



Effect of Deviations from General Relativity on Searches for Gravitational-wave Microlensing and Type II Strong Lensing

Mick Wright¹, Justin Janquart^{2,3}, and Nathan K. Johnson-McDaniel⁴

¹ SUPA, School of Physics and Astronomy, University of Glasgow, Glasgow, G12 9SU, UK

² Department of Physics, Institute for Gravitational and Subatomic Physics (GRASP), Utrecht University, Princetonplein 1, NL-3584 CC Utrecht, The Netherlands

³ Nikhef—National Institute for Subatomic Physics, Science Park, NL-1098 XG Amsterdam, The Netherlands

⁴ Department of Physics and Astronomy, The University of Mississippi, University, MS 38677, USA

Received 2024 March 26; revised 2024 December 4; accepted 2024 December 9; published 2025 March 4

Abstract

As the gravitational-wave (GW) detector network is upgraded and the sensitivity of the detectors improves, novel scientific avenues open for exploration. For example, tests of general relativity (GR) will become more accurate as smaller deviations can be probed. Additionally, the detection of lensed GWs becomes more likely. However, these new avenues could also interact with each other, and a GW event presenting deviations from GR could be mistaken for a lensed one. Here, we explore how phenomenological deviations from GR or binaries of exotic compact objects could impact those lensing searches focusing on a single event. We consider strong lensing, millilensing, and microlensing, and find that certain phenomenological deviations from GR may be mistaken for all of these types of lensing. Therefore, our study shows that future candidate lensing events would need to be carefully examined to avoid a false claim of lensing where instead a deviation from GR has been seen.

Unified Astronomy Thesaurus concepts: Gravitational lensing (670); Gravitational microlensing (672); Gravitational waves (678); General relativity (641)

1. Introduction

General relativity (GR) is the most successful description of gravity to date, able to predict the behaviors of many astrophysical phenomena with great accuracy (E. Berti et al. 2015) and passing many laboratory tests (C. M. Will 2014). One such prediction of GR is the existence of gravitational waves (GWs), which were first observed in 2015 (B. P. Abbott et al. 2016a) by the Advanced LIGO detectors (J. Aasi et al. 2015), later joined by Advanced Virgo (F. Acernese et al. 2014), and have been observed with increasing regularity ever since—with ~90 detections reported in the latest catalog published by the LIGO, Virgo, and KAGRA (LVK) collaborations, GWTC-3 (R. Abbott et al. 2023). Many more detections are anticipated during the fourth LVK observing run (O4), currently underway, and in the future, when existing detectors become more sensitive and additional ground-based detectors come online (see, e.g., M. Punturo et al. 2010; M. Evans et al. 2021).

As the detection of GWs moves from being exceptional to commonplace, the focus is shifting to investigate a range of phenomena that they may allow us to probe. In particular, two active areas of study are the search for gravitational lensing of GW signals (O. A. Hannuksela et al. 2019; R. Abbott et al. 2021a, 2024b; J. Janquart et al. 2023b) and more detailed investigations as to whether there is evidence for any deviations from the predictions of GR (B. P. Abbott et al. 2016b, 2019a, 2019b; R. Abbott et al. 2021b, 2021c) in the dynamic, highly nonlinear regime to which GW signals from compact binary mergers grant access.

To briefly describe GW lensing: Similar to light, GW signals that pass by massive objects have their propagation modified

by that object. These modifications take a variety of forms depending upon the mass scale of the lens in question. This leads to a categorization of lensing phenomena based upon these scales. In this work, we focus on three of these scales which are of interest for the lensing of GW signals and define them in terms of the phenomenology expected to occur to these signals with current ground-based detectors. These scales are *strong lensing*, which results in multiple distinct signals—called images—which each experience a particular magnification, phase shift, and time delay (Y. Wang et al. 1996; L. Dai & T. Venumadhav 2017; J. M. Ezquiaga et al. 2021); *millilensing*, in which the time delay is sufficiently short that the signals overlap, resulting in beating patterns, but which may still be treated geometrically (A. Liu et al. 2023); and *microlensing*, in which the overlap is sufficient to result in significant frequency-dependent wave-optics effects in the beating patterns of the signal (S. Deguchi & W. D. Watson 1986; R. Takahashi & T. Nakamura 2003; Z. Cao et al. 2014).

As already stated, GR has passed all of the laboratory and astrophysical tests that have been conducted so far. GW signals offer the ability to test gravity in the dynamical, highly nonlinear regime. In general, tests using GW signals have focused on placing constraints on potential deviations from GR rather than on affirmative testing of non-GR theories. This has been due to a lack of accurate waveform models for inspiral, merger, and ringdown (IMR) in these non-GR theories. Modifications to gravity will broadly affect three areas of the GW: generation, with different post-Newtonian coefficients (e.g., S. Tahura & K. Yagi 2018; L. Bernard et al. 2022; B. Shiralilou et al. 2022), as well as differences in the merger-ringdown phase (e.g., M. Okounkova et al. 2020; M. Corman et al. 2023) and quasi-normal mode (QNM) spectrum of the final black hole (e.g., M. Srivastava et al. 2021; P. Wagle et al. 2022; P. A. Cano et al. 2023); propagation (e.g., S. Mirshekari et al. 2012; V. A. Kostelecký & M. Mewes 2016; M. Mewes 2019; T. Zhu et al. 2024); and the polarization modes—with a



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

general theory of gravity allowing for up to six modes: two tensorial, two vectorial, and two scalar, compared with only the two tensorial (termed *plus* and *cross*) modes predicted by GR (D. M. Eardley et al. 1973a, 1973b).

Some specific examples of additional effects on GWs that modifications to gravity may yield would be dispersion (S. Mirshekari et al. 2012) due to a massive graviton or Lorentz violations; amplitude birefringence (M. Okounkova et al. 2022), the explicit enhancement or suppression of the left or right circular polarization modes; or amplitude dampening (E. Belgacem et al. 2018; A. Nishizawa 2018).

As of the end of the third observing run (O3) and the construction of GWTC-3, analysis of the detected GW signals has not yielded any confident detections of GW lensing (O. A. Hannuksela et al. 2019; L. Dai et al. 2020; X. Liu et al. 2021; R. Abbott et al. 2021a, 2024b; J. Janquart et al. 2023b) nor any significant support for physics beyond GR (B. P. Abbott et al. 2016b, 2019a, 2019b; R. Abbott et al. 2021b, 2021c). However, with O4 underway and the expected increase in the sensitivity of the detectors (B. P. Abbott et al. 2020), there is the potential for the detection of lensed GW events (see, e.g., K. K. Y. Ng et al. 2018) as well as the prospect of even greater sensitivity to small deviations from GR (R. Abbott et al. 2021c).

Both lensing and deviations from GR have the capacity to cause alterations to GW waveforms. Moreover, in some cases the effect of lensing can be leveraged to test deviations from GR. For instance, one can probe extra polarizations (S. Goyal et al. 2021; I. Magaña Hernandez 2022), birefringence (J. M. Ezquiaga & M. Zumalacárregui 2020; S. Goyal et al. 2023), or modified dispersion relations including the massive graviton hypothesis (A. Finke et al. 2021; H. Narola et al. 2024). It is therefore vital in the individual searches for these phenomena to be aware of the effect that each phenomenon could have on the other, and in particular to understand whether false positives may arise in one search as a result of the other alteration. Here, we investigate the effect of some phenomenological deviations from GR or proxies for signals from black hole mimickers on the searches for GW microlensing and type II strong lensing—searches which consider only a single GW event at a time. Specifically, we consider the effect of dispersive propagation from a massive graviton, a modified energy and angular momentum flux, a modified QNM spectrum in the ringdown portion of the signal, an additional scalar polarization, and numerical relativity (NR) binary neutron star (BNS) waveforms scaled to binary black hole (BBH)-like masses as a proxy for the signal from a binary of black hole mimickers. We also note that lensing could bias the result of analyses looking for deviations from GR. This complementary approach has been investigated in A. Mishra et al. (2024), where they show that in some cases microlensed events can be mistaken for deviations from GR. We refer the reader to A. Gupta et al. (2024) for a review of other potential phenomena that may be mistaken for deviations from GR.

The rest of the paper adopts the following structure. Section 2 introduces GW lensing including the three considered regimes. Section 3 details the potential deviations from GR explored in this work. Section 4 discusses the tests performed and their results. Section 5 then gives the conclusions and future prospects from our investigations. Appendix A gives technical details about the construction of the non-GR signals and illustrates the waveforms we consider.

Appendix B outlines the complete prior choices for the analyses in this work. Appendices C, D, and E give additional results from the microlensing, millilensing, and type II strong-lensing investigations, respectively.

2. Gravitational Lensing

As previously discussed, the lensing of a GW signal occurs when it passes by a massive object, which modifies its propagation. The exact nature of the modification depends upon the mass scale of the lensing object, however the effect in all cases may be described by a relatively simple relationship. A standard, unlensed GW signal may be described by its frequency-domain strain, $h^U(f)$. After undergoing the process of lensing, the resultant lensed signal strain, $h^L(f)$, is described in terms of the unlensed strain and an *amplification factor*, $F(f)$, by

$$h^L(f) = F(f) \times h^U(f). \quad (1)$$

The amplification factor depends upon the mass density profile of the lensing object. However, for any profile it may be calculated from the following general expression (P. Schneider et al. 1992; R. Takahashi & T. Nakamura 2003):

$$F(w, \mathbf{y}) = \frac{w}{2\pi i} \int d^2x \exp[iwT(\mathbf{x}, \mathbf{y})], \quad (2)$$

where w is the dimensionless frequency, $\mathbf{y}$ is the dimensionless displacement of the source from the optical axis, $\mathbf{x}$ is the dimensionless displacement of the image from the optical axis, and the function T yields the dimensionless time delay of the resulting image. The dimensionless displacements may be related to the physical positions of the *impact parameter*, ξ , and the *source position*, η , by defining two scales, one for the lens plane, ξ_0 —which is defined on a per lens model basis—and one for the source plane, $\eta_0 = \xi_0 D_{OL}/D_{OS}$, where the quantities D_{XY} represent the angular distances between the observer O , lens L , and source S . The definitions for ξ_0 for the two specific lensing models considered in this work are $\xi_0 = (4M_L D_{OL} D_{LS}/D_{OS})^{1/2}$ for an isolated point-mass lens (P. Schneider et al. 1992)—where M_L is the mass of the lens—and $\xi_0 = 4\pi\sigma_v^2 D_{OL} D_{LS}/D_{OS}$ for a singular isothermal sphere (SIS) lens (J. Binney & S. Tremaine 1987), where σ_v is the lens velocity dispersion. With these definitions, the dimensionless displacements are then given by

$$\mathbf{x} = \frac{\xi}{\xi_0}, \quad \mathbf{y} = \frac{\eta}{\eta_0}. \quad (3)$$

The conversion from frequency, f , to dimensionless frequency, w , is given by

$$w = \frac{D_{OS}}{D_{OL} D_{LS}} \xi_0^2 (1 + z_L) 2\pi f, \quad (4)$$

for a lens at redshift z_L . In the case of strong lensing, $w \gg 1$ for the frequencies at which a signal is within the detectable band of the current ground-based detector network. This allows a simplification of Equation (2) by taking the geometric optics approximation.

In the geometric optics approximation, only the stationary points of the dimensionless time delay function, $T(\mathbf{x}, \mathbf{y})$, contribute, reducing Equation (2) to a summation over these points (T. T. Nakamura & S. Deguchi 1999). The stationary

points correspond to the individual, resolvable strong-lensing images, and the j th such image will have its individual amplification given by (T. T. Nakamura & S. Deguchi 1999; R. Takahashi & T. Nakamura 2003; J. M. Ezquiaga et al. 2021)

$$F_j(f) = |\mu_j|^{1/2} \exp[2i\pi t_j f - i\pi n_j \text{sign}(f)]. \quad (5)$$

Equation (5) indicates that the effect of strong lensing is threefold. It is the combination of a magnification, μ_j , scaling the signal's amplitude; a temporal delay, t_j , of the signal due to both the increased path length and Shapiro delay; and an overall phase shift, termed the Morse phase. For the latter, the Morse factor, n_j , may take one of three specific values: 0, $1/2$, or 1 depending upon whether the stationary point corresponds to a minimum, saddle point, or maximum of the time delay function. These cases are referred to as type I, II, and III images, respectively.

Evidence for an individual signal being type II may be searched for directly, as in this case the phase shift yields a frequency-dependent time delay for different frequency components. This may be detectable in the case of a signal with sufficiently measurable higher-order modes (HOMs; J. M. Ezquiaga et al. 2021; J. Janquart et al. 2021b, 2023b; Y. Wang et al. 2021; A. Vijaykumar et al. 2023; R. Abbott et al. 2024b). An implementation of such a waveform is implemented in the GOLUM package (J. Janquart et al. 2021a, 2022, 2023).

The millilensing case is similar to the strong-lensing case in that the lens mass is still sufficiently high that the geometric optics approximation may be applied. However, the lens is decreased in scale from that of galaxies and galaxy clusters to that of subgalactic structure. Thus, the induced time delays are small enough that the resulting signals overlap, resulting in beating patterns within the waveform, with the total amplification being described by the summation of each of the images from Equation (5). Implementation of this as a phenomenological modification to a waveform with which to perform parameter-estimation-based searches was done by A. Liu et al. (2023) using the Gravelamps package (M. Wright & M. Hendry 2022; M. Wright et al. 2022).

In the case of microlensing, the mass scales of the lensing object are sufficiently low that some or all of the frequencies of the signal inside of the detectable band will not yield $w \gg 1$. This means that the geometric optics approximation will not be valid for all of the detectable frequencies with current setups, and the full wave-optics treatment must be considered. Equation (2) must then be calculated for whichever mass density profile the lens is being used to conduct the investigation.

The simplest class of profiles that may be considered for a lensing object are axially symmetric lenses. In this case, Equation (2) may be simplified to (R. Takahashi & T. Nakamura 2003)

$$F(w, y) = -iwe^{iwy^2/2} \times \int_0^\infty dx x J_0(wxy) \exp \left[iw \left(\frac{1}{2}x^2 - \psi(x) + \phi_m(y) \right) \right], \quad (6)$$

where J_0 is a Bessel function, $\psi(x)$ is the lensing potential, and $\phi_m(y)$ is the phase shift required to enforce that the minimal time delay induced by the lensing is zero.

The simplest of the axisymmetric lenses is the isolated point-mass lens. In this case, the lensing potential is given by $\psi(x) = \ln x$, and the phase term is given by $\phi_m(y) = (x_m - y)^2/2 - \ln x_m$, where the image position yielding the minimum time delay, x_m , is given by $y + \sqrt{y^2 + 4}$. Inserting these into Equation (6) yields an expression that may be analytically solved, leading to (R. Takahashi & T. Nakamura 2003)

$$F(w, y) = \exp \left[\frac{\pi w}{4} + i \frac{w}{2} \ln \left(\frac{w}{2} - 2\phi_m(y) \right) \right] \times \Gamma \left(1 - \frac{i}{2}w \right) {}_1F_1 \left(\frac{i}{2}w; 1; \frac{i}{2}wy^2 \right), \quad (7)$$

where ${}_1F_1$ is the confluent hypergeometric function of the first kind. Due to this analytic form of the amplification factor, the isolated point mass is the most computationally tractable profile with which to conduct microlensing searches, and as such has been used in such searches previously (O. A. Hannuksela et al. 2019; R. Abbott et al. 2021a, 2024b; J. Janquart et al. 2023b). Even where more realistic profiles may be deployed, the isolated point-mass profile makes for a useful first pass to determine candidates worthy of follow-up (J. Janquart et al. 2023b) to limit the need for the use of these more computationally intensive profiles, such as the SIS (J. Binney & S. Tremaine 1987) or Navarro–Frenk–White (NFW; J. F. Navarro et al. 1997) lens profiles. An implementation of these various profiles may be found in the Gravelamps package (M. Wright & M. Hendry 2022; M. Wright et al. 2022) leveraging the C arbitrary-precision library ARB (F. Johansson 2017).

3. Deviations from General Relativity and Generation of Modified Waveforms

In general, alternative theories of gravity will affect three aspects of the GW signal itself: its generation, propagation, and polarizations—such modifications may affect the detection of the GW (see, e.g., H. Narola et al. 2023; A. Sharma et al. 2024). We consider phenomenological modifications to all of these, as well as proxies for waveforms from a binary of black hole mimickers (in GR). Specifically, we consider the following:

1. Dispersion due to a massive graviton.
2. Self-consistent modifications to the energy flux and the QNM spectrum as proxies for general modifications arising from non-GR theories.
3. Addition of a scalar polarization.
4. Scaled waveforms from a BNS system as a proxy for a black hole mimicker.

The philosophy of the non-GR waveform construction is to include an interesting non-GR effect in a controlled manner, not intending to emulate any specific alternative theory, except for the massive graviton case (where this still is just a phenomenological introduction of dispersive propagation). These waveform constructions are also used in a mock data challenge being carried out by the LVK testing GR group. These phenomenologies are some of those that are possible in extended theories, but are by no means exhaustive—for instance, they do not include birefringent propagation effects. These cases were however deemed a sufficient selection for an initial investigation. We now describe exactly how these

waveforms are constructed. These waveforms are all illustrated in Appendix A.4.

All of the non-GR waveforms are constructed based on the `TEOBResumS-v3-GIOTTO` effective-one-body (EOB) GR waveform model (T. Damour & A. Nagar 2014; A. Nagar et al. 2016, 2018, 2020a; S. Akçay et al. 2021; G. Riemenschneider et al. 2021; R. Gamba et al. 2022). We use this as a model for precessing BBHs on quasi-circular orbits. It contains the dominant $(2, \pm 2)$ modes as well as the $(2, \pm 1)$, $(3, \pm 3)$, $(3, \pm 2)$, $(4, \pm 4)$, and $(5, \pm 5)$ higher modes (in the coprecessing frame) with a robust merger-ringdown model (A. Nagar et al. 2020b). All these modes are included in our injections.

3.1. Dispersion and Massive Graviton

In GR, the propagation of GWs is nondispersive and consequently follows the relation $E = p^2 c^2$ between the energy, E , and momentum, p , of the wave. In non-GR theories, the propagation may be dispersive. The following parametric relation describes the leading effect of dispersion in a number of non-GR scenarios (S. Mirshekari et al. 2012):

$$E^2 = p^2 c^2 + A_\alpha p^\alpha c^\alpha, \quad (8)$$

where the phenomenological parameters α and A_α describe the dispersion.

In the specific case of a massive graviton with no other dispersive effects, special relativity predicts (C. M. Will 1998)

$$E^2 = p^2 c^2 + m_g^2 c^4, \quad (9)$$

which is a special case of Equation (8) with $\alpha = 0$ and $\sqrt{A_0} = m_g c^2 > 0$. The effect of this dispersion is to introduce a dephasing of the entire signal in the frequency domain compared with the GR prediction (C. M. Will 1998):

$$\delta\Psi(f) = \frac{\pi c D_0}{\lambda_g^2 f}. \quad (10)$$

Here, $\lambda_g = h/(m_g c)$ is the Compton wavelength of the graviton, and D_0 is a distance measure given by (S. Mirshekari et al. 2012)

$$D_0 = \frac{1+z}{H_0} \int_0^z \frac{dz'}{(1+z')^2 \sqrt{\Omega_M(1+z')^3 + \Omega_\Lambda}}, \quad (11)$$

where z is the binary's redshift, $H_0 = 67.90 \text{ km s}^{-1} \text{ Mpc}^{-1}$ is the Hubble constant, $\Omega_M = 0.3065$ is the matter density, and $\Omega_\Lambda = 0.6935$ is the dark energy density. Here, we use the TT+lowP+lensing+ext results from P. A. R. Ade et al. (2016) as in the LVK analysis. Using all the high-significance BBH detections through the O3 observing run, the LVK Collaboration has set bounds on the possible mass of the graviton by constraining this dispersive effect, giving $m_g \leq 1.27 \times 10^{-23} \text{ eV}/c^2$ at 90% credibility (R. Abbott et al. 2021c), recently improved to $m_g \leq 9.6 \times 10^{-24} \text{ eV}/c^2$ by simultaneously modeling the astrophysical population of compact binaries (E. Payne et al. 2023).

3.2. Modifications to Energy Flux and QNM Spectrum

Non-GR theories often lead to differences in all of the GW energy and angular momentum fluxes (and also the linear momentum flux), the binary's binding energy, and the QNM

spectrum of the final black hole.⁵ However, for the purposes of these investigations, we isolate the effect of modifications of the energy flux and QNM spectrum specifically, recalling that we are not trying to emulate any specific theory. In both cases, we set the mass and spin of the final black hole used to obtain the QNM frequencies and damping times in the model iteratively, so that they satisfy energy and angular momentum balance (since the energy and angular momentum carried away by the GWs are modified in both cases).

For the modification to the energy and angular momentum flux, we generalize the construction used in A. Ghosh et al. (2016, 2017) and N. K. Johnson-McDaniel et al. (2022), which applied it to an older nonspinning EOB model, to the modern `TEOBResumS-v3-GIOTTO` model. Specifically, we modify the energy flux by multiplying those waveform modes which enter at the second post-Newtonian (2PN) order by a factor $a_2^{1/2}$, so their contribution to the energy and angular momentum fluxes is multiplied by a factor a_2 . One thus gets modifications to the GW phasing at 2PN, 3PN, and all higher PN orders. This is similar to what one expects for the modifications to the GW phasing in non-GR theories. For instance, in dynamical Chern–Simons theory, the GW phasing also starts to differ from GR at 2PN (K. Yagi et al. 2012). The modes that first enter at 2PN are the $(3, \pm 2)$, $(4, \pm 4)$, and $(4, \pm 2)$ modes. We only modify the $(3, \pm 2)$ and $(4, \pm 4)$ modes, since those are the ones that first enter the energy flux at 2PN and have a robust merger-ringdown model in `TEOBResumS-v3-GIOTTO`. The contribution from the omitted $(4, \pm 2)$ modes is much smaller than that from the $(3, \pm 2)$ and $(4, \pm 4)$ modes (only $\sim 5 \times 10^{-4}$ of the total Newtonian energy flux from this set of modes).

We consider both the case where the $(3, \pm 2)$ and $(4, \pm 4)$ modes in the GW signal are multiplied by $a_2^{1/2}$, as well as the case where they are not (so the additional energy and angular momentum is radiated in a field that does not couple to GW detectors). We refer to these as the cases with and without waveform scaling, respectively. We consider the latter case in order to have larger deviations in the PN coefficients, length of the waveform (from a given frequency), and final mass and spin (since such larger deviations were found to be necessary to obtain a detectable GR deviation in the earlier studies that just analyzed the dominant modes) without having the significantly non-GR waveform features caused by significantly increasing the amplitude of the higher modes. The waveforms are illustrated in Appendix A.4, and Appendix A.2 shows the final mass and spin obtained in this construction and with the modified QNM spectrum.

As a proxy for the modified QNM spectrum one would obtain in a non-GR theory, we modify the QNM spectrum from that of Kerr to that of Kerr–Newman, which includes a charge parameter in addition to the mass and spin. We use the results for the $(2, 2, 0)$ and $(3, 3, 0)$ QNMs computed in Ó. J. C. Dias et al. (2022b; see also G. Carullo et al. 2022). Here, the (ℓ, m, n) notation includes the overtone index n in addition to the multipolar indices ℓ and m . The additional QNMs computed in Ó. J. C. Dias et al. (2022a) were not available when these waveforms were first constructed, so we do not use them here, but instead construct the additional QNMs needed for the `TEOBResumS-v3-GIOTTO` ringdown model (the $n = 0, 1$ QNMs for each ℓ, m mode included) of these using simple

⁵ See the discussion in Section 1, where the first two deviations lead to the deviations in post-Newtonian coefficients discussed there.

expressions based on the eikonal approximation or spin expansion, as discussed in Appendix A.1. As shown in N. K. Johnson-McDaniel et al. (2022), a number of standard tests of GR applied by the LVK (in, e.g., R. Abbott et al. 2021c) can detect sufficiently large modified energy fluxes. There are also standard tests sensitive to modified QNM spectra, including dedicated ringdown analyses (see, e.g., R. Abbott et al. 2021c). None of these tests has detected a deviation from GR. However, it is not straightforward to translate either of these constraints into constraints on a_2 or the charge of the final black hole in these waveforms.

3.3. Addition of Scalar Polarization

Here, we want to add a scalar polarization to a GR waveform. We would expect any alternative theory that radiates in additional polarizations to have a different tensor waveform than in GR, due for example to the change in phasing from the extra energy radiated in the nontensor modes. However, here we leave the tensor polarization waveform unchanged, in order to isolate the effects of the additional polarization. We use the leading tensor-scalar waveform results from L. Bernard et al. (2022) to scale the tensor polarizations, specializing to the case where the sensitivities of the two objects are equal. In this case, there is no dipole radiation, which is consistent with our construction, since there is no dipole radiation present in the tensor waveform. This choice gives nonzero scalar modes for the spherical harmonic modes where $\ell + m$ is even, where the specific expressions we use are given in Appendix A.3. The contribution from scalar modes is parameterized by a ratio $\mathcal{A}$ of dominant-mode scalar to tensor contributions to the response of the detector with the largest tensor mode signal-to-noise ratio (SNR).

Additionally, we still apply this scaling to the modes where $\ell + m$ is even in the inertial frame, even for precessing systems as we consider here. This is a further simplification of the behavior one would expect in a precessing system emitting scalar radiation, where all the scalar modes in the inertial frame will be nonzero due to the binary's precession, even though only the scalar modes where $\ell + m$ is even are nonzero in the coprecessing frame. However, it is in keeping with the general philosophy of having a simple, somewhat physically motivated addition of the scalar modes to the signal.

Again, the LVK Collaboration has looked for possible extra polarizations in observed BBH events, finding no support for possible additional ones (see R. Abbott et al. 2021c, for the latest results). However, we do not try to translate those constraints into any constraints on these scalar-tensor waveforms.

3.4. Black Hole Mimickers

We now consider proxies for waveforms from black hole mimickers. Black hole mimickers can be created by additional fields in GR, such as boson stars (S. L. Liebling & C. Palenzuela 2023), or more generally dark matter stars (G. F. Giudice et al. 2016), or by more exotic non-GR phenomena, such as firewalls (A. Almheiri et al. 2013) or gravastars (P. O. Mazur & E. Mottola 2004). See V. Cardoso & P. Pani (2019) for a review. We would ideally consider waveforms from numerical simulations of binaries of black hole mimickers. However, such simulations have not yet reached the same level of sophistication as simulations of

binaries of black holes and/or neutron stars, even for binary boson stars, the best-modeled black hole mimickers (but see N. Siemonsen & W. E. East 2023, for recent work on simulations of mergers of quasi-circular binary boson stars). We thus consider BNS waveforms scaled to BBH-like masses, as in N. K. Johnson-McDaniel et al. (2020), since waveforms from binaries of black hole mimickers are also expected to have the effects of a nonzero tidal deformability (see, e.g., V. Cardoso et al. 2017) and non-Kerr spin-induced multipole moments (see, e.g., N. Uchikata et al. 2016; C. Adam et al. 2022) in the inspiral and a post-merger signal that can be different from a standard GR black hole ringdown (see, e.g., N. Siemonsen & W. E. East 2023).

We consider two specific waveforms, both from the CoRe catalog (T. Dietrich et al. 2018b; A. Gonzalez et al. 2023), and including all spherical harmonic modes through $\ell = 4$:

1. The NR simulation of a precessing BNS with masses of $1.35M_\odot$ and $1.11M_\odot$ and dimensionless spins of $(-0.077, -0.077, -0.077)$, $(-0.090, -0.090, -0.090)$ at a dimensionless frequency of $GMf/c^3 = 4.76 \times 10^{-3}$ from T. Dietrich et al. (2018a; CoRe ID BAM:0110). This simulation uses the SLy (F. Douchin & P. Haensel 2001) equation of state (EOS), so the stars have dimensionless tidal deformabilities of 392 and 1291, giving an effective dimensionless tidal deformability, $\bar{\Lambda}$ (L. Wade et al. 2014), of 724.
2. The NR simulation from M. Ujevic et al. (2022) of a nonspinning BNS with masses of $1.72M_\odot$ and $0.98M_\odot$ (CoRe ID BAM:0131). This simulation also uses the SLy EOS, so the stars have dimensionless tidal deformabilities of 66 and 2569, giving an effective dimensionless tidal deformability of 509. This case is chosen since it has a more BBH-like post-merger signal than the precessing binary does (see the plots in Appendix A.4).

An approach used to look for possible exotic compact objects is to look for echoes, repeated signals due to signal reflecting multiple times off radial potential barriers. So far, no such echoes have been detected (R. Abbott et al. 2021c). However, echoes will only be present for some binaries of black hole mimickers, and are not present for, e.g., mergers of binary boson stars or in the scaled BNS waveforms. We also note that for some particular events, some more in-depth analyses with exotic objects have been performed. For example, GW190521, an event with exceptionally high masses (R. Abbott et al. 2020a, 2020b), has been analyzed with waveforms for head-on Proca star collisions, finding a small preference for such waveforms compared to BBH waveforms (J. Calderón Bustillo et al. 2021; J. Calderón Bustillo et al. 2023). The latter reference also analyzes some other high-mass events with the Proca star collision waveforms, but does not find these waveforms to be favored for any firm detection.

4. Results

In general, if one analyzes a signal that is not described by the underlying waveform model, then that analysis will attempt to use any freedom available in the model to best fit the model to the provided data. Depending on the specific ways in which parameters affect the model, this can lead to a spurious apparent match with the analysis model. In this context, to investigate how each of the deviations from GR or the BBH hypothesis described in Section 3 would impact the single-

Table 1
Source Parameter Values Used in the GW150914-like Injections

Source Parameter	Value
Total mass, M_{tot}	$72M_{\odot}$
Mass ratio, q	0.8
Luminosity distance, d_L	452 Mpc
Inclination, ι	1.59 rad
R.A.	1.68 rad
decl.	−1.27 rad
Polarization angle, ψ	3.93 rad
Coalescence time, t_c	1126259462s
Primary dimensionless spin vector, χ_1	(−0.1, −0.1, −0.1)
Secondary dimensionless spin vector, χ_2	(−0.1, 0.0, 0.1)

image lensing searches, a series of 65 simulated signals using as their basis GW150914-like parameters (e.g., B. P. Abbott et al. 2016a; R. Abbott et al. 2024a), given in Table 1, was constructed. This total number was broken down across the deviation types as follows:

1. Massive graviton: 20 logarithmically spaced values of A_0 from $5 \times 10^{-45} \text{ eV}^2$ to $5 \times 10^{-43} \text{ eV}^2$, corresponding to $m_g \simeq 7.07 \times 10^{-23} \text{ eV}/c^2$ and $m_g \simeq 7.07 \times 10^{-22} \text{ eV}/c^2$, respectively.
2. Modified energy flux: The following values of a_2 with waveform scaling: 2, 5, 10, 20, 50, and the following values without waveform scaling: 20, 50, 75, 100, 125, 150, 175, 200.
3. Modified QNM spectrum: Values of the final dimensionless charge, Q_f , from 0.50 to 0.75 in steps of 0.05.
4. Scalar polarization: Values of $\mathcal{A}$ from 0.05 to 0.50 in steps of 0.05.
5. Precessing black hole mimicker following T. Dietrich et al. (2018a; BHM-P): Scaling the total mass from $90M_{\odot}$ to $150M_{\odot}$ in steps of $10M_{\odot}$.
6. Nonspinning black hole mimicker following M. Ujevic et al. (2022; BHM-NS): Scaling the total mass from $80M_{\odot}$ to $160M_{\odot}$ in steps of $10M_{\odot}$.

These choices for the non-GR parameters were designed to give cases where one expects the GR deviation to be easily detectable by at least some analyses as well as cases where it will be less easily detectable. In particular, the graviton masses are all well above the constraints set by the LVK, since those come from combining together the constraints from many events, while we only analyze single events here. In the black hole mimicker cases, the lower bound on the total mass is chosen to ensure sufficient higher-mode content is within the LVK band (i.e., from 20 Hz)—here the lowest mass ensures that the $|m| \leq 3$ modes are within band. The upper bound is chosen to retain a sufficient number of inspiral cycles in band for analysis. Each simulated signal was then subjected to the Gravelamps micro- and millilensing analysis pipeline and the GOLUM type II analysis, relying on the Bilby inference framework (G. Ashton et al. 2019) and using the dynesty (J. S. Speagle 2020) nested sampler and the IMRPhenomXPHM (G. Pratten et al. 2021) model for precessing BBH waveforms implemented in LALSuite (LIGO Scientific Collaboration et al. 2024). IMRPhenomXPHM includes the same higher modes in the coprecessing frame as the TEOBResumS-v3-GIOTTO model, except for the (5, ± 5) modes. The injected frames were made without noise, and were analyzed under the assumption

of the predicted O4 noise curves for the LIGO-Virgo network, using the LIGO noise curve with a 190 Mpc horizon (B. P. Abbott et al. 2020). This yielded SNRs for the individual injections between 21.81 and 65.88, which is sufficiently high for obvious detection. The configurations for the lensing analyses conform to those that were deployed in J. Janquart et al. (2023b) and R. Abbott et al. (2024b) and used the priors laid out in full in Appendix B.

We stress here that a false positive from the following analyses would not by itself constitute a claim of detection in a realistic scenario—this would require additional analyses as well as a background study to fully calibrate the significance of the results of these investigations. False positives from the following analyses, however, would be the impetus for additional investigations, such as those carried out in J. Janquart et al. (2023b).

4.1. General Relativity Baseline

As a step before the examination of non-GR waveforms, we first examine the results of our investigations on an injection of the GW150914-like parameters given in Table 1 to set a baseline from which the non-GR results may be compared. Here, we briefly summarize the results of investigations of this waveform, with the more detailed descriptions of how to interpret these results provided in the following respective analysis sections:

1. An investigation using the isolated point-mass microlensing model did not find support for the microlensing hypothesis, with a $\log_{10} \mathcal{B}_{\text{U}}^{\text{ML}}$ —the Bayes factor comparing the evidence for the microlensed and unlensed cases—of -2.5 .
2. A similar investigation with the SIS microlensing model found this model more closely aligned with the unlensed case, with a $\log_{10} \mathcal{B}_{\text{U}}^{\text{SIS}}$ of 0.0. An examination of the posteriors for this case revealed that the lens-mass posterior reflects the prior, and the source position posterior is entirely above $y=1$, which mirrors the expectations for an unlensed event.
3. An investigation using the phenomenological millilensing model did not find support for the existence of any additional millisignals, and preferred the unlensed waveform with $\log_{10} \mathcal{B}_{\text{U}}^{\text{Mil}} = -0.4$.
4. The type II strong-lensing analysis using a discrete prior on the Morse index did not find support for the type II hypothesis. The $\log_{10} \mathcal{B}_{\text{I+III}}^{\text{II}}$ in this case was -2.7 , in line with the expectations for an unlensed signal.
5. The type II strong-lensing investigation using a continuous prior on the Morse index did not find support for the type II case, with a median value of the recovered Morse index at 0.47 and the 90% credible interval (C.I.) boundaries at 0.02 and 0.91. In addition to a broad posterior, it is also important to note a significantly lower density of posterior points around the $n=0.5$ region compared to the cases for the type I and III images. In Section 4.5, we plot this posterior, overlaying it with a non-GR case which is indistinguishable from the GR case.

The conclusion would therefore be that this specific event should not be considered as a lensing candidate, as it adheres to the expectations for an unlensed event. Thus, one would expect that any suggestion of false positivity in the non-GR waveform

Table 2
Results of Performing Isolated Point-mass Microlensing Analysis Using Gravelamps on the Simulated Modified GR Events

Massive Graviton		Modified Energy Flux		Modified QNM Spectrum		BHM-P	
m_g (eV/c ²)	$\log_{10} \mathcal{B}_U^{\text{ML}}$	a_2	$\log_{10} \mathcal{B}_U^{\text{ML}}$	$\mathcal{Q}_f$	$\log_{10} \mathcal{B}_U^{\text{ML}}$	M ($M_\odot$)	$\log_{10} \mathcal{B}_U^{\text{ML}}$
7.07×10^{-23}	-2.3	2	-2.4	0.50	-2.4	80	15.1 ^{β}
7.98×10^{-23}	-2.0	5	-2.2	0.55	-2.4	90	18.8 ^{β}
9.01×10^{-23}	-2.1	10	-1.6	0.60	-2.3	100	26.8 ^{β}
1.02×10^{-22}	-2.3	20	0.9	0.65	-2.3	110	31.2 ^{β}
1.15×10^{-22}	-2.4	50	8.6 ^{β}	0.70	-2.0	120	24.3 ^{β}
1.30×10^{-22}	-2.3	20 ^a	-2.4	0.75	-2.1	130	41.2 ^{β}
1.46×10^{-22}	-2.3	50 ^a	-2.2			140	47.4 ^{β}
1.65×10^{-22}	-2.2	75 ^a	-2.1			150	34.3 ^{β}
1.86×10^{-22}	-2.3	100 ^a	-2.2	Scalar Polarization		BHM-NS	
2.10×10^{-22}	-2.2	125 ^a	-2.3	$\mathcal{A}$	$\log_{10} \mathcal{B}_U^{\text{ML}}$	M ($M_\odot$)	$\log_{10} \mathcal{B}_U^{\text{ML}}$
2.38×10^{-22}	-2.2	150 ^a	-2.2	0.05	-2.6	80	-0.7
2.68×10^{-22}	-2.0	175 ^a	-2.1	0.10	-2.5	90	0.1
3.03×10^{-22}	-2.2	200 ^a	-1.8	0.15	-2.7	100	1.3 ^{β}
3.42×10^{-22}	-2.2			0.20	-2.4	110	3.2 ^{β}
3.86×10^{-22}	-0.5			0.25	-2.6	120	5.9 ^{β}
4.35×10^{-22}	9.6 ^{β}			0.30	-3.1	130	7.3 ^{β}
4.92×10^{-22}	27.3 ^{β}			0.35	-3.0	140	7.4 ^{β}
5.55×10^{-22}	28.1 ^{β}			0.40	-3.1	150	9.2 ^{β}
6.26×10^{-22}	15.6 ^{β}			0.45	-3.2	160	11.3 ^{β}
7.07×10^{-22}	24.6 ^{β}			0.50	-3.1		

Notes. Cells with a β superscript indicate the cases with $\log_{10} \mathcal{B}_U^{\text{ML}} > 1$ that would be flagged for further investigation as microlensing candidates. As can be seen, aside from the modified energy flux with no waveform scaling, the modified QNM spectrum, and the addition of a scalar polarization, each GR deviation has at least some cases in which false positives may be identified.

^a Indicates the injections have been done without waveform scaling.

analysis based upon this system would be driven by the deviations from GR rather than the source parameter combination.

4.2. Point-mass Microlensing Analysis

The first step of microlensing investigations is an analysis using the isolated point-mass microlensing model, due to the increased computational tractability of calculating the amplification factor compared with other models.

Table 2 gives the complete results of this investigation, with the superscripted β indicating those analyses with a Bayes factor comparing the point-mass microlensing and unlensed cases, $\mathcal{B}_U^{\text{ML}}$, indicating a sufficiently high support for the microlensing hypothesis that they would be flagged for additional follow-up in the style of J. Janquart et al. (2023b). This threshold is set at $\log_{10} \mathcal{B}_U^{\text{ML}} > 1$. The choice of this value is based upon the background microlensing analysis done in R. Abbott et al. (2021a). The errors on the Bayes factor are expected to be of $\mathcal{O}(0.1)$ in this and the subsequent SIS and millilensing analyses. As can be seen within the table, modifications to the QNM spectrum, the addition of a scalar polarization, and non-waveform-scaling modifications to the energy flux do not yield any scenarios in the space tested in which a false preference for the isolated point-mass microlensing model is obtained.

The remaining cases, however, demonstrate at least one scenario in which not only is a false preference for the microlensing hypothesis obtained, but that false preference is strong—particularly in the cases of the injections with a large graviton mass and the precessing black hole mimicker waveforms based on T. Dietrich et al. (2018a; BHM-P). This latter case has two additional points of note. In the analysis of

the BHM-P waveforms, the prior on the luminosity distance needed to be extended from an upper limit of 2 Gpc to an upper limit of 5 Gpc in order to fully enclose the estimated posterior. Analysis of some of these injections also revealed a secondary weaker local maximum-likelihood point in the parameter space found by some of the completed chains, leading to significant disagreement between the chains. The combination of the chains for the results given in Table 2 reflects those with the highest preference for the investigation. This behavior is seen in both the lensed and unlensed analyses of these waveforms.

A possible reason that the scaled BNS waveforms display such strong preference for the microlensing model may lie in their morphology. These waveforms display dephasing and amplitude modification in their mergers as well as an oscillatory post-merger. These are similar to the kinds of effects produced by the oscillatory wave-optics regime amplification factor in microlensing, and thus these additional effects may be relatively well fit using these parameters as compared with the available fits from the unlensed waveform model.

One of the first-stage checks of potential microlensing candidates performed in J. Janquart et al. (2023b) was the checking of the lens parameter posteriors. In the case of an unlensed event, one would not expect to see a strong preference for a nonzero lens mass or a source position less than the upper bound of the prior. The reasoning for the latter instance is that, in all cases, the amplification factor of the point-mass model is oscillatory, but higher values of source position have decreased amplitude and so more closely align with unlensed waveforms. Seeing deviation from either of these expectations would therefore be a sign that further investigation is required, and that the signal may be identified as microlensed.

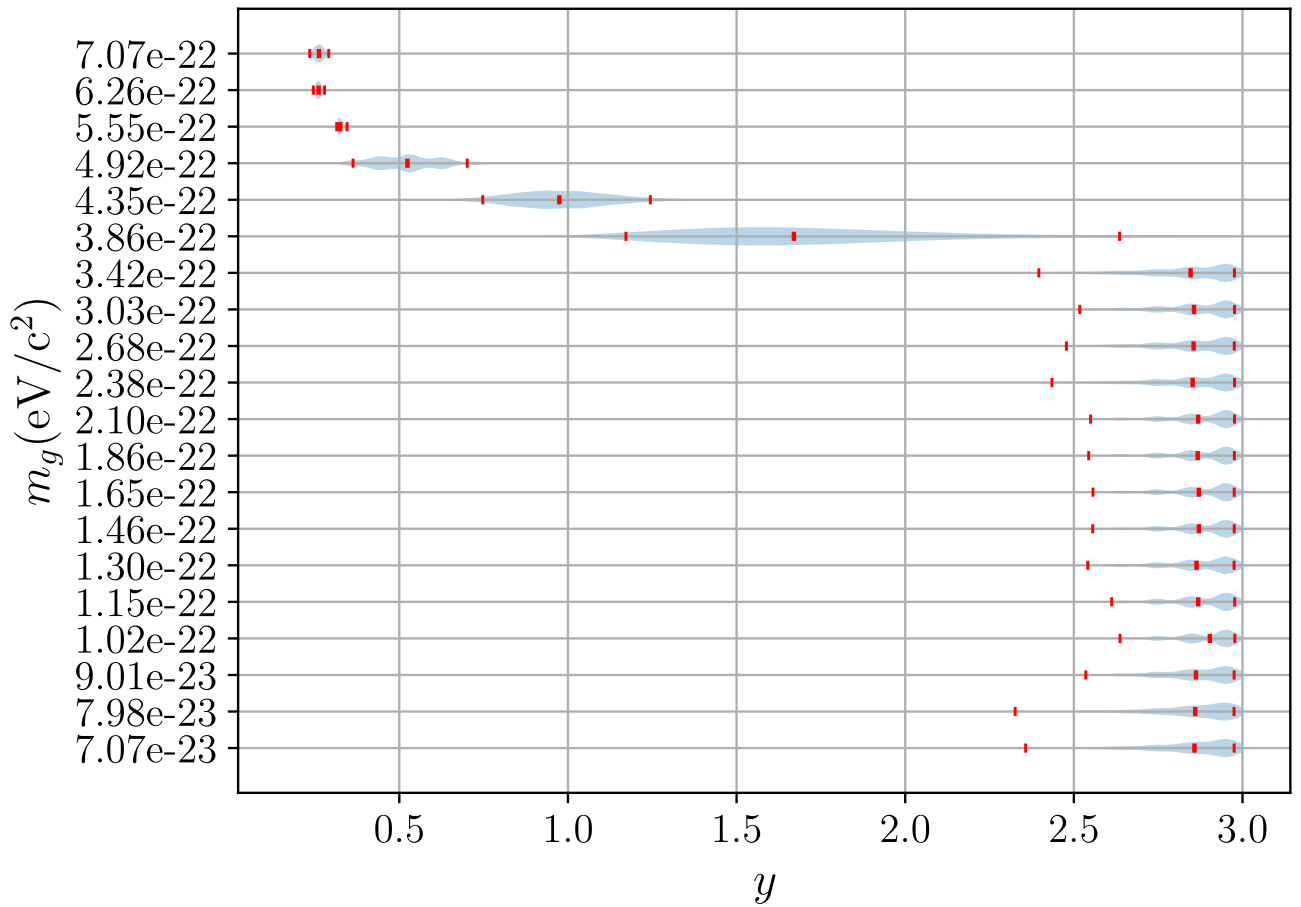


Figure 1. Posteriors on the source position for the massive graviton series of injections analyzed using the isolated point-mass microlensing model. The red ticks indicate the median and 90% C.I. Lower-graviton-mass injections display a preference for source position values toward the upper prior boundary and are railing against it, which is the expectation for unlensed events. However, the higher-graviton-mass injections display significant constraints toward a particular value of source position within the prior space, which would be the expectation for a microlensed event.

Figure 1 shows the posteriors on the dimensionless source position for the analyses performed on the massive graviton injections. For the cases with a smaller graviton mass, one finds the expected result for a nonlensed event of significant support for values around the upper prior bound on the source position. However, beginning with the injection with $m_g = 3.86 \times 10^{-22} \text{ eV}/c^2$, one instead obtained constraints within the prior range. As one can see in Table 2, the threshold case is the last of the injections that retains preference for the nonlensed hypothesis, albeit with significantly less support compared to the injections with smaller graviton masses. As such, these events would require further investigation as potential microlensing events combined with the Bayes factors shown. However, here it is worth noting that the lowest deviations leading to false positives already correspond to graviton masses above the limit set by current analyses (R. Abbott et al. 2021c).

The remaining posterior plots are shown within Appendix C.1. These results are similar to those shown for the massive graviton injections. The injections that show significant support for the microlensing hypothesis also demonstrate apparent false constraints on the lensing parameters, with those with high Bayes factor support demonstrating significant such false constraints, which would require further investigation.

4.3. SIS Microlensing Analysis

The follow-up analysis performed in J. Janquart et al. (2023b) on microlensing candidates was to investigate an additional microlensing mass density profile, specifically that of the SIS profile. It was noted in that work that the microlensing candidate event displayed increased preference for the SIS model compared to the point-mass model. However, the primary purpose of that follow-up would be to gain information about the lensing object from an actual lensed event. Nonetheless, we perform this analysis here on these injections to ascertain whether the false-positive status shown in the point-mass model is restricted to that model or is retained/increased by this additional model. We also deploy this investigation on all of the injections as opposed to only the subset that would be flagged for additional follow-up to check if additional false positives could occur.

Table 3 shows the results of performing this analysis, comparing the SIS model with the point-mass model. Cells with a superscripted β are those that displayed a $\log_{10}$ Bayes factor preference for the point-mass model over the unlensed case and for the SIS model over the point-mass model. This represents a minority of cases, specifically the highest-mass massive graviton injections and some of the nonspinning black hole mimicker cases—though many of these latter cases would also be compatible with indistinguishability. Cells with a

Table 3
Results of Performing the SIS Microlensing Analysis Using Gravelamps on the Simulated Modified GR Events

Massive Graviton		Modified Energy Flux		Modified QNM Spectrum		BHM-P	
m_g (eV/c ²)	$\log_{10} \mathcal{B}_{\text{PM}}^{\text{SIS}}$	a_2	$\log_{10} \mathcal{B}_{\text{PM}}^{\text{SIS}}$	$\mathcal{Q}_f$	$\log_{10} \mathcal{B}_{\text{PM}}^{\text{SIS}}$	M ($M_\odot$)	$\log_{10} \mathcal{B}_{\text{PM}}^{\text{SIS}}$
7.07×10^{-23}	2.4	2	2.5	0.50	2.5	80	-14.3 ^{γ}
7.98×10^{-23}	2.2	5	2.3	0.55	2.6	90	-1.3 ^{γ}
9.01×10^{-23}	2.3	10	1.8	0.60	2.4	100	-0.8 ^{γ}
1.02×10^{-22}	2.3	20	-0.7 ^{γ}	0.65	2.4	110	-4.3 ^{γ}
1.15×10^{-22}	2.4	50	-2.3 ^{γ}	0.70	2.2	120	-1.2 ^{γ}
1.30×10^{-22}	2.4	20 ^a	2.5	0.75	2.1	130	-7.6 ^{γ}
1.46×10^{-22}	2.3	50 ^a	2.4			140	-85.8
1.65×10^{-22}	2.4	75 ^a	2.2			150	-5.4 ^{γ}
1.85×10^{-22}	2.4	100 ^a	2.3	Scalar Polarization		BHM-NS	
2.10×10^{-22}	2.4	125 ^a	2.3	$\mathcal{A}$	$\log_{10} \mathcal{B}_{\text{PM}}^{\text{SIS}}$	M ($M_\odot$)	$\log_{10} \mathcal{B}_{\text{PM}}^{\text{SIS}}$
2.38×10^{-22}	2.3	150 ^a	2.4	0.05	2.7	80	-0.1
2.68×10^{-22}	2.2	175 ^a	2.3	0.10	2.8	90	0.0
3.03×10^{-22}	2.3	200 ^a	2.0	0.15	3.0	100	0.7 ^{β}
3.42×10^{-22}	2.3			0.20	2.6	110	-0.3 ^{γ}
3.86×10^{-22}	0.9			0.25	2.8	120	0.2 ^{β}
4.35×10^{-22}	-1.3 ^{γ}			0.30	3.4	130	0.2 ^{β}
4.92×10^{-22}	-3.8 ^{γ}			0.35	3.3	140	0.2 ^{β}
5.55×10^{-22}	-8.0 ^{γ}			0.40	3.1	150	-0.9 ^{γ}
6.26×10^{-22}	4.0 ^{β}			0.45	3.4	160	-0.2 ^{γ}
7.07×10^{-22}	1.3 ^{β}			0.50	3.1		

Notes. Cells with a γ superscript are those events that had a $\log_{10}$ Bayes factor in favor of the point-mass microlensing case that displayed a $\log_{10}$ Bayes factor supporting the point-mass model over the SIS, though the SIS remains preferred over the unlensed case. Cells with a β superscript are those events that had a $\log_{10}$ Bayes factor in favor of the point-mass microlensing case that displayed a $\log_{10}$ Bayes factor supporting the SIS model over the point-mass one.

^a Indicates the injections have been done without waveform scaling.

superscripted γ are those that displayed a $\log_{10}$ Bayes factor preference for the point-mass model over the unlensed model and over the SIS model.

In all but one case where the point-mass model is preferred over the SIS, the SIS would remain preferred over the unlensed case—though in the case of the $80M_\odot$ precessing black hole mimicker waveform, this preference would be just below the $\log_{10} \mathcal{B}_{\text{U}}^{\text{L}} > 1$ threshold of interest. The exception is the case of the $140M_\odot$ precessing black hole mimicker. In this case, the source parameter posterior estimates do not reflect the primary or secondary likelihood maxima found in the point-mass and unlensed analyses, and are instead in between the source parameter values corresponding to these, which may have resulted in the substantially weaker support for this model. Similar to the point-mass case, the precessing black hole mimicker investigations required extending the luminosity distance prior to 5 Gpc.

Figure 2 shows the source position posteriors for the massive graviton series of injections. Similar to the results shown in Figure 1, the posteriors for those injections that display preference for the microlensed waveform show constraints within the prior boundary. Those events that display preference for the unlensed waveform show broader posteriors than those for the point-mass case, with all support above $y = 1$. This is the expectation in the SIS case, where above this point geometric optics predicts that only one image will be produced by lensing from this profile. This results in a flat amplification factor within this parameter region, which therefore more closely resembles an unlensed waveform. This may be the reason for the favoring of the SIS model over the point-mass lens model in the cases where the unlensed waveform was preferred over the point-mass model, and why the difference

would be sufficient to bring the log Bayes factor between the SIS and unlensed waveforms into indistinguishability.

Appendix C.2 shows the remaining posteriors for the injections, which, as is the case with the point-mass results, display similar behavior to the massive graviton results. In those results with preference for the microlensing hypothesis sufficient to trigger this follow-up investigation, these posteriors are consistently below the $y = 1$ threshold, thus retaining oscillatory behavior of the amplification factor even at high frequency. Consequently, the SIS microlensing follow-up would not allow for any additional candidates to be ruled out. Some additional steps that could be deployed in order to further test these cases are discussed in J. Janquart et al. (2023b).

4.4. Millilensing Analysis

In addition to the secondary microlensing model, J. Janquart et al. (2023b) investigated the potential for the microlensing candidate to be a millilensing candidate. In this scenario, the lens is sufficiently massive that geometric optics applies and Equation (5) describes the amplification. However, the time delays are sufficiently short that the images overlap, resulting in similar beating patterns to microlensing but without wave-optics effects. We replicate this analysis using a phenomenological millilensing model introduced in A. Liu et al. (2023) and implemented within Gravelamps. We allow for up to four signals to be present within the overall signal, which is sufficient to give an indication of whether there is support for additional images but remains tractable for performing a large number of analyses.

Table 4 gives the results of the millilensing analysis showing the $\log_{10}$ Bayes factor between the millilensing and the

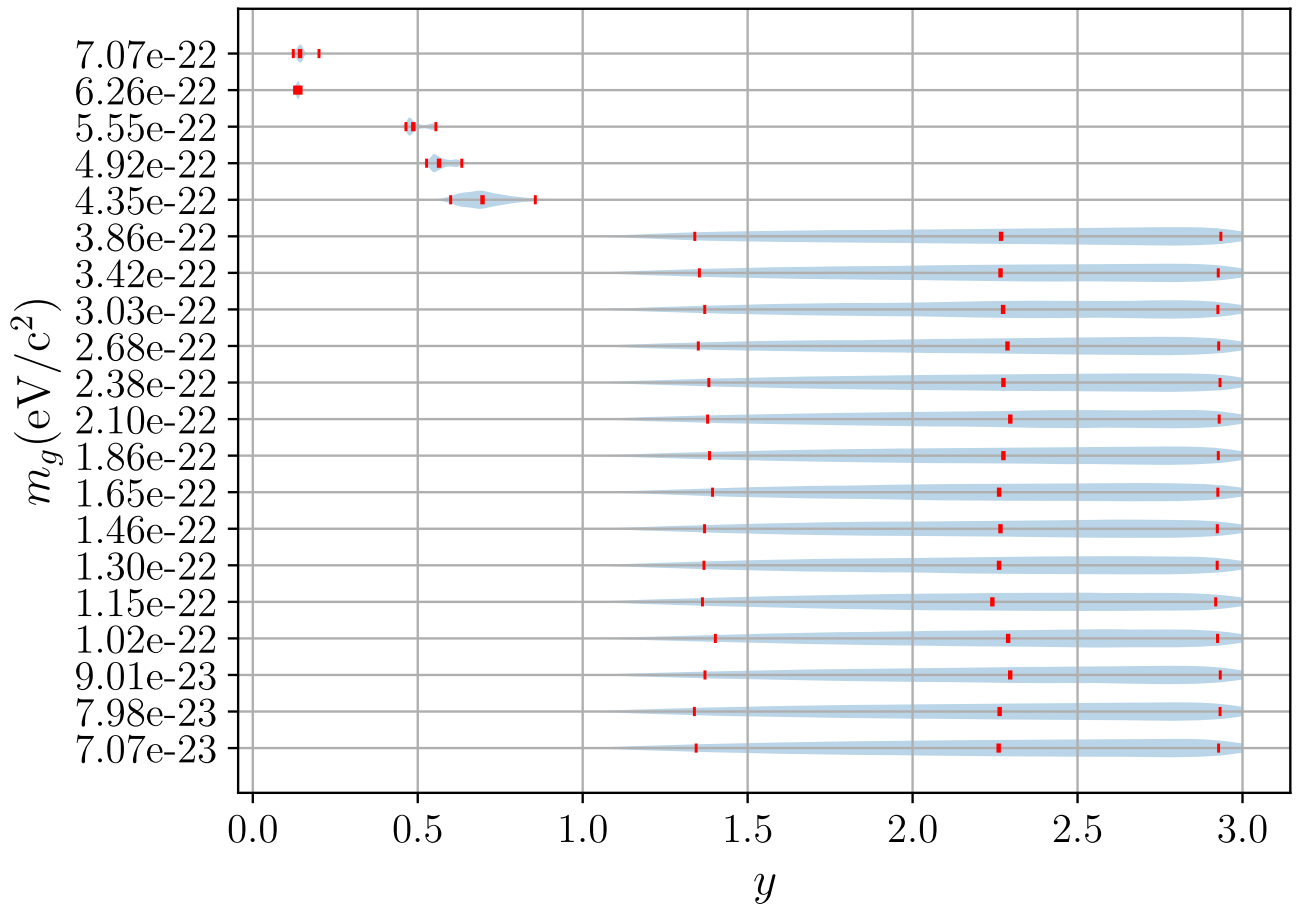


Figure 2. Recovered posteriors on the source position for the massive graviton series of injections analyzed using the SIS microlensing model. The red ticks indicate the median and 90% C.I. Lower-graviton-mass injections display support broadly across the $y > 1$ range. This is the expectation for unlensed events as in this region of parameter space only a single image is produced, resulting in a flat amplification factor that resembles an unlensed event. By contrast, the higher-mass injections display constraints toward particular values within the prior boundary and specifically below $y = 1$, i.e., the regions within which the amplification factor would be oscillatory. This is the expectation for microlensing candidates.

nonlensed waveforms. The superscripted symbol indicates the recovered number of images from the posterior from one to four. As can be seen, all cases are present. Table 4 indicates that in addition to all candidates that are flagged for investigation by the microlensing analysis, the millilensing analysis also flags two of the injections with an additional scalar polarization as false positives with a preference for a single image. Examining the posterior for the first image for the $\mathcal{A} = 0.20$ case indicates a preference for a type II image, comprising 98% of the total posterior, which is confirmed by the strong-lensing discrete type II analysis shown in Table 5. The two lowest-mass nonspinning black hole mimicker injections are additionally now found to have sufficient support for follow-up consideration in this case, with one and two image recoveries, respectively. This would be in line with the slender support for the $M = 90M_\odot$ case in the microlensing analysis, and the slender disfavoring for the $M = 80M_\odot$ case.

We perform a similar posterior consistency check to the one we did for the microlensing analysis. For nonlensed events, the posterior space for the lensing observables should reflect the respective priors, all of which were uniform. For lensed events, the posteriors should show a constraint for each of the parameter sets corresponding to images that are preferred to exist and reflect the priors for images which are not. This behavior is illustrated in Figure 3, which shows the posteriors on the effective luminosity distance of a proposed second image for the massive graviton

series of injections. As can be seen from Table 4, preference for multiple images begins with $m_g = 4.35 \times 10^{-22} \text{ eV}/c^2$, and this is the point at which the posteriors cease being uniform and display a constraint to a particular value.

Similar to the SIS analysis, the $140M_\odot$ precessing black hole mimicker waveform does not yield support for the millilensing hypothesis despite posterior support for four images similar to the $130M_\odot$ and $150M_\odot$ cases. In this instance, the source posteriors in all chains more closely resembled the local maximum found for the point-mass and unlensed analyses, and this was the case with a repeated analysis of this waveform. This may be an indication that all analyses are untrustworthy in the case of this waveform. Additionally, as the interference between the images in the millilensing case continues to produce an oscillatory amplification factor, this may be the reason that, similar to the microlensing case, the scaled BNS waveforms yield significant support for the millilensing case. The increasing number of images would also indicate that at higher masses, progressively more parameters may be used to gain a better fit to these waveforms.

Appendix D displays the rest of the effective luminosity distance and time delay posteriors for the second image of the millilensing analysis. Both of these mirror the trend of Figure 3, with broad, uninformative posteriors when a second image has no support and constrained posteriors when support exists for that image. This is repeated in the third and fourth image posteriors.

Table 4
Results of Performing the Phenomenological Millilensing Analysis Using Gravelamps on the Simulated Modified GR Events

Massive Graviton		Modified Energy Flux		Modified QNM Spectrum		BHM-P	
m_g (eV/c ²)	$\log_{10} \mathcal{B}_U^{\text{Mil}}$	a_2	$\log_{10} \mathcal{B}_U^{\text{Mil}}$	$\mathcal{Q}_f$	$\log_{10} \mathcal{B}_U^{\text{Mil}}$	M ($M_\odot$)	$\log_{10} \mathcal{B}_U^{\text{Mil}}$
7.07×10^{-23}	-0.5	2	-0.5	0.50	-0.5	80	23.1 ^{δ}
7.98×10^{-23}	-0.1	5	-0.3	0.55	-0.3	90	46.6 ^{δ}
9.01×10^{-23}	-0.3	10	-0.4	0.60	-0.4	100	69.4 ^{δ}
1.02×10^{-22}	-0.3	20	1.3 ^{δ}	0.65	-0.3	110	72.1 ^{δ}
1.15×10^{-22}	-0.1	50	22.5 ^{β}	0.70	0.1	120	28.5 ^{γ}
1.30×10^{-22}	-0.4	20 ^a	-0.3	0.75	-0.3	130	50.3 ^{β}
1.46×10^{-22}	-0.4	50 ^a	-0.3			140	-4.9
1.65×10^{-22}	-0.3	75 ^a	-0.4			150	60.2 ^{β}
1.86×10^{-22}	-0.3	100 ^a	-0.5	Scalar Polarization		BHM-NS	
2.10×10^{-22}	-0.4	125 ^a	-0.6	$\mathcal{A}$	$\log_{10} \mathcal{B}_U^{\text{Mil}}$	M ($M_\odot$)	$\log_{10} \mathcal{B}_U^{\text{Mil}}$
2.38×10^{-22}	-0.1	150 ^a	-0.3	0.05	-0.6	80	1.1 ^{ξ}
2.68×10^{-22}	-0.2	175 ^a	-0.4	0.10	-0.1	90	1.0 ^{γ}
3.03×10^{-22}	-0.3	200 ^a	0.0	0.15	-0.1	100	4.1 ^{δ}
3.42×10^{-22}	-0.3			0.20	1.3 ^{ξ}	110	9.8 ^{δ}
3.86×10^{-22}	0.0			0.25	1.7 ^{ξ}	120	15.9 ^{β}
4.35×10^{-22}	11.6 ^{γ}			0.30	-0.2	130	17.6 ^{δ}
4.92×10^{-22}	32.0 ^{γ}			0.35	-0.0	140	18.8 ^{δ}
5.55×10^{-22}	39.7 ^{γ}			0.40	-0.3	150	17.7 ^{δ}
6.26×10^{-22}	74.7 ^{β}			0.45	-0.2	160	17.9 ^{δ}
7.07×10^{-22}	76.6 ^{β}			0.50	-0.1		

Notes. Cells with superscripts indicate those events that had a $\log_{10}$ Bayes factor in favor of the millilensing case over the nonlensing case. The superscripted symbol indicates the number of images favored, with ξ corresponding to a single image, γ corresponding to two images, δ to three, and β to four.

^a Indicates the injections have been done without waveform scaling.

Table 5
Results of Performing Type II Image Searches with a Discrete Prior on the Simulated Modified GR Events

Massive Graviton		Modified Energy Flux		Modified QNM Spectrum		BHM-P	
m_g (eV/c ²)	$\log_{10} \mathcal{B}_{\text{I+III}}^{\text{II}}$	a_2	$\log_{10} \mathcal{B}_{\text{I+III}}^{\text{II}}$	$\mathcal{Q}_f$	$\log_{10} \mathcal{B}_{\text{I+III}}^{\text{II}}$	M ($M_\odot$)	$\log_{10} \mathcal{B}_{\text{I+III}}^{\text{II}}$
7.07×10^{-23}	-1.5	2	-2.3	0.50	-1.1	80	-0.1
7.98×10^{-23}	-1.2	5	-3.0	0.55	-0.9	90	-0.5
9.01×10^{-23}	-0.4	10	$-\infty$	0.60	-0.8	100	-0.1
1.02×10^{-22}	0.1	20	$-\infty$	0.65	-0.1	110	-0.1
1.15×10^{-22}	0.1	50	-3.6	0.70	-0.1	120	0.3 ^{β}
1.30×10^{-22}	0.1	20 ^a	-1.6	0.75	-0.1	130	0.2 ^{β}
1.46×10^{-22}	0.0	50 ^a	-0.6			140	-0.6
1.65×10^{-22}	0.0	75 ^a	-0.5			150	0.5 ^{β}
1.86×10^{-22}	0.0	100 ^a	-1.0	Scalar Polarization		BHM-NS	
2.10×10^{-22}	0.0	125 ^a	-0.9	$\mathcal{A}$	$\log_{10} \mathcal{B}_{\text{I+III}}^{\text{II}}$	M ($M_\odot$)	$\log_{10} \mathcal{B}_{\text{I+III}}^{\text{II}}$
2.38×10^{-22}	0.1	150 ^a	-0.6	0.05	-0.7	80	-0.1
2.68×10^{-22}	0.0	175 ^a	0.5 ^{β}	0.10	-0.2	90	0.0
3.03×10^{-22}	0.0	200 ^a	-0.1	0.15	1.0 ^{β}	100	0.0
3.42×10^{-22}	0.1			0.20	2.2 ^{β}	110	-0.3
3.86×10^{-22}	0.1			0.25	3.2 ^{β}	120	-0.2
4.35×10^{-22}	0.0			0.30	-1.0	130	-1.4
4.92×10^{-22}	0.0			0.35	-1.8	140	-1.6
5.55×10^{-22}	0.0			0.40	-1.9	150	-1.5
6.26×10^{-22}	0.0			0.45	-2.3	160	-1.3
7.07×10^{-22}	0.0			0.50	-2.8		

Notes. Cells with a β superscript indicate those analyses that would be flagged for further investigation as strongly lensed type II images. As can be seen, follow-up analyses would be performed for some events with an extra scalar polarization, one case of modified energy flux, and some cases of the precessing T. Dietrich et al. (2018a)-based black hole mimickers (BHM-P).

^a Indicates the injections have been done without waveform scaling.

The Morse-phase posteriors do not display a consistent constraint to a single value for those images that are supported, varying between support for one value, support for two of the

three values, or are unconstrained between all three. This is not necessarily unexpected for true lensing candidates and so would not be sufficient to discard the injections as such.

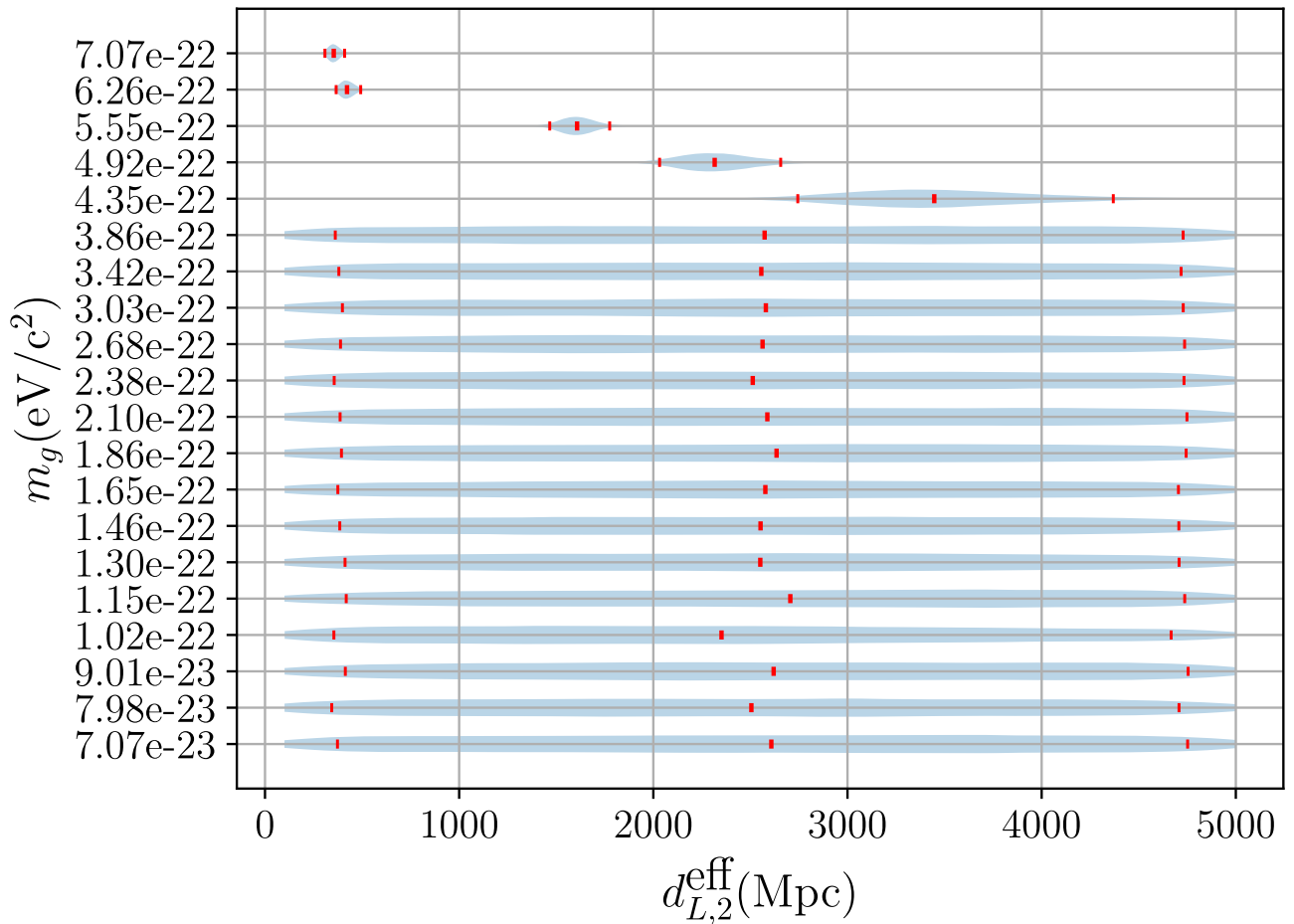


Figure 3. Recovered posteriors on the effective luminosity distance of the proposed second image for the massive graviton series of injections analyzed using the phenomenological millilensing model. The red ticks indicate the median and 90% C.I. Lower-graviton-mass injections display broad posteriors across the entire space, which is an indication of lack of support for a second image. The higher-graviton-mass injections, which favor the millilensing hypothesis with a number of images greater than two, do show constraints to particular values within this region as is expected if a millilensed signal is present.

While the luminosity distance posterior needed to be extended for the microlensing investigations into the precessing black hole mimicker waveforms, the millilensing investigations into these waveforms additionally required an increase in the maximal effective proposed luminosity distance from the baseline 5 Gpc used. This was allowed to reach 100 Gpc, with the estimates for the median effective luminosity distance varying between 5 and 12 Gpc. This reinforces the suggestion from the microlensing analyses that these waveforms would demonstrate the apparent highest magnifications. While this level of magnification may shed some doubt on these events due to it being unlikely, it is difficult to infer the absolute magnification of an unknown signal. Rather, it is only possible to infer the relative magnification between the images, which in these cases was between 1 and 10, and would not be in and of itself disqualifying when compared to expected values from current models (A. More & S. More 2022; J. Janquart et al. 2023a).

4.5. Type II Strong-lensing Analysis

When performing strongly lensed type II image searches, we typically follow a series of steps. First, we analyze the data using a discrete and uniform prior on the Morse factor, where we allow only for $n = \{0, 0.5, 1\}$, corresponding to the three physical image types. From this, we compute two Bayes

factors, comparing the type II hypothesis with the type I or type III hypotheses. This is done by reweighting the evidence under one of the image type hypotheses based on the final samples and the prior. To get the evidence for a given image type X , $\mathcal{Z}_X$, we compute a series of weights for every Morse factor sample $w_i^X = p(n_i = \text{type}_X | \text{data}) / p(n_i = \text{type}_X)$, where the numerator corresponds to the posteriors and the denominator to the prior. The evidence is then computed based on these weights as $\mathcal{Z}_X = (1/N) \sum_i w_i^X \mathcal{Z}$, where $\mathcal{Z}$ is the evidence obtained via the strong-lensing run. We then compute the Bayes factor as $\mathcal{B}_X^{\text{II}} = \mathcal{Z}_{\text{II}} / \mathcal{Z}_X$, where $X = \{\text{I, III}\}$ when comparing type I with type I or III, respectively. Finally, following the threshold used in R. Abbott et al. (2024b), an event is seen as interesting if we find $\log_{10} \mathcal{B}_I^{\text{II}} > 0.1$ or $\log_{10} \mathcal{B}_{\text{III}}^{\text{II}} > 0.1$.⁶ This is a conservative choice as one expects an error of $\mathcal{O}(0.1)$ on these Bayes factors, and therefore a value of 0.1 is still only slightly favoring the lensing hypothesis. Such a conservative choice makes sense when considering one does not want to miss a first detection of lensing.

When an event is interesting, we need to assess if the observed evidence for a type II image is genuine. Since the

⁶ In an ideal situation, the two Bayes factors should be the same since type I and type III images are degenerate (J. M. Ezquiaga et al. 2021). However, due to finite sampling effects, it can happen that the two have slightly different values. Additionally, if there are no samples supporting the type II image hypothesis, we get a null Bayes factor, hence a $\log_{10}$ value of $-\infty$.

discrete prior allows only for the physical values, there is no way of knowing if a different value of the Morse factor would be a better fit to the data. If one finds an unphysical value of the Morse factor to be preferred, the event is probably not lensed, although something different could be happening, needing further inquiries to settle whether the effect has a noise or physical origin. Thus, for interesting events, we rerun the lensing analysis, this time using a continuous prior on the Morse phase. In this work, we directly perform the two analyses, sampling over the Morse factor in a discrete and continuous way regardless of the result coming from the former analysis, to see if the second step could be used to flag anomalies in other situations than lensed candidates.

For the discrete Morse factor analyses, whether or not one detects possible strong-lensing features depends on the specific non-GR waveform considered and the value of the deviation parameter. For the massive graviton waveforms, no deviation is found. In some cases, the type II hypothesis is clearly disfavored, in line with the event having HOMs and not being a lensed type II image. For the scalar polarizations, we find either a strong support for type II images—when $\mathcal{A} \in \{0.15, 0.2, 0.25\}$ —or a strong disfavoring. While this behavior is clarified in the case of a continuous Morse factor, this would mean that, in type II image search setups, some of these deviations would be followed up with more in-depth analyses. In the case of a modified energy flux, none of the cases leads to a favoring of the type II image. For the modified QNM spectrum, none of the discrete analyses favor the presence of a type II image, and the disfavoring grows stronger as Q_f decreases. Finally, turning to the black hole mimickers, the situation is different from one model to the other. For the precessing signals following T. Dietrich et al. (2018a), we have an alternation between favoring and disfavoring for the type II hypothesis, with support for total masses of 120, 130, and $150M_\odot$. For the nonspinning M. Ujevic et al. (2022) waveform, no support for type II images is found, with only some strong disfavoring in some cases. So, for the exotic object case, follow-ups would take place or not depending on the specifics of the binary being considered and how the exotic objects’ properties affect the waveform. Table 5 gives a more quantitative overview of the results for the different non-GR/non-BBH waveforms and their corresponding parameter values.

By comparing Tables 2 and 5, one sees that the apparent hierarchy seen in microlensing is not as present in the discrete type II strong-lensing searches. The reason is clarified by the continuous runs. As will be clear in what follows, in some cases the peak value for the Morse factor posterior—which resembles a Gaussian with a relatively small width—is between the values attainable from lensing when analyzing modified GR waveforms. However, the discrete priors force the samples to be at specific values. Therefore, when deviations are detectable through the Morse-phase shift but the Morse factor corresponds to a non-GR value, the position of the peak will strongly influence the (dis-)favoring of the type II image. As an illustrative example, if the GR deviations can be mimicked with $n = 0.8$ for a given value of its deviation parameter, we will most likely disfavor the type II hypothesis. However, if for a different deviation parameter value, the mimicking happens for $n = 0.7$, then the type II image is favored. However, both deviations are expected to be seen in the continuous search as

they would manifest themselves as a Morse factor value not compatible with the lensing hypothesis.

In Table 6, we show the median values and the 90% C.I. found for the different alternative GR theories. Here more cases display evidence for non-GR values. The explanation is the one given in the previous paragraph: Some deviations correspond to unphysical Morse factors, which lead to a type II disfavoring in the discrete case. Such a case is indistinguishable from a GR event with strong HOM contribution, and therefore would probably not be followed up in usual analyses. In particular, we see that for the scalar polarization, all continuous runs would lead to questioning the lensing hypothesis based on the posterior. For the modified energy flux, except for a couple of cases, most analyses would also lead to a follow-up. For these two non-GR waveforms, it is particularly interesting to note that the smallest deviations already lead to questioning the Morse factor value, and while the recovered posteriors are not compatible with lensing, they are away enough from the GR values to warrant further scrutiny. For the modified QNMs, only the three lowest values, strongly disfavoring lensing in the discrete case, would have enough significance in their posterior to entail follow-up analyses. For the precessing black hole mimickers following T. Dietrich et al. (2018a), the events favoring type II analyses still do here. The other events do not display enough evidence to deserve a follow-up, even if a few have a distribution mildly preferring non-GR values. Finally, for the massive graviton and the nonspinning black hole mimickers following M. Ujevic et al. (2022), none of the continuous runs would lead to follow-up analyses, which is in line with the values seen in the discrete case.

Here, it is also important to stress that, in some cases, we face a disfavoring of the type II hypothesis in the discrete case which is also seen in the continuous case. However, the posterior can be such that the favored values for the Morse factor are 0 and 1, which is, again, indistinguishable from the unlensed GR case with strong HOM contribution. An illustration of such a posterior is given in Figure 4, where we overlay the GR and a non-GR case for illustration purposes. Additionally, we give a few example of intriguing posteriors, incompatible with lensing or GR, in Figure 5. One sees that some of these posteriors are still compatible with GR but have a shape such that more in-depth studies are needed. On the other hand, some posteriors simply do not include values explainable through GR and lensing. A more complete overview of the recovered posteriors for all the considered GR deviations can be found in Appendix E.

Once an event has been found with an unusual Morse factor value, the standard follow-up would be to check for noise artifacts or waveform systematics (J. Janquart et al. 2023b; R. Abbott et al. 2024b) to verify whether the observed deviation has a physical origin. What our results show is that, should an unusual Morse factor value be found—favoring or not a type II image—it would be important to verify whether GR deviations are observed or not, and which of the two hypotheses is favored.

4.6. Additional Biases on Source Parameters

The primary focus of this work is whether deviations from GR would impact single-image lensing analyses; in particular, whether any of these deviations would cause a false-positive result from the investigations. If such a false positive were mistakenly identified in a real analysis, one question would also

Table 6
Results of Performing Type II Image Searches with a Continuous Prior on the Simulated Modified GR Events

Massive Graviton		Modified Energy Flux		Modified QNM Spectrum		BHM-P	
m_g (eV/c^2)	Med. and 90% C.I.	a_2	Med. and 90% C.I.	Q_f	Med. and 90% C.I.	M ($M_\odot$)	Med. and 90% C.I.
7.07×10^{-23}	$0.17^{+0.81}_{-0.16}$	2	$0.82^{+0.12\gamma}_{-0.14}$	0.50	$0.82^{+0.14\gamma}_{-0.21}$	80	$0.45^{+0.50}_{-0.40}$
7.98×10^{-23}	$0.24^{+0.73}_{-0.22}$	5	$0.82^{+0.11\gamma}_{-0.11}$	0.55	$0.81^{+0.15\gamma}_{-0.41}$	90	$0.76^{+0.20}_{-0.71}$
9.01×10^{-23}	$0.41^{+0.53}_{-0.37}$	10	$0.83^{+0.11\gamma}_{-0.12}$	0.60	$0.79^{+0.18\gamma}_{-0.70}$	100	$0.40^{+0.46}_{-0.44}$
1.02×10^{-22}	$0.47^{+0.47}_{-0.42}$	20	$0.83^{+0.09\gamma}_{-0.09}$	0.65	$0.61^{+0.35}_{-0.55}$	110	$0.48^{+0.47}_{-0.43}$
1.15×10^{-22}	$0.50^{+0.45}_{-0.44}$	50	$0.82^{+0.07\gamma}_{-0.07}$	0.70	$0.51^{+0.44}_{-0.46}$	120	$0.36^{+0.52\beta}_{-0.30}$
1.30×10^{-22}	$0.50^{+0.45}_{-0.45}$	20 ^a	$0.81^{+0.14\gamma}_{-0.15}$	0.75	$0.49^{+0.45}_{-0.45}$	130	$0.65^{+0.29\beta}_{-0.56}$
1.46×10^{-22}	$0.49^{+0.46}_{-0.44}$	50 ^a	$0.78^{+0.15\gamma}_{-0.17}$			140	$0.34^{+0.63}_{-0.30}$
1.65×10^{-22}	$0.51^{+0.44}_{-0.46}$	75 ^a	$0.76^{+0.16\gamma}_{-0.17}$			150	$0.59^{+0.32\beta}_{-0.46}$
1.86×10^{-22}	$0.51^{+0.44}_{-0.46}$	100 ^a	$0.83^{+0.13\gamma}_{-0.20}$	Scalar Polarization		BHM-NS	
2.10×10^{-22}	$0.50^{+0.44}_{-0.45}$	125 ^a	$0.80^{+0.17\gamma}_{-0.65}$	$\mathcal{A}$	Med. and 90% C.I.	M ($M_\odot$)	Med. and 90% C.I.
2.38×10^{-22}	$0.49^{+0.45}_{-0.44}$	150 ^a	$0.73^{+0.24}_{-0.67}$	0.05	$0.78^{+0.13\gamma}_{-0.14}$	80	$0.57^{+0.38}_{-0.52}$
2.68×10^{-22}	$0.50^{+0.45}_{-0.45}$	175 ^a	$0.77^{+0.17\beta}_{-0.51}$	0.10	$0.76^{+0.13\gamma}_{-0.14}$	90	$0.50^{+0.45}_{-0.45}$
3.03×10^{-22}	$0.50^{+0.44}_{-0.45}$	200 ^a	$0.58^{+0.37}_{-0.53}$	0.15	$0.73^{+0.13\beta}_{-0.14}$	100	$0.40^{+0.46}_{-0.44}$
3.42×10^{-22}	$0.49^{+0.46}_{-0.44}$			0.20	$0.70^{+0.12\beta}_{-0.13}$	110	$0.67^{+0.29}_{-0.64}$
3.86×10^{-22}	$0.50^{+0.45}_{-0.45}$			0.25	$0.68^{+0.12\beta}_{-0.12}$	120	$0.47^{+0.49}_{-0.42}$
4.35×10^{-22}	$0.50^{+0.45}_{-0.44}$			0.30	$0.77^{+0.10\gamma}_{-0.10}$	130	$0.58^{+0.37}_{-0.53}$
4.92×10^{-22}	$0.49^{+0.46}_{-0.44}$			0.35	$0.77^{+0.10\gamma}_{-0.09}$	140	$0.81^{+0.17}_{-0.79}$
5.55×10^{-22}	$0.49^{+0.46}_{-0.44}$			0.40	$0.78^{+0.09\gamma}_{-0.09}$	150	$0.81^{+0.16}_{-0.77}$
6.26×10^{-22}	$0.45^{+0.49}_{-0.41}$			0.45	$0.78^{+0.09\gamma}_{-0.09}$	160	$0.77^{+0.21}_{-0.75}$
7.07×10^{-22}	$0.49^{+0.46}_{-0.44}$			0.50	$0.78^{+0.08\gamma}_{-0.08}$		

Notes. The values given are the median values with the 90% C.I. Cells with a superscripted β are those that would favor the type II hypothesis, and therefore lead to follow-up analyses. Those with a superscripted γ are the cases that show deviation from the GR and lensing values but that are not flagged by the discrete analyses. As can be seen, follow-up analyses would be performed for all the events with a scalar polarization, most of the events with a modified energy flux, a few with modified QNMs, and the same events as for the discrete case for the precessing T. Dietrich et al. (2018a)-based black hole mimickers.

^a Indicates the injections have been done without waveform scaling.

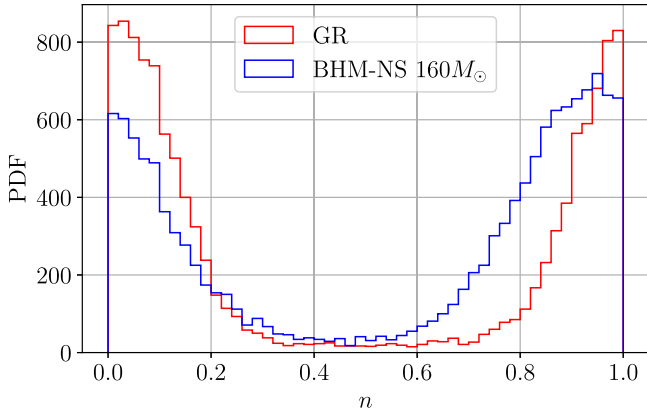


Figure 4. Example posterior for the continuous Morse factor analysis where one could not distinguish whether the event has deviation or not, since the posterior is compatible with having a Morse factor of 0 or 1, which corresponds to an unlensed GR event. This example is the posterior for the nonspinning M. Ujevic et al. (2022)-based black hole mimickers with $M = 160M_\odot$ (in blue). Overlaid in red is the results of the GR analysis. One can clearly see the two distributions are close and present the same features, meaning we cannot claim deviation from GR in the BHM-NS case.

be what additional biases are added to the source parameters due to this misinterpretation of the data. In particular, one obvious question is whether the parameters estimated by the lensing analyses are significantly degraded compared to their GR estimates.

Table 7 displays the comparison of the median and 90% C.I. boundaries between the lensed and unlensed estimates for the

chirp mass, mass ratio, and luminosity distance for the point-mass microlensing and unlensed investigations. These are the parameters that are most likely to be affected in the microlensing analysis.

The chirp mass and mass ratio estimates in the massive graviton and modified energy flux cases do not display significant signs of consistent bias as a result of the microlensing investigation. In the black hole mimicker cases, the lensed investigations are typically larger, with one or two exceptions, particularly in the nonspinning $140M_\odot$ case—in line with the peculiarities noted about the analysis of this event.

The luminosity distance estimates are significantly more affected, as would be expected, given that the posteriors demonstrate constraints in the lens parameter posteriors. In all cases, the lensed estimate is significantly higher than the unlensed case, with the estimates in the precessing black hole mimicker case comparatively higher than those of the other deviation types, again indicating that these have the highest apparent magnification.

The conclusion of this exercise in the microlensing analysis therefore is that, in addition to the false identification of the signal as a lensed signal, the source parameters under this false model would be worse in all cases than that of the already incorrect GR model.

For the type II image searches, extra bias could also be present when including lensing. However, it is expected to be less important than for the microlensing case since the extra Morse factor can only lead to some additional dephasing, while microlensing gives rise to frequency-dependent effects,

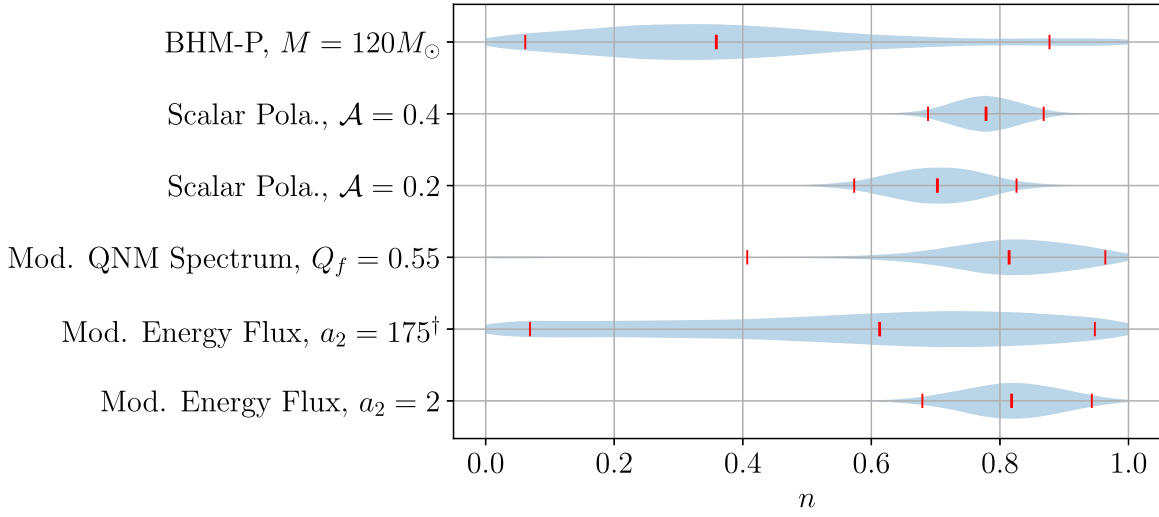


Figure 5. Examples of a few posteriors illustrating the continuous Morse factor recovery for cases in which extra follow-up would be required. The red ticks indicate the median and 90% C.I. Some of these posteriors simply favor values that are not compatible with GR and possibly type II, while others have Morse factor values excluding lensing or GR.

Table 7

Estimates on the Chirp Mass $\mathcal{M}_c$, Mass Ratio q , and Luminosity Distance d_L for the Point-mass and Nonlensed Analyses of Those Events That Would Be Identified as False Positives Under the Point-mass Microlensing Hypothesis

Injection	Nonlensed Estimate			Lensed Estimate		
	$\mathcal{M}_c$ ($M_\odot$)	q	d_L (Mpc)	$\mathcal{M}_c$ ($M_\odot$)	q	d_L (Mpc)
Massive Graviton						
$m_g = 4.35 \times 10^{-22} \text{ eV}/c^2$	$26.31^{+0.23}_{-0.03}$	$0.166^{+0.004}_{-0.003}$	1118^{+59}_{-51}	$26.29^{+0.20}_{-0.20}$	$0.167^{+0.007}_{-0.003}$	1338^{+137}_{-101}
$m_g = 4.92 \times 10^{-22} \text{ eV}/c^2$	$23.81^{+0.49}_{-0.12}$	$0.182^{+0.008}_{-0.038}$	1134^{+64}_{-67}	$24.25^{+0.12}_{-0.14}$	$0.144^{+0.004}_{-0.002}$	1644^{+279}_{-194}
$m_g = 5.55 \times 10^{-22} \text{ eV}/c^2$	$21.74^{+0.12}_{-0.13}$	$0.139^{+0.008}_{-0.005}$	1101^{+77}_{-62}	$21.65^{+0.09}_{-0.10}$	$0.137^{+0.004}_{-0.002}$	1930^{+62}_{-104}
$m_g = 6.26 \times 10^{-22} \text{ eV}/c^2$	$16.97^{+0.16}_{-0.19}$	$0.25^{+0.02}_{-0.01}$	279^{+37}_{-31}	$16.23^{+0.32}_{-0.25}$	$0.83^{+0.01}_{-0.01}$	544^{+84}_{-69}
$m_g = 7.07 \times 10^{-22} \text{ eV}/c^2$	$14.10^{+0.17}_{-0.17}$	$0.85^{+0.02}_{-0.02}$	255^{+43}_{-33}	$14.08^{+0.17}_{-0.18}$	$0.83^{+0.02}_{-0.02}$	463^{+79}_{-63}
Modified Energy Flux						
$a_2 = 50$	$31.25^{+0.76}_{-0.88}$	$0.51^{+0.07}_{-0.06}$	502^{+53}_{-56}	$31.89^{+0.73}_{-0.90}$	$0.50^{+0.06}_{-0.05}$	536^{+49}_{-48}
BHM-P						
$M = 80M_\odot$	$35.62^{+2.53}_{-1.64}$	$0.31^{+0.09}_{-0.04}$	923^{+135}_{-136}	$35.46^{+2.44}_{-1.61}$	$0.31^{+0.08}_{-0.04}$	926^{+734}_{-167}
$M = 90M_\odot$	$41.91^{+2.99}_{-2.40}$	$0.40^{+0.08}_{-0.05}$	945^{+140}_{-136}	$41.02^{+3.05}_{-2.21}$	$0.39^{+0.07}_{-0.05}$	1342^{+850}_{-487}
$M = 100M_\odot$	$47.73^{+3.02}_{-2.34}$	$0.50^{+0.07}_{-0.05}$	927^{+149}_{-151}	$50.62^{+7.25}_{-5.06}$	$0.57^{+0.19}_{-0.11}$	2152^{+796}_{-918}
$M = 110M_\odot$	$52.63^{+4.05}_{-2.71}$	$0.60^{+0.10}_{-0.06}$	841^{+170}_{-163}	$60.22^{+6.61}_{-7.89}$	$0.79^{+0.18}_{-0.20}$	2342^{+903}_{-661}
$M = 120M_\odot$	$57.37^{+11.40}_{-4.72}$	$0.70^{+0.29}_{-0.11}$	863^{+283}_{-176}	$65.67^{+3.11}_{-5.89}$	$0.93^{+0.07}_{-0.17}$	2566^{+949}_{-796}
$M = 130M_\odot$	$55.94^{+6.92}_{-4.04}$	$0.68^{+0.17}_{-0.10}$	994^{+223}_{-253}	$66.85^{+2.15}_{-3.81}$	$0.96^{+0.03}_{-0.09}$	2537^{+757}_{-713}
$M = 140M_\odot$	$54.80^{+4.86}_{-3.41}$	$0.65^{+0.10}_{-0.08}$	1073^{+226}_{-218}	$67.07^{+2.00}_{-9.86}$	$0.97^{+0.03}_{-0.20}$	2400^{+598}_{-918}
$M = 150M_\odot$	$53.72^{+4.03}_{-3.52}$	$0.62^{+0.10}_{-0.08}$	1100^{+277}_{-237}	$59.47^{+6.16}_{-5.05}$	$0.86^{+0.13}_{-0.19}$	2347^{+561}_{-650}
$M = 160M_\odot$	$50.52^{+4.63}_{-3.53}$	$0.55^{+0.11}_{-0.08}$	1053^{+287}_{-215}	$60.24^{+6.29}_{-6.34}$	$0.85^{+0.13}_{-0.20}$	2585^{+649}_{-561}
BHM-NS						
$M = 100M_\odot$	$48.02^{+1.97}_{-1.76}$	$0.50^{+0.04}_{-0.03}$	1021^{+90}_{-77}	$48.78^{+2.41}_{-2.52}$	$0.54^{+0.07}_{-0.06}$	1735^{+158}_{-331}
$M = 110M_\odot$	$47.74^{+1.77}_{-1.72}$	$0.50^{+0.04}_{-0.03}$	950^{+80}_{-72}	$48.78^{+1.89}_{-2.06}$	$0.54^{+0.06}_{-0.05}$	1612^{+136}_{-164}
$M = 120M_\odot$	$49.76^{+1.80}_{-1.70}$	$0.54^{+0.04}_{-0.04}$	944^{+85}_{-74}	$51.79^{+2.20}_{-2.06}$	$0.60^{+0.08}_{-0.05}$	1658^{+154}_{-165}
$M = 130M_\odot$	$49.63^{+1.60}_{-1.55}$	$0.53^{+0.04}_{-0.03}$	799^{+70}_{-60}	$51.58^{+1.80}_{-1.69}$	$0.58^{+0.05}_{-0.04}$	1395^{+112}_{-102}
$M = 140M_\odot$	$33.54^{+0.18}_{-0.46}$	$0.255^{+0.003}_{-0.004}$	294^{+26}_{-23}	$33.45^{+0.13}_{-0.20}$	$0.253^{+0.001}_{-0.003}$	397^{+33}_{-29}
$M = 150M_\odot$	$35.10^{+0.62}_{-0.78}$	$0.28^{+0.01}_{-0.01}$	376^{+46}_{-38}	$45.67^{+1.84}_{-1.46}$	$0.45^{+0.04}_{-0.03}$	761^{+108}_{-94}

Note. Estimates given are the median of the posterior, with the error bars indicating the boundaries of the 90% C.I. For easy reference, the true values are $\mathcal{M}_c = 31.11M_\odot$, $q = 0.8$ for the massive graviton and modified energy injections, $\mathcal{M}_c = 0.4328M$, $q = 0.82$ for the BHM-P injections, and $\mathcal{M}_c = 0.4154M$, $q = 0.57$ for the BHM-NS injections. The true value of d_L is 452 Mpc for all injections.

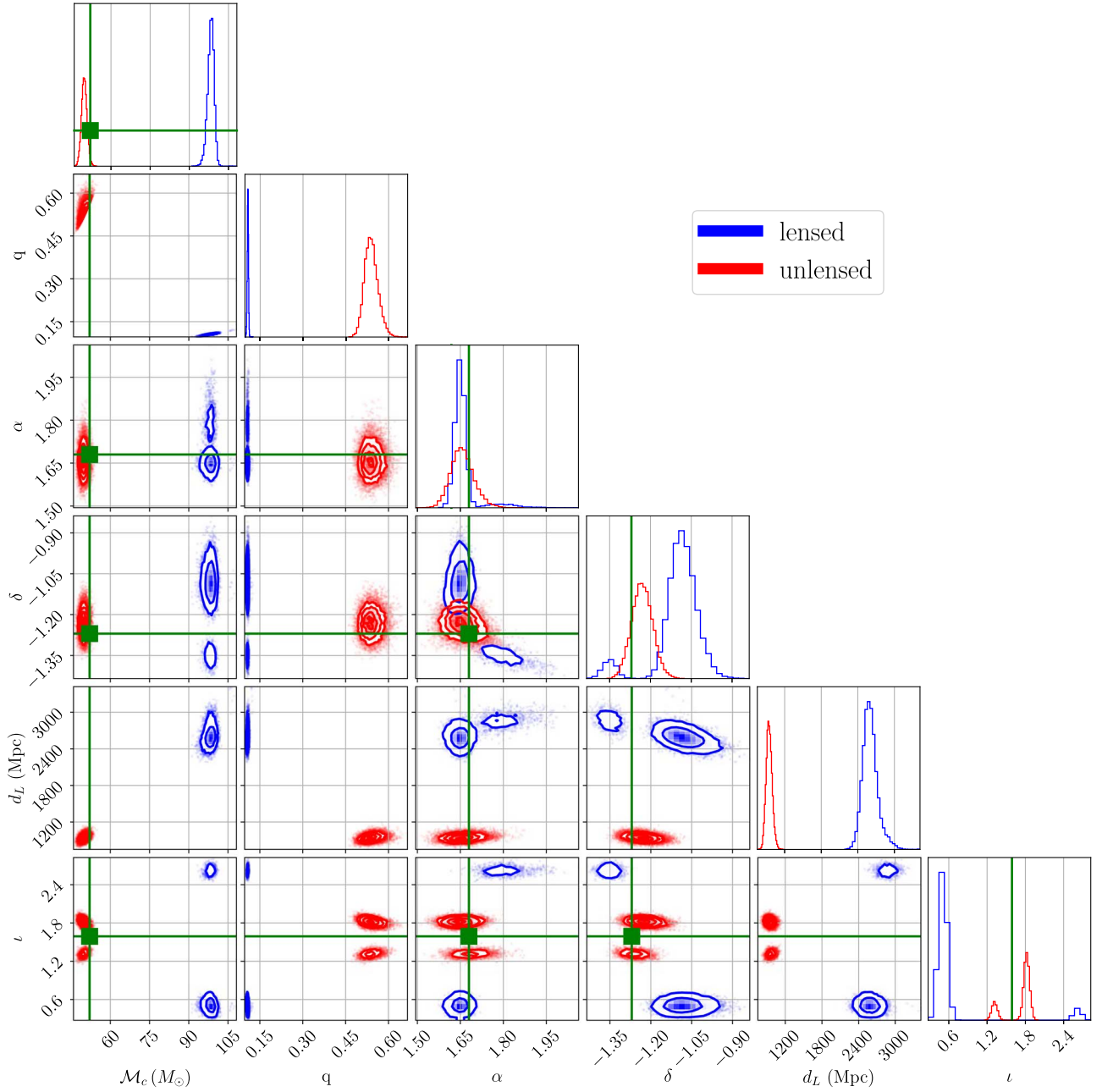


Figure 6. Representative corner plot for the parameter recovery when analyzing the $M = 120M_{\odot}$ precessing black hole mimickers waveform following T. Dietrich et al. (2018a) and searching for type II images with continuous Morse factor values. The parameters shown, from left to right, are the chirp mass, the mass ratio, the R. A., the decl., the luminosity distance, and the inclination. Both the lensed and the unlensed analyses do not recover the true parameters (in green), though the unlensed analysis is quite close for the sky location, but they are biased differently. (The true values of $q = 0.82$ and $d_L = 452$ Mpc are off of the plotted regions.)

tunable to match the effect of non-GR effects. This is confirmed by our results. For all the cases leading to follow-up analyses in the continuous Morse factor scenario except for the precessing black hole mimickers following T. Dietrich et al. (2018a), the type II results present no significant bias compared to the unlensed analysis. In the worst-case scenario, some slight shift in the peak value can be observed, but the posterior still covers the same region of the parameter space for all parameters.

Looking more into the three triggers related to black hole mimickers, we find that their posteriors are strongly biased for all the parameters. In particular, the lensed case leads to higher and more asymmetric masses, a higher spin for the first object,

and a larger luminosity distance when compared to the unlensed analysis. Moreover, the sky location that is found is completely changed. An illustrative corner plot for the $M = 120M_{\odot}$ case is given in Figure 6, where one also sees the true values. Because these black hole mimicker waveforms are significantly different from BBH waveforms, particularly in the post-inspiral region, changes in the Morse factor can impact the posterior quite a bit in some cases (J. Janquart et al. 2021b; Y. Wang et al. 2021; A. Vijaykumar et al. 2023). These changes lead to a result which better matches the injected waveforms, as confirmed by the higher SNR recovered for the posterior samples corresponding to the lensed case compared to the unlensed case, as represented in Figure 7. We see that the

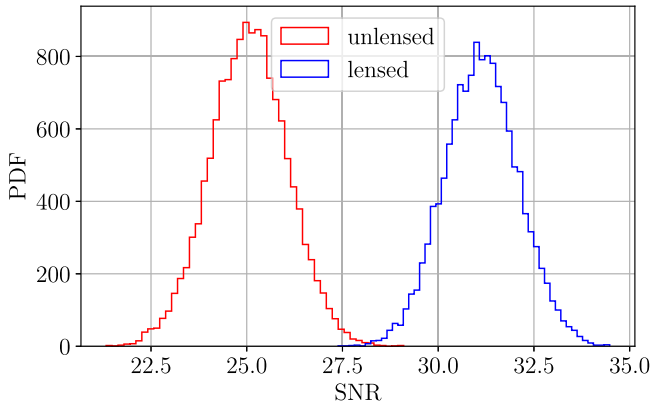


Figure 7. Representation of the recovered SNR distribution for the lensed and unlensed recovery of the $M = 120M_{\odot}$ BHM-P case. One sees that the SNR is higher for the samples corresponding to the lensed recovery, meaning that the extra phase shift offered in that case, combined with the biasing of the other parameters, better fits the data.

lensed case recovers an SNR of $31.1^{+1.6}_{-1.6}$, while the unlensed case finds $25.1^{+1.7}_{-1.7}$.

5. Conclusion

In the coming years, with continued improvements to the sensitivity of the current ground-based detector network (B. P. Abbott et al. 2020), there is a reasonable likelihood that a lensed GW signal will be observed (see, e.g., S.-S. Li et al. 2018). Should candidates for such a detection be identified, it will be important to identify any potential sources of false positivity that may mimic the signatures of lensing. In this work, we have analyzed phenomenological deviations from GR in order to understand if any such deviations would in and of themselves cause such false positivity in those searches for lensing which examine one signal at a time—the microlensing, millilensing, and type II strong-lensing searches.

To perform this investigation, we have taken several possible deviations from GR—specifically, a massive graviton, modifications to the energy flux and QNM spectrum, and an additional scalar polarization mode, as well as the possibility of exotic compact objects—and performed both the microlensing and strong-lensing type II image search as found in R. Abbott et al. (2024b) as well as the follow-up analyses performed in J. Janquart et al. (2023b) to determine to what extent any false-positive candidates would be identified, and if any could be ruled out by the follow-up analyses. Here, we have used the predicted noise curves for O4. Our results indicate that in the microlensing, millilensing, and type II image search cases there are a number of potential scenarios that could result in a false positive.

In the initial isolated point-mass microlensing analysis, a graviton with mass $\geq 4.36 \times 10^{-22} \text{ eV}/c^2$ would cause a sufficient preference for the microlensing hypothesis to warrant further follow-up investigation. This further investigation, in the form of SIS microlensing analysis and millilensing analysis, would similarly find significant support for their respective models, meaning that these would be identified as lensing candidates. However, GW observations place constraints on the mass of the graviton to well below this threshold ($< 1.27 \times 10^{-23} \text{ eV}/c^2$ at 90% C.I.). Beneath this threshold, support for the unlensed waveform was consistent. On the contrary, for type II strongly lensed images, dispersion due to a

massive graviton does not lead to support for lensing, and no false positivity is expected in this case.

Modifications to the energy flux with $a_2 = 50$ were able to produce a false positive in both the initial point-mass microlensing search and the follow-up SIS microlensing analysis. The follow-up millilensing analysis would also identify the modification with $a_2 = 20$ as a slenderly favored candidate. Such a deviation also led to strong disfavoring of type II images in the discrete type II images searches. This is indistinguishable from the results for a GR signal with significant HOM content, and would not be followed up. Still, when investigating the continuous results, we find a deviating value for the Morse factors in nearly all cases, and in particular for the smallest deviation parameters in this work. Those are not compatible with lensing Morse factors but would be sufficiently odd to warrant extra scrutiny.

Neither modifications to the QNM spectrum nor the addition of an extra scalar polarization mode yielded any situations in which the initial microlensing investigation would yield a sufficiently confident false positive to warrant additional investigation, with the unlensed model confidently preferred in all cases. However, the $\mathcal{A} = 0.20$ and $\mathcal{A} = 0.25$ additional scalar polarization cases would be slenderly preferred in a millilensing analysis as single-image type II candidates, which is consistent with the findings of the strong-lensing type II analysis in these cases. Additionally, the modified QNM spectrum leads to Morse factor values incompatible with GR or lensing but sufficiently constrained to require additional investigations. For the scalar polarization, values not found as type II images in the discrete analyses lead to a strong disfavoring, and their continuous equivalents show Morse factor posteriors that are incompatible with lensing. Again, the posteriors are sufficiently well constrained to entail extra investigations. We stress here that values of $\mathcal{A} = 0.15, 0.20, 0.25$ lead to false positives in the discrete and continuous cases and also in the millilensing analyses. Such events could lead to false identifications of lensing.

Binaries of exotic compact objects, for which we used the proxy of scaling BNS waveforms to BBH masses, resulted in significant false positivity in both the precessing and nonspinning cases. These waveforms displayed preference for all three of the point-mass, SIS, and millilensing models compared with a standard unlensed case in the majority of the analyses performed. As such objects have not been conclusively ruled out, microlensing and millilensing candidates would need to be examined carefully for any signatures of being such objects. One particular indicator in the case of the nonspinning analysis might be the significant chain disagreement when parameter estimation was performed on these waveforms. For strongly lensed type II images, the picture is somewhat different as only a few cases lead to a preference of type II images, corresponding to the precessing case using the T. Dietrich et al. (2018b) waveform with $M \in \{120, 130, 150\}M_{\odot}$. Still, those events would require extra investigations as both the discrete and continuous runs lead to a preference for the type II image hypothesis.

We also examined the recovered posteriors of the lensed waveforms compared with the unlensed waveforms to note whether false identification as a lensed waveform would further degrade the source parameter estimates compared with the already incorrect GR waveform. In the case of the microlensing analysis, this would have significant consequences on the

luminosity distance estimates, which with very few exceptions were considerably worse than the unlensed estimates. Similarly, for the type II image analyses, significant differences between the lensed and the unlensed recovery can happen, although large deviations are only present in a few scenarios.

The conclusions of this work therefore are that there are scenarios in which a deviation from GR would be the source of a false-positive result in single-event lensing searches, and therefore any candidates for detection should be investigated for potential deviations from GR as part of the identification process for lensed events. Should a misidentification of a GR deviation as a lensed event take place, the source parameter estimates would display further degradation from the truth as compared with the unlensed GR analysis. Additionally, the continuous Morse factor lensing analysis (which has the same phase modification as one of the modified dispersion testing GR analyses; J. M. Ezquiaga et al. 2022) is quite sensitive to certain GR deviations.

It will therefore be important to know if the tests of GR being applied to GW data would be able to identify the GR deviations that can be misidentified as lensed events, and whether the ultimate preference would be for the GR deviation rather than the lensing hypothesis. On the first question, it is likely that the GR deviation would be identified in at least some cases, since a smaller graviton mass than the minimum that gives significant evidence for lensing was identified as such with very high confidence in a similar injection study in N. K. Johnson-McDaniel et al. (2022). However, this is not guaranteed to be the case for all deviations, so it will be important to study the response of the lensing and testing GR analyses to waveforms in various alternative theories, once these can be calculated sufficiently accurately, and to first consider phenomenological GR deviations of the type considered here. Indeed, the LVK Testing General Relativity group is also studying the response of the tests of GR the group applies to the types of waveforms considered here.

Acknowledgments

The authors thank the Testing General Relativity group of the LIGO-Virgo-KAGRA Collaborations for useful discussions and in particular N. V. Krishnendu for internal review of the manuscript.

We also thank the authors of T. Dietrich et al. (2018a) and M. Ujevic et al. (2022) for making their waveforms available, and Archisman Ghosh for developing the code used to create the injections.

M.W. is supported by the Science and Technology Facilities Council (STFC) of the United Kingdom Doctoral Training grant No. ST/W507477/1. J.J. is supported by the research program of the Netherlands Organization for Scientific Research (NWO). N.K.J.-M. is supported by NSF grant AST-2205920.

The authors are grateful for computational resources provided by the LIGO Laboratory and supported by NSF grants PHY-0757058 and PHY-0823459. 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 acknowledge the Python community (G. van Rossum 1995) and the core set of tools that enabled this work including NumPy (C. R. Harris et al. 2020), SciPy (P. Virtanen et al. 2020), AstroPy (A. M. Price-Whelan et al. 2022),

Matplotlib (J. D. Hunter 2007), Corner (D. Foreman-Mackey 2016), Bilby-pipe (I. M. Romero-Shaw et al. 2020), and Asimov (D. Williams et al. 2023).

Appendix A

Details of the Construction of Modified Waveforms and Illustrations of the Waveforms

A.1. Scalings to Obtain Other Kerr–Newman QNMs

We now describe how we obtain approximate (complex) frequencies for the Kerr–Newman $(2, 1, n)$, $(3, 2, n)$, $(4, 4, n)$, and $(5, 5, n)$ ($n \in \{0, 1\}$) QNMs from the results for the $(2, 2, 0)$ and $(3, 3, 0)$ QNMs, for which we have tabulated results from Ó. J. C. Dias et al. (2022b). The authors also provide results for the $(2, 2, 1)$ mode, but we will not end up using this, except as a cross-check. For $n = 1$, we only need the imaginary part (equivalently the damping time), since this is all that is used in TEOBResumS-v3-GIOTTO.

We start from the eikonal approximation for the complex frequencies, which is valid for large ℓ , using the version from H. Yang et al. (2012) and P.-C. Li et al. (2021):

$$\omega_{\ell mn}^{\text{eik}} = \left(\ell + \frac{1}{2}\right)\omega_{\text{orb}} + m\omega_{\text{prec}} - i\left(n + \frac{1}{2}\right)\gamma_{\text{L}}. \quad (\text{A1})$$

Here, ω_{orb} and ω_{prec} are frequencies corresponding to the spherical photon orbit and γ_{L} is the Lyapunov exponent of that orbit. We take all of these to be free parameters, so we will refer to our final result as a *tuned eikonal* expression. Specifically, we determine ω_{orb} and ω_{prec} using the real parts of the $(2, 2, 0)$ and $(3, 3, 0)$ QNMs (denoted with a superscript “Re”), giving

$$\omega_{\text{orb}} = 6\omega_{220}^{\text{Re}} - 4\omega_{330}^{\text{Re}}, \quad (\text{A2a})$$

$$\omega_{\text{prec}} = -7\omega_{220}^{\text{Re}} + 5\omega_{330}^{\text{Re}}. \quad (\text{A2b})$$

We apply this expression to obtain the real parts of the $(4, 4, 0)$ and $(5, 5, 0)$ QNMs, where we find that it gives good accuracy for Kerr, particularly for larger spins; see Figure 8. In the application to Kerr–Newman, we compute the ratio between the Kerr–Newman and Kerr expressions and use this to scale the Kerr QNM frequencies in TEOBResumS-v3-GIOTTO, so that this reduces to the Kerr expressions used in TEOBResumS-v3-GIOTTO for zero charge.

For the real parts of the $(2, 1, 0)$ and $(3, 2, 0)$ QNMs, we find that the tuned eikonal expression is not as accurate as writing

$$\omega_{\ell, \ell-1, 0}^{\text{Re, spinning}} \simeq \omega_{\ell \ell 0}^{\text{Re, nonspinning}} + \frac{\ell-1}{\ell} \times (\omega_{\ell \ell 0}^{\text{Re, spinning}} - \omega_{\ell \ell 0}^{\text{Re, nonspinning}}). \quad (\text{A3})$$

We obtained this expression by noting that there is only a difference between the QNM frequencies for different m modes due to spin, and then using the scaling of the frequency with m from the eikonal expression, which is the same as that for the linear-in-spin corrections (see, e.g., P. Pani et al. 2013), so we refer to this as a *spin-expansion-like* expression. As before, we scale the Kerr frequencies by the ratio of this expression for Kerr–Newman and Kerr to recover the Kerr frequencies exactly in the limit of zero charge. We illustrate the accuracies of this expression for Kerr in Figure 8, using the data from E. Berti et al. (2006, 2009).

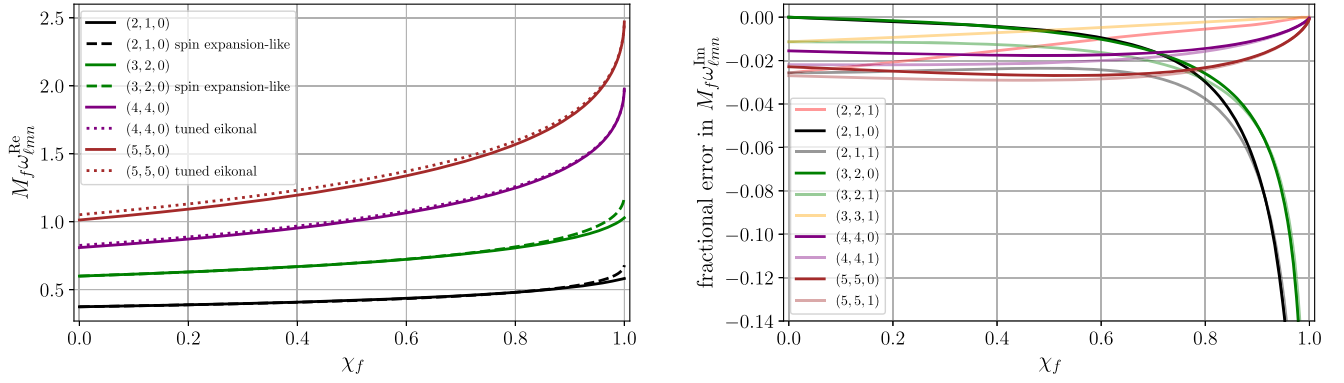


Figure 8. Illustrations for Kerr of the approximations we use for the (complex) QNM frequencies to obtain our scalings for the modes for which we do not have Kerr–Newman results. Left panel: the spin-expansion-like expression in Equation (A3) we use for the (2, 1, 0) and (3, 2, 0) QNMs and the tuned eikonal expression from Equations (A1) and (A2) we use for the (4, 4, 0) and (5, 5, 0) QNMs. Right panel: the fractional errors in the tuned eikonal expression for the imaginary QNM frequencies. Of course, as discussed in the text, we apply all these corrections to the Kerr frequencies in such a way that they reduce to Kerr exactly for zero charge. We just show their predictions for Kerr as an indication of how well we expect them to reproduce the Kerr–Newman results.

For the imaginary parts of all the modes, we take $\gamma_L/2$ to be given by the (2, 2, 0) inverse damping time for the $\ell = 2$ modes and by the (3, 3, 0) inverse damping time for all other modes. We find that for the final spins and charges we consider, there is a small difference ($<2\%$) between the values of γ_L that are inferred from the (2, 2, 0) and (3, 3, 0) modes and an even smaller difference ($<0.6\%$) between the values that would be inferred from the (2, 2, 0) and (2, 2, 1) modes, i.e., if one takes the (2, 2, 1) damping time to give $3\gamma_L/2$. These expressions give accuracies of better than 3% for Kerr for the final spin values we consider, as illustrated in Figure 8. (However, recall that since we are using these expressions to obtain a phenomenological non-GR waveform, the accuracies are not that important, since the use of the Kerr–Newman values is purely a way to obtain reasonable scalings for all the necessary QNM frequencies.) We then scale the Kerr values of the damping times used in `TEOBResumS-v3-GIOTTO` by the ratio of the Kerr–Newman value to the Kerr value for γ_L .

A.2. Illustrations of Modified EOB Final Mass and Spin

See Figure 9 for the final mass and spin one obtains self-consistently with the changes to the QNM spectrum described in the previous subsection as well as the modifications to the flux. We also compare with the final mass and spin given by the precessing NR surrogate model `NRSur7dq4` (V. Varma et al. 2019). The small differences between the surrogate values and the self-consistent GR value are likely an indication of the accuracy of the `TEOBResumS-v3-GIOTTO` model, since the surrogate model’s errors are smaller than these differences for most cases. Note that while the final spin decreases with increasing charge, the final black hole is always closer to extremal (and thus the QNMs are less damped) as the charge increases.

A.3. Expressions for Scalar Modes

Here, we give the explicit expressions we use to obtain the scalar modes we add to obtain the scalar-tensor waveforms. Recall that we take the sensitivities to be equal, so $\mathcal{S}_- = 0$ in the notation of L. Bernard et al. (2022), and the amplitude of the scalar waveform depends on a single combination of scalar-

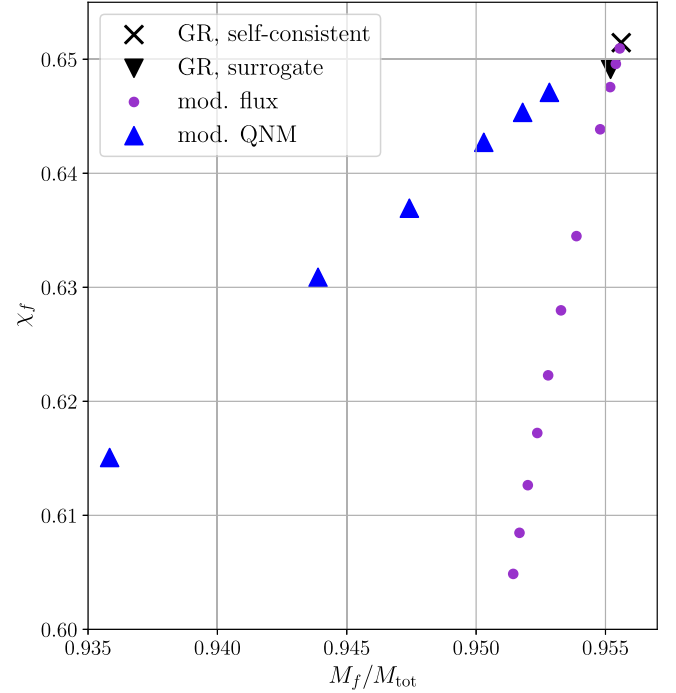


Figure 9. The self-consistent final mass and spin used for the waveforms with the modified energy flux and QNM spectrum, as well as the self-consistent results obtained for the underlying GR `TEOBResumS-v3-GIOTTO` model and the NR surrogate prediction. The modified flux points correspond to $a_2 \in \{2, 5, 10, 20, 50, 75, 100, 125, 150, 175, 200\}$ and the modified QNM points correspond to $Q_f \in \{0.5, 0.55, 0.6, 0.65, 0.7, 0.75\}$. In both cases the points move away from the GR value with increasing values of the deviation.

tensor theory parameters:

$$a_{\text{scalar}} := \frac{\sqrt{\alpha} \zeta \mathcal{S}_+}{1 - \zeta}. \quad (\text{A4})$$

Here, the parameters on the right-hand side are defined in Table I of L. Bernard et al. (2022). Using just the leading-order expression, Equation (A4) then allows the scalar waveforms to be written in terms of their tensor waveform counterparts. Absorbing all constant quantities into $\bar{a}_{\text{scalar}}$, this may be

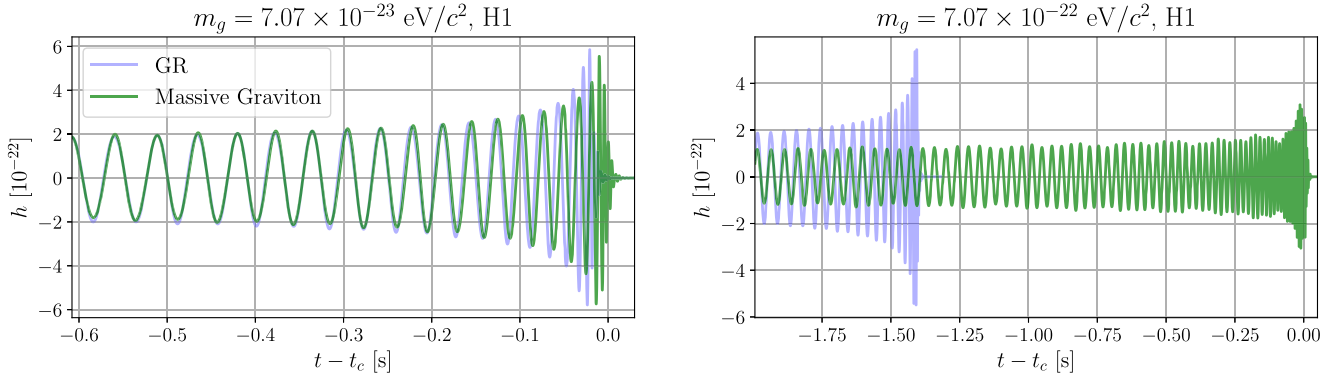


Figure 10. Illustrations of the massive graviton waveforms projected onto the H1 detector and compared with the GR waveform, aligned at 20 Hz. We show the smallest and largest graviton masses considered in this paper.

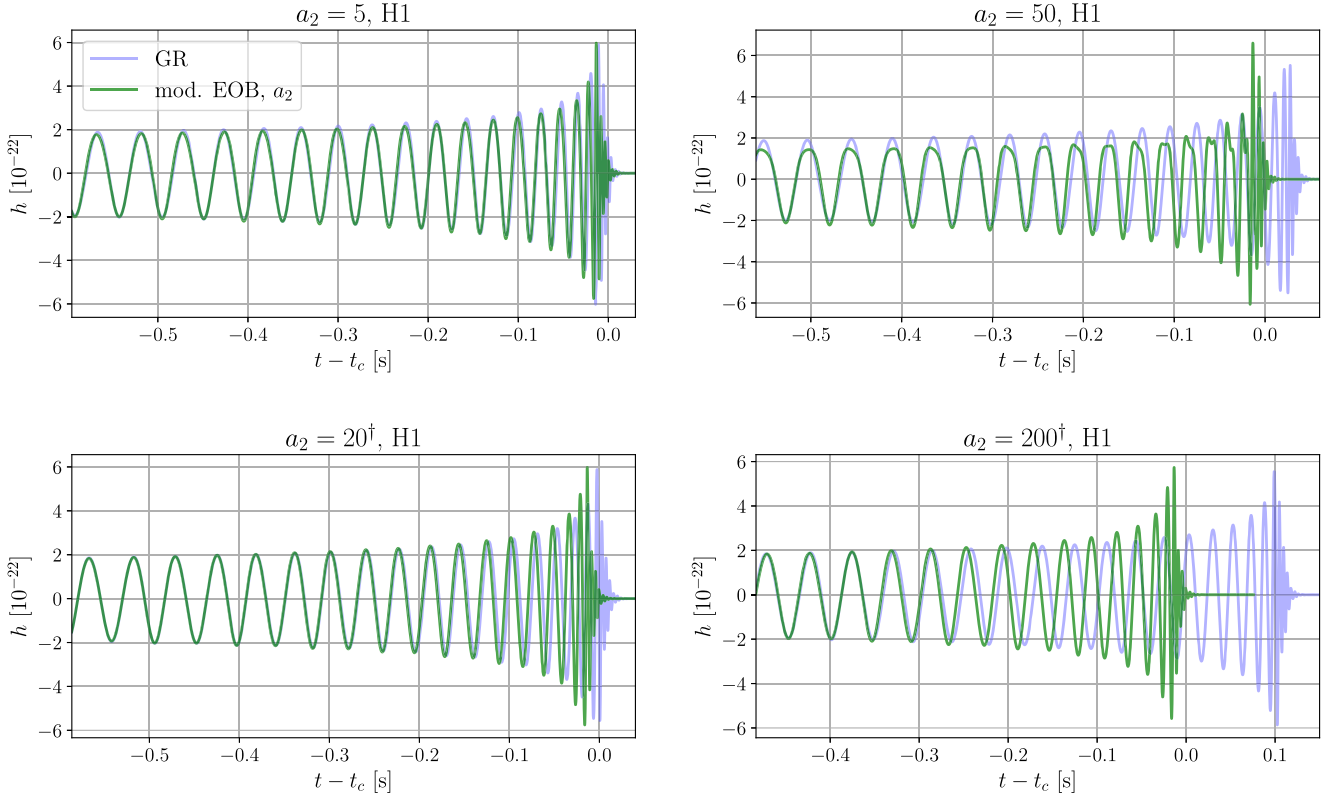


Figure 11. Illustrations of the modified EOB waveforms with modified flux, both with (upper panels) and without (lower panels) waveform scaling, compared with the GW150914-like GR waveform. The waveforms are aligned at 20 Hz and projected onto the H1 detector. We show the smallest and largest GR deviations considered in the case without waveform scaling. In the case with waveform scaling, we show the second-smallest and largest GR deviations, since the modified EOB waveform is almost indistinguishable from the GR waveform for the smallest deviation we consider in this paper.

written as

$$h_{\ell m}^s = \begin{cases} -\bar{a}_{\text{scalar}} \sqrt{\frac{\ell(\ell-1)}{(\ell+1)(\ell+2)}} h_{\ell m} & \text{for } \ell + m \text{ even,} \\ 0 & \text{otherwise.} \end{cases} \quad (\text{A5})$$

The overall scalar contribution to the detector strain is therefore

$$h^{\text{scalar}} = F_S \sum_{\ell, m} h_{\ell m}^s {}^0Y_{\ell m}(\iota, \phi_0), \quad (\text{A6})$$

where $F_S = -\frac{1}{2} \sin^2 \theta \cos 2\phi$ is the antenna response pattern for the scalar breathing mode (C. M. Will 2014), in which θ and ϕ are the spherical coordinates of the source in the

detector frame, and ${}^0Y_{\ell m}$ are the standard zero spin spherical harmonics, where ι and ϕ_0 are the binary inclination and phase at the time of coalescence, respectively. We use the implementation of F_S in the PyCBC package (A. Nitz et al. 2023).

We set the value of the $\bar{a}_{\text{scalar}}$ quantity in terms of the relative amplitude of the $(2, \pm 2)$ scalar and tensor polarizations' contribution to the response of the detector with the largest SNR for the tensor signal, here H1 (so thus includes the spherical harmonics and the detector's antenna pattern). Specifically, we fix the amplitude ratio given by

$$\mathcal{A} = \frac{\bar{a}_{\text{scalar}}}{\sqrt{6}} \frac{|F_S|}{\sqrt{F_+^2 + F_\times^2}} \frac{2 \cdot {}^0Y_{22}(\iota, 0)}{-{}^2Y_{2,2}(\iota, 0) + -{}^2Y_{2,-2}(\iota, 0)}, \quad (\text{A7})$$

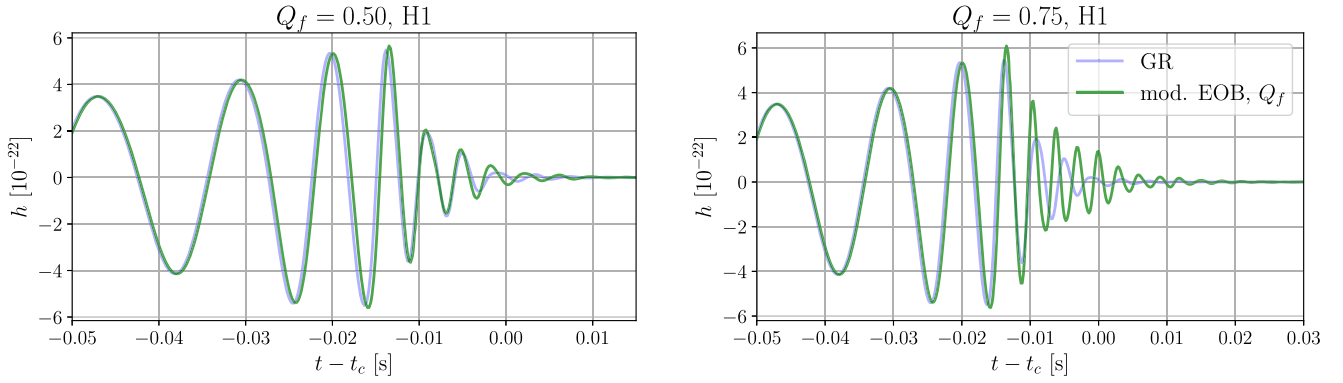


Figure 12. Illustrations of the modified EOB waveforms with modified QNM spectrum, compared with the GW150914-like GR waveform. The waveforms are aligned at 20 Hz and projected onto the H1 detector, but are zoomed in to show the merger-ringdown phase, which is the only portion of the waveform where there are any differences from GR. We show the smallest and largest GR deviations we consider in this paper.

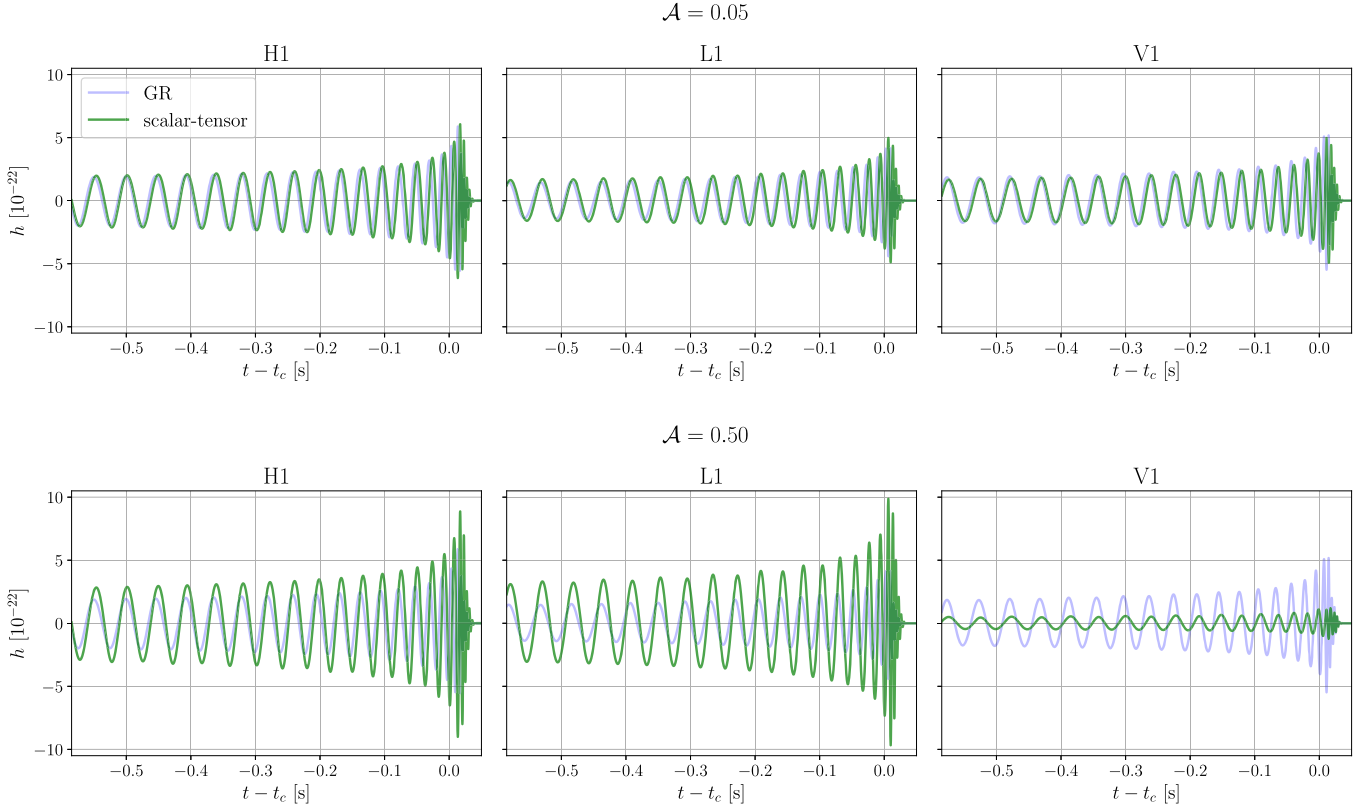


Figure 13. Illustrations of the scalar-tensor waveforms projected onto the three detectors, compared with the GW150914-like GR waveform and aligned at 20 Hz. These show the smallest and largest $\mathcal{A}$ values considered in this paper.

where F_+ and $F_\times$ are the antenna responses for the plus and cross tensor polarizations, respectively, and ${}^{-2}Y_{2,\pm 2}$ are the $(2, \pm 2)$ spin- (-2) -weighted spherical harmonics.

A.4. Illustrations of Waveforms

Here, we show representative illustrations of the non-GR and black hole mimicker (BHM) waveforms we consider, in all cases comparing with the GR BBH equivalent with the same masses and spins (computed using `TEOBResumS-v3-`

`GIOTTO`). We align the waveforms at 20 Hz using the inertial frame $(2, 2)$ mode. In all but one of the non-GR cases, we show the smallest and largest GR deviation considered. We show the representative waveforms for the massive graviton, modified flux, modified QNM spectrum, and scalar-tensor waveforms in Figures 10, 11, 12, and 13, in all cases comparing with the GW150914-like GR waveforms. We show the BHM waveforms in Figure 14, where we use the smallest total mass considered in this paper, in order to show the largest amount of

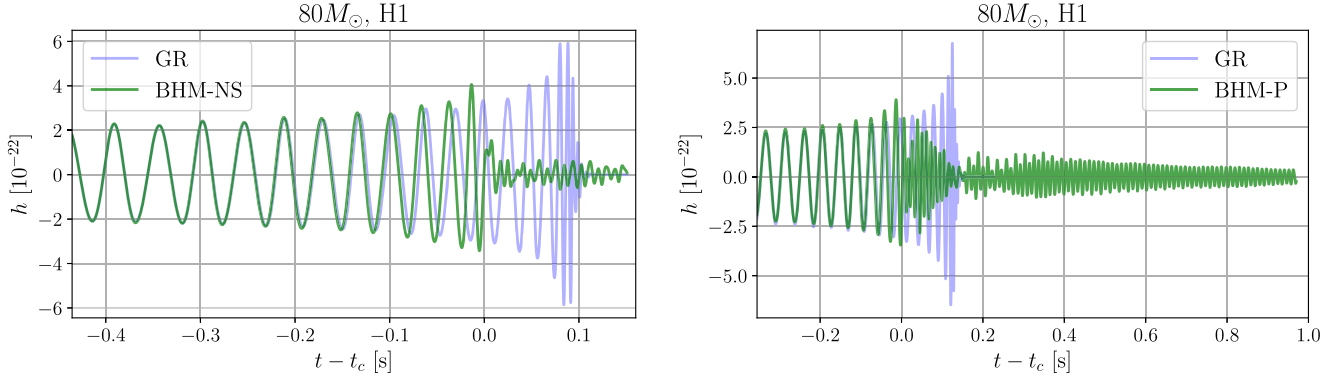


Figure 14. Illustrations of the BHM waveforms projected onto the H1 detector for a total mass of $80M_{\odot}$, compared to the GR waveforms with the same masses and spins, aligned at 20 Hz.

the waveform (since the plot starts at 20 Hz). The post-merger portions of the BHM waveforms are truncated (as they are in the injections) before they start to exhibit drifts due to numerical inaccuracies.

Appendix B

Prior Choices for the Parameter Estimation

In this appendix, specifically in Table 8, we present the full set of priors that were used for the parameter-estimation investigations. These are the defaults for all analyses. In the case of the BHM-P investigations, the luminosity distance prior maximum was increased to 5000 Mpc and the effective luminosity distance priors for the millilensing investigation were increased to 10^5 Mpc. For definitions of each of the source parameters, see Appendix E of I. M. Romero-Shaw et al. (2020).

Additionally, we note that the chirp mass prior range is not a typo. The erroneously large width of this prior was accidentally placed in the settings. While this will have made the sampling process less efficient, it should not have impacted the final results.

Finally, there is a small inconsistency between the specific PLANCK15 (P. A. R. Ade et al. 2016) cosmology used by Bilby—which corresponds to the one in astropy (A. M. Price-Whelan et al. 2022)—and the one in the LVK Algorithm Library used for the generation of the massive graviton waveforms. The differences between these implementations are minor and not expected to have impacted the final results. In particular, the cosmology is used in the generation of the injections only to determine the amount of dephasing from the graviton mass and luminosity distance.

Table 8
Base Prior Values Used in the Injection Set

Parameter	Prior Distribution	Minimum	Maximum
Source Parameters			
$\mathcal{M}_c (M_{\odot})$	Uniform in component masses	5	300
q	Uniform in component masses	0.1	1.0
$m_1 (M_{\odot})$	Constraint	10	80
$m_2 (M_{\odot})$	Constraint	10	80
d_L (Mpc)	Uniform in comoving volume and source time	100	2000
a_1	Uniform	0	0.99
a_2	Uniform	0	0.99
θ_1	Sine		
θ_2	Sine		
θ_{IN}	Sine		
ϕ_{12}	Uniform	0	2π
ϕ_{JL}	Uniform	0	2π
R.A.	Uniform	0	2π
decl.	Cosine		
ψ	Uniform	0	π
Phase	Uniform	0	2π
Microlensing Parameters			
$M_L (M_{\odot})$	Uniform	1	10^4
y	Power law ($\alpha = 1$)	0.1	3.0
d_{Lens}/d_L	Delta (0.5)		
Millilensing Parameters			
Number of images	Uniform (discrete)	1	4
n_i	Uniform (discrete)	0	1
dt_i (s)	Uniform	0.001	0.1
$dL_i^{\text{eff}} (M_{\odot})$	Uniform	100	5000

Appendix C

Additional Results from the Microlensing Investigation

C.1. Point-mass Investigation

In this appendix, we present the complete set of lens posteriors for the isolated point-mass microlensing analysis in Figures 15–20. These results follow the pattern discussed in

detail for the source position posteriors from the massive graviton series of injections in Section 4.2. For the source position posteriors, only those injections that favor the microlensing hypothesis yield any constraints within the prior boundaries, with the remainder showing significant railing toward larger source positions. These behaviors are consistent with the expectations for lensed and unlensed events, respectively. Of note are the results for the black hole mimicker

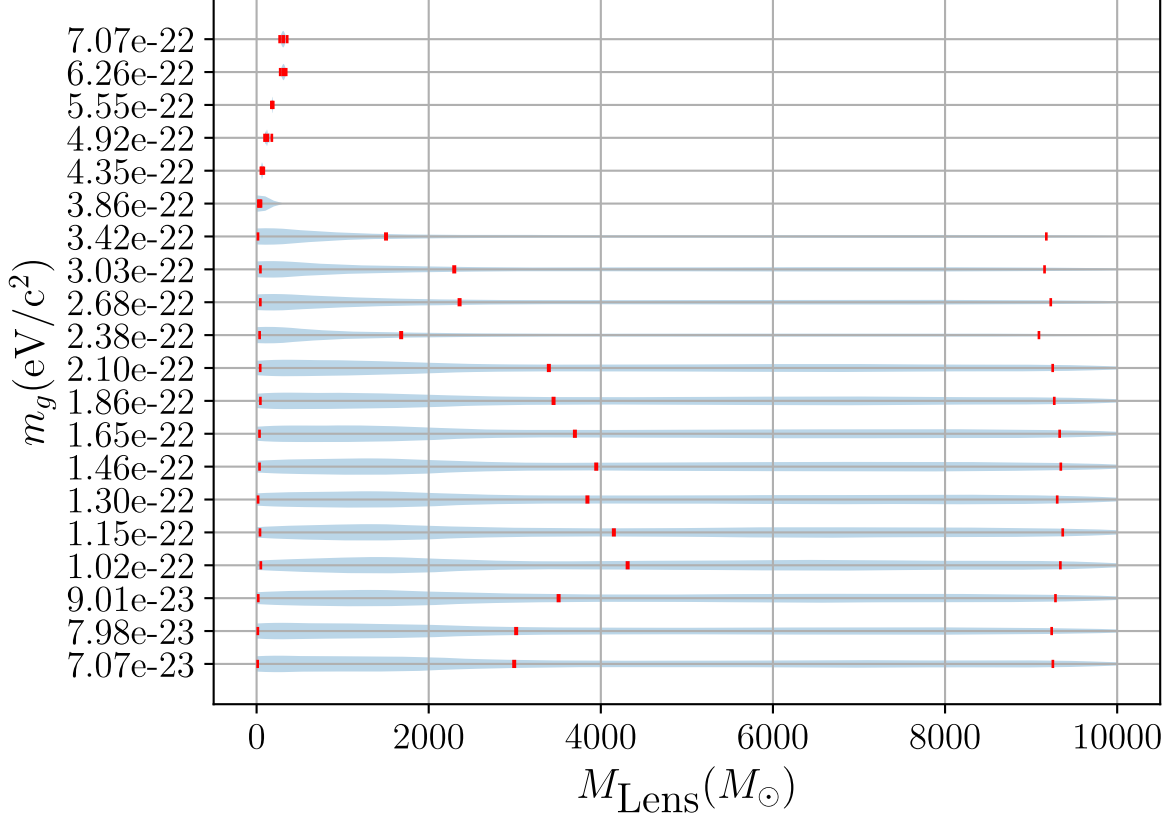


Figure 15. Recovered posteriors on the lens mass (taken to be halfway between the source and observer) from the massive graviton series of injections. The red ticks indicate the median and 90% C.I. The middle red tick indicates the median of the posterior with the other two representing the boundaries of the 90% C.I. The lower-mass results are broad, with greater support for very-low-mass lenses, in line with the expectations for unlensed events. The higher-mass graviton injections, however, display support for specific values of lens mass within the prior region, which is in line with the expectation for microlensed events.

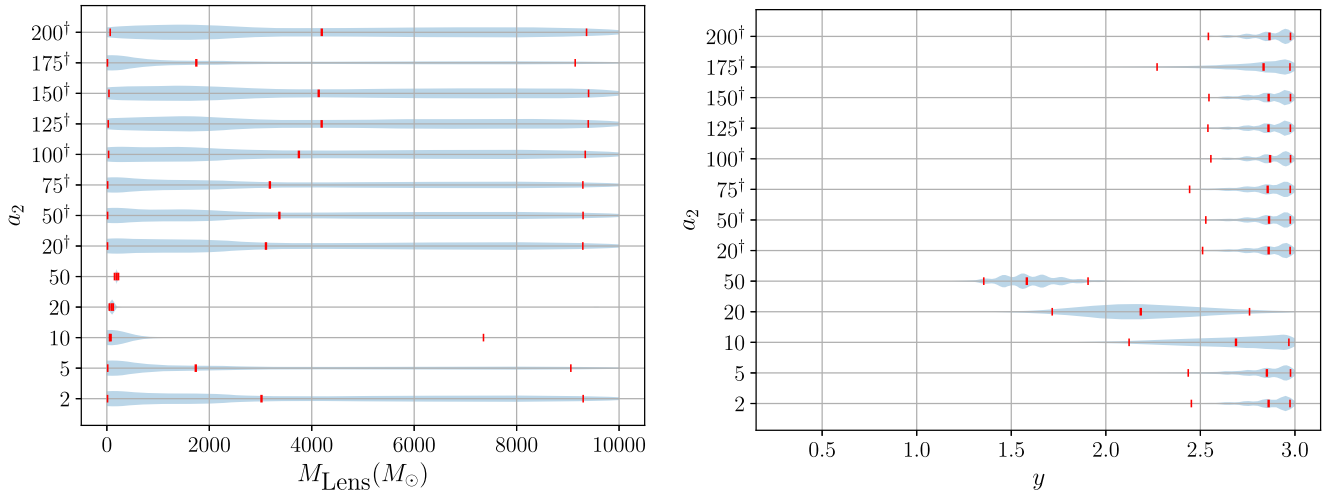


Figure 16. Recovered posteriors on the lens mass (left) and source position (right) from the isolated point-mass analysis of the modified energy flux series of injections, where the dagger denotes the injections without waveform scaling. The red ticks indicate the median and 90% C.I. As is the case with the massive graviton injections, only those injections that result in preference for the microlensing hypothesis display constraints toward specific values within the prior bounds.

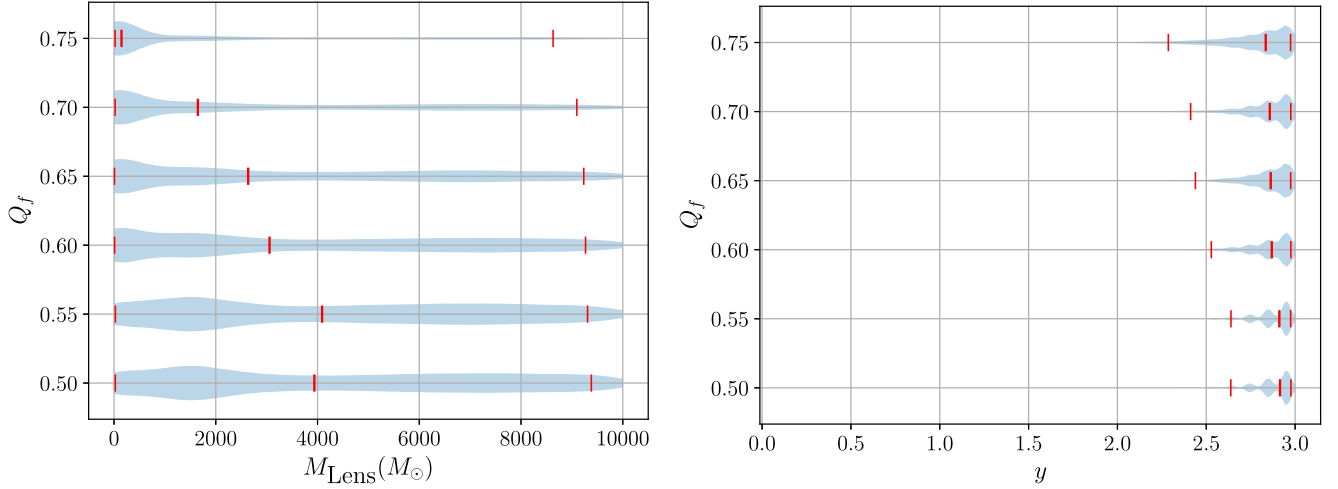


Figure 17. Recovered posteriors on the lens mass (left) and source position (right) from the isolated point-mass analysis of the modified QNM spectra series of injections. The red ticks indicate the median and 90% C.I. In agreement with the lack of any false-positive preference for the microlensing hypothesis for these injections, these posteriors are in line with the expectations for unlensed events.

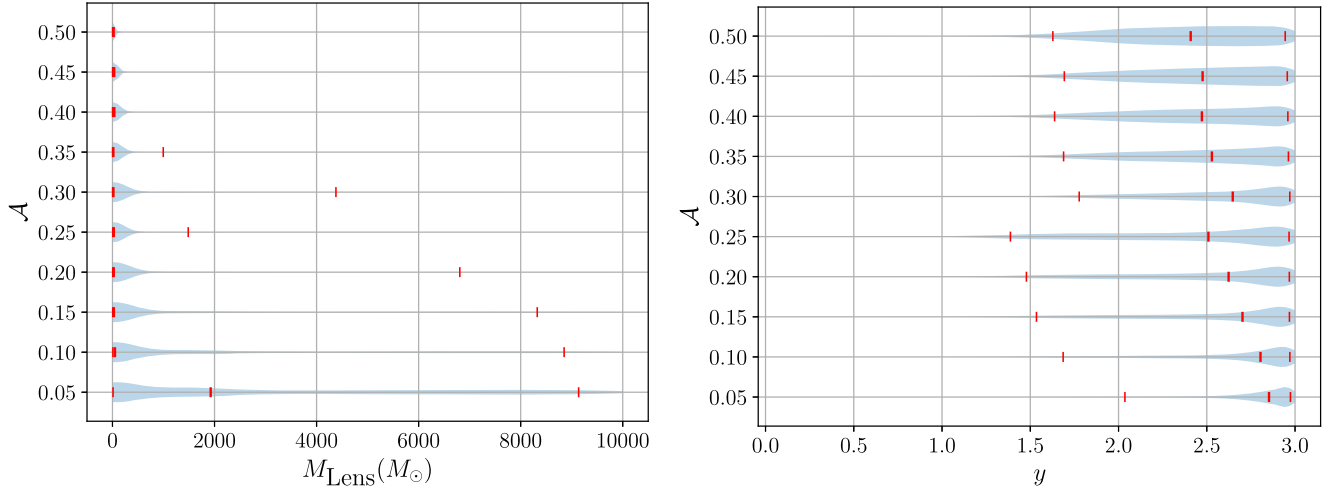


Figure 18. Recovered posteriors on the lens mass (left) and source position (right) from the isolated point-mass analysis of the additional scalar polarization series of injections. The red ticks indicate the median and 90% C.I. As for the modified QNM injections, these injections have no false-positive preference for the microlensing hypothesis, and the posteriors are within expectations for unlensed events.

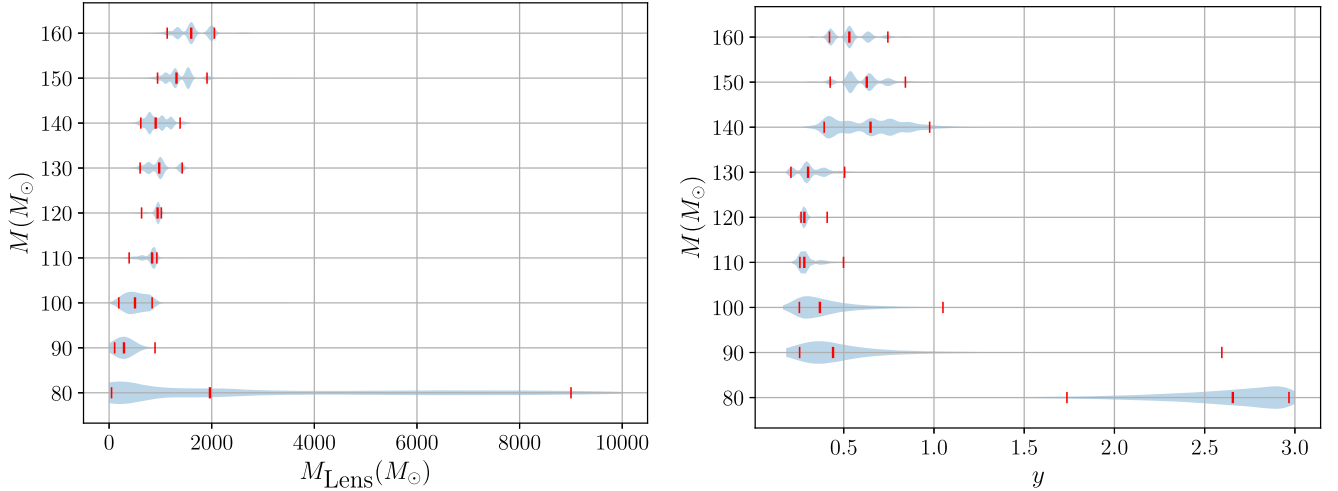


Figure 19. Recovered posteriors on the lens mass (left) and source position (right) from the isolated point-mass analysis of the black hole mimicker waveforms based on the precessing BNS waveform from T. Dietrich et al. (2018a). The red ticks indicate the median and 90% C.I. All cases here demonstrate a preference for the microlensing hypothesis, and this is reflected in the consistently constrained lens parameter posteriors. Of particular note is the apparent multimodality, which is not expected for true microlensing events.

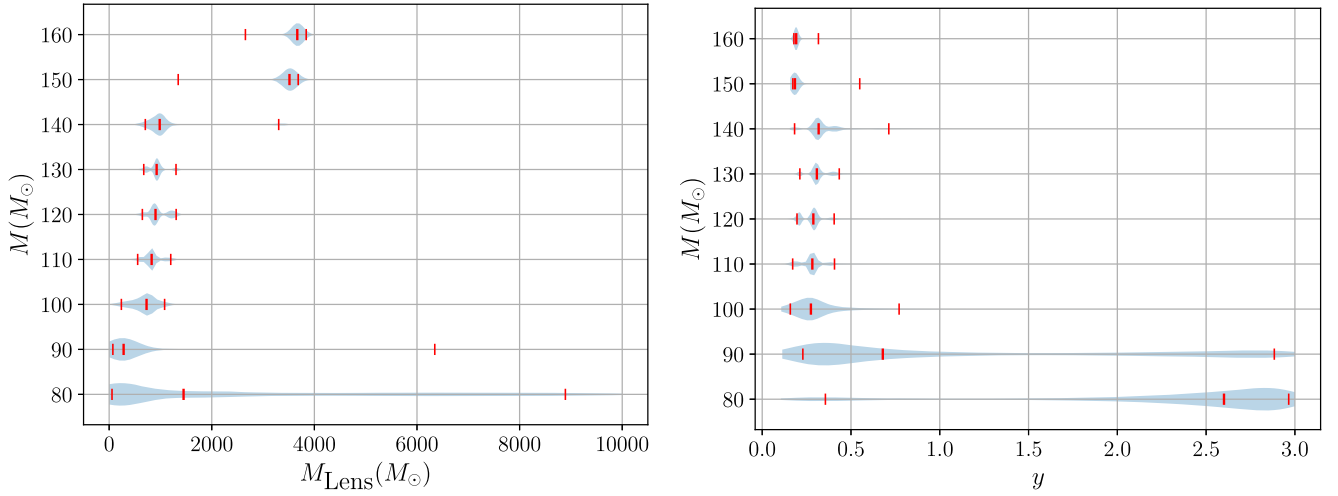


Figure 20. Recovered posteriors on the lens mass (left) and source position (right) from the isolated point-mass analysis of the black hole mimicker waveforms based on the nonspinning BNS waveform from M. Ujevic et al. (2022). The red ticks indicate the median and 90% C.I. Of note is that the $90M_{\odot}$ case displays some constraint within the prior boundary despite the microlensed case being insufficiently preferred by the Bayes factor that this event would be selected for further analysis. Some of the results display multimodality, which is not expected in true microlensing analysis.

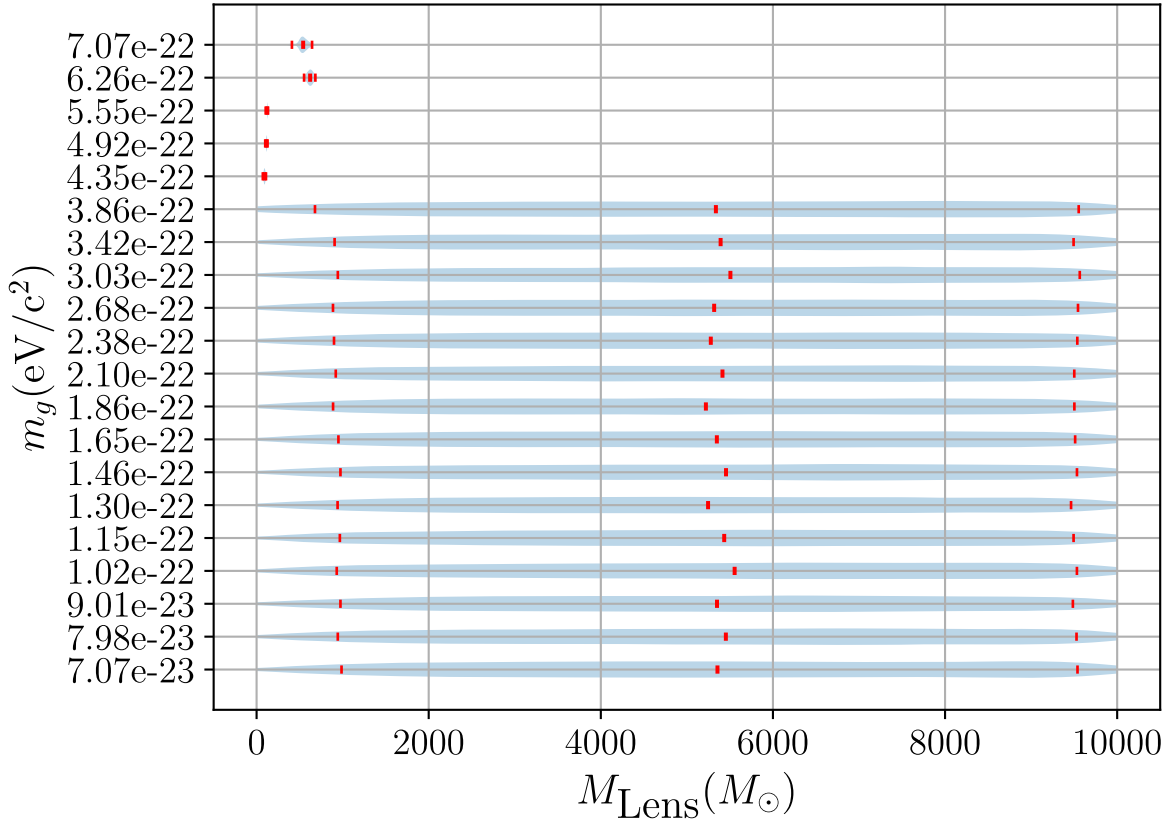


Figure 21. Recovered posteriors on the lens mass from the SIS analysis of the massive graviton series of injections. The red ticks indicate the median and 90% C.I. The lower-mass injections display broad uninformative posteriors, which is in line with the expectations for unlensed events. In combination with the source position posteriors, these would result in flat amplification factors that more closely resemble unlensed signals. The higher-mass injections, as was the case with the point-mass analysis, display constraints toward specific values, which is indicative of their false-positive status.

injections, with cases for both the precessing and nonspinning sets displaying multimodality in the posteriors. This is not expected for true microlensing candidates and would indicate that further investigation is needed on these events. However, given the Bayes factor support this would be insufficient to rule these out as microlensing candidates.

Similarly, the lens-mass posteriors (assuming the lens is placed halfway between the source and observer) reflect the

expectations of unlensed events for those events that yielded no support for the microlensing hypothesis—broad posteriors with some support for low mass values that rail against the lower prior boundary. This combination results in the smallest amplification factor oscillation, resulting in the closest match to an unlensed waveform. Those candidates that did favor the microlensing hypothesis again display constraints toward specific values within the posterior. The black hole mimicker

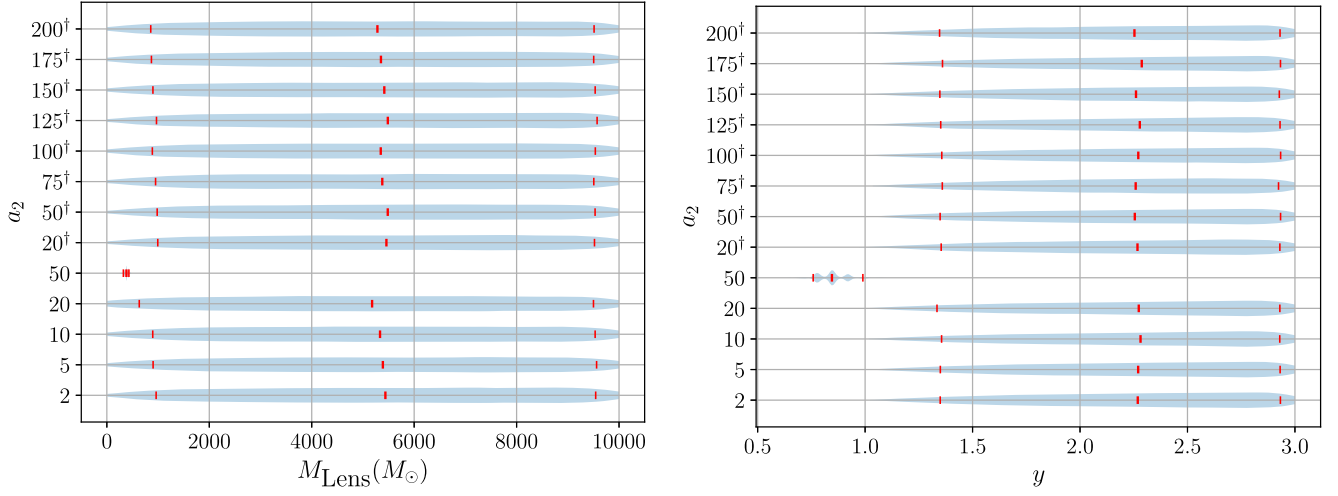


Figure 22. Recovered posteriors on the lens mass (left) and source position (right) from the SIS analysis of the modified energy flux series of injections, where the dagger denotes the injections without waveform scaling. The red ticks indicate the median and 90% C.I. In all but the $a_2 = 50$ case, these display the same broad posteriors with $y > 1$ that is expected of unlensed events—this includes the $a_2 = 20$ case, which has some support for the point-mass model.

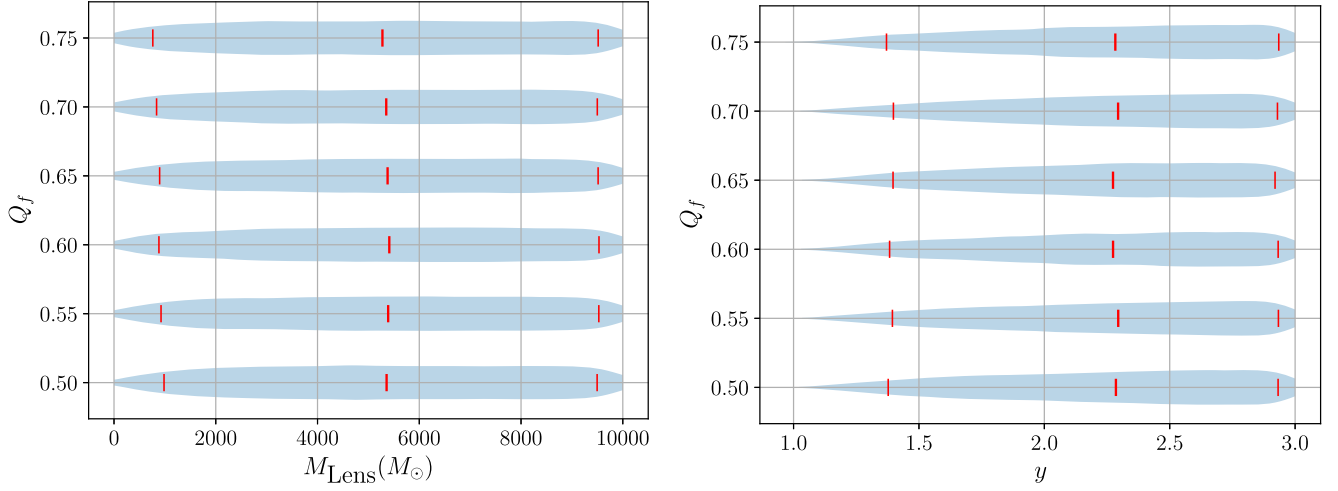


Figure 23. Recovered posteriors on the lens mass (left) and source position (right) from the SIS analysis of the modified QNM spectra series of injections. The red ticks indicate the median and 90% C.I. In all cases, these display the same broad posteriors with $y > 1$ that is expected of unlensed events.

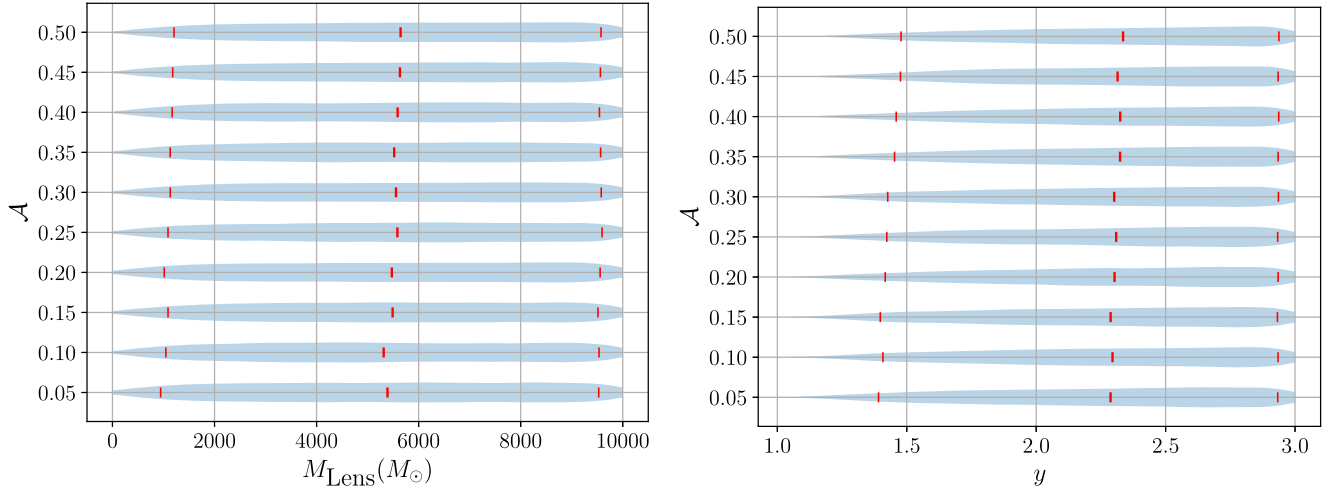


Figure 24. Recovered posteriors on the lens mass (left) and source position (right) from the SIS analysis of the additional scalar polarization series of injections. The red ticks indicate the median and 90% C.I. In all cases, these display the same broad posteriors with $y > 1$ that is expected of unlensed events.

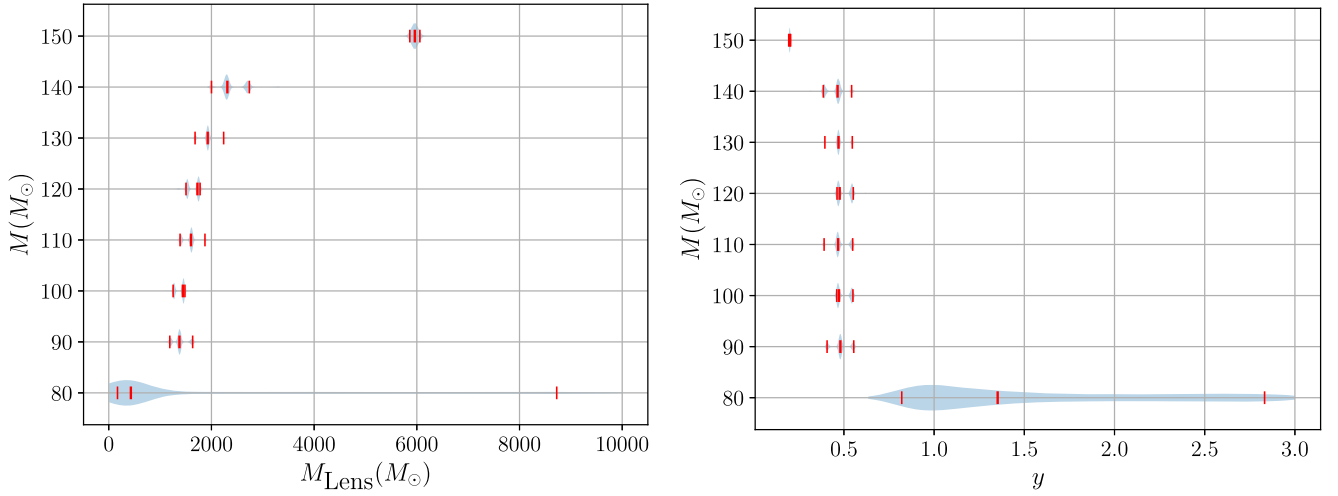


Figure 25. Recovered posteriors on the lens mass (left) and source position (right) from the SIS analysis of the precessing black hole mimicker series of injections. The red ticks indicate the median and 90% C.I. In this instance, the $M = 80 M_{\odot}$ displays posteriors that have some support for values inside of the region of parameter space that would yield an oscillatory amplification factor, alongside support for the single-image case. The other cases all yield a constraint to specific values as is expected of candidate lensing events, including the $M = 140 M_{\odot}$ case, despite the significant disfavoring in this case which may be due to finding source parameters between the local and global maximum found in the point-mass and unlensed analyses.

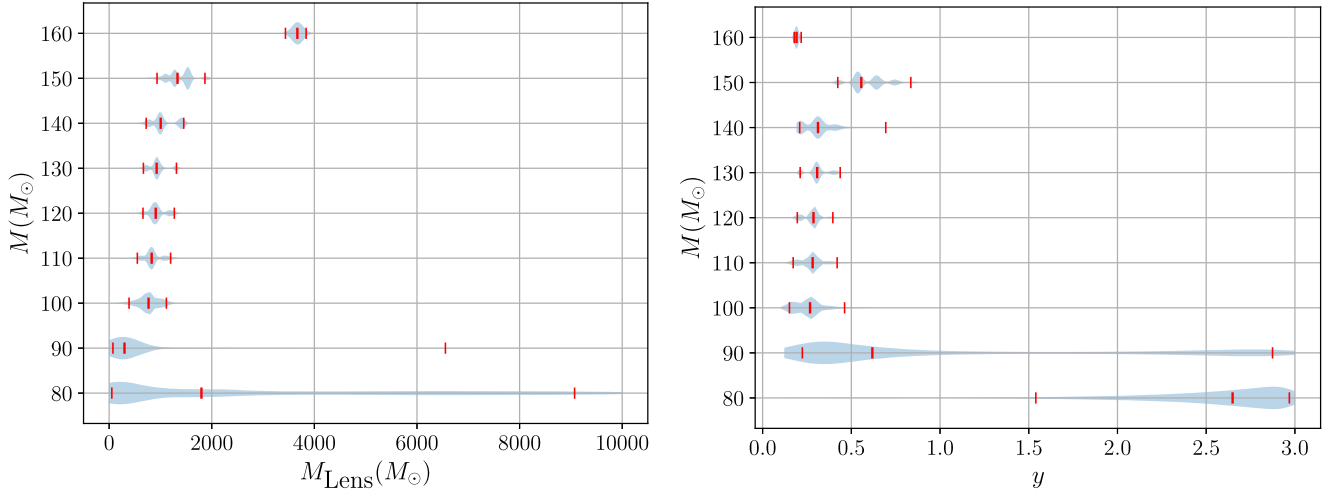


Figure 26. Recovered posteriors on the lens mass (left) and source position (right) from the SIS analysis of the nonspinning black hole mimicker series of injections. The red ticks indicate the median and 90% C.I. In this instance, only the $M = 80 M_{\odot}$ and $M = 90 M_{\odot}$ cases match the expectations for unlensed events, with a more distinct constraint in the remaining cases. Similar to the point-mass results for these, there is however some unexpected multimodality in a few of the recovered posteriors.

waveforms display a preference for significantly higher-mass lenses than those for the massive graviton or modified energy flux cases. The lens-mass posteriors in the black hole mimicker cases also display some multimodality, with the same implications as the multimodality in the source position posteriors discussed above.

C.2. SIS Investigation

Similar to Appendix C.1, in this appendix we present the complete set of recovered lens posteriors for the SIS microlensing analysis in Figures 21–26. The results largely follow the pattern discussed in Sections 4.2 and 4.3 in that those candidates that were considered false positives for microlensing largely follow the expectations for microlensed candidates and vice versa.

In particular, as is the case for unlensed events, the source position posteriors are largely uninformative beyond supporting only the $y > 1$ region in which in geometric optics only a

single image will be produced and the resultant amplification factor will be flat and thus more closely resemble an unlensed waveform. This leads to an expectation of broader lens-mass posteriors compared with the point-mass case, and this is observed in the resultant posteriors. One result worth noting is the modified energy flux, $a_2 = 20$ case. This has some support, though insufficient for further investigation, for the point-mass hypothesis. Both the Bayes factor and the posteriors in the SIS case do cast more significant doubt on this, however, as the posteriors match much more closely with the unlensed expectation for this model.

Similar to the point-mass analysis, both the precessing and nonspinning black hole mimicker injections have cases displaying multimodality in the posteriors. This would not be the expected behavior of true microlensing candidates. However, it would be insufficient to rule these candidates out given the Bayes factor support. These candidates would therefore require additional investigation to give a final determination of their status.

A result of particular note is the $140M_{\odot}$ precessing black hole mimicker waveform, which displays recovered posteriors similar to the candidate lensing events despite the significant disfavoring for this model compared to both the point-mass and unlensed models. This may indicate that the source parameters which were found between the local and global likelihood maxima in these analyses are themselves less supported than those found in the maximum for the point-mass and unlensed analyses.

Appendix D

Additional Results from the Millilensing Investigation

In this appendix, we present the set of recovered posteriors on the effective luminosity distance and time delay of the

second proposed image in the millilensing analysis for the injections. These are shown in Figures 27–32. The expectations for unlensed events would be to see broad, uninformative posteriors spanning the prior space of both parameters, with the expectations for lensing candidates to obtain constraints on these quantities for as many images as are supported. These expectations are repeated for the third and fourth proposed images.

In all cases this is borne out in the recovered posteriors, with the exception of the $140M_{\odot}$ precessing black hole mimicker case, which, similar to the SIS analysis of this case, displays posteriors in line with the expectation for lensed candidates. This pattern is repeated in the third and fourth proposed image posteriors, which are not shown here.

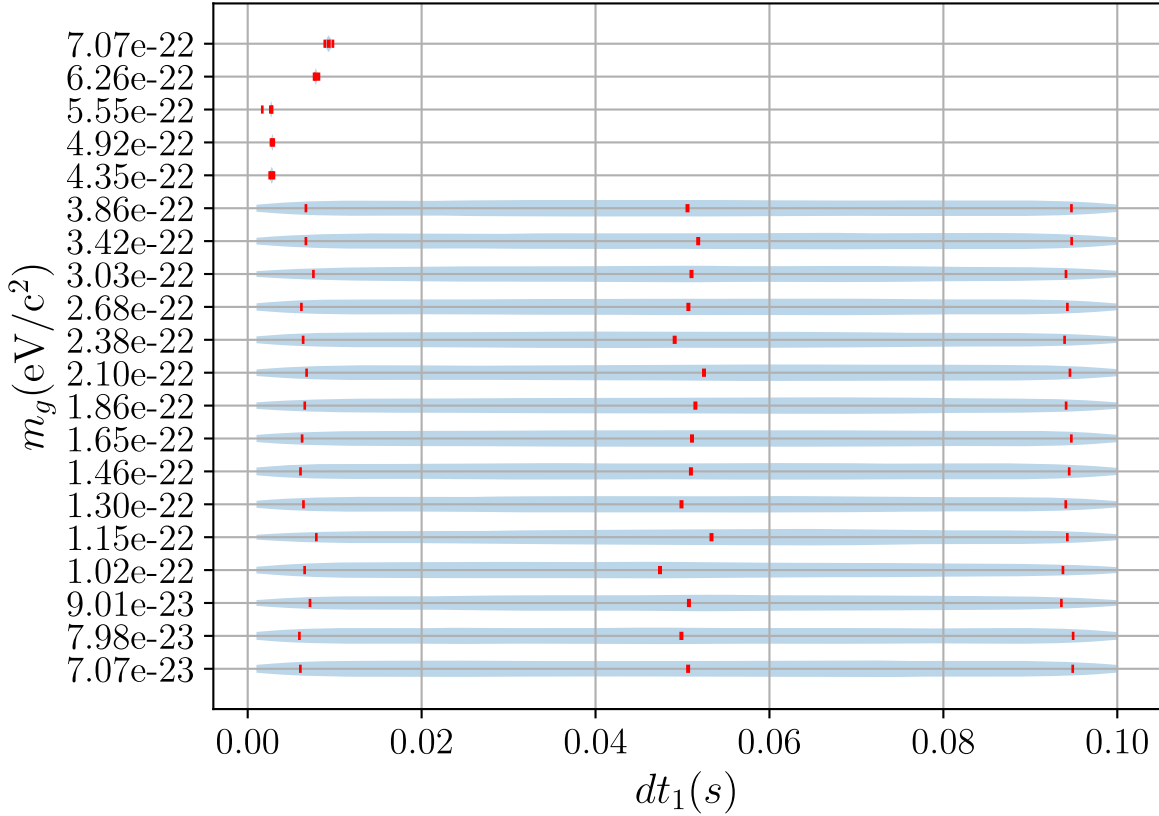


Figure 27. Recovered posteriors on the time delay between the first and second images for the massive graviton series of injections. The red ticks indicate the median and 90% C.I. The lower-mass injections reflect the expectation for unlensed events, with broad, uninformative posteriors indicative of a lack of support for the second image. The higher-mass injections display constraints toward a particular value, which is in line with the expectations for a true millilensing event.

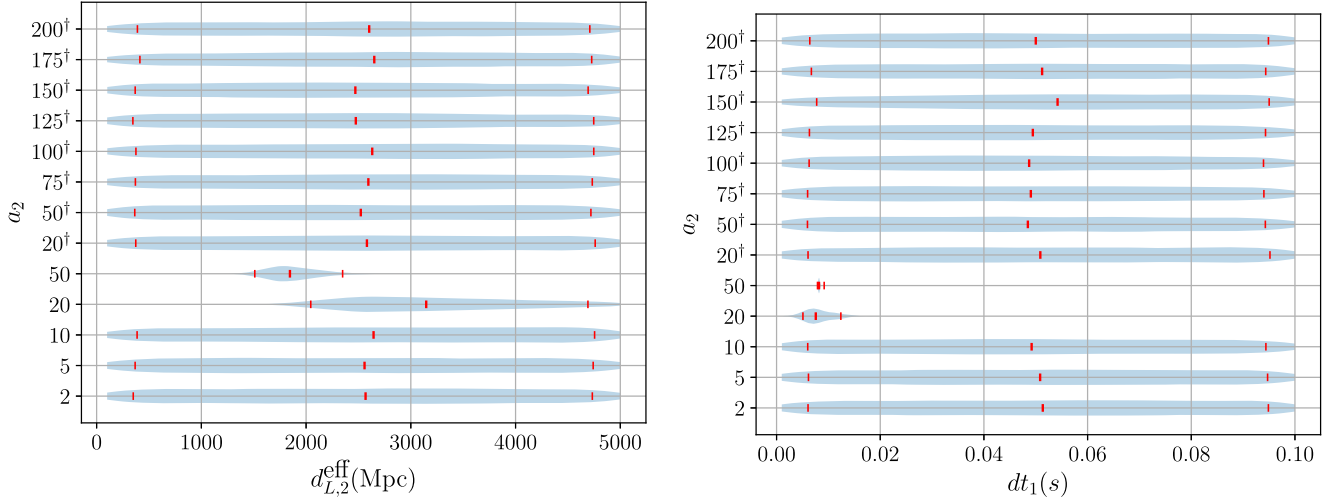


Figure 28. Recovered posteriors on the effective luminosity distance (left) and time delay (right) of the second proposed image in the millilensing analysis of the modified energy flux series of injections, where the dagger denotes the injections without waveform scaling. The red ticks indicate the median and 90% C.I. In these, only the $a_2 = 20$ and $a_2 = 50$ cases display constraints indicative of support for a second image, reflected in their support for the millilensing model.

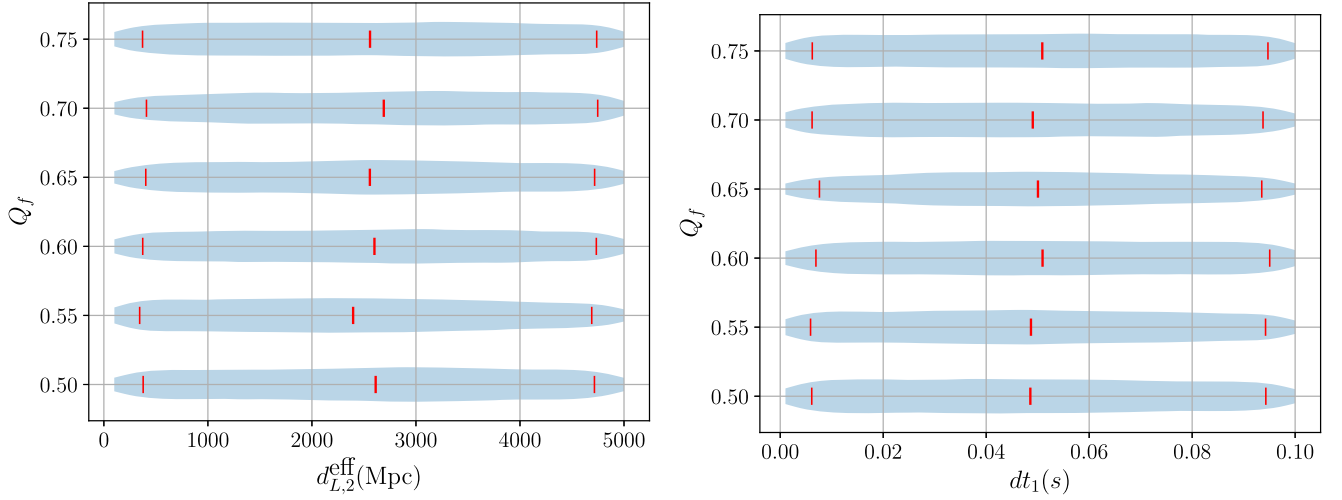


Figure 29. Recovered posteriors on the effective luminosity distance (left) and time delay (right) of the second proposed image in the millilensing analysis of the modified QNM spectra series of injections. The red ticks indicate the median and 90% C.I. In these, all cases display the broad, uninformative posteriors indicating no support for a second image.

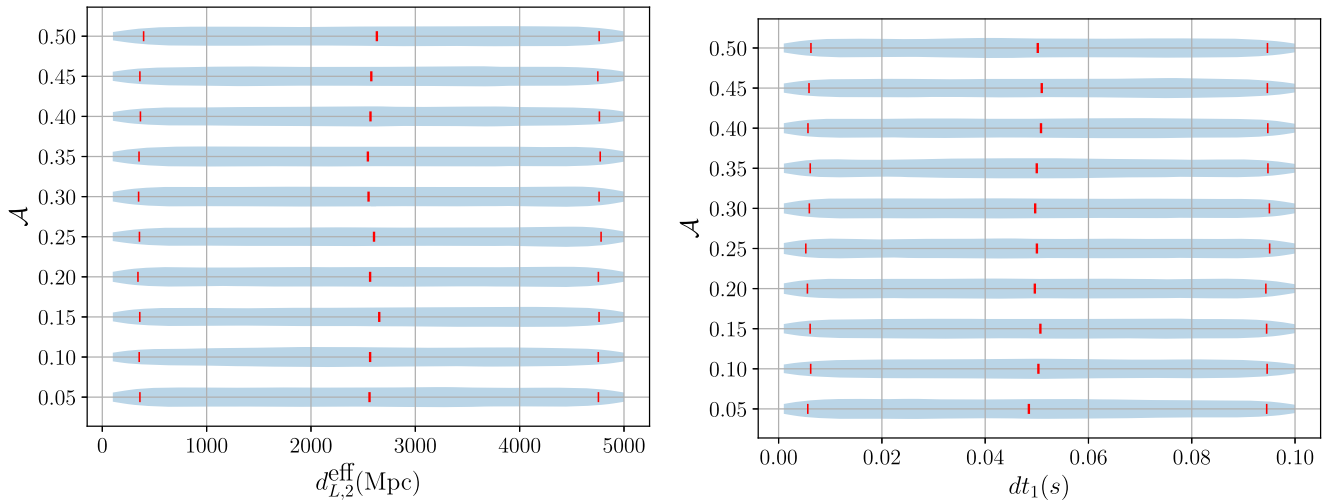


Figure 30. Recovered posteriors on the effective luminosity distance (left) and time delay (right) of the second proposed image in the millilensing analysis of the additional scalar polarization series of injections. The red ticks indicate the median and 90% C.I. In all cases, the recovered posteriors are indicative of a lack of support for a second image, including the $\mathcal{A} = 0.20$, 0.25 cases that display Bayes factor support for the millilensing model, which would indicate that support for the model is coming from the Morse factor of the first image, in line with this resulting in a false positive for the type II image search as well.

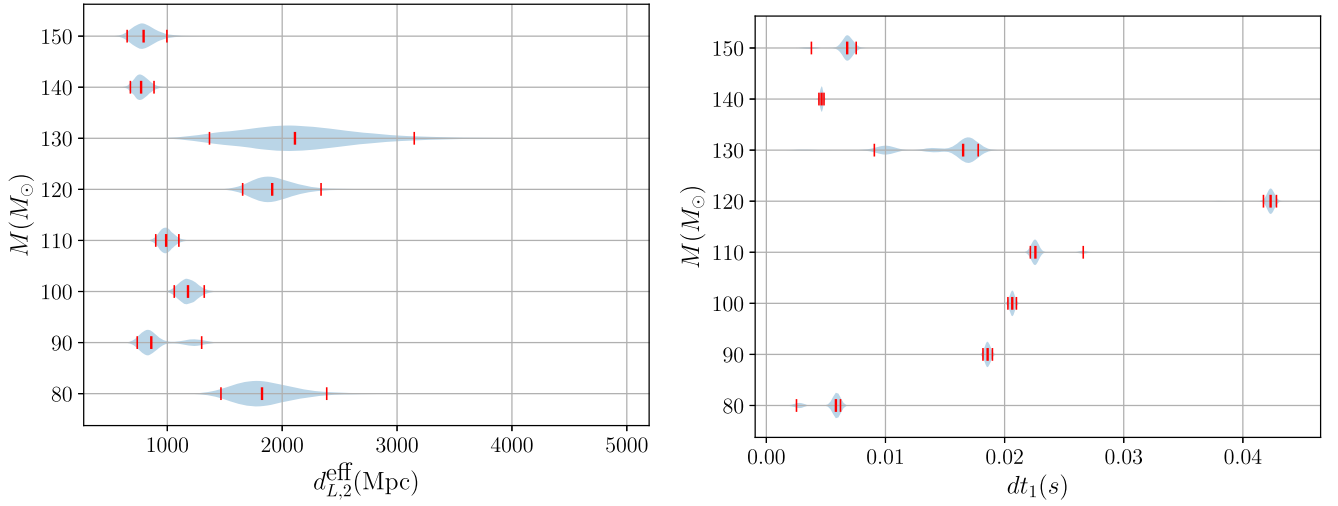


Figure 31. Recovered posteriors on the effective luminosity distance (left) and time delay (right) of the second proposed image in the millilensing analysis of the precessing black hole mimicker series of injections. The red ticks indicate the median and 90% C.I. All cases display constraints toward specific values, indicating support for a second image in this case, including the $140M_\odot$ case, which has a Bayes factor preference for the unlensed case.

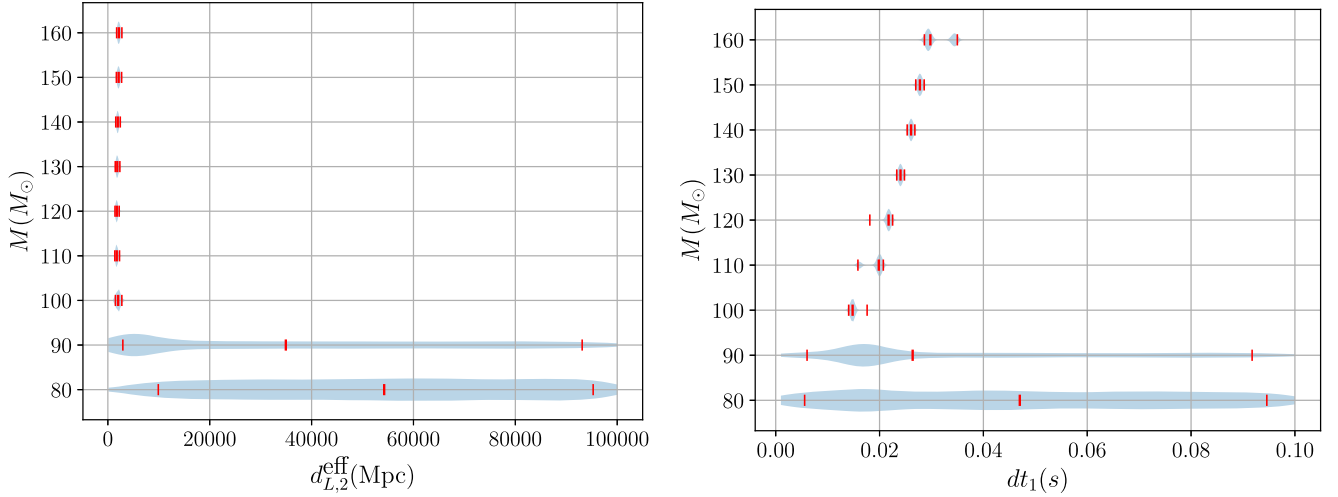


Figure 32. Recovered posteriors on the effective luminosity distance (left) and time delay (right) of the second proposed image in the millilensing analysis of the nonspinning black hole mimicker series of injections. The red ticks indicate the median and 90% C.I. In this instance, the $80M_\odot$ case displays broad posteriors that do not support a second image. The $90M_\odot$ case displays broad posteriors but with a notable clustering toward a particular value, in line with the slender support for two images seen in the Bayes factor. The remainder of the cases which all display strong support for the millilensing model with additional images have tight constraints on the posteriors.

Appendix E Additional Results from the Type II Strong-lensing Investigation

In this appendix, we present the results for all the analyses done for the continuous type II lensing searches. These results are complementary with the ones presented in Section 4.5 and, more precisely, the numbers shown in Table 6.

First, the results for the massive graviton are given in Figure 33. One sees that the posteriors are flat for most graviton masses considered. For the two lowest masses, the posteriors even disfavor the type II value, which is in line with an unlensed event containing HOMs. Therefore, none of the posteriors would lead to extra inquiries when a massive graviton is present.

Moving to the modified energy flux, looking at the results shown in Figure 34, one sees that for the four highest amplitudes the Morse factor posteriors are mostly flat. Then, when reducing the value for a_2 , we obtain more constrained

posteriors, with values generally peaking around $n = 0.8$ and no support for the lensing physical values. Because the main value is generally above 0.75, these deviations in the Morse factor are not flagged in the discrete analyses. However, should they be found in the continuous case, the results would most likely trigger follow-up analyses.

In the case of modified QNMs, for which the results are shown in Figure 35, the picture is a bit similar to the one for the modified energy flux. The highest Q_f values lead to a mostly uninformative posterior. The lower values lead to more constrained posteriors with a peak around 0.8. Here, some support is left for the lensing GR values, with $n = 1$ still being possible for the most extreme case.

For the scalar polarizations, as can be seen from Figure 36, all the values of the deviation parameter lead to the well-constrained posteriors peaking away from the GR values for lensing. The exact width and median values of the posteriors change from one $\mathcal{A}$ value to the next, but, essentially, extra inquiries would be triggered in all cases when the analysis is

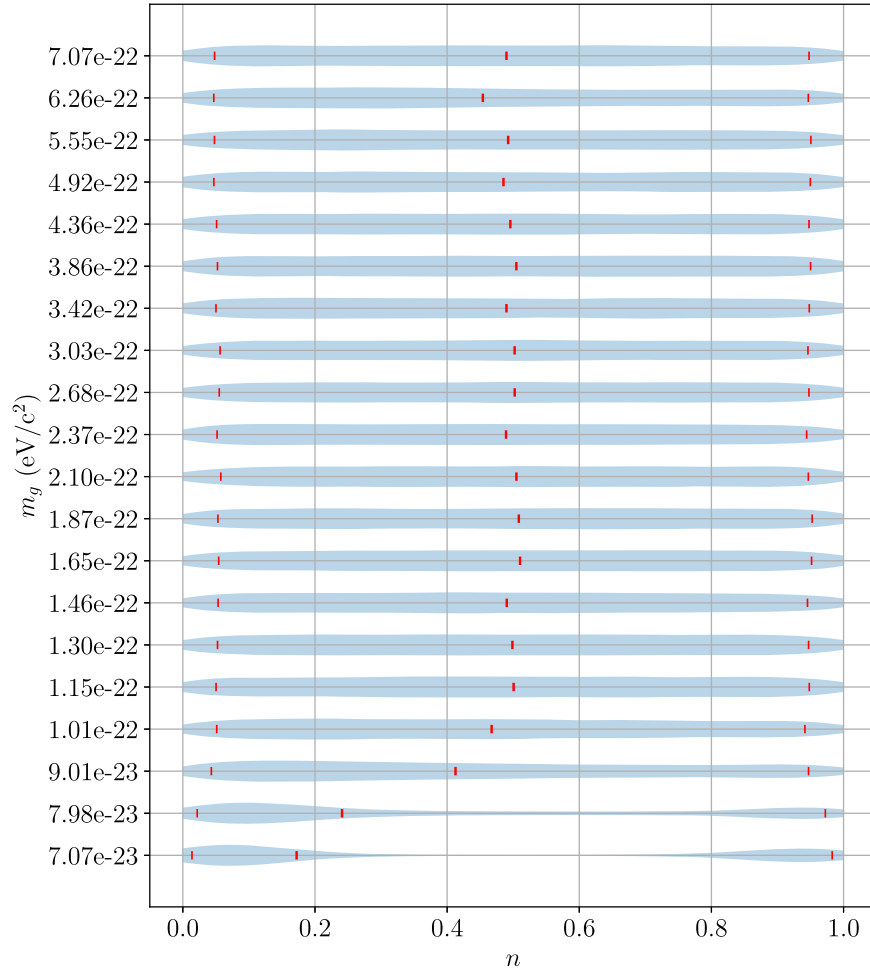


Figure 33. Representation of the recovered posterior for the Morse factor when doing the type II strong-lensing search using a continuous prior for the massive graviton case. The red ticks indicate the median and 90% C.I. Most of the posteriors are agnostic to the image type, and the ones for the two lowest masses even disfavor type II images.

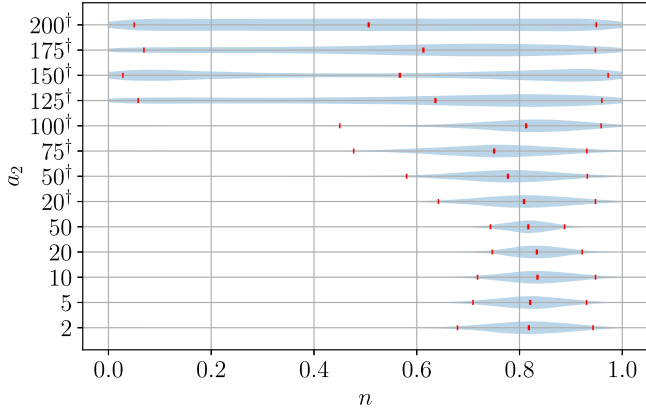


Figure 34. Recovered posteriors for the Morse factor when using a continuous prior for the case of a modified energy flux. The values with a dagger mean that the injections were done without waveform scaling. The red ticks indicate the median and 90% C.I. The four highest values are mostly uninformative, while the others all peak around 0.8, with low to no support for the physical values one obtains from lensing. Therefore, such deviations would not be detected when considering Bayes factors between different strong-lensing hypotheses, but could be detected when searching for type II images with a continuous prior.

performed with a continuous prior. We also see that in some cases most of the posterior weight is well above 0.75, which explains why the analyses done with a discrete prior are unable

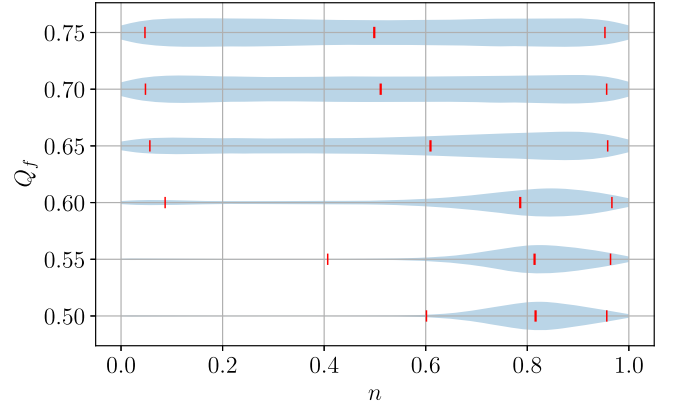


Figure 35. Morse factor posteriors for the lensing type II image searches with a continuous prior for the modified QNM spectra injections. The red ticks indicate the median and 90% C.I. For the highest parameter values, the posteriors are noninformative. However, for the lower values, one sees the posteriors become more constrained and peak away from the GR values. Still, they retain (weak) compatibility with values allowed by GR ($n = 0$ and/or 1).

to find anything particular. For those posteriors that have significant presence below 0.75, the discrete search for type II images had found possible lensing signatures; see Table 5.

For the black hole mimickers, the situation is a bit different depending on whether we consider the precessing case following

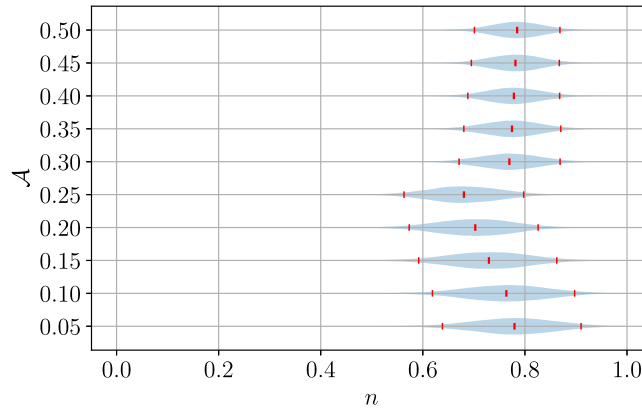


Figure 36. Morse factor posteriors for the lensing type II image searches with a continuous prior for the additional scalar polarization injections. The red ticks indicate the median and 90% C.I. All values of A lead to a deviation from the values expected for lensing in GR. In some cases, the posteriors have more support close to 0.5 than to 0 or 1, meaning that these cases would be identified as possibly lensed by the discrete type II image search.

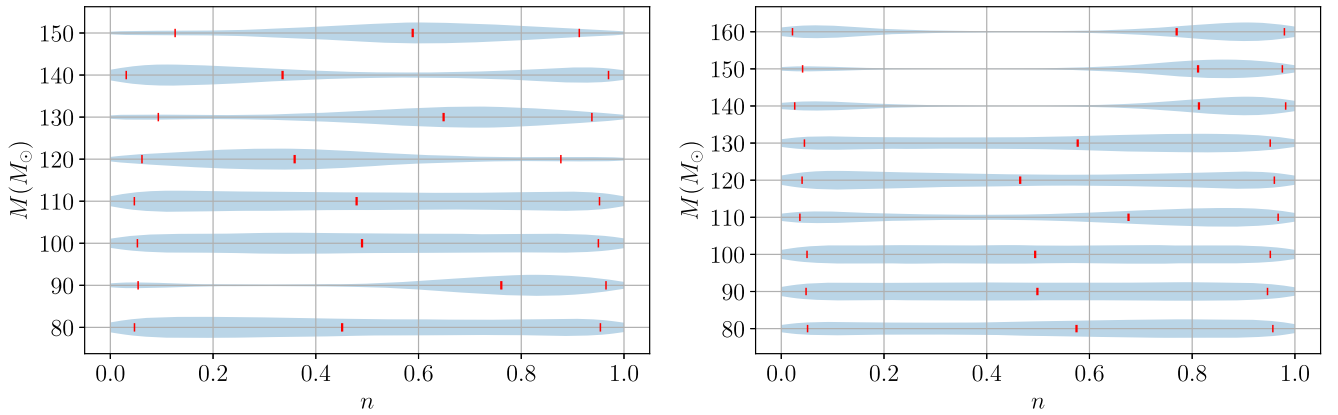


Figure 37. Recovered posteriors for the Morse factor when using a continuous prior and analyzing a binary of black hole mimickers. Left panel: the precessing case based on the simulation from T. Dietrich et al. (2018a). Right panel: the nonspinning case based on the simulation from M. Ujevic et al. (2022). The red ticks indicate the median and 90% C.I. For the precessing binary, in some cases the posteriors show a larger support for Morse factor values close to 0.5, which would be consistent with a type II image. For the nonspinning binary, the posteriors never show support for type II images or any values deviating from GR. Hence, they would not lead to any follow-up inquiries.

T. Dietrich et al. (2018a) or the nonspinning case following M. Ujevic et al. (2022). For the latter, only the three highest masses lead to informative posteriors, but they simply discard the type II hypothesis—in line with the expectations for non-type II GR events with HOMs. The corresponding posteriors are shown in the right panel of Figure 37. When considering the results for the precessing black-hole-mimicker-like waveform based on T. Dietrich et al. (2018a), the posteriors change more from one situation to the other, as can be seen in the left panel of Figure 37. In many cases they are mostly uninformative. However, for $M = 120, 130$, and $150 M_{\odot}$, there is a higher support for values close to $n = 0.5$. Therefore, such examples could potentially be mistaken for type II images and would require further investigations should they be seen.

References

- Aasi, J., Abbott, B. P., Abbott, R., et al. 2015, *CQGra*, **32**, 074001
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2020, *LRR*, **23**, 3
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016a, *PhRvL*, **116**, 061102
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016b, *PhRvL*, **116**, 221101
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2019a, *PhRvL*, **123**, 011102
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2019b, *PhRvD*, **100**, 104036
- Abbott, R., Abbott, T. D., Abraham, S., et al. 2020a, *PhRvL*, **125**, 101102
- Abbott, R., Abbott, T. D., Abraham, S., et al. 2020b, *ApJL*, **900**, L13
- Abbott, R., Abbott, T. D., Abraham, S., et al. 2021a, *ApJ*, **923**, 14
- Abbott, R., Abbott, T. D., Abraham, S., et al. 2021b, *PhRvD*, **103**, 122002
- Abbott, R., Abbott, T. D., Acernese, F., et al. 2023, *PhRvX*, **13**, 041039
- Abbott, R., Abbott, T. D., Acernese, F., et al. 2024a, *PhRvD*, **109**, 022001
- Abbott, R., Abe, H., Acernese, F., et al. 2021c, arXiv:2112.06861
- Abbott, R., Abe, H., Acernese, F., et al. 2024b, *ApJ*, **970**, 191
- Acernese, F., Agathos, M., Agatsuma, K., et al. 2014, *CQGra*, **32**, 024001
- Adam, C., Castelo, J., García Martín-Caro, A., et al. 2022, *PhRvD*, **106**, 123022
- Ade, P. A. R., Aghanim, N., Arnaud, M., et al. 2016, *A&A*, **594**, A13
- Akçay, S., Gamba, R., & Bernuzzi, S. 2021, *PhRvD*, **103**, 024014
- Almheiri, A., Marolf, D., Polchinski, J., & Sully, J. 2013, *JHEP*, **2013**, 62
- Ashton, G., Hübner, M., Lasky, P. D., et al. 2019, *ApJS*, **241**, 27
- Belgacem, E., Dirian, Y., Foffa, S., & Maggiore, M. 2018, *PhRvD*, **97**, 104066
- Bernard, L., Blanchet, L., & Trestini, D. 2022, *JCAP*, **2022**, 008
- Berti, E., Barausse, E., Cardoso, V., et al. 2015, *CQGra*, **32**, 243001
- Berti, E., Cardoso, V., & Starinets, A. O. 2009, *CQGra*, **26**, 163001
- Berti, E., Cardoso, V., & Will, C. M. 2006, *PhRvD*, **73**, 064030
- Binney, J., & Tremaine, S. 1987, *Galactic Dynamics*, Princeton Series in Astrophysics (Princeton, NJ: Princeton Univ. Press)
- Calderón Bustillo, J., Sanchis-Gual, N., Leong, S. H. W., et al. 2023, *PhRvD*, **108**, 123020
- Calderón Bustillo, J., Sanchis-Gual, N., Torres-Forné, A., et al. 2021, *PhRvL*, **126**, 081101
- Cano, P. A., Fransen, K., Hertog, T., & Maenaut, S. 2023, *PhRvD*, **108**, 024040
- Cao, Z., Li, L.-F., & Wang, Y. 2014, *PhRvD*, **90**, 062003

- Cardoso, V., Franzin, E., Maselli, A., Pani, P., & Raposo, G. 2017, *PhRvD*, **95**, 084014
- Cardoso, V., & Pani, P. 2019, *LRR*, **22**, 4
- Carullo, G., Laghi, D., Johnson-McDaniel, N. K., et al. 2022, *PhRvD*, **105**, 062009
- Corman, M., Ripley, J. L., & East, W. E. 2023, *PhRvD*, **107**, 024014
- Dai, L., & Venumadhav, T. 2017, arXiv:1702.04724
- Dai, L., Zackay, B., Venumadhav, T., Roulet, J., & Zaldarriaga, M. 2020, arXiv:2007.12709
- Damour, T., & Nagar, A. 2014, *PhRvD*, **90**, 044018
- Deguchi, S., & Watson, W. D. 1986, *PhRvD*, **34**, 1708
- Dias, Ó. J. C., Godazgar, M., & Santos, J. E. 2022a, *JHEP*, **07**, 076
- Dias, Ó. J. C., Godazgar, M., Santos, J. E., et al. 2022b, *PhRvD*, **105**, 084044
- Dietrich, T., Bernuzzi, S., Brüggmann, B., Ujevic, M., & Tichy, W. 2018a, *PhRvD*, **97**, 064002
- Dietrich, T., Radice, D., Bernuzzi, S., et al. 2018b, *CQGra*, **35**, 24LT01
- Douchin, F., & Haensel, P. 2001, *A&A*, **380**, 151
- Eardley, D. M., Lee, D. L., & Lightman, A. P. 1973a, *PhRvD*, **8**, 3308
- Eardley, D. M., Lee, D. L., Lightman, A. P., Wagoner, R. V., & Will, C. M. 1973b, *PhRvL*, **30**, 884
- Evans, M., Adhikari, R. X., Afle, C., et al. 2021, arXiv:2109.09882
- Ezquiaga, J. M., Holz, D. E., Hu, W., Lagos, M., & Wald, R. M. 2021, *PhRvD*, **103**, 064047
- Ezquiaga, J. M., Hu, W., Lagos, M., Lin, M.-X., & Xu, F. 2022, *JCAP*, **2022**, 016
- Ezquiaga, J. M., & Zumalacárregui, M. 2020, *PhRvD*, **102**, 124048
- Finke, A., Foffa, S., Iacovelli, F., Maggiore, M., & Mancarella, M. 2021, *PhRvD*, **104**, 084057
- Foreman-Mackey, D. 2016, *JOSS*, **1**, 24
- Gamba, R., Akçay, S., Bernuzzi, S., & Williams, J. 2022, *PhRvD*, **106**, 024020
- Ghosh, A., Ghosh, A., Johnson-McDaniel, N. K., et al. 2016, *PhRvD*, **94**, 021101
- Ghosh, A., Johnson-McDaniel, N. K., Ghosh, A., et al. 2017, *CQGra*, **35**, 014002
- Giudice, G. F., McCullough, M., & Urbano, A. 2016, *JCAP*, **2016**, 001
- Gonzalez, A., Zappa, F., Breschi, M., et al. 2023, *CQGra*, **40**, 085011
- Goyal, S., Haris, K., Mehta, A. K., & Ajith, P. 2021, *PhRvD*, **103**, 024038
- Goyal, S., Vijaykumar, A., Ezquiaga, J. M., & Zumalacárregui, M. 2023, *PhRvD*, **108**, 024052
- Gupta, A., Arun, K. G., Barausse, E., et al. 2024, arXiv:2405.02197
- Hannuksela, O. A., Haris, K., Ng, K. K. Y., et al. 2019, *ApJL*, **874**, L2
- Harris, C. R., et al. 2020, *Natur*, **585**, 357
- Hunter, J. D. 2007, *CSE*, **9**, 90
- Janquart, J., Hannuksela, O. A., Haris, K., & Van Den Broeck, C. 2021a, *MNRAS*, **506**, 5430
- Janquart, J., Haris, K., & Hannuksela, O. 2022, GOLUM: A Software for Rapid Strongly-lensed Gravitational Wave Parameter Estimation, <https://github.com/lemnisl2/golum>
- Janquart, J., Haris, K., Hannuksela, O. A., & Van Den Broeck, C. 2023, *MNRAS*, **526**, 3088
- Janquart, J., More, A., & Van Den Broeck, C. 2023a, *MNRAS*, **519**, 2046
- Janquart, J., Seo, E., Hannuksela, O. A., Li, T. G. F., & Van Den Broeck, C. 2021b, *ApJL*, **923**, L1
- Janquart, J., Wright, M., Goyal, S., et al. 2023b, *MNRAS*, **526**, 3832
- Johansson, F. 2017, *ITCmp*, **66**, 1281
- Johnson-McDaniel, N. K., Ghosh, A., Ghonge, S., et al. 2022, *PhRvD*, **105**, 044020
- Johnson-McDaniel, N. K., Mukherjee, A., Kashyap, R., et al. 2020, *PhRvD*, **102**, 123010
- Kostelecký, V. A., & Mewes, M. 2016, *PhLB*, **757**, 510
- Li, P.-C., Lee, T.-C., Guo, M., & Chen, B. 2021, *PhRvD*, **104**, 084044
- Li, S.-S., Mao, S., Zhao, Y., & Lu, Y. 2018, *MNRAS*, **476**, 2220
- Liebling, S. L., & Palenzuela, C. 2023, *LRR*, **26**, 1
- LIGO Scientific Collaboration, Virgo Collaboration, & KAGRA Collaboration 2024, LVK Algorithm Library - LALSuite, Free software (GPL), doi:10.7935/GT1W-FZ16
- Liu, A., Wong, I. C. F., Leong, S. H. W., et al. 2023, *MNRAS*, **525**, 4149
- Liu, X., Hernandez, I. M., & Creighton, J. 2021, *ApJ*, **908**, 97
- Magaña Hernandez, I. 2022, arXiv:2211.01272
- Mazur, P. O., & Mottola, E. 2004, *PNAS*, **101**, 9545
- Mewes, M. 2019, *PhRvD*, **99**, 104062
- Mirshakari, S., Yunes, N., & Will, C. M. 2012, *PhRvD*, **85**, 024041
- Mishra, A., Krishnendu, N. V., & Ganguly, A. 2024, *PhRvD*, **110**, 084009
- More, A., & More, S. 2022, *MNRAS*, **515**, 1044
- Nagar, A., Bernuzzi, S., Del Pozzo, W., et al. 2018, *PhRvD*, **98**, 104052
- Nagar, A., Damour, T., Reisswig, C., & Pollney, D. 2016, *PhRvD*, **93**, 044046
- Nagar, A., Pratten, G., Riemenschneider, G., & Gamba, R. 2020a, *PhRvD*, **101**, 024041
- Nagar, A., Riemenschneider, G., Pratten, G., Retegno, P., & Messina, F. 2020b, *PhRvD*, **102**, 024077
- Nakamura, T. T., & Deguchi, S. 1999, *PTHPS*, **133**, 137
- Narola, H., Janquart, J., Haegel, L., et al. 2024, *PhRvD*, **109**, 084064
- Narola, H., Roy, S., & Sengupta, A. S. 2023, *PhRvD*, **107**, 024017
- Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, *ApJ*, **490**, 493
- Ng, K. K. Y., Wong, K. W. K., Broadhurst, T., & Li, T. G. F. 2018, *PhRvD*, **97**, 023012
- Nishizawa, A. 2018, *PhRvD*, **97**, 104037
- Nitz, A., Harry, I., Brown, D., et al. 2023, gwastro/pycbc: v2.2.2 release of PyCBC, v2.2.2, Zenodo, doi:10.5281/zenodo.8340277
- Okounkova, M., Farr, W. M., Isi, M., & Stein, L. C. 2022, *PhRvD*, **106**, 044067
- Okounkova, M., Stein, L. C., Moxon, J., Scheel, M. A., & Teukolsky, S. A. 2020, *PhRvD*, **101**, 104016
- Pani, P., Berti, E., & Gualtieri, L. 2013, *PhRvL*, **110**, 241103
- Payne, E., Isi, M., Chatzioannou, K., & Farr, W. M. 2023, *PhRvD*, **108**, 124060
- Pratten, G., Garca-Quirs, C., Colleoni, M., et al. 2021, *PhRvD*, **103**, 104056
- Price-Whelan, A. M., Lim, P. L., Earl, N., et al. 2022, *ApJ*, **935**, 167
- Punturo, M., Abernathy, M., Acernese, F., et al. 2010, *CQGra*, **27**, 084007
- Riemenschneider, G., Retegno, P., Breschi, M., et al. 2021, *PhRvD*, **104**, 104045
- Romero-Shaw, I. M., Talbot, C., Biscoveanu, S., et al. 2020, *MNRAS*, **499**, 3295
- Schneider, P., Ehlers, J., & Falco, E. E. 1992, Gravitational Lenses (Berlin: Springer Science & Business Media)
- Sharma, A., Roy, S., & Sengupta, A. S. 2024, *PhRvD*, **109**, 124049
- Shiralilou, B., Hinderer, T., Nissanke, S. M., Ortiz, N., & Witek, H. 2022, *CQGra*, **39**, 035002
- Siemonsen, N., & East, W. E. 2023, *PhRvD*, **108**, 124015
- Speagle, J. S. 2020, *MNRAS*, **493**, 3132
- Srivastava, M., Chen, Y., & Shankaranarayanan, S. 2021, *PhRvD*, **104**, 064034
- Tahura, S., & Yagi, K. 2018, *PhRvD*, **98**, 084042
- Takahashi, R., & Nakamura, T. 2003, *ApJ*, **595**, 1039
- Uchikata, N., Yoshida, S., & Pani, P. 2016, *PhRvD*, **94**, 064015
- Ujevic, M., Rashti, A., Gieg, H., Tichy, W., & Dietrich, T. 2022, *PhRvD*, **106**, 023029
- van Rossum, G., 1995 Python Tutorial CS-R9526, Centrum Wiskunde & Informatica
- Varma, V., Field, S. E., Scheel, M. A., et al. 2019, *PhRvR*, **1**, 033015
- Vijaykumar, A., Mehta, A. K., & Ganguly, A. 2023, *PhRvD*, **108**, 043036
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *NatMe*, **17**, 261
- Wade, L., Creighton, J. D. E., Ochsner, E., et al. 2014, *PhRvD*, **89**, 103012
- Wagle, P., Yunes, N., & Silva, H. O. 2022, *PhRvD*, **105**, 124003
- Wang, Y., Lo, R. K. L., Li, A. K. Y., & Chen, Y. 2021, *PhRvD*, **103**, 104055
- Wang, Y., Stebbins, A., & Turner, E. L. 1996, *PhRvL*, **77**, 2875
- Will, C. M. 1998, *PhRvD*, **57**, 2061
- Will, C. M. 2014, *LRR*, **17**, 4
- Williams, D., Veitch, J., Chiofalo, M. L., et al. 2023, *JOSS*, **8**, 4170
- Wright, M., & Hendry, M. 2022, *ApJ*, **935**, 68
- Wright, M., Liu, A., Seo, E., & Wong, I. C. F. 2022, Gravelamps: Gravitational Wave Lensing Mass Profile Model Selection, <https://git.ligo.org/mick.wright/Gravelamps>
- Yagi, K., Stein, L. C., Yunes, N., & Tanaka, T. 2012, *PhRvD*, **85**, 064022
- Yang, H., Nichols, D. A., Zhang, F., et al. 2012, *PhRvD*, **86**, 104006
- Zhu, T., Zhao, W., Yan, J.-M., et al. 2024, *PhRvD*, **110**, 064044