridge, MA 02139, USA*
[14]*Department of Physics and Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA 02139, USA*
[15]*STAR Institute, Bâtiment B5, Université de Liège, Sart Tilman, B-4000 Liège, Belgium*
[16]*Instituto de Física Teórica UAM/CSIC, Universidad Autonoma de Madrid, E-28049 Madrid, Spain*
[17]*Department of Physics and Astronomy, KU Leuven, Celestijnenlaan 200D, B-3001 Leuven, Belgium*
[18]*Department of Electrical Engineering (ESAT), KU Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium*
[19]*University of Wisconsin-Milwaukee, Milwaukee, WI 53201, USA*
[20]*McWilliams Center for Cosmology, Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213, USA*
[21]*Department of Astronomy & Astrophysics, Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai 400005, India*

This paper has been typeset from a TEX/LATEX file prepared by the author.