Measuring source properties and quasi-normal-mode frequencies of heavy massive black-hole binaries with LISA

Alexandre Toubiana,[1] Lorenzo Pompili,[1] Alessandra Buonanno,[1,2] Jonathan R. Gair,[1] and Michael L. Katz[1,3]

[1]*Max Planck Institute for Gravitationsphysik (Albert Einstein Institute), Am Mühlenberg 1, 14476 Potsdam, Germany*
[2]*Department of Physics, University of Maryland, College Park, MD 20742, USA*
[3]*NASA Marshall Space Flight Center, Huntsville, Alabama 35811, USA*

The laser-interferometer space antenna (LISA) will be launched in the mid 2030s. It promises to observe the coalescence of massive black-hole (BH) binaries with signal-to-noise ratios (SNRs) reaching thousands. Crucially, it will detect some of these binaries with high SNR both in the inspiral and the merger-ringdown stages. Such signals are ideal for tests of General Relativity (GR) using information from the whole waveform. Here, we consider astrophysically motivated binary systems at the high-mass end of the population observable by LISA, and simulate their LISA signals using the newly developed parametrised, multipolar, aligned-spin effective-one-body model: `pSEOBNRv5HM`. This model includes the $(2,2)$, $(2,1)$, $(3,3)$, $(3,2)$, $(4,4)$, $(4,3)$ and $(5,5)$ harmonics. The merger-ringdown signal in this model depends on the binary properties (masses and spins), and also on parameters that describe fractional deviations from the GR quasi-normal-mode (complex) frequencies of the remnant BH. Performing full Bayesian analyses, we assess to which accuracy LISA will be able to constrain deviations from GR in the ringdown signal when using information from the whole signal. We find that these deviations can typically be constrained to within 10% and in the best cases to within 1%. We also show that with this model we can measure the binary masses and spins with great accuracy even for very massive BH systems with low SNR in the inspiral, thanks to having a consistent dependence on the binary parameters in the merger-ringdown signal. In particular, individual source-frame masses can typically be constrained to within 10% and as precisely as 1%, and individual spins can typically be constrained to within 0.1 and, in the best cases, to within 0.001. Finally, we probe the accuracy of the `SEOBNRv5HM` waveform family by performing synthetic injections of GR numerical-relativity waveforms. For the source parameters considered, when higher harmonics are included and the SNR $\gtrsim 87$, we measure erroneous deviations from GR due to systematics in the waveform model. These results confirm the need for improving waveform models to perform tests of GR with binary BHs at high SNR with LISA.

I. INTRODUCTION

We are now well into the era of gravitational-wave (GW) astronomy, with 90 observations of compact-object binaries [1, 2] by the LIGO-Virgo-KAGRA (LVK) Collaboration [3–5], and other claimed detections [6, 7]. The fourth observing run of the LVK Collaboration has just started, with the promise of many new detections thanks to improved sensitivity [8]. In addition to unveiling an otherwise hardly detectable population of binary black holes (BBHs) [9–13], constraining the equation-of-state of neutron stars [14, 15] and inferring astrophysical and cosmological information [16, 17], GWs allow us to test General Relativity (GR) [18–21] in the strong-gravity and high-velocity regime, which is not accessible to other experiments. Indeed, by comparing predictions for the GW signal of a BBH within GR to the observed data we can constrain deviations from GR.

One of the most promising approaches to probe deviations from GR with GWs are the so-called "ringdown tests". In the last stage of the coalescence of a BBH, after the two BHs have merged, the remnant BH is in a perturbed state and relaxes to a steady state configuration through GW emission. This stage is called the ringdown. In this final stage, the signal is a superposition of damped sinusoids with frequencies and damping times that depend exclusively on the properties of the remnant [22–27]. Within GR, the "no-hair" conjecture [28] tells us that those are the mass and the spin of the final BH, since astrophysical BHs are expected to carry no electric charge. Some gravity theories predict additional "hairs" for BHs, due for instance to cosmological boundary conditions or the presence of nearby matter [29–31] or to additional fields [32–39]. In any case, the exact relation between the properties of the remnant and the spectrum of quasi-normal modes (QNMs) (i.e., the sets of frequencies and damping times) is theory-dependent [40–54]. By measuring two or more QNMs we can test if the signal agrees with GR. This is the basic idea behind BH spectroscopy [55, 56]. On the other hand, the amount to which each mode is excited (i.e., its amplitude), and the relative phases between them do depend on the properties of the BHs in the binary and the binary dynamics [57–60]. Therefore, a consistent modelling of the merger-ringdown together with the inspiral can improve our ability to measure the QNMs, and to constrain deviations from GR during the ringdown. This is the approach followed in Refs. [61, 62], where the authors developed a parametrised model of the ringdown signal as part of the full inspiral-merger-ringdown (IMR) waveforms [63, 64] in the effective one-body (EOB) formalism [65, 66]. Such a model can be used to perform parametrised (or theory-agnostic) ringdown tests of GR by allowing the QNMs to deviate from their GR prediction: a departure from the Kerr spectrum would be indicative of non-GR effects. The model has also been extended to parametrise the plunge-merger stages in Ref. [67], and to carry out theory-specific tests of GR in the ringdown in Ref. [68]. Here, we employ the parametrised ringdown test, which has already been applied to analyse the GW signals observed by the LVK Collaboration, showing so far consistency with GR [20, 21, 61, 62, 67]. The precision of the test has so far been limited by the low signal-to-noise ratio (SNR) of the sources, which has led to measurement errors for the frequency and decay time of the dominant QNM on the order of 10% and 20%, respectively, when combining events in a hierarchical way [21]. More specifically, for LIGO-Virgo observations this test can be applied only when both the pre- and post-inspiral regimes have at least SNR ~ 8, which has been the case for 12 binary systems [21]. The best single-event measurement [20] has been obtained with GW150914, which has a total SNR of 24 [69]. Other approaches have also been developed to do BH spectroscopy with LIGO-Virgo data [20, 21], using a superposition of damped sinusoids [70, 71], in some cases augmented with QNM amplitudes calibrated to numerical-relativity (NR) simulations.

Scheduled for launch in the mid-2030s, the laser-interferometer space antenna (LISA) [72], will detect massive BH binaries (MBHBs) with SNRs reaching thousands, sometimes both in the inspiral and in the merger-ringdown [73–75]. MBHBs are therefore promising candidates for performing ringdown tests that use information from the full signal. Previous studies on ringdown analysis with LISA [58, 74, 76] focused on "pure" ringdown tests (i.e., using a superposition of damped sinusoids after the merger) and employed simplified methods and criteria such as the Fisher matrix formalism [77, 78] to estimate the measurement accuracy and the distinguishability between QNMs. In this paper, we simulate LISA observations of MBHBs and run full Bayesian analyses on those in order to assess to which accuracy putative deviations from GR in the ringdown could be constrained from such observations when using information from the whole signal. Furthermore, we make use of parametrised EOB waveforms developed using the state-of-the-art multipolar aligned-spin model `SEOBNRv5HM` [1] developed in Refs. [79–82]. Henceforth, we denote the parametrised model as `pSEOBNRv5HM`.

Since the expected population of MBHBs is highly uncertain [73, 83–88], here we focus on a few astrophysically realistic systems, compatible with the predictions of models where MBHs form from heavy seeds [89]. For simplicity, in this work we neglect the effect of spin-precession and eccentricity. Our study is performed by using the same waveform model for generating mock injections and for estimating the parameters of the source. However, it is crucial to assess if such tests of GR could be spoiled by the limited accuracy of our theoretical models when performed on real data. Therefore, we assess the impact of systematics in waveform modelling on ringdown tests by simulating mock LISA injections with waveforms from numerical relativity (NR) and using `pSEOBNRv5HM` waveforms to perform the Bayesian analysis. Finally, we assess to which extent a consistent modelling of the full signal allows us to measure the binary parameters in GR also for systems that are merger-ringdown dominated and have low SNR in the inspiral.

[1] The generic name `SEOBNRvnEPHM` indicates that the version `vn` of the EOB model is calibrated to NR simulations (NR), includes spin (S) and precessional (P) effects, eccentricity (E) and higher modes (HM, i.e. higher harmonics).

This paper is organised as follows. In Sec. II we present the details of our parametrised EOB model, describe how the synthetic LISA observations are generated, and lay down the basis of our Bayesian analyses. In Sec. III we summarise the astrophysical systems that we simulate. We present our results when using the `pSEOBNRv5HM` model both for injection and parameter estimation in Sec. IV, then in Sec. V we discuss the impact of systematics. Finally, we present our conclusions in Sec. VI. In the appendices, we discuss how the settings of `pSEOBNRv5HM` waveforms impact the parameter estimation, and we show that, in GR, `SEOBNRv5HM` and `IMRPhenomTHM` [90], a time-domain waveform model from the IMR phenomenological family [91–93], predict similar measurement errors for the parameters of the source, and that their measurement is little affected by adding the QNM deviation parameters in the `pSEOBNRv5HM` model. Throughout this paper we will use natural units in which $c = G = 1$.

II. METHODS

A. Parametrised waveform model

We consider a binary with BH component (detector-frame) masses m_1 and m_2 and define the mass ratio $q = m_1/m_2 \geq 1$ and the total mass $M_t = m_1 + m_2$. We limit to BHs moving on quasi-circular orbits with aligned/anti-aligned spins (aligned spins for short), and define the (dimensionless) spin variables $\chi_1 = S_1/m_1^2$ and $\chi_2 = S_2/m_2^2$, which range between -1 and 1. We denote the luminosity distance of the source as D_L and the cosmological redshift as z. We adopt the cosmology determined by the Planck mission (2018) [94]. Masses, times and frequencies are in the detector-frame, unless they carry a subscript s. Source-frame masses $m_{1,s}$ and $m_{2,s}$ are related to the detector-frame ones by $m_i = (1 + z)m_{i,s}$.

The GW polarisations can be expanded in the basis of spin-weight -2 spherical harmonics as

$$h_+(\mathbf{\Theta}, \iota, \varphi_0; t) - i h_\times(\mathbf{\Theta}, \iota, \varphi_0; t) = \frac{1}{D_L} \sum_{\ell,m} {}_{-2}Y_{\ell m}(\iota, \varphi_0)\, h_{\ell m}(\mathbf{\Theta}; t)\,, \tag{2.1}$$

where the parameters (ι, φ_0) denote the binary's inclination angle with respect to the direction perpendicular to the orbital plane and the azimuthal direction to the observer, respectively, and $\mathbf{\Theta}$ denotes the intrinsic parameters (masses and spins) of the binary. We build our parametrised model using the `SEOBNRv5HM` model [81] in GR, which includes several higher harmonics, notably the $(\ell, |m|) = (2,1), (3,3), (3,2), (4,4), (4,3)$ and $(5,5)$ harmonics, in addition to the dominant $(2,2)$ harmonic. For aligned-spin binaries $h_{\ell m} = (-1)^\ell h^*_{\ell -m}$, therefore, we focus on (ℓ, m) harmonics with $m > 0$.

In the EOB framework [66], the GW harmonics are decomposed as

$$\begin{aligned} h_{\ell m}(\mathbf{\Theta}, t) = {} & h_{\ell m}(\mathbf{\Theta}, t)^{\text{insp-plunge}}\, \theta(t^{\ell m}_{\text{match}} - t) \\ & + h_{\ell m}(\mathbf{\Theta}, t)^{\text{merger-RD}}\, \theta(t - t^{\ell m}_{\text{match}})\,, \end{aligned} \tag{2.2}$$

where $\theta(t)$ is the Heaviside step function, $h^{\text{insp-plunge}}_{\ell m}$ corresponds to the inspiral-plunge part of the waveform, while

$h_{\ell m}^{\text{merger-RD}}$ represents the merger-ringdown waveform. In particular, as explained in Ref. [81], $t_{\text{match}}^{\ell m}$ is chosen to be the peak of the $(2, 2)$ harmonic amplitude for all (ℓ, m) harmonics except $(5, 5)$, for which it is taken as the peak of the $(2, 2)$ harmonic minus $10M_t$. In the following, we suppress the $\mathbf{\Theta}$ dependence for ease of notation.

For all harmonics, except for $(\ell, |m|) = (3, 2)$ and $(4, 3)$ which exhibit post-merger oscillations due to mode-mixing [95, 96], the merger-ringdown waveform employs the following ansatz [64, 81, 97],

$$h_{\ell m}^{\text{merger-RD}}(t) = \nu\, \tilde{\mathcal{A}}_{\ell m}(t)\, e^{i\tilde{\phi}_{\ell m}(t)}\, e^{-i\sigma_{\ell m0}(t-t_{\text{match}}^{\ell m})}, \tag{2.3}$$

where $\nu = m_1 m_2/(m_1+m_2)^2$ is the symmetric mass ratio of the binary, and $\sigma_{\ell m0}$ is the complex frequency of the least-damped QNM, having overtone number zero, of the remnant BH. We define the corresponding oscillation frequency, $f_{\ell m0}$, and the damping time, $\tau_{\ell m0}$, as

$$f_{\ell m0} = \frac{1}{2\pi}\,\mathrm{Re}\,(\sigma_{\ell m0}) = -\frac{1}{2\pi}\sigma_{\ell m0}^{\mathrm{I}}, \tag{2.4a}$$

$$\tau_{\ell m0} = -\frac{1}{\mathrm{Im}\,(\sigma_{\ell m0})} = -\frac{1}{\sigma_{\ell m0}^{\mathrm{R}}}. \tag{2.4b}$$

The functions $\tilde{\mathcal{A}}_{\ell m}(t)$ and $\tilde{\phi}_{\ell m}(t)$ are given by [64, 81, 97–99]:

$$\tilde{\mathcal{A}}_{\ell m}(t) = c_{1,c}^{\ell m} \tanh[c_{1,f}^{\ell m}\,(t - t_{\text{match}}^{\ell m}) \,+\, c_{2,f}^{\ell m}] \,+\, c_{2,c}^{\ell m}, \tag{2.5a}$$

$$\tilde{\phi}_{\ell m}(t) = \phi_{\text{match}}^{\ell m} - d_{1,c}^{\ell m} \log\left[\frac{1 + d_{2,f}^{\ell m} e^{-d_{1,f}^{\ell m}(t-t_{\text{match}}^{\ell m})}}{1 + d_{2,f}^{\ell m}}\right], \tag{2.5b}$$

where $\phi_{\text{match}}^{\ell m}$ is the phase of the inspiral-plunge harmonic (ℓ, m) at $t = t_{\text{match}}^{\ell m}$.

The coefficients $d_{1,c}^{\ell m}$ and $c_{i,c}^{\ell m}$ $(i = 1, 2)$ are constrained by the requirement that the amplitude and phase of $h_{\ell m}(t)$ are continuously differentiable (C^1) at $t = t_{\text{match}}^{\ell m}$. This allows us to write the coefficients $c_{i,c}^{\ell m}$ as [64, 81]:

$$\begin{aligned} c_{1,c}^{\ell m} = &\frac{1}{c_{1,f}^{\ell m}\nu}[\partial_t |h_{\ell m}^{\text{insp-plunge}}(t_{\text{match}}^{\ell m})| \\ &- \sigma_{\ell m}^{\mathrm{R}} |h_{\ell m}^{\text{insp-plunge}}(t_{\text{match}}^{\ell m})|]\cosh^2(c_{2,f}^{\ell m}), \end{aligned} \tag{2.6a}$$

$$\begin{aligned} c_{2,c}^{\ell m} = &\frac{|h_{\ell m}^{\text{insp-plunge}}(t_{\text{match}}^{\ell m})|}{\nu} - \frac{1}{c_{1,f}^{\ell m}\nu}[\partial_t |h_{\ell m}^{\text{insp-plunge}}(t_{\text{match}}^{\ell m})| \\ &- \sigma_{\ell m}^{\mathrm{R}} |h_{\ell m}^{\text{insp-plunge}}(t_{\text{match}}^{\ell m})|]\cosh(c_{2,f}^{\ell m})\sinh(c_{2,f}^{\ell m}), \end{aligned} \tag{2.6b}$$

and $d_{1,c}^{\ell m}$ as

$$d_{1,c}^{\ell m} = \left[\omega_{\ell m}^{\text{insp-plunge}}(t_{\text{match}}^{\ell m}) - \sigma_{\ell m}^{\mathrm{I}}\right]\frac{1 + d_{2,f}^{\ell m}}{d_{1,f}^{\ell m} d_{2,f}^{\ell m}}, \tag{2.7}$$

where $\omega_{\ell m}^{\text{insp-plunge}}(t)$ is the frequency of the inspiral-plunge EOB harmonic. The coefficients $c_{i,f}^{\ell m}$ and $d_{i,f}^{\ell m}$ are obtained through fits to a large set of NR waveforms ($\sim$ 440), spanning mass ratios up to 20 and spins up to 0.998, and BH perturbation-theory merger-ringdown waveforms for mass ratio 1000. Crucially, the fits depend on the binary's masses and spins $\mathbf{\Theta}$, and can be found in Appendix D of Ref. [81]. As an example, we illustrate in Fig. 1 how the GW amplitude and frequency of the (2,2) harmonic changes, during the late inspiral, merger and ringdown, as the component spins are varied, for a binary with mass ratio 2 and equal spins.

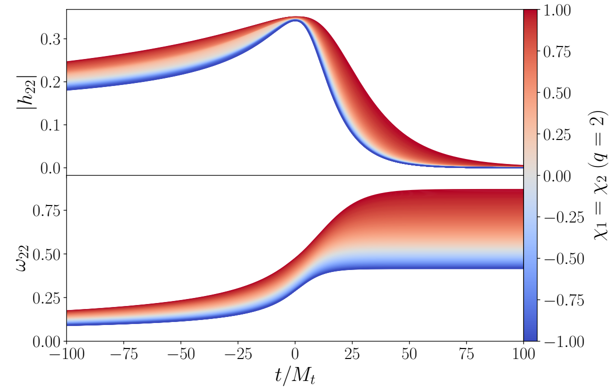


FIG. 1. Time evolution near the merger, which occurs at $t = 0$, of the `SEOBNRv5HM` (2, 2) harmonic amplitude (top panel) and instantaneous frequency (bottom panel) when varying the spin components of a binary with mass ratio 2, assuming equal and aligned spins.

For the $(3, 2)$ and $(4, 3)$ harmonics, the mode-mixing behaviour is modelled by applying the previous construction to the spheroidal harmonics [100] $(3, 2, 0)$ and $(4, 3, 0)$, which feature a monotonic amplitude and frequency evolution [101]. The spheroidal $(3, 2, 0)$ and $(4, 3, 0)$ harmonics can be related to the spherical harmonics by [81]

$$^{S}h_{320}(t) \simeq \frac{h_{32}(t)\,\mu_{2220}^* - h_{22}(t)\,\mu_{2320}^*}{\mu_{2330}^*\,\mu_{2220}^*}, \tag{2.8a}$$

$$^{S}h_{430}(t) \simeq \frac{h_{43}(t)\,\mu_{3440}^* - h_{33}(t)\,\mu_{3430}^*}{\mu_{3330}^*\,\mu_{3440}^*}, \tag{2.8b}$$

where $\mu_{m\ell\ell'n}$ are harmonic mixing coefficients, obtained using fits from Ref. [102]. Thus, the $(3, 2)$ and $(4, 3)$ harmonics are obtained by combining the $(3, 2, 0)$ and $(4, 3, 0)$ harmonics with the $(2, 2)$ and $(3, 3)$ harmonics, inverting Eqs. (2.8a) and (2.8b).

In the `SEOBNRv5HM` model constructed in Ref. [81], the complex QNM frequencies in GR are obtained for each (ℓ, m) harmonic as a function of the BH's final mass and spin using the `qnm` Python package [103]. The BH's mass and spin are in turn computed using the fitting formulas of Refs. [104] and [105], respectively. In this work, following the strategy of Refs. [61, 62, 67, 71, 106–108], we introduce parametrised fractional deviations to the QNM frequencies, which are free parameters of the model (see Ref. [68] where the deviations were mapped to specific gravity theories alternative to GR).

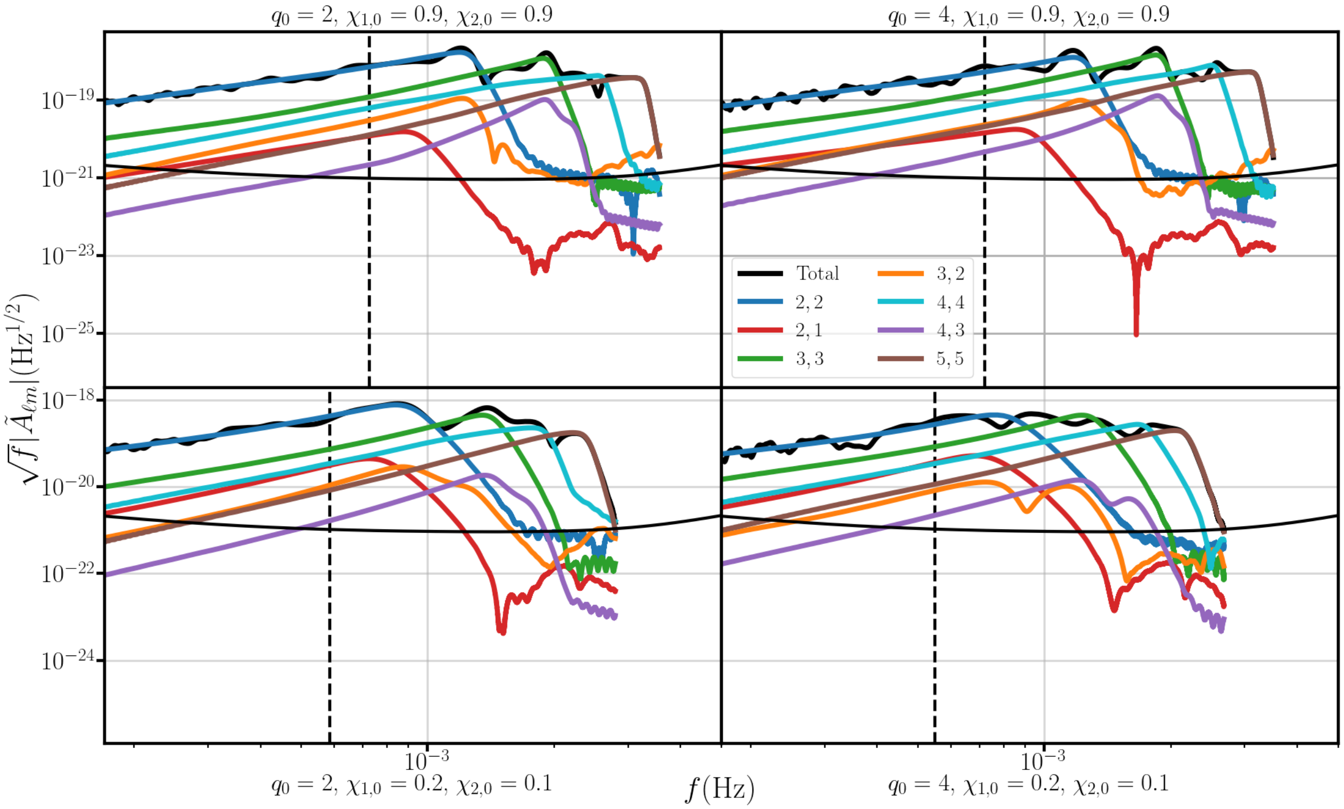


FIG. 2. Amplitude in frequency domain in the TDI channel A broken into the contributions of each of the harmonics included in the pSEOBNRv5HM model, $\tilde{A}_{\ell m}$. The latter is computed using the harmonic decomposition of Eq. (2.1) in the TDI equations (2.10a). The thin black line shows the LISA PSD and the black-dashed lines the GW frequency at t_c (i.e., the separation between the inspiral and merger-ringdown regimes). We recall that t_c is defined from the peak of h_{22}, and that the computation of the TDI variables involves second derivatives of the GW polarisation, leading to an offset between the maximum of h_{22} and those of A and E. Moreover, these quantities are defined in time-domain, whereas we show here the frequency-domain amplitudes, and the time-domain peak typically corresponds to lower frequencies than the frequency-domain one. We plot the amplitudes only for $M_{t,0} = 2 \times 10^7 M_\odot$ and $z_0 = 3$. For systems at $z_0 = 5$, one should simply rescale the amplitudes by $25,924/47,647 \simeq 0.54$, and for systems with $M_{t,0} = 2 \times 10^8 M_\odot$, one should multiply the amplitudes by 10^2 and the frequencies by 10^{-1}. As expected, the (2,2) harmonic is the loudest, followed by the (3,3), (4,4) and (5,5) harmonics. For high-spins systems (upper row), the (3,2) and (4,3) harmonics are louder than the (2,1) harmonic, whereas the opposite is true for low-spins systems (lower row).

More explicitly, we perform the substitutions

$$f_{\ell m0} \to f_{\ell m0}\,(1 + \delta f_{\ell m}), \tag{2.9a}$$
$$\tau_{\ell m0} \to \tau_{\ell m0}\,(1 + \delta \tau_{\ell m}), \tag{2.9b}$$

where for ease of notation we have dropped the zero overtone subscript in the deviation parameters. We shall denote this parametrised model as pSEOBNRv5HM. We note that allowing $\sigma_{\ell m0}$ to vary freely also modifies the $c_{i,c}^{\ell m}$ and $d_{1,c}^{\ell m}$ coefficients in Eqs. (2.6a), (2.6b), and (2.7), which enter the amplitude and phase functions $\tilde{\mathcal{A}}_{\ell m}(t)$ and $\tilde{\phi}_{\ell m}(t)$. As a consequence, such a modification can lead to deviations from the GR prediction in the ringdown signal starting soon after the merger. The plunge-merger stage of the waveform could be, in principle, also modified, as done for example in Ref. [67], by introducing deviations with respect to the GR predictions to the time at which the amplitude peaks, and to the value of the amplitude and frequency at this instant, for each waveform harmonic.

Finally, the inspiral-plunge EOB waveforms (2.2) are computed based on the two-body dynamics that are computed by solving Hamilton's equations with a suitable EOB Hamiltonian and radiation-reaction force (see Refs. [80, 81] for details).

B. Generation of LISA signals

We use the long-wavelength approximation [109] to compute the response of LISA to an incoming GW, which is valid when the GW wavelength is much larger than the LISA arm length L (i.e., in terms of the GW frequency, when $2\pi f L/c \ll 1$). Given that $L = 2.5 \times 10^8$m, this condition is satisfied for sources reaching maximum frequencies of $\sim 10^{-3}$Hz, such as the MBHBs we consider in this work. Under this approximation, LISA is somewhat similar to two LIGO/Virgo-type detectors rotated with respect to each other by $\pi/4$, and with angles of $\pi/3$ between the arms.

Transforming Eq. (47) of Ref. [110] to the time domain, we find that under the long-wavelength approximation the time-delay-interferometry (TDI) variables A, E and T [111] (which, for an interferometer with equal arms and equal noise levels in each optical link, provide three noise-uncorrelated datasets) are given by:

$$\begin{aligned}
A &= -3\sqrt{2}\left(\frac{L}{c}\right)^2\left[F_+(\lambda,\beta,\psi)\ddot{h}_+ + F_\times(\lambda,\beta,\psi)\ddot{h}_\times\right],\\
E &= -3\sqrt{2}\left(\frac{L}{c}\right)^2\left[F_+(\lambda+\pi/4,\beta,\psi)\ddot{h}_+ + F_\times(\lambda+\pi/4,\beta,\psi)\ddot{h}_\times\right],\\
T &= 0,
\end{aligned} \tag{2.10a}$$

where F_+ and $F_\times$ are the antenna pattern functions:

$$F_+(\lambda,\beta,\psi) = \cos(2\psi)F_{+,0}(\lambda,\beta) + \sin(2\psi)F_{\times,0}(\lambda,\beta), \tag{2.11a}$$

$$F_\times(\lambda,\beta,\psi) = -\sin(2\psi)F_{+,0}(\lambda,\beta) + \cos(2\psi)F_{\times,0}(\lambda,\beta), \tag{2.11b}$$

$$F_{+,0}(\lambda,\beta) = \frac{1}{2}\left(1+\sin^2\beta\right)\cos(2\lambda-\pi/3), \tag{2.11c}$$

$$F_{\times,0}(\lambda,\beta) = \sin^2\beta\sin(2\lambda-\pi/3). \tag{2.11d}$$

In the above equations, λ, β and ψ are the longitude, latitude and polarisation in the LISA frame. We refer to Ref. [110] for the relation between the angles in the LISA frame and those in the solar-system-barycentre frame. As anticipated, this is similar to the response of ground-based detectors. The main difference is that it is the second derivatives of the waveform polarisations that enter Eq. (2.10a). This comes as a consequence of taking waveform differences in order to perform TDI, with a time-step that goes to zero in the long-wavelength approximation. Finally, because most of the SNR of the signals we consider is accumulated in the last stages of the evolution (i.e., from a few hours to a few days), we do not take into account the motion of LISA about the Sun. Therefore, λ, β and ψ in Eq. (2.10a) are not varying with time. Henceforth, we will use $\tilde{A}$ and $\tilde{E}$ to denote the Fourier transform of A and E.

We generate EOB waveforms from the frequency $\mathrm{f_{gen}} = 5\times10^{-5}[M_{t,0}/(2\times10^7 M_\odot)]$Hz until the end of the signal. After transforming to the frequency-domain, we keep the portion of the signal between $f_{\rm min} = 2\times10^{-4}[M_{t,0}/(2\times10^7 M_\odot)]$Hz and $f_{\rm max}$, to eliminate spurious features due to Fourier transform. The maximum frequency is chosen such that the frequency-domain amplitude is 1% of its maximum value. We verified that our choice of $f_{\rm min}$ leads to a loss in SNR of less than 2%.

The time and phase alignment of the signals is done in the following way. We define the time to coalescence, t_c, as the moment the amplitude of the (2,2) harmonic reaches its peak, and define the phase of coalescence, φ_c, as the phase of the (2,2) harmonic contribution to the total waveform at t_c. In practice, the last step is done by choosing the azimuthal angle φ_0 such that the phase of ${}_{-2}Y_{22}(\iota,\varphi_0)\,h_{22}(t_c)$ is φ_c. This choice of φ_0 is then propagated consistently to other harmonics. We use t_c to split between the inspiral and merger-ringdown regimes. We note that t_c for EOB waveforms coincides with $t^{\ell m}_{\rm match}$ in Eq. (2.2), for all (ℓ,m) harmonics except $(5,5)$.

C. Bayesian analysis

We define the noise-weighted inner product between two data streams, d_1 and d_2, as:

$$(d_1|d_2) = 4\mathcal{R}e\left[\int_0^{+\infty}\frac{d_1(f)d_2^*(f)}{S_n(f)}df\right], \tag{2.12}$$

where $S_n(f)$ is the power spectral density (PSD). In this work, we use the SciRDv1 noise curve [112], which corresponds to the scientific requirement for the LISA mission, and defines pessimistic noise levels compared to current predictions. For a given choice of the PSD, the SNR of a signal h is defined as $\mathrm{SNR} = \sqrt{(h|h)}$.

To quantify the precision with which LISA observations will estimate the parameters of a source, we work in a Bayesian framework and compute the posterior distribution on the source parameters, θ, given an observed dataset, d, using Bayes' theorem:

$$p(\theta|d) = \frac{p(d|\theta)p(\theta)}{p(d)}, \tag{2.13}$$

where $p(d|\theta)$ is the likelihood, $p(\theta)$ is the prior and $p(d)$ is the evidence. As long as we are not interested in model selection, the latter acts as a normalisation constant, and thus can be discarded. We take the prior to be flat in the (detector-frame) total mass, M_t, the mass-ratio, q, the spins, $-1\le\chi_1\le1$ and $-1\le\chi_2\le1$, the time to coalescence, t_c, and the phase at coalescence, φ_c. For the systems we consider here, the intrinsic parameters M_t, q, χ_1 and χ_2 are typically well measured, so that the actual priors have little importance. We take a flat prior on ψ, $\cos(\iota)$ and $\log_{10}(D_L)$ and fix the sky location (λ, β) to its true value to facilitate the convergence of the chains. Those parameters are not expected to correlate strongly with intrinsic parameters [110], at least for aligned-spin binaries, and so this simplification should not significantly affect our conclusions. Finally, we take a flat prior between -1 and 1 for the fractional deviations to the QNMs, $\delta f_{\ell m}$ and $\delta\tau_{\ell m}$. Assuming noise to be stationary and Gaussian, the likelihood reads:

$$p(d|\theta) \propto \prod_{c\in[A,E]}\exp\left[-\frac{1}{2}(d_c - s_c(\theta)|d_c - s_c(\theta))\right]. \tag{2.14}$$

The posterior distribution is then sampled via a Markov-chain Monte-Carlo algorithm (MCMC). We use the Eryn sampler [113, 114] for this purpose.

III. ASTROPHYSICAL BINARY SYSTEMS

We consider a set of 16 binary systems, defined from all possible combinations of the following choices of parameters:

- $M_{t,0} = 2\times10^7 M_\odot$ or $M_{t,0} = 2\times10^8 M_\odot$,
- $q_0 = 2$ or $q_0 = 4$,
- $\chi_{1,0} = \chi_{2,0} = 0.9$ or $\chi_{1,0} = 0.2$, $\chi_{2,0} = 0.1$,

		$M_{t,0} = 2 \times 10^7 M_\odot$			$M_{t,0} = 2 \times 10^8 M_\odot$		
		IMR	Merger-Ringdown	Inspiral	IMR	Merger-Ringdown	Inspiral
$\chi_{1,0} = \chi_{2,0} = 0.9$	$q_0 = 2$	3892	3715	1162	441	435	69
	$q_0 = 4$	3659	3547	897	47	472	53
$\chi_{1,0} = 0.2, \chi_{2,0} = 0.1$	$q_0 = 2$	1894	1809	562	171	168	28
	$q_0 = 4$	1296	1231	404	118	117	20

TABLE I. Full IMR SNR and its decomposition into the contribution of the merger-ringdown and inspiral stages for the systems considered in this work at $z_0 = 3$. Values at $z_0 = 5$ can be obtained by rescaling by $25,924/47,647 \simeq 0.54$. We recall that the inspiral and merger-ringdown SNRs add quadratically. All the systems we consider are merger-ringdown dominated, although some of them have high SNR in the inspiral as well, up to thousands.

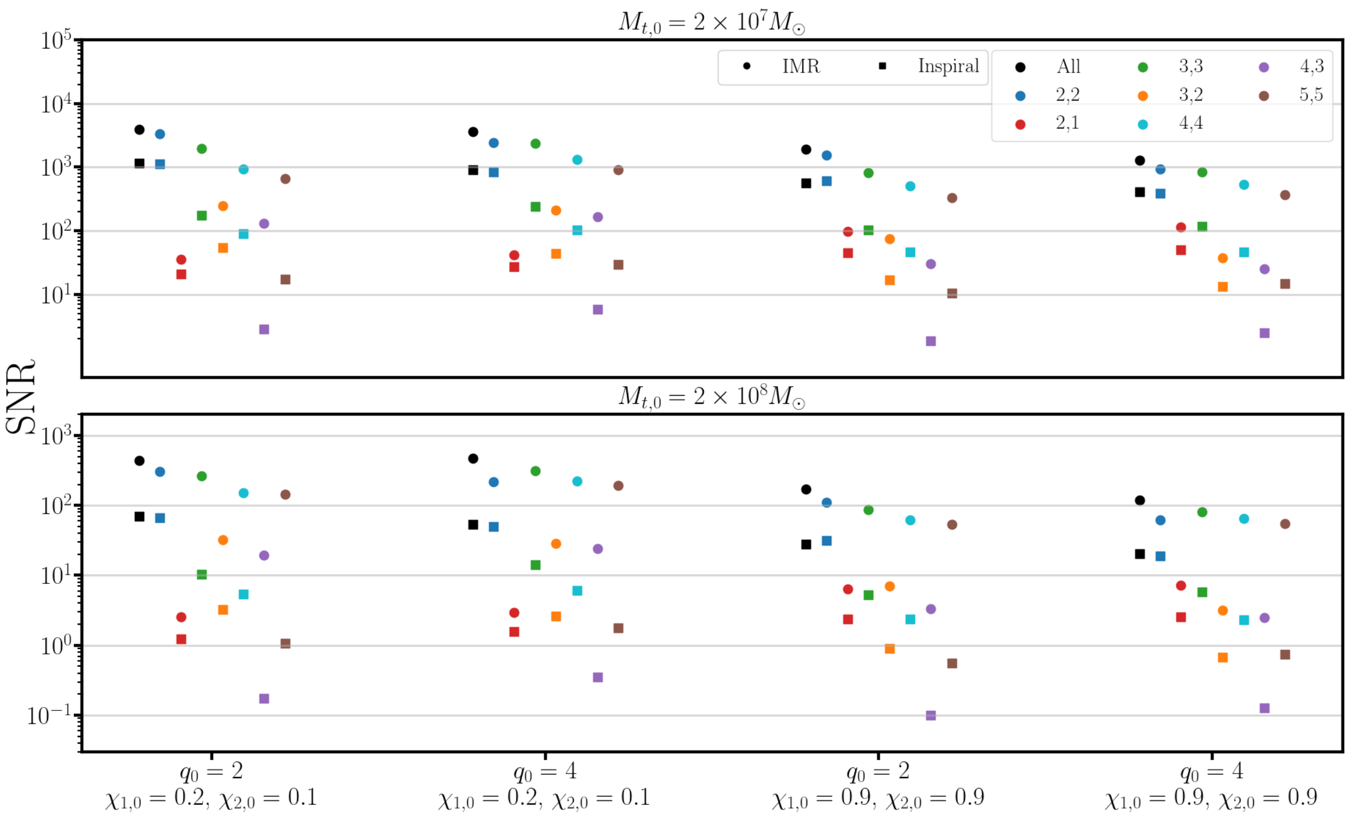


FIG. 3. Total SNR (black) and SNR in the individual harmonics (colours) for all the systems at $z_0 = 3$. The values at $z_0 = 5$ can be obtained by rescaling by $25,924/47,647 \simeq 0.54$. Circles show the full IMR SNR and squares the SNR in the inspiral stage.

- $z_0 = 3$ ($D_{L,0} = 25,924$ Mpc) or $z_0 = 5$ ($D_{L,0} = 47,647$ Mpc).

Subscripts 0 indicate the true value of the parameter used to generate the synthetic injections. This set of systems lies in the high-mass end of predictions for the population visible to LISA, as predicted from semi-analytic models of MBH populations that use heavy seeds for the MBH progenitors [73, 83–88]. Different heavy seed scenarios have been proposed, such as the collapse of proto-galactic disks as a result of bar instabilities [115], the run-away collision of stars at the centre of galaxies [116], or the direct collapse of gas at the centre of galaxies [117] (see Ref. [89] for a review). Such heavy systems are the ones expected to have higher SNR in the merger-ringdown [74, 75], and are therefore the most relevant ones to our analysis. In particular, very heavy systems ($M_{t,0} = 2 \times 10^8 M_\odot$) are expected to have very little SNR in the inspiral, and it is interesting to assess how well the parameters of such systems can be measured. Focusing on such massive systems is even more well-motivated following the latest results from pulsar-timing–array observations [118–121]. If the apparent signal in the pulsar-timing–array data is generated by massive black hole inspirals, it indicates that MBHs might be more massive than originally expected. Semi-analytical models predict a wide range of values for the mass-ratio, but the vast majority of systems are predicted to have comparable masses. This is also the domain where our current IMR models are the most reliable. The spin of a MBH typically depends on the amount of gas surrounding it and on how it has acquired mass and angular momentum through accretion [122].

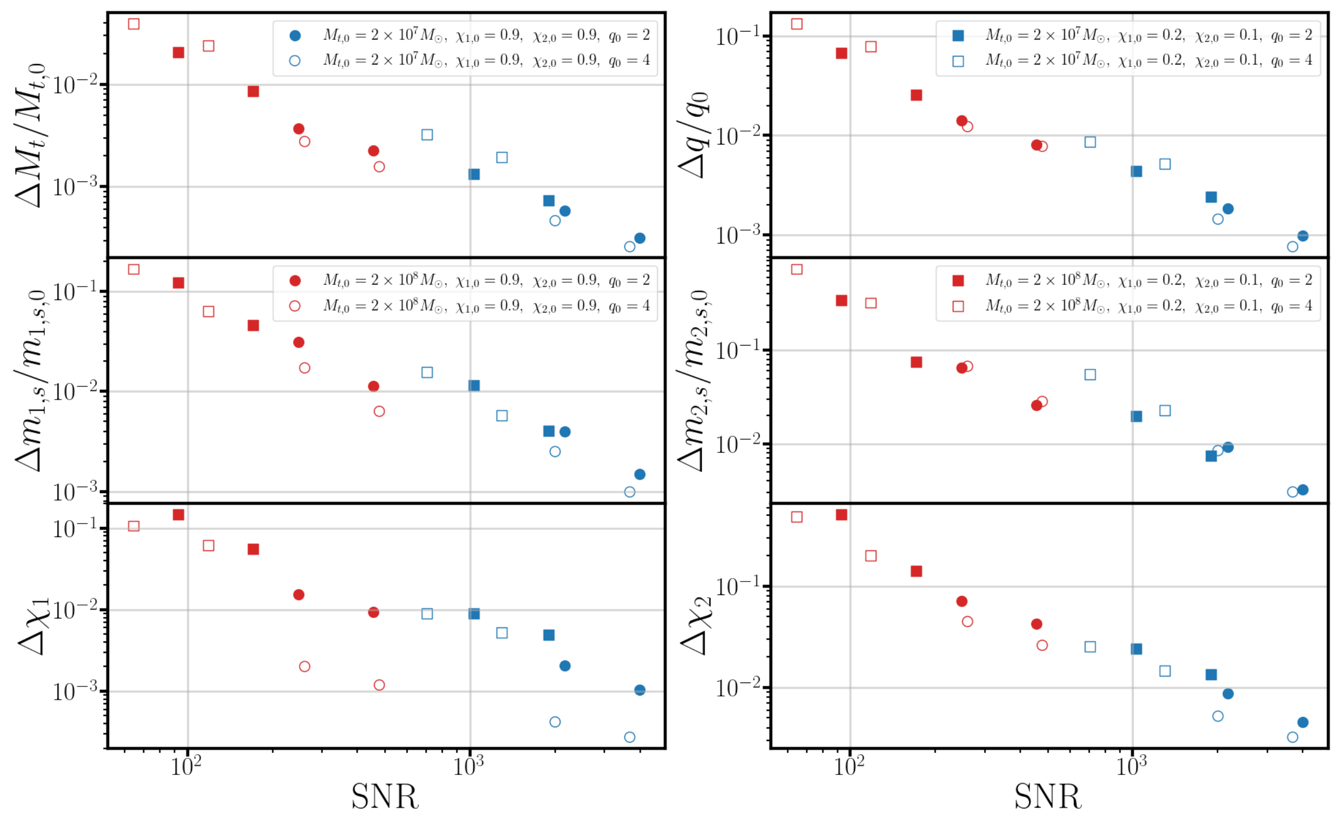


FIG. 4. Measurement error (90% confidence interval centred around the median) of intrinsic parameters. For mass parameters, we show the errors relative to the injection value. We inject signals within GR and recover them with GR templates (i.e., we employ `SEOBNRv5HM`). All points are doubled because of the two redshift possibilities, $z_0 = 3$ ($z_0 = 5$) yields larger (lower) SNR and lower (larger) errors. We show the detector-frame total mass and the mass ratio in the top row, the source-frame individual masses in the middle one, and the dimensionless spins in the bottom row. At leading order, measurement errors follow the $1/\mathrm{SNR}$ trend.

We consider two possibilities in order to cover both the case where MBHs are efficiently-spun up and the case they are not. The relative alignment of the spins in a MBHB also depends on the presence of gas around the binary. Mergers happening in a gas-rich environment tend to have aligned-spins due to the Bardeen-Peterson effect [123, 124]. Binaries formed through triplet interactions can also have misaligned spins, in addition to having high eccentricity [85]. As discussed, we neglect eccentricity and spin-precession here for simplicity, and focus on quasi-circular binaries. Exploring different values for q, χ_1 and χ_2 is interesting because they affect how much higher-harmonics are excited, and therefore how well it is possible to constrain the QNMs other than the fundamental one. Finally, the redshifts at which MBHBs coalesce depend on when MBH seeds form and on the different time-scales at play during the hardening of the binary [125–127]. In heavy-seed models, MBHBs are expected to merge dominantly at late times (i.e., low redshift). Since M_t is the detector-frame total mass, changing the redshift only affects the SNR of the system. Based on the predictions of semi-analytical models [73, 83–88], we expect to observe up to a few tens of systems similar to the ones we defined during the nominal mission duration (four and a half years). For all binary systems we take $\iota_0 = \pi/3$, $\psi_0 = \pi/3$, $\lambda_0 = \pi/3$, $\beta_0 = \pi/3$ and $\varphi_{c,0} = 0$.

In Fig. 2 we plot the frequency-domain amplitude of the TDI variable A for the systems with $M_{t,0} = 2 \times 10^7 M_\odot$ and $z_0 = 3$, and the four combinations of mass ratio and spins. We show the contribution of each harmonic, $\tilde{A}_{\ell m}$, using Eq. (2.1) when computing the TDI variables (see Eq (2.10a)). The black-dashed lines indicate the GW frequency at t_c, which we choose as the separation between the inspiral and merger-ringdown regimes. As expected, the (2,2) harmonic is the loudest, but higher-harmonics are also important, in particular the (3,3), (4,4) and (5,5). This is better quantified in Fig. 3, where we show the total SNR and the contribution of each harmonic. We observe that the relative importance of the subdominant (2,1), (3,2) and (4,3) harmonics depends primarily on the spins: (2,1) is more dominant for low-spins systems, whereas (3,2) and (4,3) are more dominant for high-spins systems. We note the very high SNR of some of these systems, reaching a few thousands. Figure 3 also shows the inspiral SNR, defined by using the GW frequency at t_c as the upper limit in the integral of Eq. (2.12). As anticipated, the signals of the systems we consider are merger-ringdown dominated. In Table I, we give the IMR, merger-ringdown and inspiral SNR of the systems at $z_0 = 3$. Although systems with $M_{t,0} = 2 \times 10^7 M_\odot$ (upper panel) still have high SNR also in the inspiral, ~ 1000, it is not the case for the very massive systems ($M_{t,0} = 2 \times 10^8 M_\odot$, lower panel), with inspiral SNRs as low as ~ 10.

IV. MEASURING SOURCE PROPERTIES AND QNMS WITH MBHB OBSERVATIONS

We work with zero-noise injections, as these are well suited to the goals of understanding systematics and measurement uncertainties [128], and perform three types of analyses using the SEOBNRv5HM and pSEOBNRv5HM models as synthetic-signal injections and templates:

1. we inject a synthetic signal *without* deviations from GR (i.e., SEOBNRv5HM) and use templates in the Bayesian analyses *not allowing* for deviations from GR (i.e., SEOBNRv5HM),
2. we inject a synthetic signal *without* deviations from GR (i.e., SEOBNRv5HM) and use templates in the Bayesian analyses *allowing* for deviations from GR (i.e., pSEOBNRv5HM),
3. we inject a synthetic signal *with* deviations from GR (i.e., pSEOBNRv5HM) and use templates in the Bayesian analyses *allowing* for deviations from GR (i.e., pSEOBNRv5HM).

The first type of analysis will estimate how well the parameters of MBHBs can be constrained assuming GR is correct. It is the first study of this kind using EOB waveforms. The second will tell us how well the deviation parameters of QNMs can be constrained, and the third for which values of the deviation parameters we can detect non-GR effects in the ringdown. We perform these mock injections for all MBHBs described in Sec. III.

A. Measurement of source parameters in GR

We show in Fig. 4 the width of the 90% confidence interval centred around the median for the intrinsic parameters (i.e., the masses and spins) as a function of the SNR of the system. The colour, shape and filling of the point indicate respectively the total mass, spin and mass ratio of the system, indicated in the legends. Each point is doubled because of the two redshifts used: $z_0 = 3$ ($z_0 = 5$) corresponds to the largest (smallest) SNR and the smallest (largest) measurement error. Note that we show the detector-frame total mass in the top row and the source-frame individual masses in the middle one. For all systems the parameters are well constrained, and we find that the error (or relative error for the mass parameters) goes as $1/\mathrm{SNR}$, as expected in the high SNR regime [77, 78, 129]. This relation is more scattered for the spin parameters, especially for χ_1: systems with $q_0 = 4$ have better spin measurement, in agreement with [75]. This is because for such systems (non-filled points), higher harmonics become more dominant (see Figs. 2 and 3) and help improve the measurement of the spins. It is remarkable that even for very massive systems (red points), which usually have low SNR in the inspiral, we get tight constraints on their parameters, similarly to what [130] found. This is the benefit of using a fully consistent modelling of the IMR signal, since in our model the merger-ringdown signal also informs us on the parameters of the component BHs in the binary.

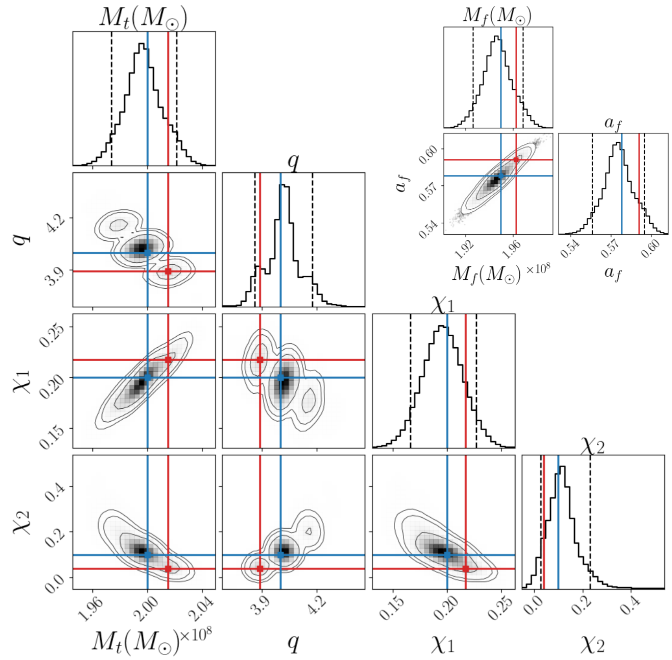


FIG. 5. Corner plot for the system with $M_{t,0} = 2\times10^8 M_\odot, \chi_{1,0} = 0.2$, $\chi_{2,0} = 0.1$, $q_0 = 4$, $z_0 = 3$, SNR=118. The blues lines show the injection and the red ones a secondary mode. Contours show the 68%, 90%, and 95% confidence intervals, and dashed lines the 0.05 and 0.95 quantiles. On the upper-right part, we show the inferred properties of the remnant. The multi-modality observed in the binary parameters is due to combinations of intrinsic parameters that yield remnant properties compatible with the true values within the measurement uncertainty, as can be seen from the absence of clear multi-modality in M_f and a_f.

For very massive systems with low SNR, we find a multi-modality in intrinsic parameters, as illustrated in Fig. 5. Secondary modes arise from combinations of parameters that yield remnant parameters compatible with the true ones within the measurement uncertainty, as shown in the upper-right part of Fig. 5. As a consequence, the merger-ringdown signal remains quasi-identical to the synthetic injection. This can be seen in Fig. 6, where we compare the waveform of the injected parameters (in blue in Fig. 5) to the one with parameters from one of the secondary maxima (in red in Fig. 5). The early part of the inspiral signal is fairly different, but this has little importance because the system is merger-ringdown dominated, and the SNR in the inspiral is very small (~ 10, see Fig. 3 and Table I).

In Appendix A we discuss the impact of the tolerance of the integrator used to solve the Hamilton equations and compute the EOB waveforms. In Appendix B we compare the measurement errors obtained with pSEOBNRv5HM to

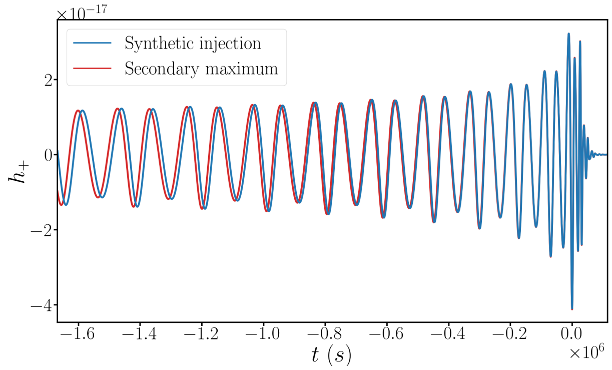


FIG. 6. Comparison between synthetic-injection waveform and the one at a secondary maximum (same colour code as in Fig. 5). The late inspiral and merger-ringdown signals match almost perfectly. The early inspiral signals are dephased, but the SNR in the inspiral is so low that it makes little difference.

the ones obtained with the IMR phenomenological model `IMRPhenomTHM` [90], finding comparable results.

B. Measurement of QNMs and possible deviations from GR

We now turn our attention to QNM measurements and constraints on deviations from GR using MBHB observations with LISA.

1. *GR injections*

First, we consider the case where the injected signal is compatible with GR, and allow for non-zero deviations when running the Bayesian analysis. In Figs. 7 and 8, we show the width of the 90% confidence interval centred around the median on the deviation parameters (i.e., $\Delta\delta f_{\ell m}$ and $\Delta\delta\tau_{\ell m}$). We find that deviations to the frequency are generally better constrained than those to the damping time. As a consequence of the higher SNR of the dominant $(2,2)$, $(3,3)$, $(4,4)$ and $(5,5)$ harmonics, fractional deviations to their QNMs are better constrained than those to the sub-dominant $(2,1)$, $(3,2)$ and $(4,3)$ harmonics. For the former, $\delta f_{\ell m}$ and $\delta\tau_{\ell m}$ are typically constrained within 0.1 and even within 0.01 for the systems with $M_{t,0} = 2\times10^7 M_\odot$ (blue points), and follow the $1/\mathrm{SNR}$ trend, with some scatter for higher harmonics (especially the $(3,3)$ and $(5,5)$ harmonics) that depends on the mass ratio. Here again, the reason for this is that for systems with $q_0 = 4$ (non-filled points), higher harmonics are more excited (see Figs. 2 and 3), so we are able to better constrain deviations in their QNMs, in agreement with [74]. Deviations in sub-dominant harmonics are poorly constrained for very massive systems (red points), $\Delta\delta f_{\ell m}$ and $\Delta\delta\tau_{\ell m} \sim 1$, which given our prior range, translates into uninformative measurements. This is due to the low SNR of these harmonics. However, we get rather good constrains, within 0.1, for systems with $M_{t,0} = 2\times10^7 M_\odot$ (blue points). We also find that for low-spin systems (circles) deviations to the $(2,1)$ harmonic are better constrained than the ones to the $(3,2)$ and $(4,3)$ harmonics, whereas the opposite happens for high-spin systems (squares), in agreement with our remark on their relative predominance in Sec. III. We show the impact of allowing for deviations from GR on the measurement of intrinsic parameters in Appendix C.

We translate these constraints on the fractional deviations into measurements of the QNMs in Fig. 9, where we show some representative examples of 90% confidence regions on the QNMs of all the harmonics included in `pSEOBNRv5HM`. The upper-left panel shows a best-case scenario, where all QNMs can be perfectly distinguished. It illustrates that the Rayleigh criterion often used to decide on the distinguishability of QNMs [58, 59, 74, 131, 132] is actually too stringent. Indeed, as can be seen from the upper-left panel of Fig. 9, the one-dimensional projection of the 90% confidence regions onto the y-axis (damping time) can overlap (e.g., for the $(4,4)$ and $(4,3)$ QNMs), although the two-dimensional regions are well separated. Thus, one should really consider the two-dimensional regions in order to decide on the distinguishability of QNMs, as pointed out in Ref. [71]. The upper-right and lower-left panel illustrate cases where not all QNMs can be resolved. As expected from our comments above, for a high-spin system such as the one shown in the upper-right panel, deviations to the $(2,1)$ QNM are poorly constrained, so its measurement uncertainty contour encloses that of the $(2,2)$ mode. Similarly, the lower-left panel shows a low-spin system, for which the $(4,3)$ QNM measurement uncertainty contour contains the $(4,4)$ one. Finally, the lower-left panel shows a worst-case scenario where the QNMs cannot be distinguished due to the large uncertainty on the sub-dominant harmonics. Note that the system in the lower-left panel has a total mass of $2\times10^7 M_\odot$, illustrating that the confidence regions of QNMs are not always all well-separated for systems with $M_{t,0} = 2\times10^7 M_\odot$, although they do tend to yield better results than for systems with $M_{t,0} = 2\times10^8 M_\odot$, as illustrated in Figs. 7 and 8.

We find some cases of multi-modality in the deviation parameters, as illustrated in Fig. 10. They can be understood by looking at the corresponding values of QNMs. Indeed, the frequency of the secondary mode in the (2,1) QNM matches the frequency of the (3,2) QNM. Because their damping times are poorly constrained, they are fairly compatible. Thus, this multi-modality can be understood as the sub-dominant harmonics trying to “match” each other.

2. *Non-GR injections*

We now consider non-GR injections, and we generate synthetic signals with non-zero deviations to the QNMs. Deviations to GR in the QNM frequencies have been derived in non-GR theories and typically vary in the range 0.01–0.1 or even smaller. In the spherically symmetric case (i.e., non-spinning BHs), they were computed in theories such as Einstein-Maxwell-dilaton [133], dynamical Chern-Simons gravity [42], Einstein-dilaton-Gauss-Bonnet gravity [40, 41, 134], higher-curvature gravity theories [53], and for some

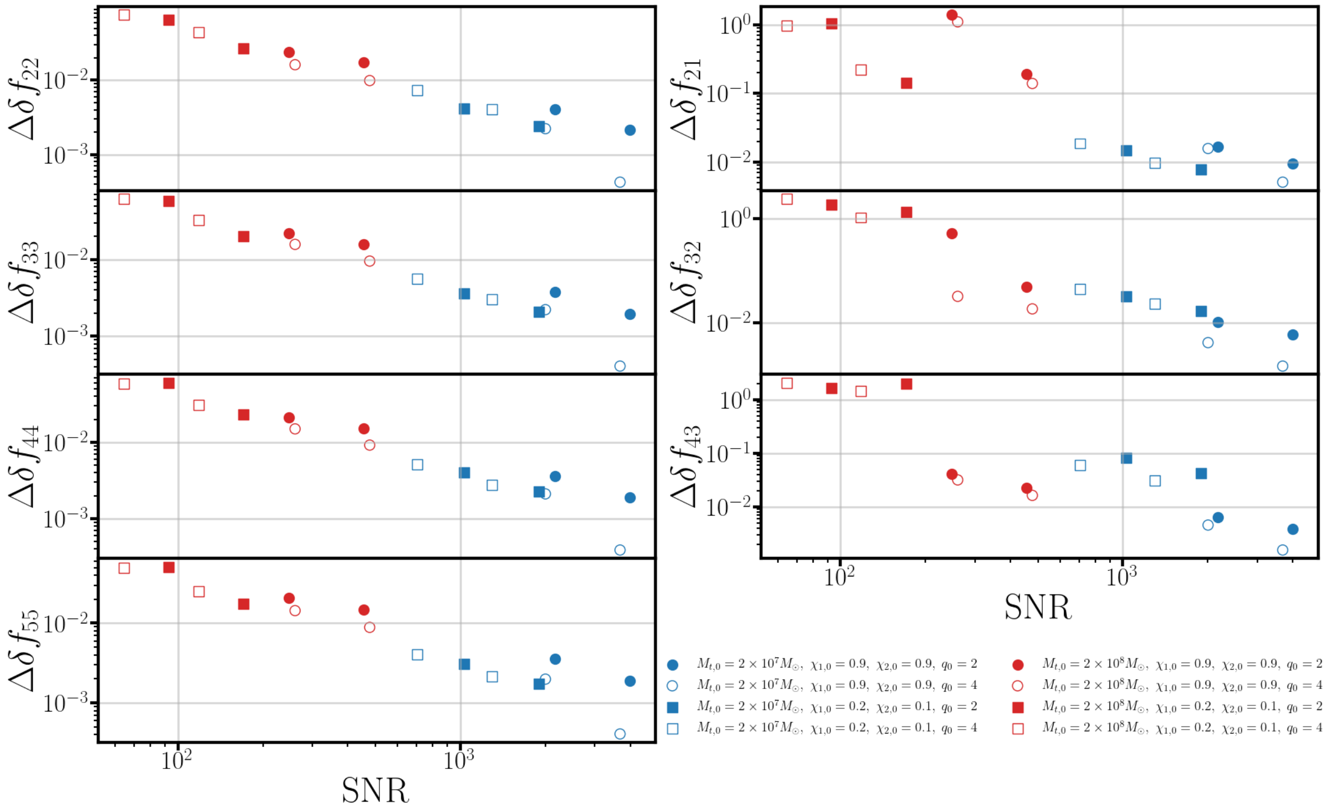


FIG. 7. Width of the 90% confidence interval on $\delta f_{\ell m}$ centred around the median for GR injections.

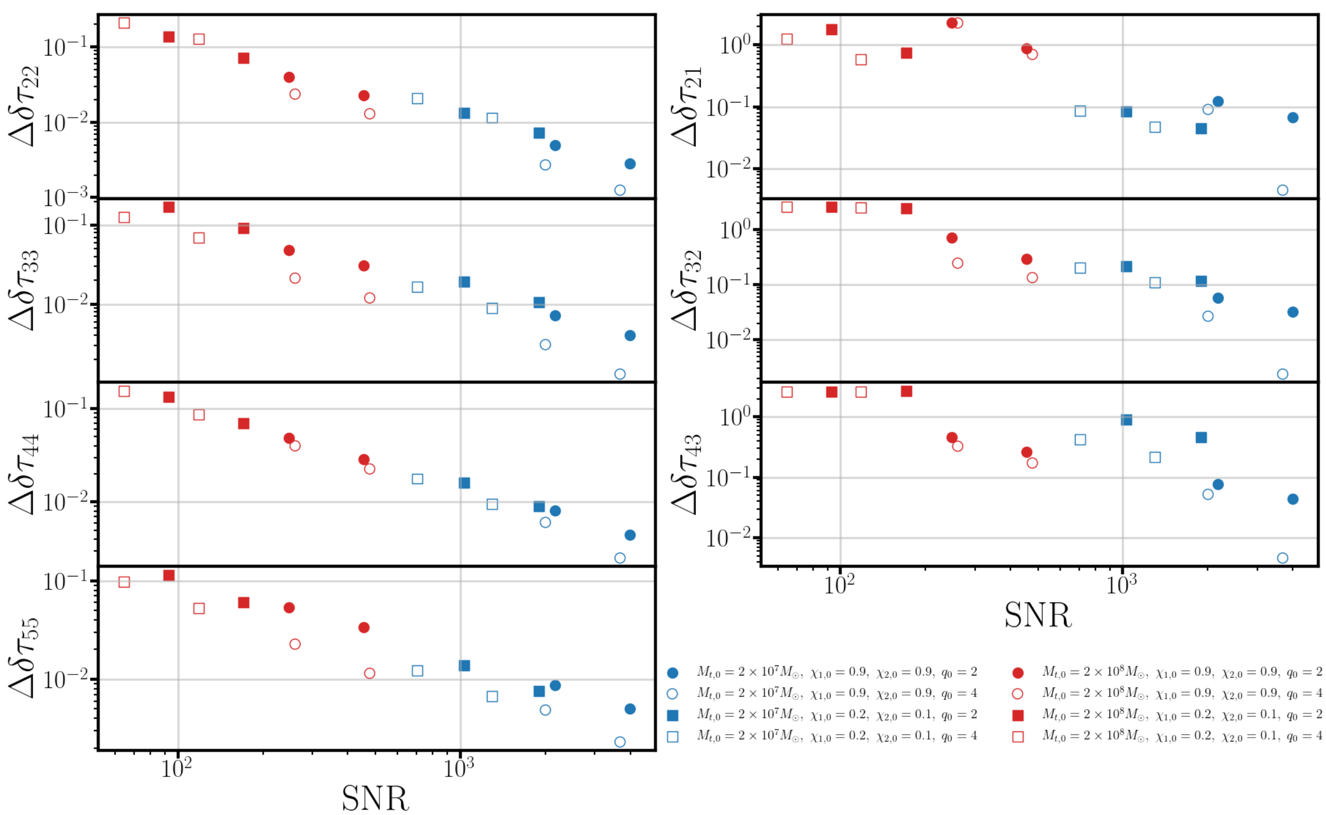
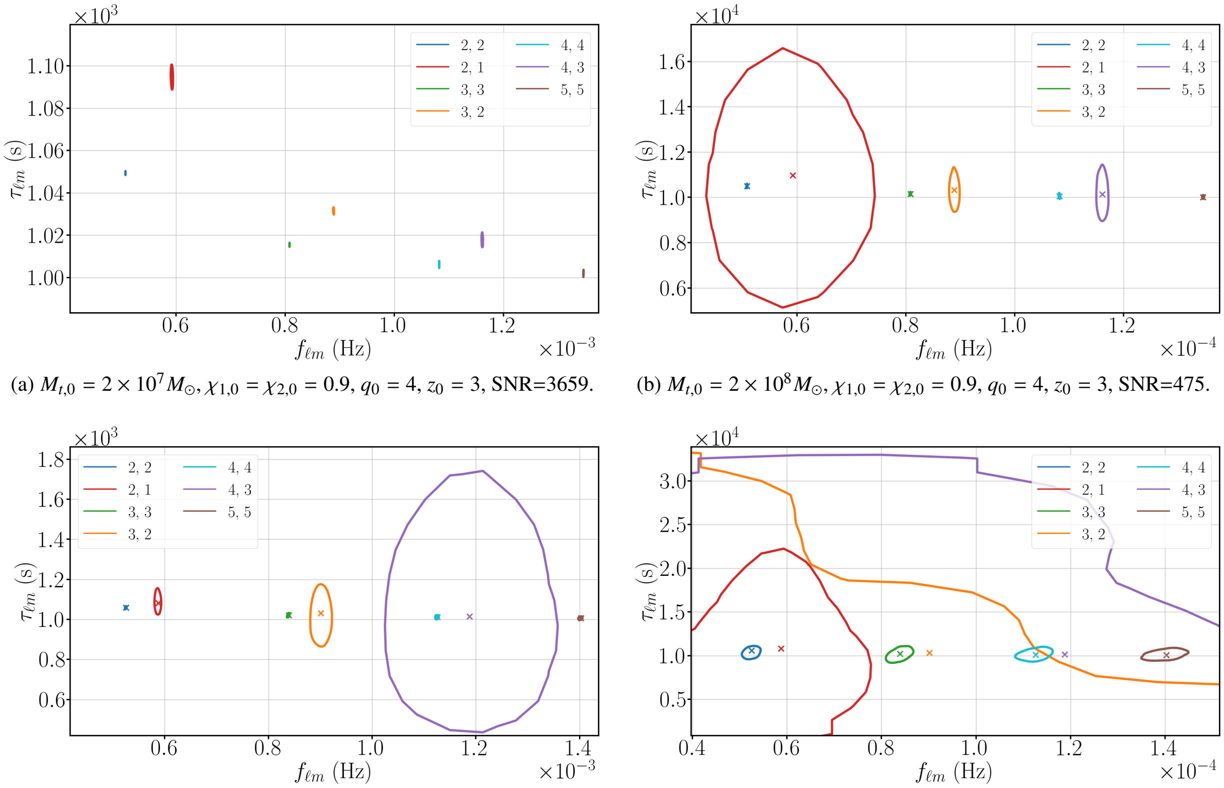


FIG. 8. Width of the 90% confidence interval on $\delta\tau_{\ell m}$ centred around the median for GR injections.



(a) $M_{t,0} = 2 \times 10^7 M_\odot, \chi_{1,0} = \chi_{2,0} = 0.9, q_0 = 4, z_0 = 3$, SNR=3659.

(b) $M_{t,0} = 2 \times 10^8 M_\odot, \chi_{1,0} = \chi_{2,0} = 0.9, q_0 = 4, z_0 = 3$, SNR=475.

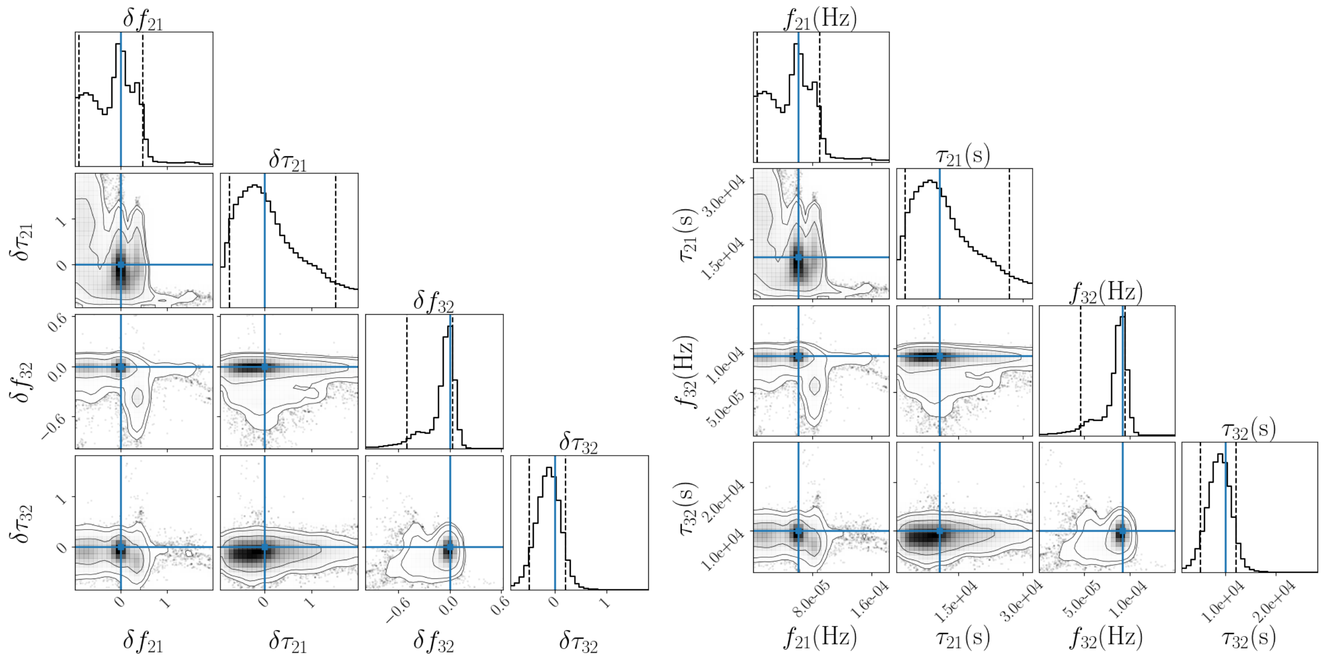
(c) $M_{t,0} = 2 \times 10^7 M_\odot, \chi_{1,0} = 0.2, \chi_{2,0} = 0.1, q_0 = 2, z_0 = 5$, SNR=1030.

(d) $M_{t,0} = 2 \times 10^8 M_\odot, \chi_{1,0} = 0.2, \chi_{2,0} = 0.1, q_0 = 2, z_0 = 5$, SNR=93.

FIG. 9. 90% confidence interval on the QNMs for four different systems. Crosses indicate the true values. Between columns only the total mass varies, and between rows mass ratio, spins, and redshift vary, keeping the same total mass. In every case, the synthetic injection is GR. The top-left panel illustrates why the usual Rayleigh criterion for resolvability is too stringent, although the one-dimensional damping time posteriors overlap, the two-dimensional ones are well-separated. That case corresponds to a "best-case scenario". The other panels show cases where not all 7 QNMs can be distinguished. The (2,1), (3,2) and (4,3) modes are typically less-well constrained due to their lower SNR (see also Figs. 7 and 8).

solutions in massive (bi)gravity [135–137]. Recently, the computation of QNMs for spinning BHs in non-GR theories has received much attention, since the remnant BHs we are observing with LIGO and Virgo have typically spins of about 0.7. They include the Kerr-Newman case in Einstein-Maxwell theory [44–47], Einstein scalar Gauss-Bonnet gravity [50, 51], higher-curvature gravity theories [52, 53] and dynamical Chern-Simnons theory [54]. Estimates for QNMs of spinning BHs in non-GR theories have also used the connection between the light ring and QNMs [41, 138–140], which is formally only valid in the eikonal $\ell \to \infty$ limit, and are known to fail to describe some families of QNMs when additional degrees of freedom are present [41].

Here, we assume the fractional deviation to GR to be 0.01 for all harmonics, both for frequencies and damping times. In Figs. 11 and 12, we show the width of the 90% confidence interval centred around the median for $\delta f_{\ell m}$ and $\delta \tau_{\ell m}$. As a rule of thumb, we consider that a deviation can be measured when it is larger than the measurement error. Graphically, this corresponds to the points that are below the black-dashed lines on Figs. 11 and 12. For this value of the deviation (i.e., 0.01), we find that it could be detected in the frequency of the (2,2), (3,3), (4,4) and (5,5) harmonics of systems with $M_{t,0} = 2 \times 10^7 M_\odot$, and for the higher SNR ones we could also detect this deviation in their damping time. We note that the errors shown in Figs. 11 and 12 are very similar to the ones we find when injecting GR signals (see Figs. 7 and 8). We also perform injections with deviations of 0.1 and 0.001 (not shown here), and find again similar errors. Therefore, we can extrapolate the results presented here and read from those figures which values of the deviations would be needed to detect them. For instance, a deviation of 0.1 could be detected for almost all systems presented here, both in the frequency and the damping time of the dominant harmonics (left column), and for the higher SNR systems, even of the sub-dominant harmonics (right column). Detecting a deviation in several harmonics, preferably both in the frequency and the damping time, would reinforce our confidence that we are truly observing effects in gravity theories alternative to GR.



(a) Deviation parameters

(b) QNMs

FIG. 10. Reduced corner plot of deviation parameters and QNMs for a GR injection of the system $M_{t,0} = 2 \times 10^8$, $\chi_{1,0} = \chi_{2,0} = 0.9$, $q_0 = 2$, $z_0 = 5$, SNR=240. We find a multi-modality in the δf_{21} posterior, which is the less well measured harmonic for this system. This leads to a secondary mode in the f_{21} posterior at frequencies that match f_{32}. Given that the damping times of the two harmonics are poorly measured and their posterior fairly compatible, this suggests that the (2,1) harmonic is trying to "match" the (3,2) harmonic and vice versa.

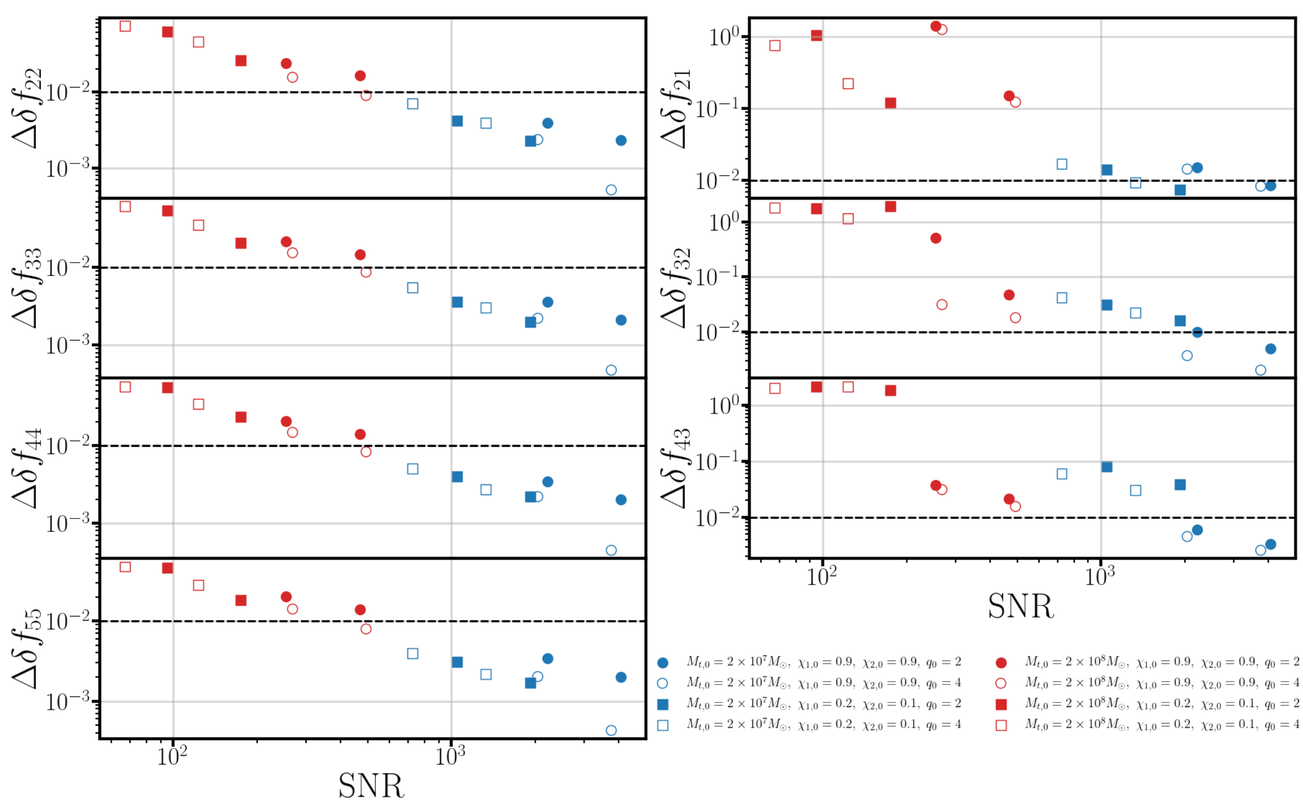


FIG. 11. Width of the 90% confidence interval centred around the median for $\delta f_{\ell m}$ for injections with $\delta f_{\ell m} = \delta\tau_{\ell m} = 0.01$, as indicated by the black-dashed lines.

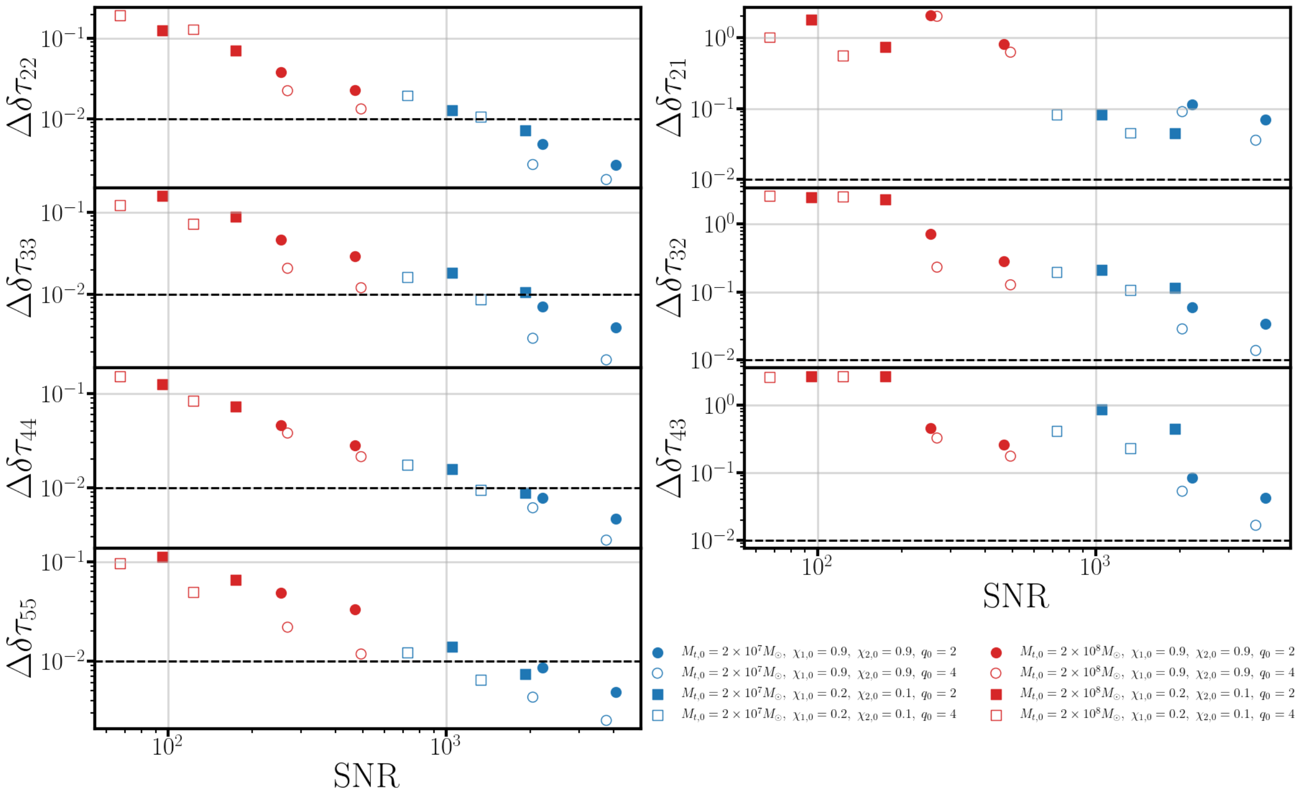


FIG. 12. Width of the 90% confidence interval centred around the median for $\delta\tau_{\ell m}$ for injections with $\delta f_{\ell m} = \delta\tau_{\ell m} = 0.01$, as indicated by the black-dashed lines.

V. IMPACT OF SYSTEMATICS

In order to assess if the lack of accuracy in waveform modelling could spoil ringdown tests of GR, we generate synthetic injections with NR waveforms and recover them with `pSEOBNRv5HM` templates. We use the waveform SXS:BBH:2125 from the Simulating eXtreme Spacetimes Collaboration [141] at the highest available resolution. It provides the signal of a BBH with mass ratio 2 and aligned spins of magnitude 0.3. We perform injections for the four total mass and redshift combinations detailed in Sec. III. When running our Bayesian analysis, we allow for deviations from GR in the QNMs.

First, we consider the case where the injected signal contains only the dominant (2,2) harmonic, and include in the `pSEOBNRv5HM` templates the same harmonic content. We compare in Fig. 13 the posteriors obtained for injections with different total masses, both at $z_0 = 5$, giving SNR = 67 and 927, respectively. The injected values (blue lines) are well within the 90% confidence regions for the heavier system (black), having total mass $M_{t,0} = 2 \times 10^8 M_\odot$. However, due to its much higher SNR, the parameter estimation of the lighter system (red), having total mass $M_{t,0} = 2 \times 10^7 M_\odot$, is strongly biased. In particular, a deviation from GR in the frequency of the (2,2) QNM is erroneously detected with high confidence. Next, we inject NR signals containing all the harmonics included in `pSEOBNRv5HM` and use templates with full harmonic content in the Bayesian analysis. As can be seen in Fig. 14, once we include higher harmonics (red), even for the more massive system the posterior shifts further from the true values while getting narrower, making it barely compatible with the true parameters. In particular, the true values of the GR deviation parameters lie at the edge of the 90% confidence regions. We stress that the worsening for the system shown in Fig. 14 is less due to the moderate increase in SNR than to the inclusion of higher harmonics, as we have verified by comparing to the results of the same system at $z_0 = 3$ (not shown here). The latter has an SNR of 123 with only the (2,2) harmonic, and for that system the parameter estimation is less biased when including only the (2,2) harmonic as for the system in Fig. 14 when including all harmonics. We have also verified (not shown here) that, when including higher harmonics for systems with $M_{t,0} = 2 \times 10^7 M_\odot$, GR is excluded at higher confidence than in the (2,2) only case shown in red in Fig. 13.

This is not surprising since the accuracy of the `SEOBNRv5` model degrades when including higher harmonics, as comprehensive comparisons to ~ 440 NR waveforms and NR surrogate models have shown [81]. We note that in Ref. [62], the authors performed a similar study for LIGO detectors at design sensitivity, and found the GR value of QNMs to be well within the 90% confidence interval even when injecting an NR waveform with an SNR of 75. The main difference to our analysis is that whereas the SNR in Ref. [62] was equally spread between inspiral and merger-ringdown stages, here all the binary systems we consider are merger-ringdown domi-

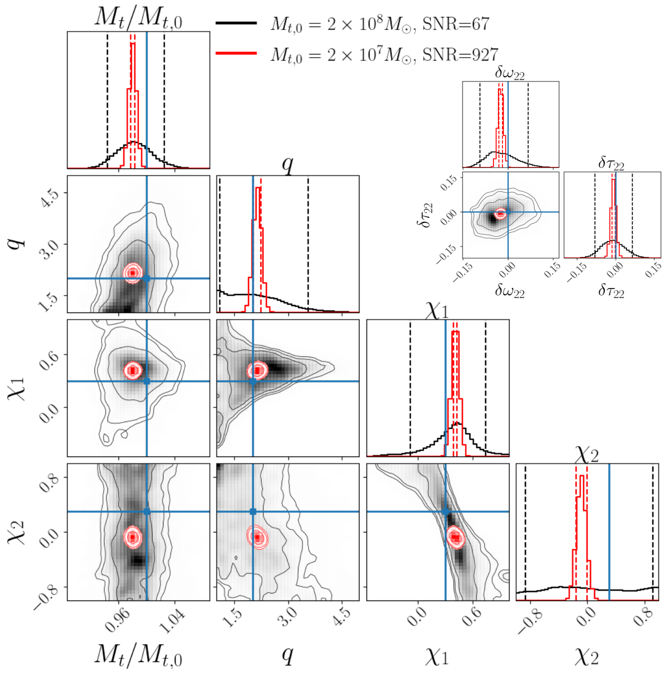


FIG. 13. Corner plot of intrinsic and deviation parameters for NR injections recovered with pSEOBNRv5HM templates. Contours show the 68%, 90%, and 95% confidence intervals, and dashed lines the 0.05 and 0.95 quantiles. We compare the results for $M_{t,0} = 2 \times 10^8 M_\odot$ (black) and $M_{t,0} = 2 \times 10^7 M_\odot$ (red), both have $q_0 = 2$, $\chi_{1,0} = \chi_{2,0} = 0.3$ and are placed at $z_0 = 5$. To ease comparison, we have rescaled the total mass to the injected value. In each case, both injection and templates contain only the (2,2) harmonic. The heavier system has lower SNR, and the impact of mismodelling is small when including only the (2,2) harmonic, so that the injected values (in blue) are well within the 90% confidence regions. It is nonetheless relevant for the lighter system, due to its much higher SNR. In the latter case, one would be misled into thinking that the frequency of the (2,2) QNM departs from the Kerr prediction. The spins are poorly measured for the heavy system due to the absence of higher harmonics.

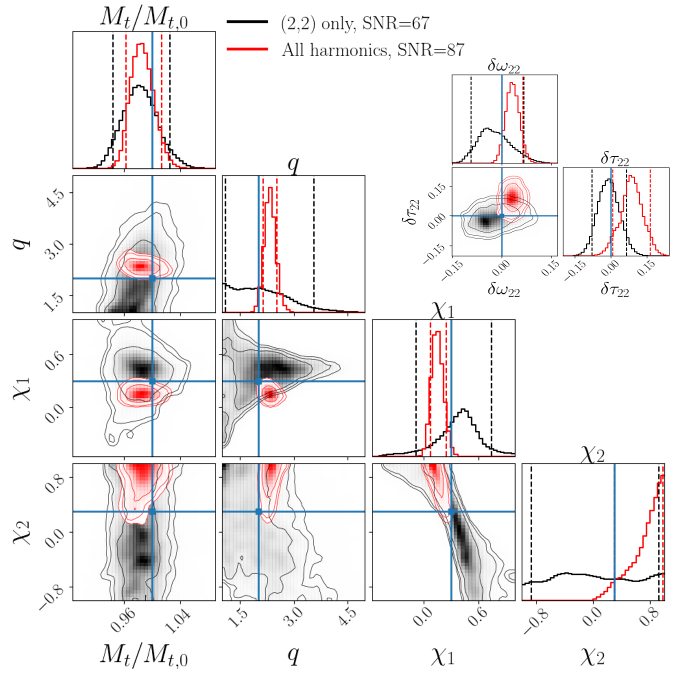


FIG. 14. Same as Fig. 13, but now comparing results when including only the (2,2) harmonic (black) versus when including all harmonics of pSEOBNRv5HM (red), for $M_{t,0} = 2 \times 10^8 M_\odot$ and $z_0 = 5$. The inclusion of higher harmonics worsens the match between NR and pSEOBNRv5HM waveforms, leading to significant biases in all parameters, in particular the GR deviation ones. Note that when including all harmonics, the deviation parameters for all QNMs are allowed to vary, and we find similar biases for all of them, but we show only the posterior for deviations in the (2,2) harmonic because of space limitation.

nated. This is the regime where our templates are less reliable, leading to larger biases.

Furthermore, in the high-SNR limit, the presence of biases in parameter estimation can also be predicted from a simpler indistinguishability criterion [142–146], based on the mismatch between two waveforms h_1 and h_2

$$\mathcal{M} = 1 - \frac{(h_1 \mid h_2)}{\sqrt{(h_1 \mid h_1)\,(h_2 \mid h_2)}}. \tag{5.1}$$

Specifically, if two waveforms fulfill the condition

$$\mathcal{M} < \frac{D}{2\,\mathrm{SNR}^2}, \tag{5.2}$$

for a given PSD and SNR, they are considered indistinguishable, and systematic errors from waveform inaccuracies are expected to be smaller than statistical errors. The prefactor D is not known precisely, but it can be estimated as the number of intrinsic parameters whose measurability is affected by model inaccuracy [145], or can be tuned by considering synthetic injections at increasing SNR [146]. Being sufficient, but not necessary, the criterion is generally too conservative, and, if it is violated, biases do not necessarily arise (e.g., see [81, 147]. Nonetheless, we can check whether the biases we observe are consistent with such a criterion. Taking $D = 6$, the criterion predicts mismatches of 6.7×10^{-4} ($M_{t,0} = 2 \times 10^8 M_\odot$ system including only the (2,2) harmonic), 3.5×10^{-6} ($M_{t,0} = 2 \times 10^7 M_\odot$ system including only the (2,2) harmonic) and 4.0×10^{-5} ($M_{t,0} = 2 \times 10^8 M_\odot$ system including all higher harmonics), below which systematic errors are expected to be subdominant.

For the $M_{t,0} = 2 \times 10^8 M_\odot$ system including only the (2,2) harmonic the mismatch [2] of the SEOBNRv5HM model against the NR simulation SXS:BBH:2125 is 5.9×10^{-5}, and one would not expect biases, as we also observe in the synthetic injection we perform. On the other hand, for the $M_{t,0} = 2 \times 10^7 M_\odot$ system including only the (2,2) harmonic and for the $M_{t,0} = 2 \times 10^8 M_\odot$ system including all higher harmonics, the values of the mismatch are 7.3×10^{-5} and 7.5×10^{-3} respectively, and are both above the indistinguishability threshold,

[2] We consider specifically the sky-and-polarisation-averaged, SNR-weighted mismatch defined as in Sec. V of Ref. [81].

therefore the presence of the biases we observe could have been predicted.

These studies show that because of the high SNRs that we will reach with LISA, the accuracy requirement for waveforms is much more stringent than for current ground-based detectors. Moreover, our results suggest that particular attention should be drawn to the modelling of higher harmonics, as their inclusion increases biases. Let us stress that even the NR waveforms we currently use are not accurate enough for SNRs of thousands [81, 146] and would need to be improved by at least one order of magnitude.

VI. CONCLUSION

Gravitational-wave observations have provided us with brand-new opportunities to test GR. In particular, ringdown tests are one of the most promising possibilities to detect deviations from GR. In this work, we have assessed how a fully consistent modelling of the IMR signal will allow us to perform high-precision ringdown tests with MBHB observations by LISA. To do so, we have performed synthetic injections of astrophysically realistic systems, and analysed them with templates in GR (`SEOBNRv5HM`) and with parameterised deviations from GR (`pSEOBNRv5HM`), using the newly released `SEOBNRv5HM` waveform family [79–82, 148]. More specifically, the `pSEOBNRv5HM` templates allow for deviations to the QNMs (frequency and damping time) of all the harmonics included in the model. All our analyses have been done in a fully Bayesian framework.

First, we have considered the case where we use the GR `SEOBNRv5HM` templates both for the synthetic injection and the Bayesian analysis. We find that having a consistent modelling of the whole signal allows us to measure the parameters of the binary accurately, even for signals with very little SNR in the inspiral (e.g., very massive MBHBs with total mass $\sim 10^8 M_\odot$). Source-frame masses can typically be measured within 10% and even within 1% for systems with total mass $\sim 10^7 M_\odot$. Spins can be measured within 0.1 and down to 0.001 for the primary BH. Second, we have shown that deviations to the QNMs of the dominant harmonics (i.e., the (2,2), (3,3), (4,4) and (5,5) harmonics), can be constrained within 10% and down to 1% for systems with total mass of the order $10^7 M_\odot$. Those are also the magnitude of deviations that we could measure in a non-GR signal. Converting the measurement of fractional deviations into measurements of QNM frequencies, we find that for most systems we could accurately measure and distinguish the QNMs of several harmonics, up to all 7 seven included in the `pSEOBNRv5HM` model in the most favourable cases (i.e., total mass $10^7 M_\odot$, mass ratio $\sim 2-4$, highly spinning and at $z \sim 3$).

Then, we have assessed the impact of systematics on ringdown tests by using NR waveforms to perform synthetic injections and analysing them with `pSEOBNRv5HM`. We have found that, in particular when higher harmonics are included, parameter estimation is significantly biased, leading to the erroneous detection of deviations from GR in high SNR signals. The results we have obtained when using `pSEOBNRv5HM` for the injection and the Bayesian analysis give us a sense of the incredibly high precision to which we will be able to perform ringdown tests with LISA. However, in order not to jeopardise those tests, the accuracy of our waveform models needs to be improved far beyond current standards, which is one of the major challenges facing the GW community over the next few years.

In this work, we have focused on MBHB systems that are in the high-mass end of predictions for LISA observations, typically produced in astrophysical models where MBHs form from the evolution of heavy seeds [89]. This is because those are the ones for which we expect the highest SNR in the merger-ringdown [74, 75]. However, it would be interesting to assess how observations of lighter systems, which might be more numerous, could be used to detect deviations from GR in the ringdown. Also, we have neglected the effect of spin-precession and eccentricity, which might not be appropriate, in particular if MBHBs harden through triplet interactions [85]. Finally, the waveform model we used in this work allows for deviations from GR only in the ringdown, whereas, if deviations are present, we should expect them to affect the whole signal. Different theory-agnostic formalisms have been proposed to account for deviations in the inspiral [149–151], typically by modifying the post-Newtonian expansion of the GW phase [152], and progress has recently been made to account for deviations in the plunge-merger stage [67]. It would be interesting to assess how to link modifications in different parts of the signal, or at least to assess how the constraints change when accounting for all possible modifications. We leave these studies for future work.

ACKNOWLEDGMENTS

It is our pleasure to thank Serguei Ossokine for providing assistance in the usage of publicly available LALSimulation [153] codes and getting access to NR waveforms.

The computational work for this manuscript was carried out on the `Hypatia` compute cluster at the Max Planck Institute for Gravitational Physics in Potsdam.

The waveform model `SEOBNRv5HM` is publicly available through the Python Package `pySEOBNR` `git.ligo.org/waveforms/software/pyseobnr`. Stable versions of `pySEOBNR` are published through the Python Package Index (PyPI), and can be installed via `pip install pyseobnr`.

Appendix A: Settings for the SEOBNR waveforms

`SEOBNR` waveforms are generated by numerically integrating the EOB equations of motion (e.g., see Ref. [81]). They are computed up to a given accuracy that depends on the tolerance used in the integration. As a consequence, at fixed tolerance, the waveform function is not smooth on the manifold of waveform parameters, and the inner product between waveforms (Eq. (2.12)) is an oscillating function on that manifold. For low SNR systems, these oscillations are negligible, since they correspond to very small changes in the likelihood between neighbouring points, but for high SNRs these oscillations become important. This is illustrated in Fig. 15, where we plot the log-likelihood as a function of χ_2 for "standard-tolerance" and "low-tolerance" waveforms, keeping all the other parameters at their true value. When using the standard tolerance, the likelihood shows many local extrema, and can reach very small values even close to the synthetic injection (indicated by the blue line). For comparison, the minimum log-likelihood in the parameter estimation runs done for this paper are typically ~ -15. The smoothness of the log-likelihood improves significantly when using low-tolerance waveforms. The counterpart of this improvement is a slowing down of the waveform computation. Our low-tolerance waveforms are ~ 5 times slower to compute than the ones with the standard `SEOBNRv5HM` configuration. We stress that the effect of these oscillations is exaggerated by looking at one slice of the parameter space (i.e., keeping all the other parameters fixed). Variations in other parameters compensate for these oscillations, and make these local extrema less pronounced. However, as we show in Fig. 16, the existence of several local extrema makes the posterior non-Gaussian and the marginal one-dimensional distributions can peak away from the true value. This is similar to the effect discussed in Sec. V.C of Ref. [110] in the context of sky localisation with LISA. For the system shown in Fig. 16, this apparent bias disappears when using low-tolerance waveforms. Let us stress that this apparent bias is an effect of projecting a non-Gaussian posterior onto one-dimensional posteriors. As indicated by the black and red lines, in each case, the maximum-posterior points found by our sampler are close to the injection point, as expected when using flat priors. All the results presented in the main body of this paper were obtained using low-tolerance waveforms.

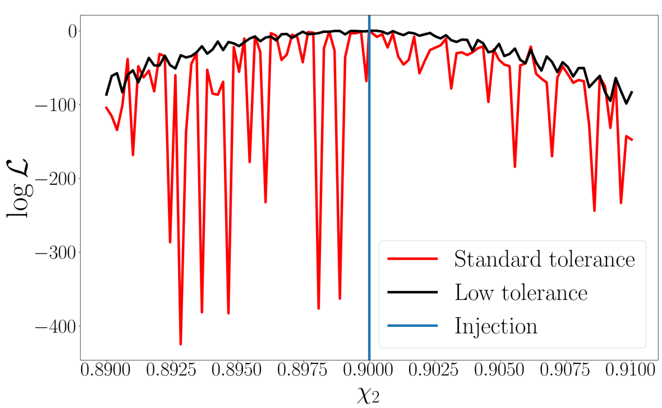


FIG. 15. Log-likelihood as a function of χ_2 when using two different tolerances, when integrating the EOB equations of motion, for a system with $M_{t,0} = 2 \times 10^8 M_\odot$, $\chi_{1,0} = \chi_{2,0} = 0.9$, $q_0 = 2$, $z_0 = 3$, SNR=441. When using waveforms computed with standard tolerance, the likelihood can be a very non-smooth function of the binary parameters, inducing secondary maxima, and yielding very small values even close to the injected value.

For some systems, this apparent bias persists even when using low-tolerance waveforms, as shown in Fig. 17. In order to further validate the argument that this is caused by the non-smoothness of the waveform across parameter space, we compare to the results obtained using the `IMRPhenomTHM` waveform model [90]. This is an IMR time-domain approximant, built from a phenomenological approach, in the spirit of the frequency-domain approximants of the phenomenological family of templates [91–93, 154–156]. It is based on post-Newtonian expressions [152] augmented by phenomenological terms fitted against NR simulations, and also calibrated to `SEOBNR` waveforms (where NR data are not available). It includes the same harmonics as `pSEOBNRv5HM`, except for the sub-dominant (3,2) and (4,3) harmonics. For non-precessing systems, the waveform is an analytic smooth function of the waveform parameters. Thus, the likelihood function is smooth across the parameter space, and we observe no apparent bias, even when looking at one-dimensional projections, as can be seen from Fig. 17. We stress that after the waveform has been generated in the time domain, the steps to compute the likelihood and perform parameter estimation are exactly the same for the two models.

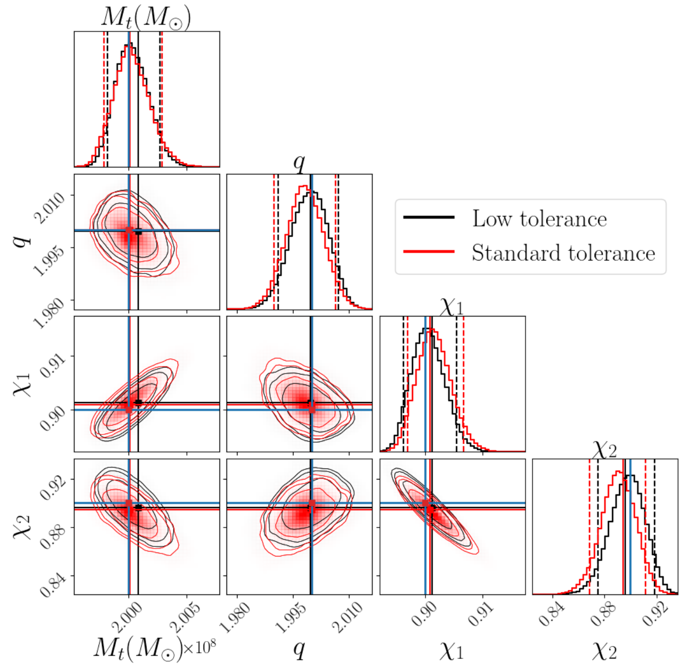


FIG. 16. Comparison between parameter estimation results when using low-tolerance waveforms (black) and standard ones (red) for the same system as in Fig. 15. Blue lines indicate the injection point, black and red symbols indicate the maximum-posterior point of the low-tolerance and standard-tolerance runs, respectively. The non-smoothness of the likelihood function illustrated in Fig. 15 induces secondary maxima over the parameter space. Those lead to very non-Gaussian distributions and induce an apparent bias when projecting onto one-dimensional posteriors. However, as indicated by the black and red lines, the maximum-posterior point is close to the synthetic injection, as expected when using flat priors.

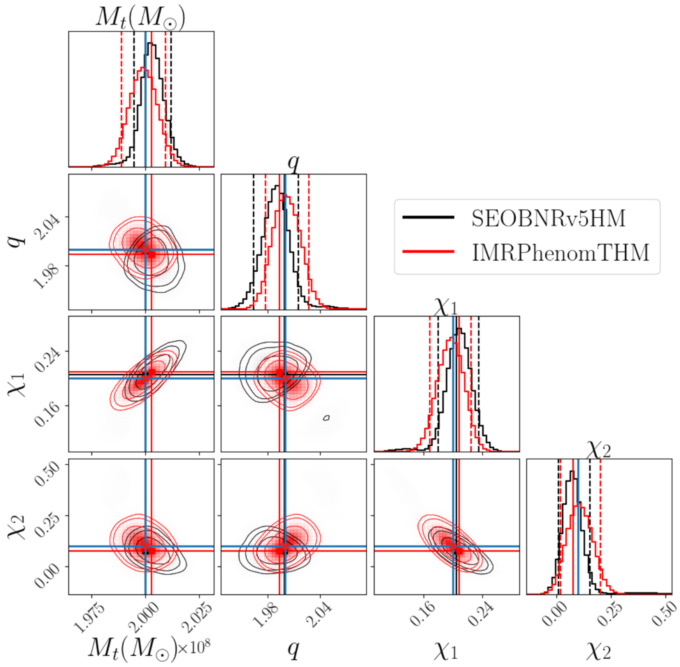


FIG. 17. Comparison of the parameter estimation results obtained using low-tolerance `SEOBNRv5HM` waveforms (black) and `IMRPhenomTHM` (red) waveforms for a system with $M_{t,0} = 2{\times}10^8 M_\odot$, $\chi_{1,0} = 0.2$, $\chi_{2,0} = 0.1$, $q_0 = 2$, $z_0 = 3$, SNR=171. The latter shows no apparent bias when looking at one-dimensional marginalised posteriors, because the `IMRPhenomTHM` waveform is a smooth function across the parameter space, and so is the likelihood, and for high SNRs the posterior is fairly Gaussian.

Appendix B: Comparison of measurements to IMRPhenomTHM

Here, we run Bayesian analyses for all the binary systems described in Sec. III using the `IMRPhenomTHM` waveform, both for the synthetic injection and parameter estimation. We restrict ourselves to the GR case. In Fig. 18, we show how the measurement errors on intrinsic parameters compare when using the `IMRPhenomTHM` and the `SEOBNRv5HM` waveforms. Black lines corresponds to $y = x$. We see that the estimates are in good agreement, with a slight discrepancy for the error on spins, in particular in the case of high-spin systems (circles). In this regime, our current waveforms are less accurate, so the agreement between them is worse (see also comparisons between these two waveform models in Ref. [81]).

Appendix C: Errors on intrinsic parameters when allowing for deviations to GR

We show in Fig. 19 the error on intrinsic parameters in the case we inject a GR signal and allow for deviations from GR when performing the Bayesian analysis, complementing the results of Sec. IV B 1. Black lines indicate $y = x$. All measurements worsen due to the higher number of parameters, but remain comparable to the pure GR case.

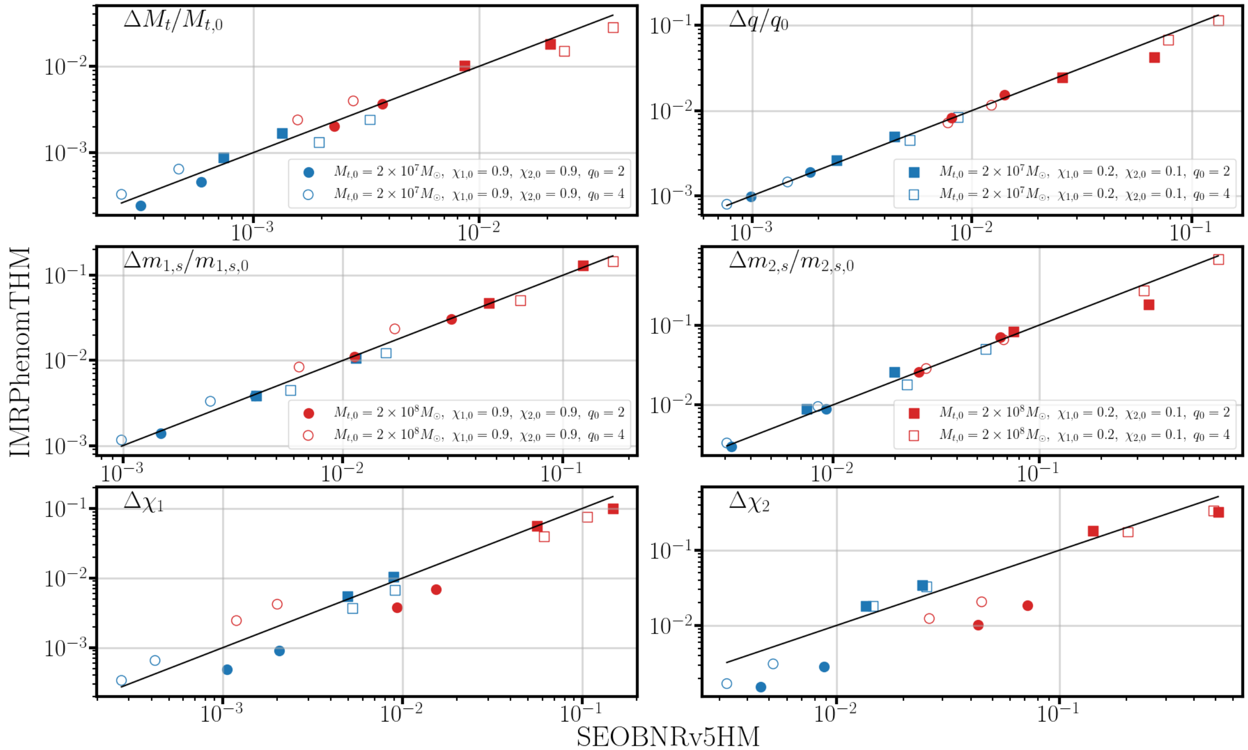


FIG. 18. Comparison between the measurement errors with `SEOBNRv5HM` (x-axes) and with `IMRPhenomTHM` (y-axes). Black lines show $y = x$. Errors are similar, with a slightly higher discrepancy for the spins, in particular for high-spin systems (circles).

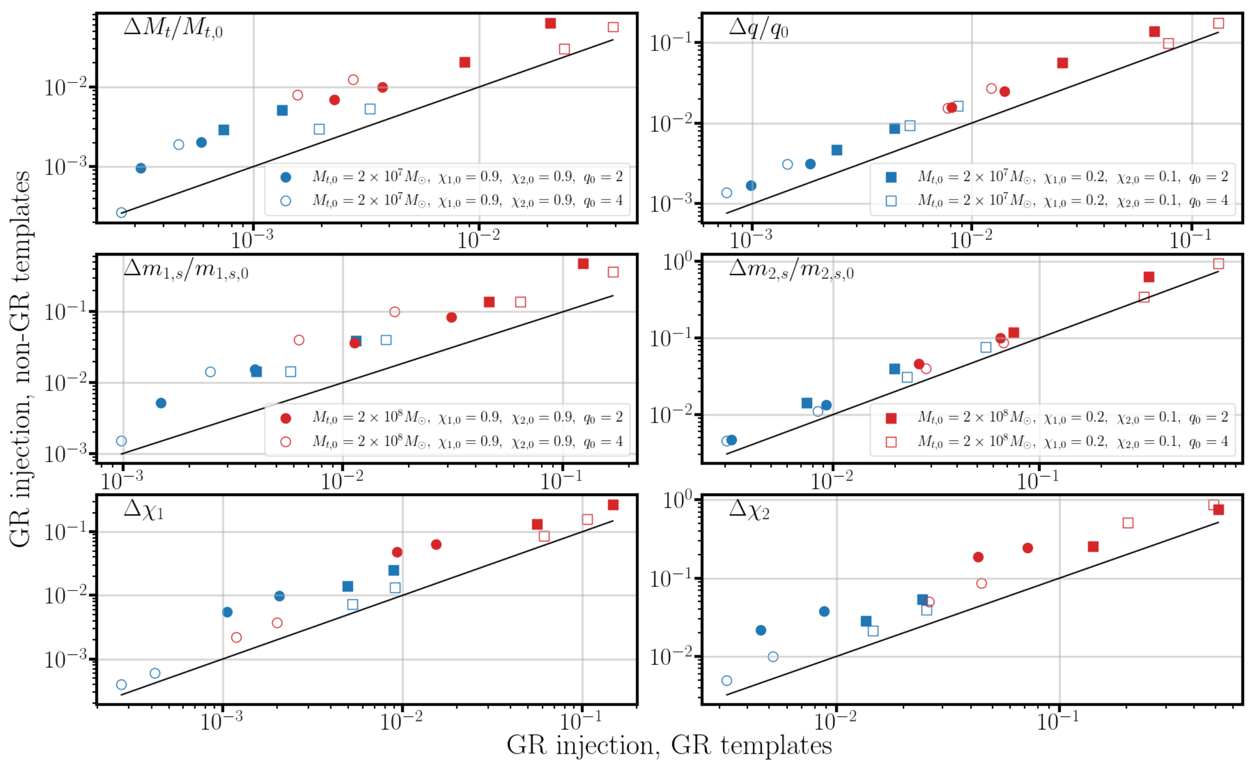


FIG. 19. Comparison of the measurement error on intrinsic parameters when setting deviation parameters to 0 (x-axes) and when letting them vary (y-axes) for a GR injection. Black lines show $y = x$. Measurements worsen when allowing for deviations from GR to vary, but remain comparable to the GR case.

[1] R. Abbott *et al.* (LIGO Scientific, VIRGO), (2021), arXiv:2108.01045 [gr-qc].
[2] R. Abbott *et al.* (LIGO Scientific, VIRGO, KAGRA), (2021), arXiv:2111.03606 [gr-qc].
[3] J. Aasi *et al.* (LIGO Scientific), Class. Quant. Grav. **32**, 074001 (2015), arXiv:1411.4547 [gr-qc].
[4] F. Acernese *et al.* (VIRGO), Class. Quant. Grav. **32**, 024001 (2015), arXiv:1408.3978 [gr-qc].
[5] T. Akutsu *et al.* (KAGRA), PTEP **2021**, 05A101 (2021), arXiv:2005.05574 [physics.ins-det].
[6] A. H. Nitz, S. Kumar, Y.-F. Wang, S. Kastha, S. Wu, M. Schäfer, R. Dhurkunde, and C. D. Capano, Astrophys. J. **946**, 59 (2023), arXiv:2112.06878 [astro-ph.HE].
[7] S. Olsen, T. Venumadhav, J. Mushkin, J. Roulet, B. Zackay, and M. Zaldarriaga, Phys. Rev. D **106**, 043009 (2022), arXiv:2201.02252 [astro-ph.HE].
[8] B. P. Abbott *et al.* (KAGRA, LIGO Scientific, Virgo, VIRGO), Living Rev. Rel. **21**, 3 (2018), arXiv:1304.0670 [gr-qc].
[9] B. P. Abbott *et al.* (LIGO Scientific, Virgo), Astrophys. J. Lett. **818**, L22 (2016), arXiv:1602.03846 [astro-ph.HE].
[10] B. P. Abbott *et al.* (LIGO Scientific, Virgo), Astrophys. J. Lett. **882**, L24 (2019), arXiv:1811.12940 [astro-ph.HE].
[11] R. Abbott *et al.* (LIGO Scientific, Virgo), Astrophys. J. Lett. **913**, L7 (2021), arXiv:2010.14533 [astro-ph.HE].
[12] R. Abbott *et al.* (KAGRA, VIRGO, LIGO Scientific), Phys. Rev. X **13**, 011048 (2023), arXiv:2111.03634 [astro-ph.HE].
[13] R. Abbott *et al.* (LIGO Scientific, Virgo), Astrophys. J. Lett. **900**, L13 (2020), arXiv:2009.01190 [astro-ph.HE].
[14] B. P. Abbott *et al.* (LIGO Scientific, Virgo), Phys. Rev. Lett. **119**, 161101 (2017), arXiv:1710.05832 [gr-qc].
[15] B. P. Abbott *et al.* (LIGO Scientific, Virgo), Phys. Rev. Lett. **121**, 161101 (2018), arXiv:1805.11581 [gr-qc].
[16] B. P. Abbott *et al.* (LIGO Scientific, Virgo, 1M2H, Dark Energy Camera GW-E, DES, DLT40, Las Cumbres Observatory, VINROUGE, MASTER), Nature **551**, 85 (2017), arXiv:1710.05835 [astro-ph.CO].
[17] B. P. Abbott *et al.* (LIGO Scientific, Virgo, Fermi GBM, INTEGRAL, IceCube, AstroSat Cadmium Zinc Telluride Imager Team, IPN, Insight-Hxmt, ANTARES, Swift, AGILE Team, 1M2H Team, Dark Energy Camera GW-EM, DES, DLT40, GRAWITA, Fermi-LAT, ATCA, ASKAP, Las Cumbres Observatory Group, OzGrav, DWF (Deeper Wider Faster Program), AST3, CAASTRO, VINROUGE, MASTER, J-GEM, GROWTH, JAGWAR, CaltechNRAO, TTU-NRAO, NuSTAR, Pan-STARRS, MAXI Team, TZAC Consortium, KU, Nordic Optical Telescope, ePESSTO, GROND, Texas Tech University, SALT Group, TOROS, BOOTES, MWA, CALET, IKI-GW Follow-up, H.E.S.S., LOFAR, LWA, HAWC, Pierre Auger, ALMA, Euro VLBI Team, Pi of Sky, Chandra Team at McGill University, DFN, ATLAS Telescopes, High Time Resolution Universe Survey, RIMAS, RATIR, SKA South Africa/MeerKAT), Astrophys. J. Lett. **848**, L12 (2017), arXiv:1710.05833 [astro-ph.HE].
[18] B. P. Abbott *et al.* (LIGO Scientific, Virgo), Phys. Rev. Lett. **116**, 221101 (2016), [Erratum: Phys.Rev.Lett. 121, 129902 (2018)], arXiv:1602.03841 [gr-qc].
[19] B. P. Abbott *et al.* (LIGO Scientific, Virgo), Phys. Rev. D **100**, 104036 (2019), arXiv:1903.04467 [gr-qc].
[20] R. Abbott *et al.* (LIGO Scientific, Virgo), Phys. Rev. D **103**, 122002 (2021), arXiv:2010.14529 [gr-qc].
[21] R. Abbott *et al.* (LIGO Scientific, VIRGO, KAGRA), (2021), arXiv:2112.06861 [gr-qc].
[22] C. V. Vishveshwara, Nature **227**, 936 (1970).
[23] W. H. Press and S. A. Teukolsky, Astrophys. J. **185**, 649 (1973).
[24] S. Chandrasekhar and S. L. Detweiler, Proc. Roy. Soc. Lond. A **344**, 441 (1975).
[25] K. D. Kokkotas and B. G. Schmidt, Living Rev. Rel. **2**, 2 (1999), arXiv:gr-qc/9909058.
[26] V. Ferrari and L. Gualtieri, Gen. Rel. Grav. **40**, 945 (2008), arXiv:0709.0657 [gr-qc].
[27] E. Berti, V. Cardoso, and A. O. Starinets, Class. Quant. Grav. **26**, 163001 (2009), arXiv:0905.2975 [gr-qc].
[28] B. Carter, Phys. Rev. Lett. **26**, 331 (1971).
[29] J. Healy, T. Bode, R. Haas, E. Pazos, P. Laguna, D. Shoemaker, and N. Yunes, Class. Quant. Grav. **29**, 232002 (2012), arXiv:1112.3928 [gr-qc].
[30] M. W. Horbatsch and C. P. Burgess, JCAP **05**, 010 (2012), arXiv:1111.4009 [gr-qc].
[31] E. Berti, V. Cardoso, L. Gualtieri, M. Horbatsch, and U. Sperhake, Phys. Rev. D **87**, 124020 (2013), arXiv:1304.2836 [gr-qc].
[32] T. P. Sotiriou and S.-Y. Zhou, Phys. Rev. Lett. **112**, 251102 (2014), arXiv:1312.3622 [gr-qc].
[33] K. Yagi, L. C. Stein, and N. Yunes, Phys. Rev. D **93**, 024010 (2016), arXiv:1510.02152 [gr-qc].
[34] E. Barausse and K. Yagi, Phys. Rev. Lett. **115**, 211105 (2015), arXiv:1509.04539 [gr-qc].
[35] H. O. Silva, J. Sakstein, L. Gualtieri, T. P. Sotiriou, and E. Berti, Phys. Rev. Lett. **120**, 131104 (2018), arXiv:1711.02080 [gr-qc].
[36] C. A. R. Herdeiro, E. Radu, N. Sanchis-Gual, and J. A. Font, Phys. Rev. Lett. **121**, 101102 (2018), arXiv:1806.05190 [gr-qc].
[37] F.-L. Julié, JCAP **10**, 033 (2018), arXiv:1809.05041 [gr-qc].
[38] F.-L. Julié and E. Berti, Phys. Rev. D **100**, 104061 (2019), arXiv:1909.05258 [gr-qc].
[39] F.-L. Julié, H. O. Silva, E. Berti, and N. Yunes, Phys. Rev. D **105**, 124031 (2022), arXiv:2202.01329 [gr-qc].
[40] J. L. Blázquez-Salcedo, F. S. Khoo, and J. Kunz, Phys. Rev. D **96**, 064008 (2017), arXiv:1706.03262 [gr-qc].
[41] J. L. Blázquez-Salcedo, C. F. B. Macedo, V. Cardoso, V. Ferrari, L. Gualtieri, F. S. Khoo, J. Kunz, and P. Pani, Phys. Rev. D **94**, 104024 (2016), arXiv:1609.01286 [gr-qc].
[42] C. Molina, P. Pani, V. Cardoso, and L. Gualtieri, Phys. Rev. D **81**, 124021 (2010), arXiv:1004.4007 [gr-qc].
[43] V. Cardoso and L. Gualtieri, Phys. Rev. D **80**, 064008 (2009), [Erratum: Phys.Rev.D 81, 089903 (2010)], arXiv:0907.5008 [gr-qc].
[44] P. Pani, E. Berti, and L. Gualtieri, Phys. Rev. Lett. **110**, 241103 (2013), arXiv:1304.1160 [gr-qc].
[45] P. Pani, E. Berti, and L. Gualtieri, Phys. Rev. **D88**, 064048 (2013), arXiv:1307.7315 [gr-qc].
[46] Z. Mark, H. Yang, A. Zimmerman, and Y. Chen, Phys. Rev. **D91**, 044025 (2015), arXiv:1409.5800 [gr-qc].
[47] O. J. C. Dias, M. Godazgar, and J. E. Santos, Phys. Rev. Lett. **114**, 151101 (2015), arXiv:1501.04625 [gr-qc].
[48] V. Cardoso, M. Kimura, A. Maselli, E. Berti, C. F. B. Macedo, and R. McManus, Phys. Rev. D **99**, 104077 (2019), arXiv:1901.01265 [gr-qc].

[49] R. McManus, E. Berti, C. F. B. Macedo, M. Kimura, A. Maselli, and V. Cardoso, Phys. Rev. D **100**, 044061 (2019), arXiv:1906.05155 [gr-qc].
[50] L. Pierini and L. Gualtieri, Phys. Rev. D **103**, 124017 (2021), arXiv:2103.09870 [gr-qc].
[51] L. Pierini and L. Gualtieri, Phys. Rev. D **106**, 104009 (2022), arXiv:2207.11267 [gr-qc].
[52] P. A. Cano, K. Fransen, T. Hertog, and S. Maenaut, (2023), arXiv:2304.02663 [gr-qc].
[53] P. A. Cano, K. Fransen, T. Hertog, and S. Maenaut, (2023), arXiv:2307.07431 [gr-qc].
[54] P. Wagle, N. Yunes, and H. O. Silva, Phys. Rev. D **105**, 124003 (2022), arXiv:2103.09913 [gr-qc].
[55] S. L. Detweiler, Astrophys. J. **239**, 292 (1980).
[56] O. Dreyer, B. J. Kelly, B. Krishnan, L. S. Finn, D. Garrison, and R. Lopez-Aleman, Class. Quant. Grav. **21**, 787 (2004), arXiv:gr-qc/0309007.
[57] E. E. Flanagan and S. A. Hughes, Phys. Rev. D **57**, 4535 (1998), arXiv:gr-qc/9701039.
[58] E. Berti, V. Cardoso, and C. M. Will, Phys. Rev. D **73**, 064030 (2006), arXiv:gr-qc/0512160.
[59] X. Jiménez Forteza, S. Bhagwat, P. Pani, and V. Ferrari, Phys. Rev. D **102**, 044053 (2020), arXiv:2005.03260 [gr-qc].
[60] X. J. Forteza, S. Bhagwat, S. Kumar, and P. Pani, Phys. Rev. Lett. **130**, 021001 (2023), arXiv:2205.14910 [gr-qc].
[61] R. Brito, A. Buonanno, and V. Raymond, Phys. Rev. D **98**, 084038 (2018), arXiv:1805.00293 [gr-qc].
[62] A. Ghosh, R. Brito, and A. Buonanno, Phys. Rev. D **103**, 124041 (2021), arXiv:2104.01906 [gr-qc].
[63] Y. Pan, A. Buonanno, M. Boyle, L. T. Buchman, L. E. Kidder, H. P. Pfeiffer, and M. A. Scheel, Phys. Rev. D **84**, 124052 (2011), arXiv:1106.1021 [gr-qc].
[64] R. Cotesta, A. Buonanno, A. Bohé, A. Taracchini, I. Hinder, and S. Ossokine, Phys. Rev. D **98**, 084028 (2018), arXiv:1803.10701 [gr-qc].
[65] A. Buonanno and T. Damour, Phys. Rev. D **59**, 084006 (1999), arXiv:gr-qc/9811091.
[66] A. Buonanno and T. Damour, Phys. Rev. D **62**, 064015 (2000), arXiv:gr-qc/0001013.
[67] E. Maggio, H. O. Silva, A. Buonanno, and A. Ghosh, (2022), arXiv:2212.09655 [gr-qc].
[68] H. O. Silva, A. Ghosh, and A. Buonanno, Phys. Rev. D **107**, 044030 (2023), arXiv:2205.05132 [gr-qc].
[69] B. P. Abbott *et al.* (LIGO Scientific, Virgo), Phys. Rev. Lett. **116**, 061102 (2016), arXiv:1602.03837 [gr-qc].
[70] G. Carullo, W. Del Pozzo, and J. Veitch, Phys. Rev. D **99**, 123029 (2019), [Erratum: Phys.Rev.D 100, 089903 (2019)], arXiv:1902.07527 [gr-qc].
[71] M. Isi and W. M. Farr, (2021), arXiv:2107.05609 [gr-qc].
[72] P. Amaro-Seoane *et al.* (LISA), (2017), arXiv:1702.00786 [astro-ph.IM].
[73] A. Klein *et al.*, Phys. Rev. **D93**, 024003 (2016), arXiv:1511.05581 [gr-qc].
[74] S. Bhagwat, C. Pacilio, E. Barausse, and P. Pani, Phys. Rev. D **105**, 124063 (2022), arXiv:2201.00023 [gr-qc].
[75] R. Cotesta, S. Marsat, A. Toubiana, and B. Emanuele, (2023).
[76] E. Berti, A. Sesana, E. Barausse, V. Cardoso, and K. Belczynski, Phys. Rev. Lett. **117**, 101102 (2016), arXiv:1605.09286 [gr-qc].
[77] L. S. Finn and D. F. Chernoff, Phys. Rev. D **47**, 2198 (1993), arXiv:gr-qc/9301003.
[78] M. Vallisneri, Phys. Rev. **D77**, 042001 (2008), arXiv:gr-qc/0703086 [GR-QC].
[79] D. P. Mihaylov, S. Ossokine, A. Buonanno, H. Estelles, L. Pompili, M. Pürrer, and A. Ramos-Buades, (2023), arXiv:2303.18203 [gr-qc].
[80] M. Khalil, A. Buonanno, H. Estelles, D. P. Mihaylov, S. Ossokine, L. Pompili, and A. Ramos-Buades, (2023), arXiv:2303.18143 [gr-qc].
[81] L. Pompili *et al.*, (2023), arXiv:2303.18039 [gr-qc].
[82] M. van de Meent, A. Buonanno, D. P. Mihaylov, S. Ossokine, L. Pompili, N. Warburton, A. Pound, B. Wardell, L. Durkan, and J. Miller, (2023), arXiv:2303.18026 [gr-qc].
[83] A. Sesana, M. Volonteri, and F. Haardt, Mon. Not. Roy. Astron. Soc. **377**, 1711 (2007), arXiv:astro-ph/0701556.
[84] A. Sesana, J. Gair, E. Berti, and M. Volonteri, Phys. Rev. D **83**, 044036 (2011), arXiv:1011.5893 [astro-ph.CO].
[85] M. Bonetti, A. Sesana, F. Haardt, E. Barausse, and M. Colpi, Mon. Not. Roy. Astron. Soc. **486**, 4044 (2019), arXiv:1812.01011 [astro-ph.GA].
[86] P. Dayal, E. M. Rossi, B. Shiralilou, O. Piana, T. R. Choudhury, and M. Volonteri, Mon. Not. Roy. Astron. Soc. **486**, 2336 (2019), arXiv:1810.11033 [astro-ph.GA].
[87] E. Barausse, I. Dvorkin, M. Tremmel, M. Volonteri, and M. Bonetti, Astrophys. J. **904**, 16 (2020), arXiv:2006.03065 [astro-ph.GA].
[88] E. Barausse and A. Lapi, (2020), arXiv:2011.01994 [astro-ph.GA].
[89] M. A. Latif and A. Ferrara, Publ. Astron. Soc. Austral. **33**, e051 (2016), arXiv:1605.07391 [astro-ph.GA].
[90] H. Estellés, S. Husa, M. Colleoni, D. Keitel, M. Mateu-Lucena, C. García-Quirós, A. Ramos-Buades, and A. Borchers, Phys. Rev. D **105**, 084039 (2022), arXiv:2012.11923 [gr-qc].
[91] P. Ajith *et al.*, Phys. Rev. D **77**, 104017 (2008), [Erratum: Phys.Rev.D 79, 129901 (2009)], arXiv:0710.2335 [gr-qc].
[92] P. Ajith *et al.*, Phys. Rev. Lett. **106**, 241101 (2011), arXiv:0909.2867 [gr-qc].
[93] L. Santamaria *et al.*, Phys. Rev. D **82**, 064016 (2010), arXiv:1005.3306 [gr-qc].
[94] N. Aghanim *et al.* (Planck), Astron. Astrophys. **641**, A6 (2020), [Erratum: Astron.Astrophys. 652, C4 (2021)], arXiv:1807.06209 [astro-ph.CO].
[95] A. Buonanno, G. B. Cook, and F. Pretorius, Phys. Rev. D **75**, 124018 (2007), arXiv:gr-qc/0610122.
[96] B. J. Kelly and J. G. Baker, Phys. Rev. D **87**, 084004 (2013), arXiv:1212.5553 [gr-qc].
[97] A. Bohé *et al.*, Phys. Rev. D **95**, 044028 (2017), arXiv:1611.03703 [gr-qc].
[98] J. G. Baker, W. D. Boggs, J. Centrella, B. J. Kelly, S. T. McWilliams, and J. R. van Meter, Phys. Rev. D **78**, 044046 (2008), arXiv:0805.1428 [gr-qc].
[99] T. Damour and A. Nagar, Phys. Rev. D **90**, 044018 (2014), arXiv:1406.6913 [gr-qc].
[100] E. Berti, V. Cardoso, and M. Casals, Phys. Rev. D **73**, 024013 (2006), [Erratum: Phys.Rev.D 73, 109902 (2006)], arXiv:gr-qc/0511111.
[101] A. Kumar Mehta, P. Tiwari, N. K. Johnson-McDaniel, C. K. Mishra, V. Varma, and P. Ajith, Phys. Rev. D **100**, 024032 (2019), arXiv:1902.02731 [gr-qc].
[102] E. Berti and A. Klein, Phys. Rev. D **90**, 064012 (2014), arXiv:1408.1860 [gr-qc].
[103] L. C. Stein, J. Open Source Softw. **4**, 1683 (2019), arXiv:1908.10377 [gr-qc].
[104] X. Jiménez-Forteza, D. Keitel, S. Husa, M. Hannam, S. Khan, and M. Pürrer, Phys. Rev. D **95**, 064024 (2017), arXiv:1611.00332 [gr-qc].

[105] F. Hofmann, E. Barausse, and L. Rezzolla, Astrophys. J. Lett. **825**, L19 (2016), arXiv:1605.01938 [gr-qc].

[106] S. Gossan, J. Veitch, and B. S. Sathyaprakash, Phys. Rev. D **85**, 124056 (2012), arXiv:1111.5819 [gr-qc].

[107] J. Meidam, M. Agathos, C. Van Den Broeck, J. Veitch, and B. S. Sathyaprakash, Phys. Rev. D **90**, 064009 (2014), arXiv:1406.3201 [gr-qc].

[108] M. Isi, M. Giesler, W. M. Farr, M. A. Scheel, and S. A. Teukolsky, Phys. Rev. Lett. **123**, 111102 (2019), arXiv:1905.00869 [gr-qc].

[109] C. Cutler, Phys. Rev. D **57**, 7089 (1998), arXiv:gr-qc/9703068.

[110] S. Marsat, J. G. Baker, and T. Dal Canton, Phys. Rev. D **103**, 083011 (2021), arXiv:2003.00357 [gr-qc].

[111] M. Tinto and S. V. Dhurandhar, Living Rev. Rel. **8**, 4 (2005), arXiv:gr-qc/0409034 [gr-qc].

[112] LISA Science Study Team, "LISA Science Requirements Document," `https://www.cosmos.esa.int/documents/678316/1700384/SciRD.pdf/25831f6b-3c01-e215-5916-4ac6e4b306fb?t=1526479841000` (2018).

[113] N. Karnesis, M. L. Katz, N. Korsakova, J. R. Gair, and N. Stergioulas, (2023), arXiv:2303.02164 [astro-ph.IM].

[114] N. Karnesis, M. L. Katz, N. Korsakova, J. R. Gair, and N. Stergioulas, "Eryn," https://github.com/mikekatz04/Eryn, (2023).

[115] M. Volonteri, G. Lodato, and P. Natarajan, Mon. Not. Roy. Astron. Soc. **383**, 1079 (2008), arXiv:0709.0529 [astro-ph].

[116] J. Binney and S. Tremaine, *Galactic dynamics* (1987).

[117] M. J. Rees, Ann. Rev. Astron. Astrophys. **22**, 471 (1984).

[118] J. Antoniadis *et al.*, (2023), arXiv:2306.16214 [astro-ph.HE].

[119] G. Agazie *et al.* (NANOGrav), Astrophys. J. Lett. **951**, L8 (2023), arXiv:2306.16213 [astro-ph.HE].

[120] D. J. Reardon *et al.*, Astrophys. J. Lett. **951**, L6 (2023), arXiv:2306.16215 [astro-ph.HE].

[121] H. Xu *et al.*, Res. Astron. Astrophys. **23**, 075024 (2023), arXiv:2306.16216 [astro-ph.HE].

[122] A. Sesana, E. Barausse, M. Dotti, and E. M. Rossi, Astrophys. J. **794**, 104 (2014), arXiv:1402.7088 [astro-ph.CO].

[123] J. M. Bardeen and J. A. Petterson, Astrophys. J. Lett. **195**, L65 (1975).

[124] T. Bogdanovic, C. S. Reynolds, and M. C. Miller, Astrophys. J. Lett. **661**, L147 (2007), arXiv:astro-ph/0703054.

[125] D. Merritt and M. Milosavljevic, Living Rev. Rel. **8**, 8 (2005), arXiv:astro-ph/0410364.

[126] M. Colpi, Space Sci. Rev. **183**, 189 (2014), arXiv:1407.3102 [astro-ph.GA].

[127] M. Tremmel, F. Governato, M. Volonteri, T. R. Quinn, and A. Pontzen, "Mon. Not. Roy. Astron. Soc." **475**, 4967 (2018), arXiv:1708.07126.

[128] C. L. Rodriguez, B. Farr, V. Raymond, W. M. Farr, T. B. Littenberg, D. Fazi, and V. Kalogera, Astrophys. J. **784**, 119 (2014), arXiv:1309.3273 [astro-ph.HE].

[129] C. Cutler and E. E. Flanagan, Phys. Rev. D **49**, 2658 (1994), arXiv:gr-qc/9402014.

[130] V. Baibhav, E. Berti, and V. Cardoso, Phys. Rev. D **101**, 084053 (2020), arXiv:2001.10011 [gr-qc].

[131] S. Bhagwat, X. J. Forteza, P. Pani, and V. Ferrari, Phys. Rev. D **101**, 044033 (2020), arXiv:1910.08708 [gr-qc].

[132] I. Ota and C. Chirenti, Phys. Rev. D **101**, 104005 (2020), arXiv:1911.00440 [gr-qc].

[133] V. Ferrari, M. Pauri, and F. Piazza, Phys. Rev. **D63**, 064009 (2001), arXiv:gr-qc/0005125 [gr-qc].

[134] P. Pani and V. Cardoso, Phys. Rev. **D79**, 084031 (2009), arXiv:0902.1569 [gr-qc].

[135] R. Brito, V. Cardoso, and P. Pani, Phys. Rev. **D88**, 023514 (2013), arXiv:1304.6725 [gr-qc].

[136] R. Brito, V. Cardoso, and P. Pani, Phys. Rev. **D87**, 124024 (2013), arXiv:1306.0908 [gr-qc].

[137] E. Babichev, R. Brito, and P. Pani, Phys. Rev. **D93**, 044041 (2016), arXiv:1512.04058 [gr-qc].

[138] K. Glampedakis, G. Pappas, H. O. Silva, and E. Berti, Phys. Rev. **D96**, 064054 (2017), arXiv:1706.07658 [gr-qc].

[139] P. Jai-akson, A. Chatrabhuti, O. Evnin, and L. Lehner, Phys. Rev. **D96**, 044031 (2017), arXiv:1706.06519 [gr-qc].

[140] K. Glampedakis and G. Pappas, Phys. Rev. **D97**, 041502 (2018), arXiv:1710.02136 [gr-qc].

[141] M. Boyle *et al.*, Class. Quant. Grav. **36**, 195006 (2019), arXiv:1904.04831 [gr-qc].

[142] E. E. Flanagan and S. A. Hughes, Phys. Rev. D **57**, 4566 (1998), arXiv:gr-qc/9710129.

[143] L. Lindblom, B. J. Owen, and D. A. Brown, Phys. Rev. D **78**, 124020 (2008), arXiv:0809.3844 [gr-qc].

[144] S. T. McWilliams, B. J. Kelly, and J. G. Baker, Phys. Rev. D **82**, 024014 (2010), arXiv:1004.0961 [gr-qc].

[145] K. Chatziioannou, A. Klein, N. Yunes, and N. Cornish, Phys. Rev. D **95**, 104004 (2017), arXiv:1703.03967 [gr-qc].

[146] M. Pürrer and C.-J. Haster, Phys. Rev. Res. **2**, 023151 (2020), arXiv:1912.10055 [gr-qc].

[147] S. Ossokine *et al.*, Phys. Rev. D **102**, 044055 (2020), arXiv:2004.09442 [gr-qc].

[148] A. Ramos-Buades, A. Buonanno, H. Estellés, M. Khalil, D. P. Mihaylov, S. Ossokine, L. Pompili, and M. Shiferaw, (2023), arXiv:2303.18046 [gr-qc].

[149] N. Yunes and F. Pretorius, Phys. Rev. D **80**, 122003 (2009), arXiv:0909.3328 [gr-qc].

[150] M. Agathos, W. Del Pozzo, T. G. F. Li, C. Van Den Broeck, J. Veitch, and S. Vitale, Phys. Rev. D **89**, 082001 (2014), arXiv:1311.0420 [gr-qc].

[151] A. K. Mehta, A. Buonanno, R. Cotesta, A. Ghosh, N. Sennett, and J. Steinhoff, Phys. Rev. D **107**, 044020 (2023), arXiv:2203.13937 [gr-qc].

[152] L. Blanchet, Living Rev. Rel. **17**, 2 (2014), arXiv:1310.1528 [gr-qc].

[153] LIGO Scientific Collaboration, "LIGO Algorithm Library - LALSuite," free software (GPL) (2018).

[154] S. Husa, S. Khan, M. Hannam, M. Pürrer, F. Ohme, X. Jiménez Forteza, and A. Bohé, Phys. Rev. D **93**, 044006 (2016), arXiv:1508.07250 [gr-qc].

[155] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. Jiménez Forteza, and A. Bohé, Phys. Rev. D **93**, 044007 (2016), arXiv:1508.07253 [gr-qc].

[156] C. García-Quirós, M. Colleoni, S. Husa, H. Estellés, G. Pratten, A. Ramos-Buades, M. Mateu-Lucena, and R. Jaume, Phys. Rev. D **102**, 064002 (2020), arXiv:2001.10914 [gr-qc].