Probing lens-induced gravitational-wave birefringence as a test of general relativity

Srashti Goyal,[1,*] Aditya Vijaykumar,[1,2,†] Jose María Ezquiaga,[3,‡] and Miguel Zumalacárregui[4,§]

[1]*International Centre for Theoretical Science, Tata Institute of Fundamental Research, Bangalore-560089, India*

[2]*Department of Physics, The University of Chicago, 5640 South Ellis Avenue, Chicago, Illinois 60637, USA*

[3]*Niels Bohr International Academy, Niels Bohr Institute, Blegdamsvej 17, DK-2100 Copenhagen, Denmark*

[4]*Max Planck Institute for Gravitational Physics (Albert Einstein Institute) Am Mühlenberg 1, D-14476 Potsdam-Golm, Germany*

(Received 26 January 2023; accepted 5 July 2023; published 21 July 2023)

Theories beyond general relativity (GR) modify the propagation of gravitational waves (GWs). In some, inhomogeneities (aka, gravitational lenses) allow interactions between the metric and additional fields to cause lens-induced birefringence (LIB): a different speed of the two linear GW polarizations (+ and ×). Inhomogeneities then act as nonisotropic crystals, splitting the GW signal into two components whose relative time delay depends on the theory and lens parameters. Here we study the observational prospects for GW scrambling, i.e. when the time delay between both GW polarizations is smaller than the signal's duration and the waveform recorded by a detector is distorted. We analyze the latest LIGO-Virgo-KAGRA catalog, GWTC-3, and find no conclusive evidence for LIB. The highest log Bayes factor that we find in favor of LIB is 3.21 for GW190521, a particularly loud but short event. However, when accounting for false alarms due to (Gaussian) noise fluctuations, this evidence is below 1σ. The tightest constraint on the time delay is < 0.51 ms at 90% confidence level (CL) from GW200311_115853. From the nonobservation of GW scrambling, we constrain the optical depth for LIB, accounting for the chance of randomly distributed lenses (e.g. galaxies) along the line of sight. Our LIB constraints on a (quartic) scalar-tensor Horndeski theory are more stringent than Solar System tests for a wide parameter range and comparable to GW170817 in some limits. Interpreting GW190521 as an active galactic nucleus (AGN) binary (i.e. taking an AGN flare as a counterpart) allows even more stringent constraints. Our results demonstrate the potential and high sensitivity achievable by tests of GR, based on GW lensing.

DOI: 10.1103/PhysRevD.108.024052

I. INTRODUCTION

The detection of gravitational waves (GWs) using the LIGO-Virgo-KAGRA (LVK) detectors [1–3] from mergers of compact objects [4–11] has enabled precision tests of general relativity (GR) in the strong-field regime [12–15]. Far away from the source, GR predicts that GWs are well described as linear perturbations of the background Friedmann-Robertson-Walker (FRW) metric [16]. Existing propagation tests hence typically consider modifications over the FRW background and its effect on the GW signals as measured at the detectors [17–19].

GR also dictates that GWs have only two tensor polarizations (+, ×) which propagate independently at the speed of light. However, in alternative theories of gravity, extra degrees of freedom (tensor, vector, scalar) [20] can mix with GWs as they propagate, producing phenomena similar to neutrino oscillations (i.e. due to interactions between different neutrino flavors [21]). In Lorentz-invariant theories, the symmetries of the FRW metric restrict mixing effects to tensor degrees of freedom, either fundamental (e.g. in bigravity) or composite (e.g. multiple vector fields) [22–26]. However, inhomogeneities spontaneously break Lorentz symmetry, allowing interaction between GWs and scalar or vector degrees of freedom [27,28]. This leads to new, testable predictions, and opens new opportunities to probe the gravitational sector beyond the FRW limit.

*srashti.goyal@icts.res.in
†aditya.vijaykumar@icts.res.in
‡jose.ezquiaga@nbi.ku.dk
§miguel.zumalacarregui@aei.mpg.de

The evolution of GWs on an inhomogeneous background is described via propagation eigenstates: linear combinations of the interaction eigenstates ($h_+, h_\times$ and perturbations of additional polarizations) with a well-defined dispersion relation (analogous to massive neutrinos). As the relation between interaction and propagation eigenstates and their speed depends on position and direction, an inhomogeneous region of space splits the original signal into several components, each arriving with a relative time delay [27]. Moreover, if deviations from GR are small, two eigenstates correspond to *mostly tensorial* polarizations (linear combinations of $h_+, h_\times$ plus a negligible correction distinguishing both), with a very small speed difference.[1]

We will refer to the difference in propagation speed between the $+, \times$ polarizations as *lens-induced birefringence* (LIB). LIB is analogous to the way a nonisotropic crystal, such as calcite, splits light into two beams. This splitting is caused by a difference in the refractive index of the linear electromagnetic polarizations, which depends on the alignment of the polarization vector with the crystal structure. In our case birefringence is caused not by anisotropies in a crystal, but by the background configuration of additional, non-GR fields which spontaneously break Lorentz symmetry. Moreover, LIB splitting is independent of the frequency (in the high-frequency approximation assumed), which would correspond to a perfectly isochromatic birefringent crystal. Because GW detectors have excellent time resolution and bad sky localization, our main observable will be the time delay between split signals and not their angular separation.

If the arrival time difference between the mostly tensorial polarizations is larger than the duration of the binary merger signal then we would see only one polarization at a time, mimicking a binary with either a zero inclination angle (face-on binary) or a 90° inclination angle (edge-on binary), appearing as GW echoes. Since the detectors are more sensitive to face-on binaries, one can expect an excess of near-zero inclinations in the case of birefringence for the population of binaries. If the delay between the polarizations is larger than typical observing runs or the amplitude of one of the polarizations decays faster than the other, e.g. in Chern-Simons gravity [29], one would also expect an anisotropic inclination distribution. Current observations though show that the orientation distribution is consistent with being isotropic [30].

However, when the time delay is less than the duration the signal, the GW waveform would be distorted or "scrambled" due to the interference of both polarizations. Note that this effect is frequency independent, and hence distinguishable from a different dispersion relation for the $+$ and $\times$ modes or the circularly polarized combinations (L-R), as predicted in GR [31–34] and alternative theories [35–37]. As it is not suppressed by the frequency, LIB is the dominant effect in the high-frequency limit for theories in which this effect is present.

Our study analyzes for the first time the arrival time difference (Δt_{12}) between the two polarization states due to different propagation speeds (frequency-independent dispersion relations) as a result of LIB. This is a new, model-independent test of a basic prediction of GR. We use these generic results to constrain GW lensing effects beyond GR, for example in scalar-tensor theories with derivative interactions [27].

LIB signatures are not linked to a specific regime of gravitational lensing in GR, such as strongly magnified or multiple images. The scale on which LIB can be observed is very sensitive to the theory parameters and independent of the Einstein radius R_E, which characterizes the regimes of gravitational lensing. Hence, for sufficiently strong deviations from GR, LIB can be detected for impact parameters much larger than R_E, typically associated with weak lensing. Therefore, LIB tests can be applied to all the GW detections. In addition, LIB can be important for lenses very close to the source or the observer, for which R_E vanishes. This is particularly interesting for sources merging near massive objects [e.g. a supermassive black hole (SMBH)] since the background configuration of the additional fields enhances LIB.

The rest of the paper is organized as follows. In Sec. II we describe our LIB waveform model and methods for data analysis and introduce parametrized LIB observation probabilities. In Sec. III, we perform the birefringence test over a set of simulated GW events, and then over real events using the Bayesian model selection framework. In Sec. IV, we study the implications of the results in constraining LIB probabilities and beyond-GR theories. Finally, in Sec. V we summarize the main results and discuss future prospects.

II. METHOD

In GR, GWs have only two polarizations $(+, \times)$ which propagate independently at the speed of light over the background FRW metric. A given ground-based detector I measures the GW signal, $h(t)$ as a linear combination of these polarizations [38],

$$h_I(t) = F_I^+ h_+(t) + F_I^\times h_\times(t) \tag{1}$$

where, $F_I^+, F_I^\times$ are the detector antenna pattern functions. In the case of compact binary coalescence (CBC), the relative amplitude of the polarization modes depends on the inclination and polarization angles of the binary with respect to the line of sight, and also depends on the sensitivity of the detector for the source location at the

[1]We will ignore the remaining eigenstates (perturbations of beyond-GR fields plus negligible corrections) because 1) their emission needs to be suppressed to avoid dipolar radiation and 2) their speed can be substantially different, making an association with the mostly tensorial part of the signal difficult [27].

time of arrival. The overall amplitude of the signal is inversely proportional to the luminosity distance of the source. The masses and spins of the source dictate the frequency evolution of the signal and its amplitude.

A. Parameterized lens-induced birefringence waveforms

When there is any inhomogeneity along the travel path of a GW, e.g. an intervening galaxy, the GW can be gravitationally lensed. Gravitational lensing of a GW can produce multiple images of the original signal (strong lensing) or cause distortions (microlensing) [39], but, in GR, both polarizations are affected in the same way, i.e. the polarization rotation is negligible for any sensible astrophysical lens [40–42]. However, in alternative theories of gravity the additional fields can couple with the tensor polarizations around the lens and modify the GW propagation eigenstates. These eigenstates are a linear combination of original GW polarizations that evolve independently, each with a different speed, thus reaching the detectors at different times. We will assume spherically symmetric lenses, focus on the limit of small deviations from GR, so the mostly metric propagation eigenstates correspond to linear combinations of $h_+, h_\times$ (depending on the projected angle between the lens and the source), and neglect the additional eigenstates (see Ref. [27] and Footnote 1).

This class of LIB of GWs can be captured in a phenomenological manner as proposed in Ref. [27]. After diagonalizing the propagation equations, the propagation eigenstates can be computed and one gets the transformation matrix $\mathcal{S}$ relating the polarization amplitudes in GR and after the LIB:

$$[h_+, h_\times]^T_{\rm LIB} = \mathcal{S}[h_+, h_\times]^T_{\rm GR} \tag{2}$$

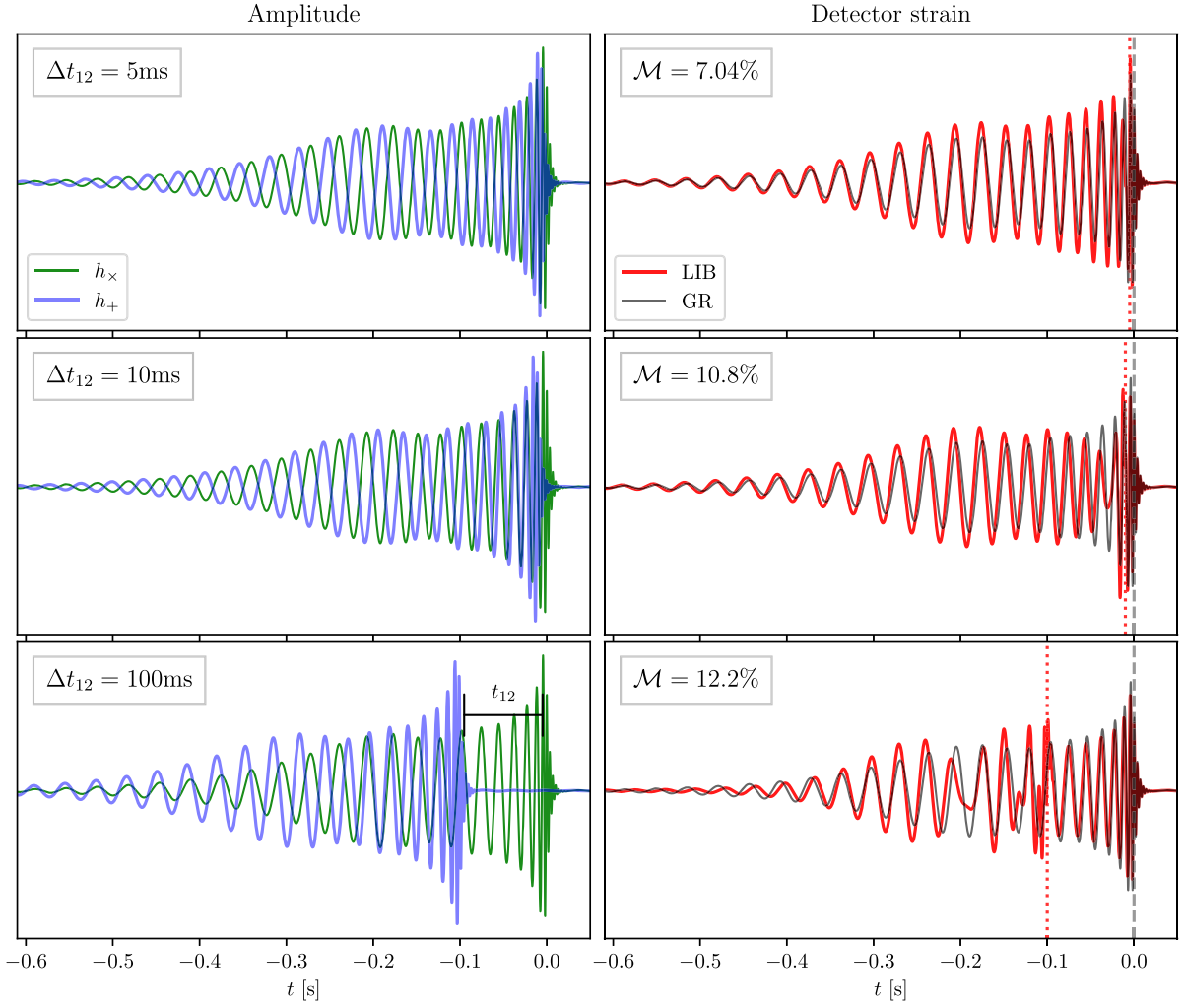


FIG. 1. GW polarizations (left) and detector strain (right) for a CBC $(30 + 30)M_\odot$ with birefringent time delays $\Delta t_{12} = 5, 10, 100$ ms (top to bottom). The sky localization and detector orientation correspond to $F^+ = -0.38$, $F^\times = 0.71$ and the LIB strain is given by Eq. (5).

where

$$\mathcal{S} = \hat{\mathcal{M}}\text{diag}(1, \Delta)\hat{\mathcal{M}}^{-1}, \tag{3}$$

$$\hat{\mathcal{M}} = \begin{bmatrix} -\sin(2\phi_{\text{lens}}) & \cos(2\phi_{\text{lens}}) \\ \cos(2\phi_{\text{lens}}) & \sin(2\phi_{\text{lens}}) \end{bmatrix}, \tag{4}$$

and $\Delta = e^{-i\omega\Delta t_{12}}$ with Δt_{12} is the time delay between the polarizations and ϕ_{lens} is the angle between the lens and the source, relative to the direction of GW propagation that dictates the polarization mixing.

It is easy to note that for $\phi_{\text{lens}} = \pi/2$, $\mathcal{S} = \text{diag}(1, \Delta)$, and hence the signal observed by the detectors will just be a superposition of $(+, \times)$ arriving at different times.

$$h_I^{\text{LIB}}(t) = F_I^+ h_+(t) + F_I^\times h_\times(t - \Delta t_{12}) \tag{5}$$

whereas, if $\Delta t_{12} = 0$ the LIB waveform morphology will be identical to the GR one, independent of ϕ_{lens}. Figure 1 compares GR and LIB waveform polarizations and the detector strains for various values of Δt_{12}. Under LIB, the polarizations interfere leading to waveform distortions.

Since lensing is an environmental effect that can occur through any local inhomogeneity in the path of GWs, the parameters Δt_{12} and ϕ_{lens} are expected to vary between GW events. The time delay distribution depends on the theory and the (usually unknown) lens properties and the configuration relative to the source. In general, one can only predict the probability of the birefringence parameters given a gravitational theory and matter distribution (unless further information or assumptions are employed about the source's location or the signal's trajectory); see Sec. II D. This is in stark contrast to other tests of GW propagation (that are done with individual GW events) in which deviations represent a fundamental property of gravity (e.g. massive graviton dispersion relations) and are thus the same across all events and only depend on their distance [20].

B. Template mismatch studies

In order to quantify distortions due to GW birefringence, we calculate the mismatch between the GR and LIB waveforms as seen by the LIGO-Virgo detectors. At each detector (I), the mismatch between the injected waveform (h_I^{inj}) and the recovery waveform (h_I^{rec}) is given by

$$M_I = 1 - \frac{(h_I^{\text{inj}}|h_I^{\text{rec}})}{||h_I^{\text{inj}}||.||h_I^{\text{rec}}||} \tag{6}$$

where, $(\cdot|\cdot)$ symbolizes the noise-weighted inner product:

$$(a|b) \equiv 2\int_{f_{\text{min}}}^{f_{\text{max}}} \frac{\tilde{a}(f)\tilde{b}^*(f)}{S_n(f)} df. \tag{7}$$

Here, $\tilde{a}, \tilde{b}$ represent the Fourier transform of the time series $a(t)$, $b(t)$, $[f_{\text{min}}, f_{\text{max}}]$ is the frequency range over which the inner product is evaluated, * represents complex conjugation and $S_n(f)$ is the colored Gaussian noise power spectral density (PSD) at the detector. The norm $||h|| = \sqrt{(h|h)}$ is the optimal signal-to-noise ratio (SNR) of a waveform. We define the total mismatch ($\mathcal{M}$) for a network of detectors as the SNR (ρ_I) squared weighted average of individual detector matches,

$$\mathcal{M} = \frac{\sum_I \rho_I^2 M_I}{\sum_I \rho_I^2}. \tag{8}$$

Note that the mismatch is a normalized quantity and is maximized over time and phase shifts. Thus, the mismatch quantifies differences in morphology between the signals, whereas during the parameter estimation (PE) from GW signals both the mismatch and SNR play a role. The log-likelihood can be approximated as, $\approx \frac{1}{2}\sum_I (h_I^{\text{inj}} - h_I^{\text{rec}}|h_I^{\text{inj}} - h_I^{\text{rec}}) \sim \sum_I \rho_I^2 M_I \sim \mathcal{M}\sum_I \rho_I^2$ around the maximum likelihood parameters (especially time and phase) where the injected and recovery waveforms have similar optimal SNRs while assuming zero noise realization in the injection. We first wish to quantify the overall detectability of the birefringence. Later, we will estimate parameters using Bayesian inference, accounting for correlations between all parameters.

Figure 2 shows frequency domain LIB and GR waveforms for a GW150914-like CBC. The waveforms are generated using the approximant IMRPhenomXPHM [43], as implemented in the LALSimulation module of the LALSuite software package [44]. The waveforms are then projected onto the LIGO and Virgo detectors using their antenna pattern functions, as implemented in the Bilby [45] software package. The LIB waveforms have additional frequency modulations which depend on the two parameters: Δt_{12} and ϕ_{lens} (see Sec. II A). Therefore we calculate the mismatch between the GR and LIB waveforms, keeping all the other parameters identical and fixed for the two waveforms. In practice, we calculate M_I using the pycbc.filter module [46]. The detector noise is generated using the zero-detuned high-power PSDs of Advanced LIGO and Advanced Virgo at their design sensitivities [47,48].

We consider two systems of binaries: one whose parameters resemble that of the first CBC detection GW150914 and one whose parameters resemble that of the higher mass-ratio CBC GW190814 where the presence of higher-order modes (HoMs) of GWs are significant. For both the systems, we inject a GR waveform and a LIB waveform and recover with the LIB waveform to calculate the mismatch. The parameters for both of the CBCs are mentioned in Appendix A.

Figure 3 shows mismatches for a GW150914-like CBC (top) and GW190814-like CBC (bottom). As expected for a GR injection i.e. $\Delta t_{12}^{\text{inj}} = 0$ and $\phi_{\text{lens}}^{\text{inj}} = 0$, the mismatch is a

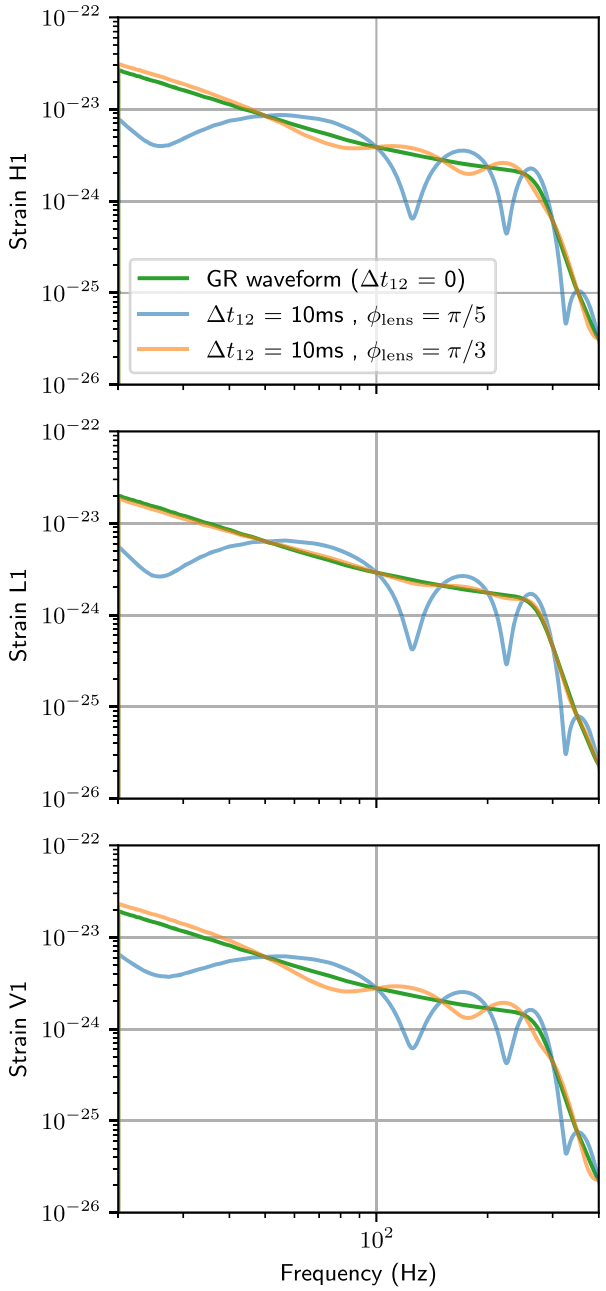


FIG. 2. GR and LIB detector frame waveform amplitudes in the frequency domain of a GW150914-like CBC. The birefringence leads to additional frequency modulations and distorts the GR waveform. The magnitude of these distortions are however dependent on the two parameters: Δt_{12} and $\phi_{\rm lens}$.

minimum for $\Delta t_{12}^{\rm rec} \simeq 0$, for all $\phi_{\rm lens}$ as expected from Eq. (3). Additionally, the local minimum of mismatch is at $\phi_{\rm lens}^{\rm rec} \simeq \pi/3$, which could be because of vanishing polarization (+ or ×) as seen at the detectors which further makes the mismatch independent of the time delay Δt_{12}. The waveform plots in Fig. 2 confirm this as the $\phi_{\rm lens}^{\rm rec} \simeq \pi/3$ waveform resembles the GR ones more as compared to the $\phi_{\rm lens}^{\rm rec} \simeq \pi/5$, especially in the Livingston (L1) detector for a GW150914-like CBC.

We also check the mismatch for the LIB injections (right panel Fig. 3) with $\Delta t_{12}^{\rm inj} = 10$ ms and $\phi_{\rm lens}^{\rm inj} = \pi/5$ and the mismatch is a minimum at, $\Delta t_{12}^{\rm rec} \simeq \pm 10$ ms and $\phi_{\rm lens}^{\rm rec} \simeq \pi/5, \pi/4 + \pi/5$. We can infer the degeneracy between Δt_{12} and the coalescence time (t_c) as follows: from Eqs. (1)–(4) if $\phi_{\rm lens} \to \phi_{\rm lens} + \pi/4$ and $\Delta \to 1/\Delta$ then, one finds $S \to S/\Delta$, which implies that the transformation matrix $\mathcal{S}$ in Eq. (3) is invariant when $(\Delta t_{12}, \phi_{\rm lens}, t_c)$ is transformed to $(-\Delta t_{12}, \phi_{\rm lens} + \pi/4, t_c + \Delta t_{12})$. This degeneracy stems from the fact that we do not know the composition of $h_{+,\times}$ before it encounters the lens. Higher harmonics of the waveform can provide this additional information, as the amplitude of the +, × polarizations for each harmonic has a different dependence on the inclination angle (cf. Fig. 1 in Ref. [49]). Sources with a high mass ratio may thus distinguish the sign of Δt_{12}. We leave these investigations for the future.

C. Bayesian inference

Bayesian model selection allows us to assign evidences for various hypotheses pertaining to the observed data, and also derive posterior probability distributions of the model parameters conditioned on individual hypotheses. Given the set of data $\{d\}$ from a network of detectors, the marginalized likelihood (or, Bayesian evidence) of the hypothesis $\mathcal{H}_\mathcal{A}$ can be computed by

$$P(\{d\}|\mathcal{H}_\mathcal{A}) = \int d\theta P(\theta|\mathcal{H}_\mathcal{A}) P(\{d\}|\theta, \mathcal{H}_\mathcal{A}), \qquad (9)$$

where θ is a set of parameters that describe the signal under hypothesis $\mathcal{H}_\mathcal{A}$ (including the masses and spins of the compact objects in the binary, the location and orientation of the binary and the arrival time and phase of the signal), $P(\theta|\mathcal{H}_\mathcal{A})$ is the prior distribution of θ under hypothesis $\mathcal{H}_\mathcal{A}$, and $P(\{d\}|\theta, \mathcal{H}_\mathcal{A})$ is the likelihood of the data $\{d\}$, given the parameter vector θ and hypothesis $\mathcal{H}_\mathcal{A}$. Given the hypothesis $\mathcal{H}_\mathcal{A}$ and data $\{d\}$, we can sample and marginalize the likelihood over the parameter space using an appropriate stochastic sampling technique such as nested sampling [50].

Bayesian model selection allows us to compare multiple hypotheses. For e.g., the odds ratio $\mathcal{O}_{\rm GR}^{\rm LIB}$ is the ratio of the posterior probabilities of the two hypotheses LIB and GR. When $\mathcal{O}_{\rm GR}^{\rm LIB}$ is greater than one then hypothesis LIB is preferred over GR and vice versa. Using Bayes theorem, the odds ratio can also be written as the product of the ratio of the prior odds $P_{\rm GR}^{\rm LIB}$ of the hypotheses and the likelihood ratio, or Bayes factor $\mathcal{B}_{\rm GR}^{\rm LIB}$:

$$\mathcal{O}_{\rm GR}^{\rm LIB} := \frac{P(\mathcal{H}^{\rm LIB}|\{d\})}{P(\mathcal{H}^{\rm GR}|\{d\})} = \frac{P(\mathcal{H}^{\rm LIB})}{P(\mathcal{H}^{\rm GR})} \times \frac{P(\{d\}|\mathcal{H}^{\rm LIB})}{P(\{d\}|\mathcal{H}^{\rm GR})} \qquad (10)$$

$$= \mathcal{P}_{\rm GR}^{\rm LIB} \times \mathcal{B}_{\rm GR}^{\rm LIB}. \qquad (11)$$

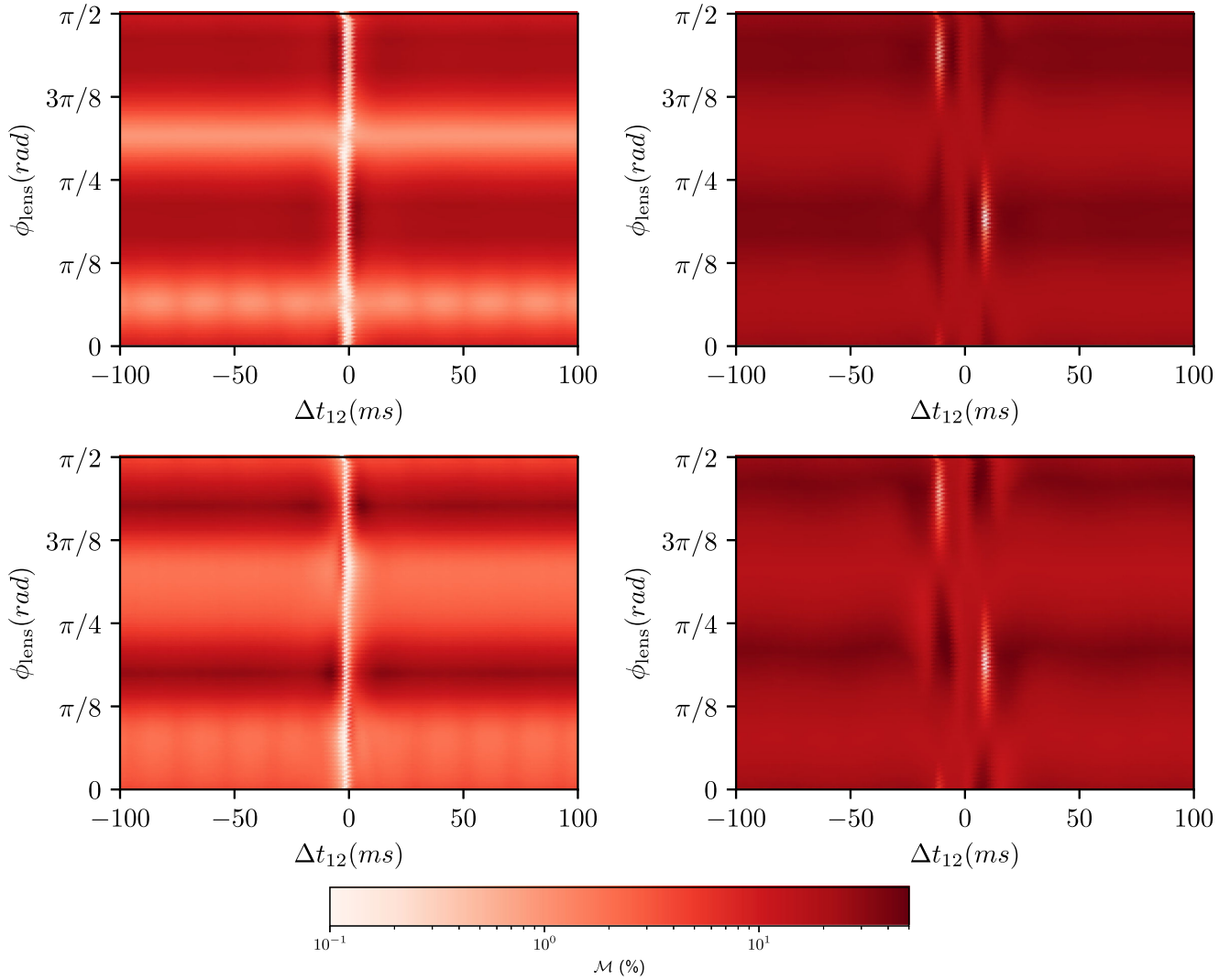


FIG. 3. Mismatch between GR and LIB waveforms for a GW150914-like CBC (top) and GW190814-like CBC (bottom). Left panel: GR injection i.e. $\Delta t_{12}^{\rm inj} = 0$ and $\phi_{\rm lens}^{\rm inj} = 0$. The mismatch is a minimum for $\Delta t_{12}^{\rm rec} \simeq 0$. Right panel: a LIB injection with $\Delta t_{12}^{\rm inj} = 10$ ms and $\phi_{\rm lens}^{\rm inj} = \pi/5$. The mismatch is a minimum at $\Delta t_{12}^{\rm rec} \simeq \pm 10$ ms and $\phi_{\rm lens}^{\rm rec} \simeq \pi/5, \pi/4 + \pi/5$.

Since GR has been tested well in a variety of settings, our prior odds are going to be highly biased towards it, i.e., $\mathcal{P}_{\rm GR}^{\rm LIB} \ll 1$. Hence, in order to claim evidence of birefringence the corresponding Bayes factor supporting the LIB hypothesis has to be very large. Since the Bayes factor is the only quantity that is derived from data, for the rest of the paper, we focus on the Bayes factor, i.e. the ratio of evidences under the two hypotheses.

The waveforms under the GR and LIB hypotheses at each detector are the same as those described in Sec. II B. We use the standard Gaussian likelihood model for estimating the posteriors of the parameters under different hypotheses (see, e.g., Ref. [38]). We use uniform priors in redshifted component masses of the binary, isotropic sky location (uniform in $\alpha, \sin\delta$) and orientation (uniform in $\cos\iota, \phi_0$), uniform in polarization angle ψ, and a prior $\propto d_L^2$ on the luminosity distance. Additionally for the LIB hypothesis, we choose the priors on Δt_{12} as uniform $\in [-100, 100]$ ms and $\phi_{\rm lens}$ as uniform $\in [0, \pi/2]$. To estimate the posterior distribution and evidences for the GR and LIB hypotheses, we use the open-source parameter estimation package bilby [45] coupled with the dynamical nested sampler dynesty [51].

D. Lensing probabilities

The (non)observation of birefringence can help us put constraints on theories beyond GR that predict LIB. According to GR, the strong lensing of GWs caused by galaxies occurs when sources lie inside the Einstein radius of the lens, which depends on the lens mass and profile. This is the relevant scale determining the probability of lensing. However, birefringence beyond GR is in principle independent of the ratio between the impact parameter and

the Einstein radius, changing the probability of observing LIB compared to strong lensing. It is thus possible to have LIB time delays without multiple images, but birefringence could also occur for strongly lensed GWs, in which case it applies to each image separately, as typical time delays between images will be larger than Δt_{12} [27].

Assuming that the lenses are randomly distributed, birefringence detection is described by Poisson statistics. A series of observations with L lensed and U unlensed GW events has an associated probability,

$$P = \exp\left(-\sum_i^U \lambda_i\right) \prod_j^L (1 - e^{-\lambda_j}). \tag{12}$$

The result depends on the LIB rate for the ith event: $\lambda_i = \int dz_s d\vec{p}_L d\vec{p}_S \tau(z_s, \vec{p}_L) \mathcal{P}_i(z_s, \vec{p}_S) P(\vec{p}_S, \vec{p}_L)$. Here S, L denote parameters corresponding to the source and lens/theory (i.e. beyond GR), $\mathcal{P}_i$ is the posterior distribution of the source parameters and P is the prior, which includes relations between parameters (i.e. the measured Δt_{12} as a function of lens mass and beyond-GR parameters).[2] The *birefringence optical depth*, $\tau(z_s, \vec{p}_L)$ is the fraction of the sky for which LIB is detectable for sources at a redshift z_s. Hereafter we will assume the posterior to be sharply peaked at the mean source redshift z_s and include the integration on the lens model parameters ($\vec{p}_L$) in the definition of the optical depth, so $\lambda_i \approx \tau(z_{s,i})$. If birefringence is excluded in all events, the probability only depends on the total optical depth $\tau_{\rm tot} \approx \sum_{i=1\ldots N} \tau_i \approx \sum_{i=1\ldots N} \lambda_i$. In Appendix B we comment on opportunities to study GW birefringence beyond the Poisson statistics.

The lensing optical depth $\tau(z_s)$ depends on the angular cross section $\hat{\sigma}(z_s, \vec{p}_L)$ and the density of lenses $\hat{n}(\vec{p}_L)$ [53]. In the following we will explicitly write the lens redshift z_L and let $\vec{p}_{L'}$ denote the remaining properties (i.e. lens mass and theory parameters). The total density of lenses at redshift z_L is then $\int \hat{n}(z_L, \vec{p}_{L'}) d\vec{p}_{L'}$. The optical depth is computed directly by adding up the cross sections weighted by the density at different redshifts, i.e.

$$\tau(z_s) = \int_0^{z_s} dz_L \int d\vec{p}_{L'} \frac{dV_c}{\delta\Omega dz_L} \hat{n}(z_L, \vec{p}_{L'}) \hat{\sigma}(z_s, z_L, \vec{p}_{L'}) \tag{13}$$

where $dV_c = \delta\Omega D_L^2 \frac{dz}{(1+z)H(z)}$ is the physical volume given the solid angle $\delta\Omega$, angular diameter distance to the lens D_L and the Hubble parameter $H(z)$. For simplicity, we will assume point-mass lenses of mass M throughout. In GR, the lensing cross section is $\sigma = \pi\theta_E^2$, where $\theta_E = R_E/D_L = \left(\frac{4GMD_{LS}}{c^2 D_L D_S}\right)^{1/2}$ is the Einstein angle, D_S is the distance to the source from the Earth and D_{LS} is the distance between the lens and source.

The relation between the LIB-time delays, the theory parameters and the configuration of the lensed system is complex (see Sec. VI of Ref. [27] for a worked-out example in a viable Horndeski theory). For this reason, we will first consider two phenomenological models of lensing cross section. As a first example, we will assume that the relevant LIB scale is proportional to the Einstein angle, $\theta_X^E = \alpha_X \theta_E$, so that the cross section becomes

$$\sigma_X^E = \pi \alpha_X^2 \theta_E^2. \tag{14}$$

Then the optical depth is given by Eq. (74) in Ref. [27] where the lenses were assumed to be point-like.

Under these assumptions, lensing probabilities are independent of the mass function.[3]

In our second example we assume that the relevant LIB scale is given by a constant physical scale associated with each halo. Moreover, we assume that this scale depends on the halo mass as a power law. Accordingly, the cross section reads,

$$\sigma_{\rm ph} = \pi \frac{R_{12}^2}{D_L^2} \left(\frac{M}{10^{12} M_\odot}\right)^{2n}. \tag{15}$$

The scale R_{12} fixes the probability of lensing for halos with $M = 10^{12} M_\odot$, while n allows us to extrapolate to different halo masses. Below we will discuss some cases of interest.

We now generalize the expression for the optical depth presented in Ref. [27] [Eq. (76)] to include a realistic halo mass function. The optical depth from Eq. (15) is given by

$$\tau^{\rm ph}(z_s, n) = \Omega_M h \left(\frac{R_{12}}{22\ {\rm kpc}}\right)^2 \hat{\tau}(z_s, n), \tag{16}$$

where

$$\hat{\tau}(z_s, n) = \int_0^{z_s} dz \frac{(1+z)^2}{H(z)/H_0} \int d\log(M) \times \left(\frac{M}{10^{12} M_\odot}\right)^{2n-1} f(M, z). \tag{17}$$

Here $f(M, z) = \frac{M^2}{\rho_0} \frac{d\hat{n}}{dM}$ is the scaled differential mass function (dimensionless) with ρ_0 being the matter density

[2] A more complete treatment should account for the selection function [52]. In terms of gravitational lensing, we expect that Δt_{12} correlates with magnification especially for sizable impact parameters (e.g. single image regime, first magnified image if multiple images are formed), which dominate the lensing cross section. Then events with larger Δt_{12} are more likely to be observed and neglecting this correlation is conservative.

[3] The mass independence also appears for the strong-lensing cross section for a distribution of point lenses [54,55]. This differs from the strong-lensing cross section for extended lenses, where the lens mass affects the formation of multiple images [53].

of the Universe at $z = 0$. We will use the Tinker *et al.* form [56] as implemented in the Colossus package [57]. We integrate in the mass range $M \in [10^7, 10^{17}]M_\odot$. The upper limit includes the most massive halos, whose number is exponentially suppressed well below our upper limit. The lower limit has a negligible impact if $n > 1/2$, as the integral is dominated by $M \sim 10^{11}$–$10^{14}M_\odot$ (see the strong-lensing curves in Fig. 9 in Ref. [58], which have the same mass dependence as $n = 1/2$). The lower limit is important when $n < 1/2$, as we will discuss below.

As our approach is phenomenological, we assume a Planck ΛCDM cosmology [59]. The true optical depth of a consistent LIB model will typically depend more strongly on the theory parameters [e.g. entering Eq. (15) via R_{12}] than on the precise values of $H(z), f(M, z)$ of the underlying LIB cosmology, including the effects of deviations from GR in cosmological expansion and structure formation. This is the case for the example theory discussed in Sec. IV B: GW lensing effects are orders of magnitude more sensitive than Solar System tests (cf. Fig. 8), which are in turn more stringent than current cosmological observations [60,61] (for theories without a screening mechanism).

In addition to the Einstein radius scaling, Eq. (14), we will consider three cases of interest:

(1) $n = 1$: The physical scale is proportional to the total halo mass, much like the Schwarzschild radius. The rates are dominated by large masses and saturate at $z_s \gtrsim 1$, as the more massive halos are exponentially suppressed at early times. This case captures the dependence of the time delay in a Horndeski model (Sec. IV B).
(2) $n = 1/2$: The scale has the same mass scaling as the Einstein radius and leads to rates independent of M. However, the overall redshift dependence is different, as R_E depends also on D_S, D_{LS}.
(3) $n = 1/3$: This mass scaling favors lighter halos and thus grows very rapidly with redshift. It is motivated by the mass dependence of the Vainshtein radius R_V, i.e. the classical strong-coupling scale [62]. For $n = 1/3$ the contribution from lighter halos diverges and a low mass cutoff needs to be included (we will take $M > 10^7 M_\odot$). We will see this mass dependence when considering a binary merging near an active galactic nucleus in a Horndeski theory (Sec. IV C).

The optical depths for each of the cases as functions of the source redshift are plotted in Fig. 7. Note that these phenomenological models assume that the cross section is independent of the LIB-time delays, and is common for all the analyzed events. Dependence on the time delay can be included, e.g. by multiplying Eqs. (14) and (15) by a factor $(\Delta t_{12}/10\ \mathrm{ms})^{-k}$. For the sake of simplicity, we will not include this dependence and instead interpret the obtained values of α_X, $R_{12}(n)$ at the median 95% CL from all analyzed events.

III. RESULTS

In order to test our method and understand the observing capabilities, we first apply our pipeline to injections. We then proceed to analyze the latest GW catalog (GWTC-3).

A. Injections

We inject GW150914-like signals in simulated Gaussian noise with $\Delta t_{12}^{\mathrm{inj}} \in \{0, 1, 3, 10, 30\}$ ms and $\phi_{\mathrm{lens}}^{\mathrm{inj}} = \pi/5$ rad and recover them by running the parameter estimation routines under the GR and LIB hypotheses, as mentioned in Sec. II C. This allows us to compute the Bayes factors $\mathcal{B}^{\mathrm{LIB}}_{\mathrm{GR}}$ to compare the two hypotheses for each injection.

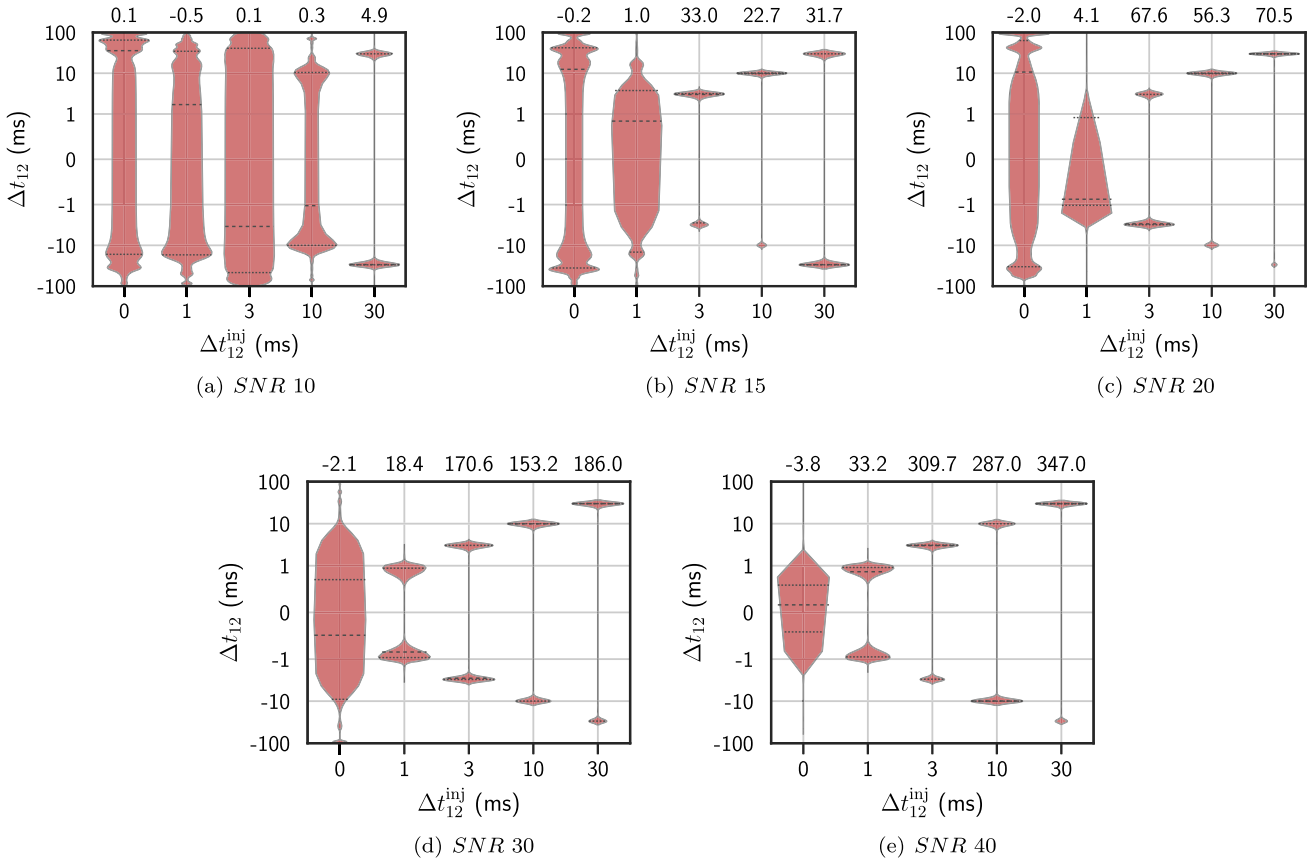
The injections are set to have $\mathrm{SNR} \in \{10, 15, 20, 30, 40\}$. These SNRs are achieved by inversely scaling the luminosity distance (d_L) of the injections. Figure 4 shows the violin plots and the log Bayes factors for these injections. The posteriors on ϕ_{lens} are uninformative in all the cases and hence not shown in the figure. This is because ϕ_{lens} is highly correlated with other parameters like ψ, ϕ, t; see Fig. 11 in Appendix A. As seen from Fig. 4, LIB time delay (Δt_{12}) as small as 1 ms is measurable from the SNR 30 and 40 signals, on the other hand for SNR 10 signals only the time delay larger than at least 10 ms is measurable. As one would expect, only with $\Delta t_{12}^{\mathrm{inj}} = 0$, i.e. GR injection the $\log \mathcal{B}^{\mathrm{LIB}}_{\mathrm{GR}} < 0$, i.e. consistent with the GR hypothesis, except for the SNR 10 case where $\log \mathcal{B}^{\mathrm{LIB}}_{\mathrm{GR}} = 0.1$ is within the intrinsic sampling error on the calculation of evidence. For $\Delta t_{12}^{\mathrm{inj}} \in \{1, 3, 10, 30\}$ ms we find that $\log \mathcal{B}^{\mathrm{LIB}}_{\mathrm{GR}} > 0$, i.e. consistent with the LIB hypothesis for all the SNRs except 10. Hence, both model selection and sensitivity to measuring the time delays improve with an increase in SNR.

We also note that the time delays are measurable up to a symmetry around $\Delta t_{12} = 0$. This is because the LIB waveform, Eq. (3), is identical at $(\Delta t_{12}, \phi_{\mathrm{lens}}, t_c)$ and $(-\Delta t_{12}, \phi_{\mathrm{lens}} + \pi/4, t_c + \Delta t_{12})$, which we also saw during mismatch studies with LIB injections; see right panel of Fig. 3. It is possible that for asymmetric and inclined binaries with significant HoMs a better measurement of ϕ_{lens} could break the Δt_{12} parity as well; however, this needs to be investigated further and is left for future studies.

Overall, as the sensitivity of the detectors improves we shall be able to measure birefringence time delays as small as 1 ms. On the other hand, in the absence of birefringence we expect to see $\Delta t_{12}(\mathrm{s})$ posteriors that are consistent with the GR value, i.e. zero and Bayes factors that favor the GR hypothesis. Most events in GWTC-3 have $\mathrm{SNR} < 30$. The time delay posteriors are hence expected to be broad; however the Bayes factors should already indicate whether LIB is present or not.

B. GWTC-3 events

We now analyze 43 CBC events from the GWTC-3, that have low detection false alarm rates, $\mathrm{FAR} \lesssim 10^{-3}\ \mathrm{yr}^{-1}$.



(a) SNR 10 (b) SNR 15 (c) SNR 20

(d) SNR 30 (e) SNR 40

FIG. 4. SNR dependence of Δt_{12}(ms) posteriors and the $\log \mathcal{B}^{\rm LIB}_{\rm GR}$ (upper x axis) for the GW150914-like injections with different values of $\Delta t_{12}^{\rm inj}$ (lower x axis) and $\phi_{\rm lens}^{\rm inj} = \pi/5$. Time delays ($\Delta t_{12}$) as small as 1 ms are recovered well with SNR 30 and 40 signals, and for SNR 10 signals time delays < 30 ms are not measurable. Both model selection and time delay measurements (without the symmetry around $\Delta t_{12} = 0$) improve as the SNR increases.

These are also the events that are considered for other tests of GR performed previously [12–15].

Figure 5 shows the Δt_{12} posteriors and the log Bayes factors for the real events. We find that for almost all the events the Δt_{12} posteriors are broad containing zero, i.e. consistent with GR. This is mostly due to the low SNRs of the events, as seen in our injection study. We also find the tightest 90% credible bounds on $|\Delta t_{12}| \lesssim 0.51$ ms coming from the event GW200311_115853 which has a reasonably high SNR ($\simeq$17.8) and moderate redshift ($z \sim 0.23$) as compared to other events. As expected $\phi_{\rm lens}$ posteriors are uninformative for almost all the events.

Thirty-eight out of 43 events result in $\log \mathcal{B}^{\rm LIB}_{\rm GR} < 0$, and hence are consistent with the GR hypothesis. Only a few events show a preference for the LIB hypothesis ($\log \mathcal{B}^{\rm LIB}_{\rm GR} > 0$), with the highest one for GW190521 (3.21) followed by GW190910_112807 (0.8), GW170823 (0.8), GW191109_010717 (0.7) and GW191129_134029 (0.1).

The Bayes factors are known to be prior dependent and their values do not signify the confidence in preferring one hypothesis over the other, but rather the preference of one hypothesis over the other given a set of prior assumptions. The model with extra parameters (LIB) could be fitting either the noise or the signal, and therefore we take a frequentist approach to determine the significance by considering different realizations of noise. We focus on the event with the highest Bayes factors (GW190521) and estimate its significance. We generate the background distribution of Bayes factors by injecting GR signals in Gaussian noise using the power spectral density around the trigger time. To calculate the false alarm probability corresponding to the observed Bayes factor for the event GW190521, we simulate 100 GR injections, whose parameters are taken from the posteriors of the GW190521 event for the GR hypothesis. Figure 6 shows the background distribution of the Bayes factors and the corresponding false alarm probability (FAP). The FAP corresponding to each $\mathcal{B}^{\rm LIB}_{\rm GR} = \kappa$ is calculated as the fraction of the background events having $\mathcal{B}^{\rm LIB}_{\rm GR} > \kappa$. We find that for the observed $\log \mathcal{B}^{\rm LIB}_{\rm GR} = 3.2$ for GW190521 the FAP is 0.48, i.e. its significance is less than 1σ.

It is to be noted that GW190521 is a remarkably loud but short (< 100 ms) signal, being easily fit by widely different hypotheses such as the head-on collision of a boson star [63] or L-R, frequency-dependent birefringence [35]. For the interested reader, in Appendix C (Fig. 10), we also plot

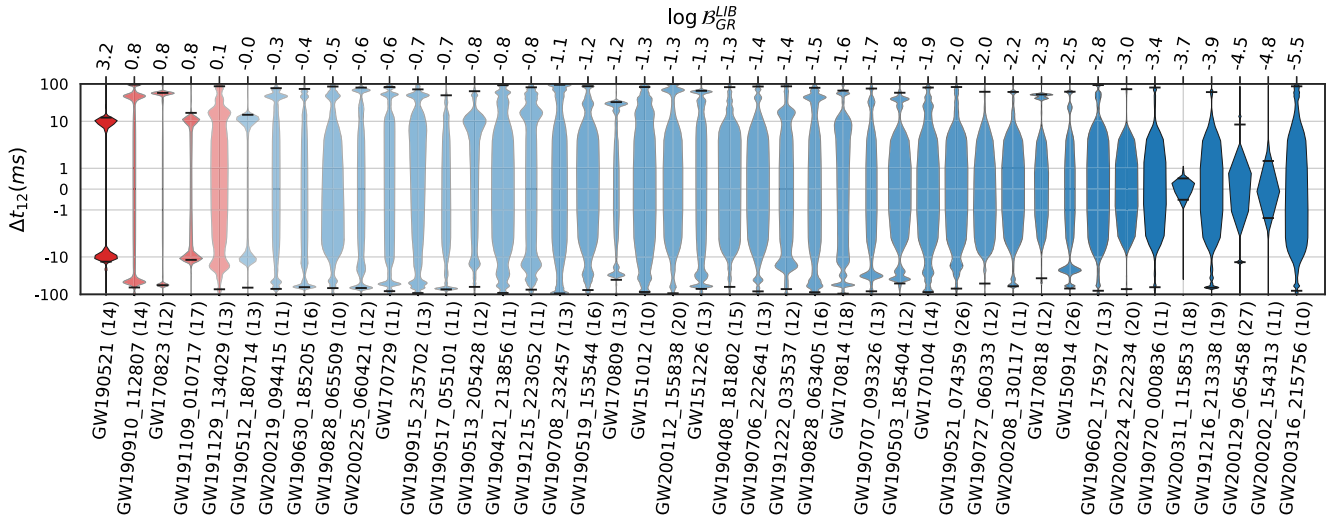


FIG. 5. LIB test of GWTC-3 events [8]. We show the posteriors on $\Delta t_{12}(\mathrm{ms})$ and Bayes factors $\log \mathcal{B}^{\mathrm{LIB}}_{\mathrm{GR}}$ (upper x axis). Events with positive Bayes factors are highlighted in red. The SNR corresponding to each event is shown inside the brackets (lower x axis, rounded off to the nearest integer).

posteriors and the waveforms corresponding to the maximum *a posteriori* parameters for both the hypotheses, over the whitened signals observed at each detector. The plots show that the LIB hypothesis is also fitting the noise, and might therefore give a high Bayes factor, $\mathcal{B}^{\mathrm{LIB}}_{\mathrm{GR}}$. The other events with $\mathcal{B}^{\mathrm{LIB}}_{\mathrm{GR}} > 0$ also show similar behavior and as their preference for LIB is marginal, we conclude that none of the events have any significant Bayes factor and find *no strong evidence* for birefringence.

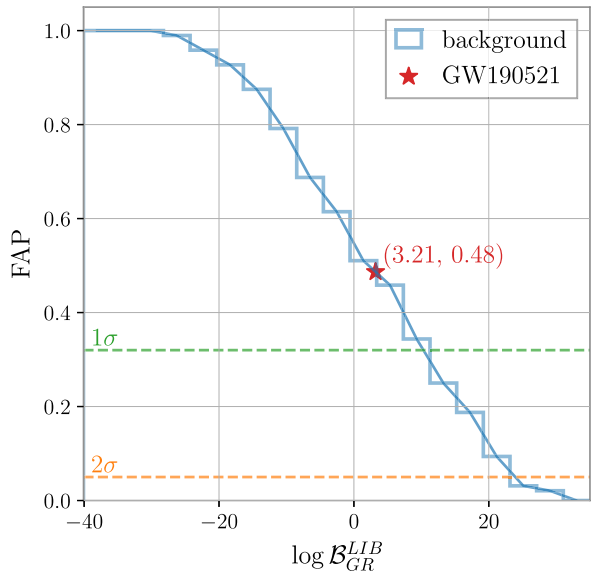


FIG. 6. Bayes factor distributions for a GW190521-like CBC, calculated by doing PE with both the hypotheses, for ∼100 GR injections from the GW190521 posteriors in different realizations of Gaussian noise. The FAP for the observed $\log \mathcal{B}^{\mathrm{LIB}}_{\mathrm{GR}} = 3.2$ is found to be 0.48.

IV. IMPLICATIONS

In our analysis of the latest GW catalog, we have found that the majority of the events disfavor birefringence. For a subset of them (most notably GW190521) while the Bayesian inference prefers the LIB hypothesis, a follow-up background study indicates that most simulated GR signals give comparable Bayes factors. In the following, we present the implications of these results. First, we consider the implications for generic LIB. Then, we study the constraints on a specific scalar-tensor theory that predicts LIB. Finally, we entertain the possibility that GW190521 was emitted in an active galactic nucleus (AGN) and is displaying evidence of birefringence.

A. Constraints on generic LIB

From the nonobservation of birefringence in the 43 events from GWTC-3 and using their median redshift values [8], we estimate the total optical depth for the LIB models discussed in Sec. II D. The nonobservation of birefringence translates to constraints on the phenomenological model parameters, as summarized in Table I. For reference, we also show the constraints obtained from the full GWTC-3 (90 events).

TABLE I. Constraints on the phenomenological models (14) and (15), assuming no birefringence detected for analyzed (all) GWTC-3 events.

$\sigma^{1/2}_{\mathrm{LIB}}$	95% CL	Comment
$\propto M$	$R_{12} < 4.4(2.9)$ kpc	Sec. IV B
$\propto R_E$	$\alpha_E < 3.0(1.6)$	
$\propto M^{1/2}$	$R_{12} < 20(12)$ kpc	
$\propto M^{1/3}$	$R_{12} < 12(6.9)$ kpc	

TABLE II. GW150914-like (top) and GW190814-like (bottom) CBC parameters used during mismatch calculations in Sec. II B and PE injection studies in Sec. III.

m_1	m_2	δ	α	ι	χ_1	χ_2	ψ	ϕ_c	t_c
38.3	33.19	−1.2	2.3	2.9	0.3	0.27	1.6	1.9	1126259462.414
24.4	2.7	−0.4	0.2	0.5	0.06	0.46	1.5	4.4	1249852256.99

The higher-redshift events have higher optical depth. Nonobservation of birefringence in distant sources leads to more stringent constraints, although the SNR scales with the inverse luminosity distance; hence some of the highest-redshift events will not be considered because of our FAR threshold. The final results depend strongly on the model via the source redshift and halo mass function. Figure 7 shows the redshift dependence of the optical depth for the parametrizations discussed, adopting the 95% CL values found by our analysis along with the observed GWTC-3 redshift distribution.

In future, with the increase in detector sensitivities and events from the high redshifts, the LIB probabilities would be better constrained. We would also be able to measure smaller LIB time delays and probe larger parameter space of alternative theories of gravity. Higher-redshift observations above our FAR threshold will be especially valuable to constrain α_E and R_{12} for $n = 1/3, 1/2$ (see Fig. 7).

B. GW birefringence in Horndeski theories

Let us now use our results in a specific theory that predicts LIB. We will present the theory and translate the constraints of the phenomenological model (Table I) into fundamental theory parameters. In the next subsection we will interpret a tentative detection of LIB in GW190521 as an AGN binary within the same theory. We will focus on a particular scalar-tensor theory within the Horndeski class [64], whose LIB

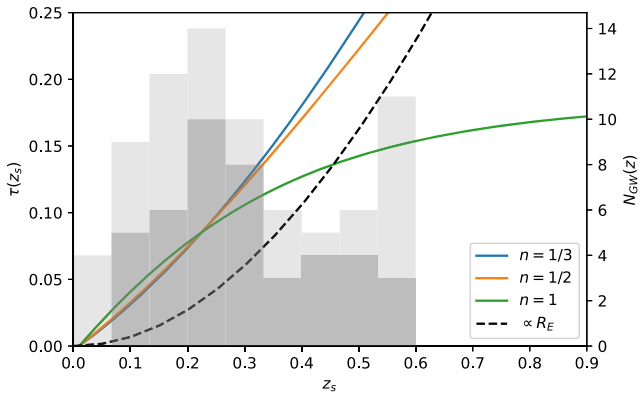


FIG. 7. Birefringence optical depth for the phenomenological models considered here, using the parameters corresponding to the 95% CL limit compatible with the nonobservation of LIB. The dark (light) gray shaded histograms show the binned redshift distribution of analyzed (all) GWTC-3 events. See Sec. II D for details.

predictions have been analyzed in detail, cf. Sec. VI in Ref. [27]. The model is described by two parameters describing couplings between the Ricci scalar (R) and the new field ϕ: a linear coupling $p_{4\phi}$ and a derivative coupling suppressed by an energy scale Λ_4. The Lagrangian of this theory can be written as [65,66]

$$\mathcal{L} \sim -\frac{1}{2}(\nabla\phi)^2 + \frac{M_P^2}{2}\left(1 + \frac{p_{4\phi}\phi}{M_P}\right)R + \frac{\phi}{\Lambda_4^2}\nabla_\mu\nabla_\nu\phi G^{\mu\nu}, \tag{18}$$

where R is the Ricci scalar, $G_{\mu\nu}$ is the Einstein tensor, M_P is the Planck mass in units of $c = h = 1$, and ∇ is the covariant derivative. The GR limit corresponds to $p_{4\phi} \to 0, \Lambda_4 \to \infty$. The parameters of this model are stringently constrained by the speed of GWs on the homogeneous FRW metric [67–70] (see also Refs. [71–73]), as observed by the near-coincident arrival of GW170817 and its associated counterpart [74]: $|c_g/c - 1| \lesssim 10^{-15}$. Compliance with this limit requires [27]

$$p_{4\phi} \lesssim 10^{-8}\Lambda_4/H_0 \quad \text{(GW170817)}. \tag{19}$$

While this constraint is extremely stringent, LIB allows comparable limits.

Specifying a model allows one to derive concrete predictions. The dependence of the time-delay contributions (Shapiro, geometric) on the lens and theory parameters is complex. Nonetheless, we observe that the time delay decreases monotonically with the impact parameter. Moreover, its slope changes and becomes very sharp beyond the Vainshtein radius

$$r_V = 1.2\ \text{Mpc}\, p_{4\phi}^{1/3}\left(\frac{M}{10^{12}M_\odot}\right)^{1/3}\left(\frac{H_0}{\Lambda_4}\right)^{2/3}. \tag{20}$$

r_V represents the scale at which the scalar field has a strong self-coupling near a massive object [62].[4] In many scalar-tensor theories this leads to *screening*: a suppression of scalar field fluctuations for $r < r_V$, allowing the theory to approximately recover GR around massive bodies. However, screening is not necessary in this model given the stringent constraint from GW170817 [Eq. (19)]. In this case, the strong interaction within r_V represents a large

[4]For extended lenses one needs to consider the effective Vainshtein radius, such that $r_V(M(r_V^{\text{eff}}) = r_V^{\text{eff}}$ [see Eq. (186) and Fig. 14 in Ref. [27]].

coupling between the scalar field and the Riemann tensor, the kind of interaction producing LIB.

For simplicity, we will focus on the Shapiro time delay. The geometric time delay is usually dominant for massive halos at intermediate distances (Fig. 12 in Ref. [27]). It is proportional to the Einstein radius, and it could thus be captured by generalizing Eq. (14) to extended lenses. Neglecting the geometric time delay is conservative but reasonable, since our constraints involve events at relatively low redshift ($z \lesssim 0.6$).

The LIB predictions have a simple dependence on the lens mass and theory parameters. We verified that $\Delta t_{12} \propto M\Lambda_4^{-4/3}$. The proportionality to the mass comes from the scaling of the impact parameter, Δt_{12} and the time spent by the GW in the region of sizable birefringence with the Vainstein radius (all are $\propto r_V$). It allows us to directly connect the theory parameters to R_{12} with $n = 1$, as constrained in the phenomenological model (15). The scaling with Λ_4 allows us then to find R_{12} by equating $\Delta t_{12}(R_{12})$ to the constrained value for different $p_{4\phi}$, but keeping $M = 10^{12} M_\odot$, Λ_4 fixed. For simplicity, we take a sensitivity of $\Delta t_{12} \sim 10$ ms to define R_{12}. Using the actual posteriors on Δt_{12} for each of the GWTC-3 events analyzed will not qualitatively affect these constraints in any significant manner.

The excluded region is shown in Fig. 8, along with constraints from the GW speed on FRW and lunar laser ranging (no screening, $p_{4\phi} \ll 1$; see Sec. V B 3 in Ref. [60]). The change in slope at low Λ corresponds to a transition in which R_{12} surpasses the Vainshtein radius (20). For $\Lambda_4 \ll H_0$ the birefringence constraints approach those of the GW speed; this happens when r_V is so large that most GWs are effectively behind a lens. Then the constraints are satisfied in the limit $c_{\rm GW} \to c$, equivalent to Eq. (19). For the sensitivity of GWTC-3 this happens for $\Lambda_4 \lesssim H_0$, where LVK frequencies lie beyond the validity of our framework as a classical effective field theory [75]. At increasing Λ_4 the constraints degrade, since the probability becomes very suppressed [Eq. (20)]. For $\Lambda^4 \gtrsim 10^2 H_0$ Solar System constraints become more efficient than birefringence.

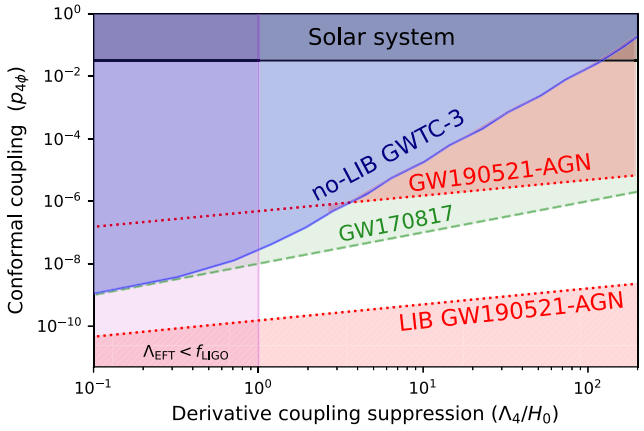


FIG. 8. 95% CL constraints on the parameters of a quartic Horndeski theory [27] using the LIB test. Shaded regions are excluded according to GWTC-3 (this work; blue solid), GW170817 [67–70] (green dashed) and GW190521 assuming an AGN binary [76] (red dotted; see Fig. 9). The GR limit corresponds to $p_{4\phi} \to 0, \Lambda_4 \to \infty$, when the scalar field is decoupled from gravity and its derivative interactions are suppressed. See Secs. IV B and IV C for details. If GW190521 is associated with an AGN, the upper shaded region improves the overall GWTC-3 constraints for $\Lambda_4 \gtrsim 3H_0$. If we further assume a detection of LIB, then the bottom red shaded region excludes GR. For reference, we also indicate Solar System constraints (gray horizontal) and the region where the GW frequencies at LIGO-Virgo detectors are larger than the (nonlinear) energy scale of the effective field theory (magenta vertical).

C. GW190521 as an AGN binary

Let us now discuss the implications of a possible birefringence detection associated with GW190521. Given the constraints from the speed of GWs (19) on our example Horndeski theory, the chances of birefringence being caused by a lens in the line of sight are very small. We will instead interpret our result, $\Delta t_{12} \gtrsim 9.5$ ms as being due to an environmental effect near the source. We will follow the scenario outlined in Ref. [76], where a candidate electromagnetic counterpart from an AGN J124942.3 + 344929, observed 34 days after the GW signal, suggested that the binary merged in the environment of an SMBH. Note that there are important uncertainties, regarding both the counterpart association (given large GW localization uncertainties [77]), and the significance of LIB detection (given our analysis of random noise realizations; see Fig. 6). This discussion is therefore not a statement on the status of GR. Instead, it proves the potential of identifying environments of GW sources to test gravity theories.

Following Ref. [76], we assume an AGN binary scenario where the mass of the SMBH is $M_{\rm SMBH} \sim 10^8 M_\odot$ and the source is located in a migration trap at $r \sim 700 G M_{\rm SMBH}$. Then, using the framework of Ref. [27] allows us to compute the time delay as a function of the angle between the observer and the source, relative to the SMBH. The results are shown in Fig. 9 for $p_{4\phi} = 10^{-8}, \Lambda_4 = 10 H_0$, compatible with GW170817 [Eq. (19)], and different distances to the SMBH (the dependence on $M_{\rm SMBH}$ is less pronounced; see below). The birefringent time delay becomes very large as $\theta \to 0$.[5] Ultimately, the maximum time delay is limited by the existence of the horizon, $\theta_s \approx 2GM/r$. The birefringence also vanishes as $\theta \to \pi$ because of geometric cancellations in spherical symmetry.

[5]Our calculation relies on small deviations from a straight trajectory. This assumption breaks down for small angles, where one needs to consider the geodesics of the SMBH space-time instead. However, our results are conservative since actual trajectories will bend toward the SMBH, thus increasing Δt_{12} relative to the straight propagation.

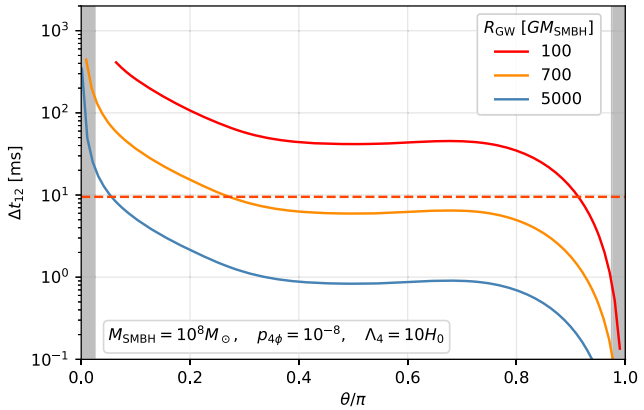


FIG. 9. Birefringent time delay for a source near a SMBH as a function of the angle of the observer, relative to the SMBH. Each line corresponds to a different source distance, for model parameters compatible with GW170817 [see. Eq. (19)]. The horizontal line corresponds to the lower bound on $\Delta t_{12} = 9.5$ ms from the analysis of GW190521. The region between the shaded areas encompasses the 95% probability for a random observer. The lowest θ represents trajectories passing at 10 Schwarzschild radii of the SMBH.

We will translate these predictions into theory parameters and include the comparison to GW190521. We take the values of $M_{\rm SMBH}$ and the source radius as fixed, and consider the credible intervals as being determined by the angle θ, cf. Eq. (B1) and Appendix B. As we do not know the emission angle, we assume a flat prior on the sphere $P(\theta) = \sin(\theta)$, and take the upper/lower 95% CL values based on $P(\theta)$ (excluding the shaded regions in Fig. 9). Limits on the theory parameters can be derived by noting that $\Delta t_{12}(\theta) \propto p_{4\phi}^{4/3} \Lambda_4^{-2/3} M_{\rm SMBH}^{1/3}$ including different assumptions about the SMBH mass. Note that $M_{\rm SMBH}$ enters with a different scaling than the lens mass in Sec. IV B, due to the source being at a fixed distance from the SMBH and within its Vainshtein radius, rather than randomly located.

The implications of GW190521 for the example theory (18) are shown in Fig. 8. The orange regions are excluded if we assume the AGN scenario as discussed above. The lower region excludes the GR limit $p_{4\phi} \to 0, \Lambda_4 \to \infty$ and relies on trusting the measured birefringence $\Delta t_{12} \gtrsim 9.5$ ms to be due to new gravitational physics. Even if the result is interpreted as noise (e.g. Fig. 6), assuming the AGN scenario leads to the exclusion of the upper orange region (assuming sensitivity to $\Delta t_{12} \lesssim 9.5$ ms). Because of the different scaling with the theory parameters, the detection of an AGN binary becomes even more constraining than GW170817 for high Λ_4. The beyond-GR interpretation can be further probed not only by AGN events but also by high-redshift multimessenger observations. In this case, the time delay between GWs and electromagnetic counterparts scales as $\approx 1s(10^8 p_{4\phi} \frac{H_0}{\Lambda_4})^2 \frac{D}{40\ \rm Mpc}$ and can be probed by distant neutron-star mergers.

V. SUMMARY AND OUTLOOK

In this paper, we explored LIB as a test of GR using observations of GWs. LIB produces a difference in the arrival times of the GW polarizations in signals from the binary mergers, predicted by some alternatives to GR. Using the Bayesian model selection framework, we can identify the signatures of birefringence and measure the time delay between the arrival of both polarizations (Δt_{12}). We showed that this difference can be measured with high accuracy, of the order of a few milliseconds with existing events and is likely to improve in the future following detector upgrades.

Using the latest GW catalog, GWTC-3, we found no strong evidence for the observation of birefringence, with the highest $\log \mathcal{B}^{\rm LIB}_{\rm GR} = 3.21$ for the heaviest binary black holes so far, GW190521. However, after simulating similar events under different noise realizations, we determined that there is a FAP of 48%. This event has been associated with an AGN flare, possibly indicating that the merger occurred near an SMBH. This AGN scenario is especially favorable for the observation of LIB since the SMBH would act as a strong source of LIB. However, the AGN flare–GW association has been disputed; see e.g. Ref. [77]. Moreover, the loudness and shortness of this event makes it susceptible to different astrophysical and fundamental physics interpretations. It has also been found to violate many tests of GR and mimic many exotic scenarios of compact binary such as the head-on collision of a boson star [63] or L-R, frequency-dependent birefringence [35]. The latter effect is related to our flavor of LIB, with two important differences: first, L-R birefringence is defined in the basis of circularly polarized waves (left vs right, rather than + vs ×), and second, it depends on the GW frequency. Both features also appear in the gravitational spin Hall effect in GR, although the L-R time delay is very suppressed [32,33].

Of the 43 analyzed events, we found that the tightest bounds on the time delay between the two polarizations is $\Delta t_{12} \sim 0.51$ ms at 90% credible intervals coming from the GW200311_115853 merger event, while the median is $\Delta t_{12} \simeq 80$ ms. From the nonobservation of LIB, we constrained the lensing optical depths in a phenomenological parametrization in which the lensing cross section is proportional to the Einstein radius or a fixed physical radius with a power-law scaling in the halo mass.

Our constraints can be translated to gravitational theories that predict LIB. As an example, we presented novel constraints on a Horndeski scalar-tensor theory featuring a new dynamical field and two free parameters. The theory is stringently constrained by the speed of GWs on the homogeneous FRW background following GW170817. Nevertheless, the lack of observed LIB places stringent bounds, which can be orders of magnitude better than Solar System tests and in some limits as tight as the GW speed bound. As a proof of principle of LIB due to a known

inhomogeneity, we interpreted GW190521 as an AGN binary (assuming that the signal originated in close proximity to an SMBH [76]) in terms of our example theory. Then, the large curvature is able to generate detectable LIB even when deviations from GR are minute. Our $|\Delta t_{12}| \gtrsim 9.5$ ms results would then exclude GR and place lower bounds on the parameters of the alternative theories of gravity. When interpreting this result as a fluctuation and GR to be correct, the AGN hypothesis is still able to produce very stringent bounds, that can even overcome those of the GW speed on FRW.

In the future, the methods we developed here could be useful for studying new classes of events. Of particular interest will be signals where the merger is either near an SMBH or is known to have a lensed counterpart due to strong lensing. In such cases, the information about the lens may improve the constraints substantially, along the lines of the AGN scenario we discussed. The increase in detection rate and a growing chance of strongly lensed GW identification makes LIB tests also relevant for future runs of LVK detectors and upcoming GW detectors such as the Einstein Telescope, Cosmic Explorer and LISA [78–81]. Last, the addition of ground-based detectors such as LIGO-India and KAGRA can allow us to measure extra linear combinations of the GW polarizations and construct a null stream [82] to extract each of the polarizations individually. The extracted polarizations can then be used to test their consistency with GR or other theories of gravity directly.

Strongly lensed GW signals may allow us to measure additional linear combinations of the same GW polarizations and hence improve various tests of GR [83], including the one proposed here. Ultimately, developing LIB predictions for other alternative theories and generalizing the model-independent parametrizations presented here will allow our results to further test the landscape of theories beyond GR.

ACKNOWLEDGMENTS

We are grateful to A. K. Mehta, P. Ajith, G. Brando, S. Savastano, G. Tambalo, Y. Wang, H. Villarrubia-Rojo and J. Tasson for fruitful discussions. S. G. and A. V. are supported by the Department of Atomic Energy, Government of India, under Project No. RTI4001. A. V. is also supported by a Fulbright Program grant under the Fulbright-Nehru Doctoral Research Fellowship, sponsored by the Bureau of Educational and Cultural Affairs of the United States Department of State and administered by the Institute of International Education and the United States-India Educational Foundation. J. M. E. is supported by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Grant agreement No. 847523 INTERACTIONS, by VILLUM FONDEN (Grant No. 37766), by the Danish Research Foundation, and under the European Union's H2020 ERC Advanced Grant "Black holes: gravitational engines of discovery" Grant agreement No. Gravitas–101052587. The numerical calculations reported in the paper were performed on the Alice computing cluster at ICTS-TIFR, with the aid of the LALSuite [44], BILBY [45], PyCBC [46] and COLOSSUS. [57] software packages. This material is based upon work supported by the NSF's LIGO Laboratory which is a major facility fully funded by the National Science Foundation.

APPENDIX A: INJECTIONS

Here we list the injection parameters for the mismatch and the parameter estimation studies. Note that the luminosity distances are scaled as per the SNRs and hence are not mentioned in the table below.

To demonstrate the correlations in the measurement of various parameters that may be leading to the uninformative $\phi_{\rm lens}$ posteriors in our PE injection studies, we show a corner plot of the posteriors for a GW150914-like binary black hole injection with SNR $= 30$, $\Delta t_{12}^{\rm inj} = 10$ ms, and $\phi_{\rm lens}^{\rm inj} = \pi/5$ rad in Fig. 11.

APPENDIX B: BEYOND-POISSON STATISTICS

The independent lens assumption fails to capture two circumstances that potentially enhance the detection of birefringence: the source environment and lensing by known objects. This situation is qualitatively different from strong lensing probabilities, which are weighted by the Einstein radius, which vanishes when $D_L \to D_S$ (near the source) or $D_L \to 0$ (near the detector). In contrast, birefringence probabilities do not suffer such suppression and can be sizable for objects near the source or the observer. Our optical-depth framework (Sec. II D) does not consider this possibility.

The source environment may play a role for LIB, as GW sources will generally be located in regions denser than the cosmic average. In this case, the host galaxy (or objects within it) would have a much larger density compared to the cosmological average used in, e.g. Eq. (16). In addition, the projected cross section $\propto 1/D^2$ will be larger for nearby objects, thus enhancing the probabilities. Given a distribution of GW sources near an object, the posterior on the theory parameters $\vec{p}$ can be obtained as

$$P(\vec{p}) = \int dr d\theta P_{\rm s}(r) \sin(\theta) P(\Delta t_{12}(r, \theta, \vec{p})). \tag{B1}$$

Here we assume a symmetric r-dependent distribution. The θ dependence corresponds to a uniform prior on the sphere. This simple dependence could be used to model the effect of the source's galaxy or nearby objects.

An extreme case of environmental enhancement is given by a binary merging in an AGN near an SMBH, as discussed in Sec. IV C, taking the multimessenger scenario of GW190521 and its implications for the example Horndeski theory. Estimates for the rate of such

events are uncertain. Nonetheless, in some cases it might be possible to associate an event with an SMBH thanks to an electromagnetic counterpart [76], multiple images due to strong lensing [84–87] or strong-field propagation effects [33].

Another potential way to improve the quoted result is by correlating the GW arrival direction with known lenses, which is relevant in cases where the Milky Way (or perhaps even the Sun) may imprint an observable birefringence. Adding information on the GW direction, relative to known objects, will allow better constraints on those scenarios more effectively than assuming randomly located lenses. For instance, if stellar-scale lenses are relevant in a given theory and the cross section scales as the physical radius (allowing nearby lenses to contribute), sources behind the Milky Way can probe a much larger effective cross section than that given by Eq. (17).

Finally, any confident detection of a lensed GW can be used to refine constraints within a given model. This would follow either through the identification of several GW detections as images of the same underlying source or through waveform distortions (millilensing). Both cases allow information about the lens mass and impact parameter to be recovered, at least when assuming a lens model [81,88,89]. That information can then place constraints within a specific theory of gravity.

APPENDIX C: GW190521 POSTERIORS UNDER LIB AND GR

In Fig. 10 we show the posteriors of the GW190521 event which has the highest log Bayes factor ($\ln \mathcal{B}^{\rm LIB}_{\rm GR} = 3.21$) from the PE runs of the LIB and GR hypotheses. The two posteriors are consistent with each other with LIB favoring a slightly higher luminosity distance (d_L) and chirp mass (M_c). Additionally, posteriors under LIB are marginally narrower as compared to GR, which might be a reason for its $\ln \mathcal{B}^{\rm LIB}_{\rm GR} > 0$. It is worth noticing that Δt_{12} is degenerate with t_c, which is itself poorly measured due to low SNR in Virgo. We also plot the waveforms using maximum *a posteriori* (MaP) parameters along with the whitened time series data [90] as observed in the Hanford (H1), Livingston (L1) and Virgo (V1) detectors. It is easy to see that the signal duration is small and the two MaP waveforms are not very different from each other except for the tiny modulations in the LIB one. It can thus be concluded that the model selection favors the LIB hypothesis because it better fits the random noise at the detectors during the event GW190521.

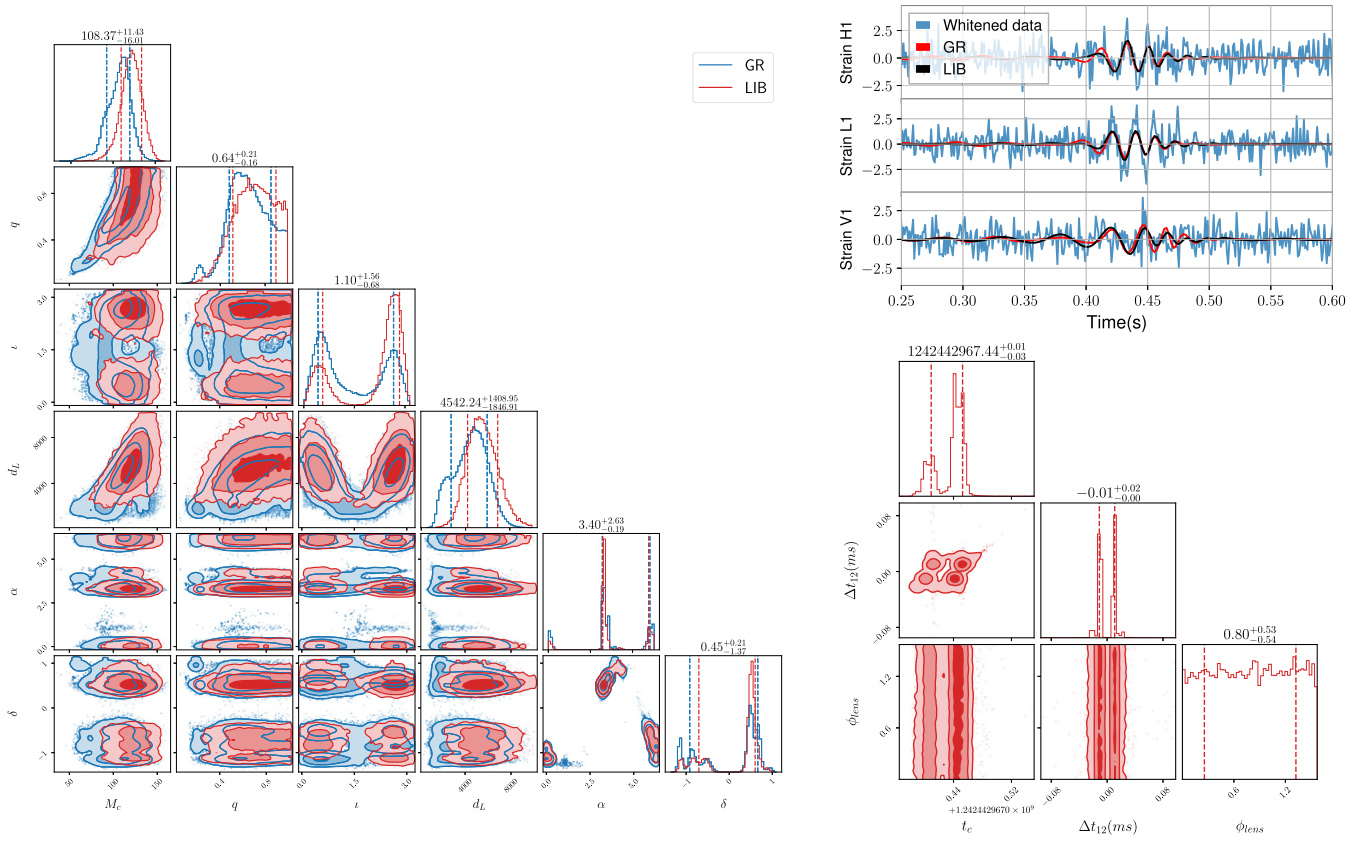


FIG. 10. GW190521 ($\log \mathcal{B}^{\rm LIB}_{\rm GR} = 3.21$) GR vs LIB posteriors, using MaP waveforms under GR and LIB ($\Delta t_{12} = 9.51$ ms, $\phi_{\rm lens} = 0.06$ rad) hypotheses with the whitened strain as observed at the LIGO-Virgo detectors.

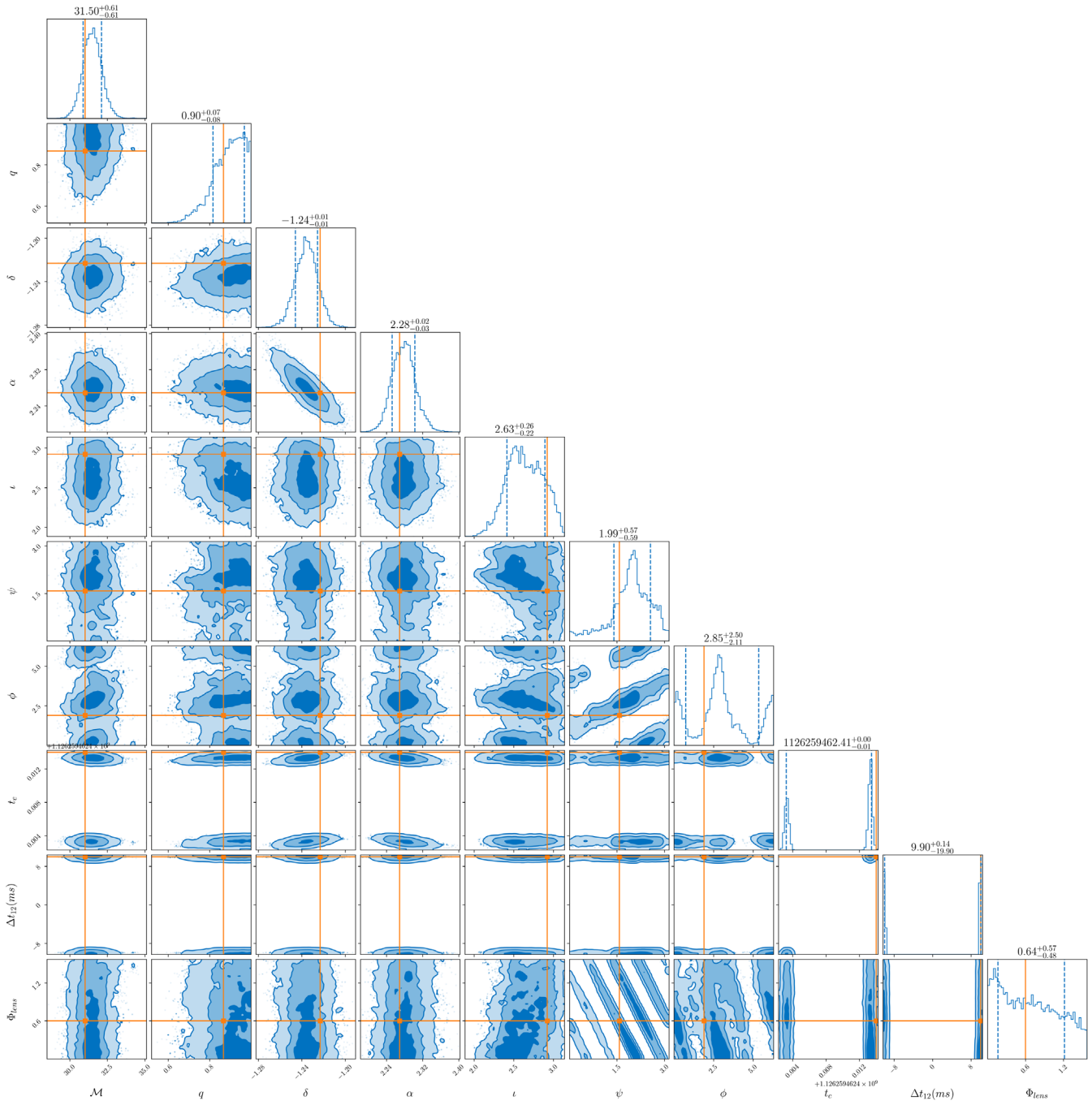


FIG. 11. Corner plot of posteriors from LIB recovery of a GW150914-like binary black hole injection with SNR $= 30$, $\Delta t_{12}^{\rm inj} = 10$ ms, and $\phi_{\rm lens}^{\rm inj} = \pi/5$ rad. The orange lines correspond to the injection parameter values. The $\phi_{\rm lens}$ posteriors are visibly correlated with ψ, ι and ϕ, leading to uninformative 1D posteriors of $\phi_{\rm lens}$.

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

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

[3] Yoichi Aso, Yuta Michimura, Kentaro Somiya, Masaki Ando, Osamu Miyakawa, Takanori Sekiguchi, Daisuke Tatsumi, and Hiroaki Yamamoto (KAGRA Collaboration), Interferometer design of the KAGRA gravitational wave detector, Phys. Rev. D **88**, 043007 (2013).

[4] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Observation of Gravitational Waves from a Binary Black Hole Merger, Phys. Rev. Lett. **116**, 061102 (2016).

[5] B. P. Abbott *et al.* (Virgo and LIGO Scientific Collaborations), GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral, Phys. Rev. Lett. **119**, 161101 (2017).

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

[7] R. Abbott *et al.* (LIGO Scientific, Virgo Collaborations), GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run, Phys. Rev. X **11**, 021053 (2021).

[8] R. Abbott *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), GWTC-3: Compact binary coalescences observed by LIGO and Virgo during the second part of the third observing run, arXiv:2111.03606.

[9] Alexander H. Nitz, Thomas Dent, Gareth S. Davies, Sumit Kumar, Collin D. Capano, Ian Harry, Simone Mozzon, Laura Nuttall, Andrew Lundgren, and Márton Tápai, 2-OGC: Open gravitational-wave catalog of binary mergers from analysis of public Advanced LIGO and Virgo data, Astrophys. J. **891**, 123 (2020).

[10] Alexander H. Nitz, Sumit Kumar, Yi-Fan Wang, Shilpa Kastha, Shichao Wu, Marlin Schäfer, Rahul Dhurkunde, and Collin D. Capano, 4-OGC: Catalog of gravitational waves from compact-binary mergers, Astrophys. J. **946**, 59 (2023).

[11] Tejaswi Venumadhav, Barak Zackay, Javier Roulet, Liang Dai, and Matias Zaldarriaga, New binary black hole mergers in the second observing run of Advanced LIGO and Advanced Virgo, Phys. Rev. D **101**, 083030 (2020).

[12] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Tests of General Relativity with GW150914, Phys. Rev. Lett. **116**, 221101 (2016).

[13] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Tests of general relativity with the binary black hole signals from the LIGO-Virgo catalog GWTC-1, Phys. Rev. D **100**, 104036 (2019).

[14] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Tests of general relativity with binary black holes from the second LIGO-Virgo gravitational-wave transient catalog, Phys. Rev. D **103**, 122002 (2021).

[15] R. Abbott *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), Tests of general relativity with GWTC-3, arXiv:2112.06861.

[16] Clifford M. Will, Bounding the mass of the graviton using gravitational wave observations of inspiralling compact binaries, Phys. Rev. D **57**, 2061 (1998).

[17] Jose María Ezquiaga and Miguel Zumalacárregui, Dark energy in light of multi-messenger gravitational-wave astronomy, Front. Astron. Space Sci. **5**, 44 (2018).

[18] Shun Arai and Atsushi Nishizawa, Generalized framework for testing gravity with gravitational-wave propagation. ii. constraints on horndeski theory, Phys. Rev. D **97**, 104038 (2018).

[19] Atsushi Nishizawa, Generalized framework for testing gravity with gravitational-wave propagation. i. formulation, Phys. Rev. D **97**, 104037 (2018).

[20] Clifford M. Will, The confrontation between general relativity and experiment, Living Rev. Relativity **17**, 4 (2014).

[21] R. L. Workman *et al.* (Particle Data Group), Review of particle physics, Prog. Theor. Exp. Phys. **2022**, 083C01 (2022).

[22] Kevin Max, Moritz Platscher, and Juri Smirnov, Gravitational Wave Oscillations in Bigravity, Phys. Rev. Lett. **119**, 111101 (2017).

[23] Kevin Max, Moritz Platscher, and Juri Smirnov, Decoherence of gravitational wave oscillations in bigravity, Phys. Rev. D **97**, 064009 (2018).

[24] Enis Belgacem *et al.* (LISA Cosmology Working Group), Testing modified gravity at cosmological distances with LISA standard sirens, J. Cosmol. Astropart. Phys. 07 (2019) 024.

[25] Jose Beltrán Jiménez, Jose María Ezquiaga, and Lavinia Heisenberg, Probing cosmological fields with gravitational wave oscillations, J. Cosmol. Astropart. Phys. 04 (2020) 027.

[26] Jose Maria Ezquiaga, Wayne Hu, Macarena Lagos, and Meng-Xiang Lin, Gravitational wave propagation beyond general relativity: Waveform distortions and echoes, J. Cosmol. Astropart. Phys. 11 (2021) 048.

[27] Jose María Ezquiaga and Miguel Zumalacárregui, Gravitational wave lensing beyond general relativity: Birefringence, echoes and shadows, Phys. Rev. D **102**, 124048 (2020).

[28] Charles Dalang, Pierre Fleury, and Lucas Lombriser, Scalar and tensor gravitational waves, Phys. Rev. D **103**, 064075 (2021).

[29] Maria Okounkova, Will M. Farr, Maximiliano Isi, and Leo C. Stein, Constraining gravitational wave amplitude birefringence and Chern-Simons gravity with GWTC-2, Phys. Rev. D **106**, 044067 (2022).

[30] Salvatore Vitale, Sylvia Biscoveanu, and Colm Talbot, The orientations of the binary black holes in GWTC-3, arXiv:2204.00968.

[31] Claudia de Rham and Andrew J. Tolley, Speed of gravity, Phys. Rev. D **101**, 063518 (2020).

[32] Lars Andersson, Jérémie Joudioux, Marius A. Oancea, and Ayush Raj, Propagation of polarized gravitational waves, Phys. Rev. D **103**, 044053 (2021).

[33] Marius A. Oancea, Richard Stiskalek, and Miguel Zumalacárregui, From the gates of the abyss: Frequency- and polarization-dependent lensing of gravitational waves in strong gravitational fields, arXiv:2209.06459.

[34] Antonio Enea Romano, A unified effective approach to cosmological perturbations, arXiv:2301.05679.

[35] Yi-Fan Wang, Stephanie M. Brown, Lijing Shao, and Wen Zhao, Tests of gravitational-wave birefringence with the open gravitational-wave catalog, Phys. Rev. D **106**, 084005 (2022).

[36] Leïla Haegel, Kellie O'Neal-Ault, Quentin G. Bailey, Jay D. Tasson, Malachy Bloom, and Lijing Shao, Search for anisotropic, birefringent spacetime-symmetry breaking in gravitational wave propagation from GWTC-3, Phys. Rev. D **107**, 064031 (2023).

[37] Flavio Bombacigno, Fabio Moretti, Simon Boudet, and Gonzalo J. Olmo, Landau damping for gravitational waves in parity-violating theories, J. Cosmol. Astropart. Phys. 02 (2023) 009.

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

[39] Ashish Kumar Meena and Jasjeet Singh Bagla, Gravitational lensing of gravitational waves: Wave nature and prospects for detection, Mon. Not. R. Astron. Soc. **492**, 1127 (2019).

[40] Jose María Ezquiaga, Daniel E. Holz, Wayne Hu, Macarena Lagos, and Robert M. Wald, Phase effects from strong gravitational lensing of gravitational waves, Phys. Rev. D **103**, 064047 (2021).

[41] Giulia Cusin and Macarena Lagos, Gravitational wave propagation beyond geometric optics, Phys. Rev. D **101**, 044041 (2020).

[42] Zhao Li, Jin Qiao, Wen Zhao, and Xinzhong Er, Gravitational faraday rotation of gravitational waves by a Kerr black hole, J. Cosmol. Astropart. Phys. 10 (2022) 095.

[43] Geraint Pratten, Cecilio García-Quirós, Marta Colleoni, Antoni Ramos-Buades, Héctor Estellés, Maite Mateu-Lucena, Rafel Jaume, Maria Haney, David Keitel, Jonathan E. Thompson, and Sascha Husa, Computationally efficient models for the dominant and subdominant harmonic modes of precessing binary black holes, Phys. Rev. D **103**, 104056 (2021).

[44] LIGO Scientific Collaboration, LIGO Algorithm Library—LALSuite, free software (GPL) (2018).

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

[46] Alex Nitz *et al.*, gwastro/pycbc: Pycbc release v1.16.8 (2020), 10.5281/zenodo.3985815.

[47] The Virgo Collaboration, Advanced Virgo sensitivity curve study, Technical Report No. VIR-0073D-12, 2012.

[48] The updated Advanced LIGO design curve, Technical Report No. LIGO-T1800044-v5, LIGO Document Control Center, 2018.

[49] Cameron Mills and Stephen Fairhurst, Measuring gravitational-wave higher-order multipoles, Phys. Rev. D **103**, 024042 (2021).

[50] John Skilling, Nested sampling for general bayesian computation, Bayesian Anal. **1**, 833 (2006).

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

[52] Salvatore Vitale, Davide Gerosa, Will M. Farr, and Stephen R. Taylor, Inferring the properties of a population of compact binaries in presence of selection effects, arXiv:2007.05579.

[53] Fei Xu, Jose Marí a Ezquiaga, and Daniel E. Holz, Please repeat: Strong lensing of gravitational waves as a probe of compact binary and galaxy populations, Astrophys. J. **929**, 9 (2022).

[54] Y. Pei, Probability of lensing magnification by cosmologically-distributed point masses., Astrophys. J. **404**, 436 (1993).

[55] Miguel Zumalacarregui and Uros Seljak, Limits on Stellar-Mass Compact Objects as Dark Matter from Gravitational Lensing of Type Ia Supernovae, Phys. Rev. Lett. **121**, 141101 (2018).

[56] Jeremy L. Tinker, Andrey V. Kravtsov, Anatoly Klypin, Kevork Abazajian, Michael S. Warren, Gustavo Yepes, Stefan Gottlober, and Daniel E. Holz, Toward a halo mass function for precision cosmology: The limits of universality, Astrophys. J. **688**, 709 (2008).

[57] Benedikt Diemer, COLOSSUS: A python toolkit for cosmology, large-scale structure, and dark matter halos, Astrophys. J. Suppl. Ser. **239**, 35 (2018).

[58] Stefano Savastano, Giovanni Tambalo, Hector Villarrubia-Rojo, and Miguel Zumalacarregui, Weakly lensed gravitational waves: Probing cosmic structures with wave-optics features, arXiv:2306.05282.

[59] N. Aghanim *et al.* (Planck Collaboration), Planck 2018 results. VI. Cosmological parameters, Astron. Astrophys. **641**, A6 (2020); **652**, C4(E) (2021).

[60] Miguel Zumalacarregui, Gravity in the era of equality: Towards solutions to the Hubble problem without fine-tuned initial conditions, Phys. Rev. D **102**, 023523 (2020).

[61] David Alonso, Emilio Bellini, Pedro G. Ferreira, and Miguel Zumalacárregui, Observational future of cosmological scalar-tensor theories, Phys. Rev. D **95**, 063502 (2017).

[62] A. I. Vainshtein, To the problem of nonvanishing gravitation mass, Phys. Lett. **39B**, 393 (1972).

[63] Juan Calderón Bustillo, Nicolas Sanchis-Gual, Alejandro Torres-Forné, José A. Font, Avi Vajpeyi, Rory Smith, Carlos Herdeiro, Eugen Radu, and Samson H. W. Leong, GW190521 as a Merger of Proca Stars: A Potential New Vector Boson of 8.7×10^{-13} eV, Phys. Rev. Lett. **126**, 081101 (2021).

[64] Gregory Walter Horndeski, Second-order scalar-tensor field equations in a four-dimensional space, Int. J. Theor. Phys. **10**, 363 (1974).

[65] Jose María Ezquiaga, Juan García-Bellido, and Miguel Zumalacárregui, Towards the most general scalar-tensor theories of gravity: A unified approach in the language of differential forms, Phys. Rev. D **94**, 024005 (2016).

[66] Jose María Ezquiaga, Juan García-Bellido, and Miguel Zumalacárregui, Field redefinitions in theories beyond Einstein gravity using the language of differential forms, Phys. Rev. D **95**, 084039 (2017).

[67] Jose María Ezquiaga and Miguel Zumalacárregui, Dark Energy After GW170817: Dead Ends and the Road Ahead, Phys. Rev. Lett. **119**, 251304 (2017).

[68] Paolo Creminelli and Filippo Vernizzi, Dark Energy after GW170817 and GRB170817A, Phys. Rev. Lett. **119**, 251302 (2017).

[69] T. Baker, E. Bellini, P. G. Ferreira, M. Lagos, J. Noller, and I. Sawicki, Strong Constraints on Cosmological Gravity from GW170817 and GRB 170817A, Phys. Rev. Lett. **119**, 251301 (2017).
[70] Jeremy Sakstein and Bhuvnesh Jain, Implications of the Neutron Star Merger GW170817 for Cosmological Scalar-Tensor Theories, Phys. Rev. Lett. **119**, 251303 (2017).
[71] Philippe Brax, Clare Burrage, and Anne-Christine Davis, The speed of Galileon gravity, J. Cosmol. Astropart. Phys. 03 (2016) 004.
[72] Lucas Lombriser and Andy Taylor, Breaking a dark degeneracy with gravitational waves, J. Cosmol. Astropart. Phys. 03 (2016) 031.
[73] Dario Bettoni, Jose María Ezquiaga, Kurt Hinterbichler, and Miguel Zumalacárregui, Speed of gravitational waves and the fate of scalar-tensor gravity, Phys. Rev. D **95**, 084029 (2017).
[74] B. P. Abbott *et al.* (Virgo, Fermi-GBM, INTEGRAL, and LIGO Scientific Collaborations), Gravitational waves and gamma-rays from a binary neutron star merger: GW170817 and GRB 170817A, Astrophys. J. **848**, L13 (2017).
[75] Claudia de Rham and Scott Melville, Gravitational Rainbows: LIGO and Dark Energy at its Cutoff, Phys. Rev. Lett. **121**, 221101 (2018).
[76] M. J. Graham *et al.*, Candidate Electromagnetic Counterpart to the Binary Black Hole Merger Gravitational Wave Event S190521g, Phys. Rev. Lett. **124**, 251102 (2020).
[77] Antonella Palmese, Maya Fishbach, Colin J. Burke, James T. Annis, and Xin Liu, Do LIGO/Virgo black hole mergers produce AGN flares? The case of GW190521 and prospects for reaching a confident association, Astrophys. J. Lett. **914**, L34 (2021).
[78] Vicky Kalogera *et al.*, The next generation global gravitational wave observatory: The science book, arXiv:2111.06990.
[79] B. Sathyaprakash *et al.*, Scientific objectives of Einstein telescope, Classical Quantum Gravity **29**, 124013 (2012); **30**, 079501(E) (2013).
[80] Xuheng Ding, Marek Biesiada, and Zong-Hong Zhu, Strongly lensed gravitational waves from intrinsically faint double compact binaries—prediction for the Einstein Telescope, J. Cosmol. Astropart. Phys. 12 (2015) 006.
[81] Mesut Çalışkan, Lingyuan Ji, Roberto Cotesta, Emanuele Berti, Marc Kamionkowski, and Sylvain Marsat, Observability of lensing of gravitational waves from massive black hole binaries with LISA, Phys. Rev. D **107**, 043029 (2023).
[82] Katerina Chatziioannou, Nicolas Yunes, and Neil Cornish, Model-independent test of general relativity: An extended post-Einsteinian framework with complete polarization content, Phys. Rev. D **86**, 022004 (2012); **95**, 129901(E) (2017).
[83] Srashti Goyal, K. Haris, Ajit Kumar Mehta, and Parameswaran Ajith, Testing the nature of gravitational-wave polarizations using strongly lensed signals, Phys. Rev. D **103**, 024038 (2021).
[84] Ryan M. O'Leary, Bence Kocsis, and Abraham Loeb, Gravitational waves from scattering of stellar-mass black holes in galactic nuclei, Mon. Not. R. Astron. Soc. **395**, 2127 (2009).
[85] Bence Kocsis and Janna Levin, Repeated bursts from relativistic scattering of compact objects in galactic nuclei, Phys. Rev. D **85**, 123005 (2012).
[86] Daniel J. D'Orazio and Abraham Loeb, Repeated gravitational lensing of gravitational waves in hierarchical black hole triples, Phys. Rev. D **101**, 083031 (2020).
[87] László Gondán and Bence Kocsis, Astrophysical gravitational-wave echoes from galactic nuclei, Mon. Not. R. Astron. Soc. **515**, 3299 (2022).
[88] Ryuichi Takahashi and Takashi Nakamura, Wave effects in gravitational lensing of gravitational waves from chirping binaries, Astrophys. J. **595**, 1039 (2003).
[89] Giovanni Tambalo, Miguel Zumalacárregui, Liang Dai, and Mark Ho-Yeuk Cheung, Gravitational wave lensing as a probe of halo properties and dark matter, arXiv:2212.11960.
[90] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), GW190521: A Binary Black Hole Merger with a Total Mass of $150M_{\odot}$, Phys. Rev. Lett. **125**, 101102 (2020).