

# Measuring source properties and quasinormal mode frequencies of heavy massive black-hole binaries with LISA

Alexandre Toubiana<sup>1</sup>, Lorenzo Pompili<sup>1</sup>, Alessandra Buonanno,<sup>1,2</sup> Jonathan R. Gair,<sup>1</sup> and Michael L. Katz<sup>1,3</sup>

<sup>1</sup>Max Planck Institute for Gravitationsphysik (Albert Einstein Institute),  
Am Mühlenberg 1, 14476 Potsdam, Germany

<sup>2</sup>Department of Physics, University of Maryland, College Park, Maryland 20742, USA

<sup>3</sup>NASA Marshall Space Flight Center, Huntsville, Alabama 35811, USA



(Received 2 August 2023; accepted 30 January 2024; published 6 May 2024)

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 in both 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 signals using the newly developed multipolar effective-one-body model: pSEOBNRv5HM. 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 quasinormal 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. 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. We also probe the accuracy of the SEOBNRv5HM waveform family by performing synthetic injections of GR numerical-relativity waveforms. Using a novel method that we develop here to quantify the impact of systematic errors, we show that, already for sources with SNR  $\mathcal{O}(100)$ , we would measure erroneous deviations from GR due to waveform model inaccuracies. One of the main sources of error is the mismodeling of the relative alignment between harmonics. These results confirm the need for improving waveform models to perform tests of GR with binary BHs observed at high SNR by LISA.

DOI: [10.1103/PhysRevD.109.104019](https://doi.org/10.1103/PhysRevD.109.104019)

## 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

---

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Open access publication funded by the Max Planck Society.

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 quasinormal 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 modeling 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 parametrized 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 parametrized (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 parametrize 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 parametrized ringdown test, which has already been applied to analyze 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 postinspiral regimes have at least  $\text{SNR} \sim 8$ , 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 parametrized EOB waveforms developed using the state-of-the-art multipolar aligned-spin model SEOBNRv5HM<sup>1</sup> developed in Refs. [79–82]. Henceforth, we denote the parametrized model as pSEOBNRv5HM. Crucially, this model includes higher harmonics, which are expected to play an important role in LISA parameter estimation [83].

Since the expected population of MBHBs is highly uncertain [73,84–89], here we focus on a few astrophysically realistic systems, compatible with the predictions of models where MBHBs form from heavy seeds [90]. 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 modeling on ringdown tests by simulating mock LISA injections with waveforms from NR and using pSEOBNRv5HM waveforms to perform the Bayesian analysis. Developing a new approach, we estimate from which SNR we expect to erroneously measure deviations from GR due to systematic effects. Finally, we assess to which extent a consistent modeling 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.

This paper is organized as follows. In Sec. II, we present the details of our parametrized EOB model, describe how the synthetic LISA observations are generated, and lay down the basis of our Bayesian analyses. In Sec. III, we summarize the astrophysical systems that we simulate. We present our results when using the pSEOBNRv5HM model for both injection and parameter estimation in Sec. IV; then in Sec. V we discuss the impact of systematics and using a novel method, we show that our models are not yet accurate enough for the signals we expect with LISA. Finally, we present our conclusions in Sec. VI. In the appendixes, we discuss how the settings of pSEOBNRv5HM waveforms impact the parameter estimation, and we show that, in GR,

<sup>1</sup>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).

SEOBNRv5HM and IMRPhenomTHM [91], a time-domain waveform model from the IMR phenomenological family [92–94], predict similar measurement errors for the parameters of the source, that their measurement is little affected by adding the QNM deviation parameters in the pSEOBNRv5HM model, and additional results of our study of systematic effects. Throughout this paper, we will use natural units in which  $c = G = 1$ .

## II. METHODS

### A. Parametrized 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 quasicircular orbits with aligned or antialigned 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) [95]. 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 polarizations can be expanded in the basis of spin-weight  $-2$  spherical harmonics as

$$\begin{aligned} & h_+(\Theta, \iota, \varphi_0; t) - ih_\times(\Theta, \iota, \varphi_0; t) \\ &= \frac{1}{D_L} \sum_{\ell, m} {}_{-2}Y_{\ell m}(\iota, \varphi_0) h_{\ell m}(\Theta; t), \end{aligned} \quad (2.1)$$

where the parameters  $(\iota, \varphi_0)$  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  $\Theta$  denotes the intrinsic parameters (masses and spins) of the binary. We build our parametrized 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  $(\ell, m)$  harmonics with  $m > 0$ .

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

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

where  $\theta(t)$  is the Heaviside step function,  $h_{\ell m}^{\text{insp-plunge}}$  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  $(\ell, m)$  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  $\Theta$  dependence for ease of notation.

For all harmonics, except for  $(\ell, |m|) = (3, 2)$  and  $(4, 3)$  which exhibit postmerger oscillations due to mode mixing [96,97], the merger-ringdown waveform employs the following ansatz [64,81,98]:

$$h_{\ell m}^{\text{merger-RD}}(t) = \nu \tilde{\mathcal{A}}_{\ell m}(t) e^{i\tilde{\phi}_{\ell m}(t)} e^{-i\sigma_{\ell m 0}(t - t_{\text{match}}^{\ell m})}, \quad (2.3)$$

where  $\nu = m_1 m_2 / (m_1 + m_2)^2$  is the symmetric mass ratio of the binary and  $\sigma_{\ell m 0}$  is the complex frequency of the least-damped QNM, having overtone number zero, of the remnant BH. We define the corresponding oscillation frequency  $f_{\ell m 0}$  and the damping time  $\tau_{\ell m 0}$ , respectively, as

$$f_{\ell m 0} = \frac{1}{2\pi} \text{Re}(\sigma_{\ell m 0}) = -\frac{1}{2\pi} \sigma_{\ell m 0}^{\text{I}}, \quad (2.4a)$$

$$\tau_{\ell m 0} = -\frac{1}{\text{Im}(\sigma_{\ell m 0})} = -\frac{1}{\sigma_{\ell m 0}^{\text{R}}}. \quad (2.4b)$$

The functions  $\tilde{\mathcal{A}}_{\ell m}(t)$  and  $\tilde{\phi}_{\ell m}(t)$  are given by [64,81,98–100]

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

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

where  $\phi_{\text{match}}^{\ell m}$  is the phase of the inspiral-plunge harmonic  $(\ell, m)$  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]

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

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

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

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

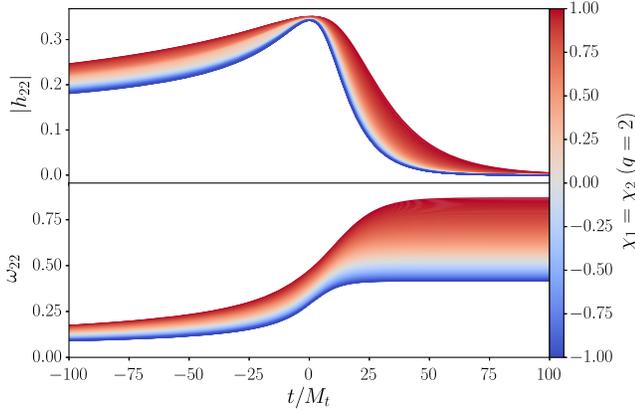


FIG. 1. Time evolution near the merger, which occurs at  $t = 0$ , of the SEOBv5NRHM (2, 2) harmonic amplitude (top panel) and instantaneous frequency (bottom panel) when varying the spin components of a binary with mass ratio 2, assuming equal and aligned spins.

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  $\Theta$  and can be found in Appendix D in 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.

For the (3, 2) and (4, 3) harmonics, the mode-mixing behavior is modeled by applying the previous construction to the spheroidal harmonics [101] (3, 2, 0) and (4, 3, 0), which feature a monotonic amplitude and frequency evolution [102]. The spheroidal (3, 2, 0) and (4, 3, 0) harmonics can be related to the spherical harmonics by [81]

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

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

where  $\mu_{m\ell\ell'n}$  are harmonic mixing coefficients, obtained using fits from Ref. [103]. 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).

Though overtones are not explicitly included in the merger-ringdown signal ansatz, unlike older versions of the model [63,104], their effect should be captured by the functions  $\tilde{\mathcal{A}}_{\ell m}(t)$  and  $\tilde{\phi}_{\ell m}(t)$ . They contain free coefficients fitted against NR simulations and allow our ansatz to

be more than a simple damped sinusoid with damping time and frequency given by the fundamental QNM. Moreover, we do not expect the linear perturbation description used in the ringdown to be valid right at the time used to transition from inspiral to merger-ringdown [105], which for most modes is the peak of the (2,2) harmonic. Thus, the choice of modeling in SEOBv5NRHM allows one to better capture such nonlinearities without having to include a transition phase.

In the SEOBv5NRHM model constructed in Ref. [81], the complex QNM frequencies in GR are obtained for each  $(\ell, m)$  harmonic as a function of the BH's final mass and spin using the QNM PYTHON package [106]. The BH's mass and spin are, in turn, computed using the fitting formulas of Refs. [107,108], respectively. In this work, following the strategy of Refs. [61,62,67,71,109–111], we introduce parametrized 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). More explicitly, we perform the substitutions

$$f_{\ell m 0} \rightarrow f_{\ell m 0}(1 + \delta f_{\ell m}), \quad (2.9a)$$

$$\tau_{\ell m 0} \rightarrow \tau_{\ell m 0}(1 + \delta \tau_{\ell m}), \quad (2.9b)$$

where for ease of notation we have dropped the zero overtone subscript in the deviation parameters. We shall denote this parametrized model as pSEOBv5NRHM. We note that allowing  $\sigma_{\ell m 0}$  to vary freely also modifies the  $c_{i,c}^{\ell m}$  and  $d_{i,c}^{\ell m}$  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 [112] 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  $2\pi fL/c \ll 1$ ). 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- or Virgo-type detectors rotated with respect to each other by  $\pi/4$  and with angles of  $\pi/3$  between the arms.

Transforming Eq. (47) in Ref. [113] to the time domain, we find that under the long-wavelength approximation the time-delay-interferometry (TDI) variables  $A$ ,  $E$ , and  $T$  [114] (which, for an interferometer with equal arms and equal noise levels in each optical link, provide three noise-uncorrelated datasets) are given by

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

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

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

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

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

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

In the above equations,  $\lambda$ ,  $\beta$ , and  $\psi$  are the longitude, latitude, and polarization in the LISA frame, respectively. We refer to Ref. [113] for the relation between the angles in the LISA frame and those in the solar-system-barycenter 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 polarizations 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,  $\lambda$ ,  $\beta$ , and  $\psi$  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  $f_{\text{gen}} = 5 \times 10^{-5}[M_{r,0}/(2 \times 10^7 M_\odot)]$  Hz until the end of the signal. After transforming to the frequency domain, we keep the

portion of the signal between  $f_{\text{min}} = 2 \times 10^{-4}[M_{r,0}/(2 \times 10^7 M_\odot)]$  Hz and  $f_{\text{max}}$ , 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  $f_{\text{min}}$  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  $\varphi_c$  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  $\varphi_0$  such that the phase of  ${}_{-2}Y_{22}(t, \varphi_0)h_{22}(t_c)$  is  $\varphi_c$ . This choice of  $\varphi_0$  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  $t_{\text{match}}^{\ell m}$  in Eq. (2.2), for all  $(\ell, m)$  harmonics except (5, 5).

### C. Bayesian analysis

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

$$(d_1|d_2) = 4\text{Re}\left[\int_0^{+\infty} \frac{d_1(f)d_2^*(f)}{S_n(f)} df\right], \quad (2.12)$$

where  $S_n(f)$  is the power spectral density (PSD). In this work, we use the SciRDv1 noise curve [115], 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  $\text{SNR} = \sqrt{(h|h)}$ .

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,  $\theta$ , given an observed dataset  $d$ , using Bayes' theorem:

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

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 normalization constant and, thus, can be discarded. We take the prior to be flat in the (detector-frame) total mass  $M_r$ , the mass ratio  $q$ , the spins  $-1 \leq \chi_1 \leq 1$  and  $-1 \leq \chi_2 \leq 1$ , the time to coalescence  $t_c$ , and the phase at coalescence  $\varphi_c$ . For the systems we consider here, the intrinsic parameters  $M_r$ ,  $q$ ,  $\chi_1$ , and  $\chi_2$  are typically well measured, so that the actual priors have little importance. We take a flat prior on  $\psi$ ,  $\cos(i)$ , and  $\log_{10}(D_L)$  and fix the sky location  $(\lambda, \beta)$  to its true value to facilitate the convergence of the chains. Those parameters are not expected to correlate strongly with intrinsic

parameters [113], 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

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

The posterior distribution is then sampled via a Markov-chain Monte Carlo algorithm (MCMC). We use the Eryn sampler [116,117] 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:

- (i)  $M_{t,0} = 2 \times 10^7 M_\odot$  or  $M_{t,0} = 2 \times 10^8 M_\odot$ ,
- (ii)  $q_0 = 2$  or  $q_0 = 4$ ,
- (iii)  $\chi_{1,0} = \chi_{2,0} = 0.9$  or  $\chi_{1,0} = 0.2, \chi_{2,0} = 0.1$ , and
- (iv)  $z_0 = 2.2$  ( $D_{L,0} = 18\,331$  Mpc) or  $z_0 = 3.7$  ( $D_{L,0} = 33\,691$  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 semianalytic models of MBH populations that use heavy seeds for the MBH progenitors [73,84–89]. Different heavy-seed scenarios have been proposed, such as the collapse of protogalactic disks as a result of bar instabilities [118], the runaway collision of stars at the center of galaxies [119], or the direct collapse of gas at the center of galaxies [120] (see Ref. [90] 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 [121–124]. 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. Semianalytical 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 [125].

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 [126,127]. Binaries formed through triplet interactions can also have misaligned spins, in addition to having high eccentricity [86]. As discussed, we neglect eccentricity and spin precession here for simplicity and focus on quasicircular binaries. Exploring different values for  $q$ ,  $\chi_1$ , and  $\chi_2$  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 timescales at play during the hardening of the binary [128–130]. 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 affects only the SNR of the system. Based on the predictions of semianalytical models [73,84–89], 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  $z_0 = 2.2$  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-spin systems, whereas (3, 2) and (4, 3) are more dominant for high-spin 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 in 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  $z_0 = 2.2$ . Although systems with  $M_{t,0} = 2 \times 10^7 M_\odot$  (upper panel) still have high SNR also in the inspiral,  $\sim 1000$ , 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  $\sim 30$ .

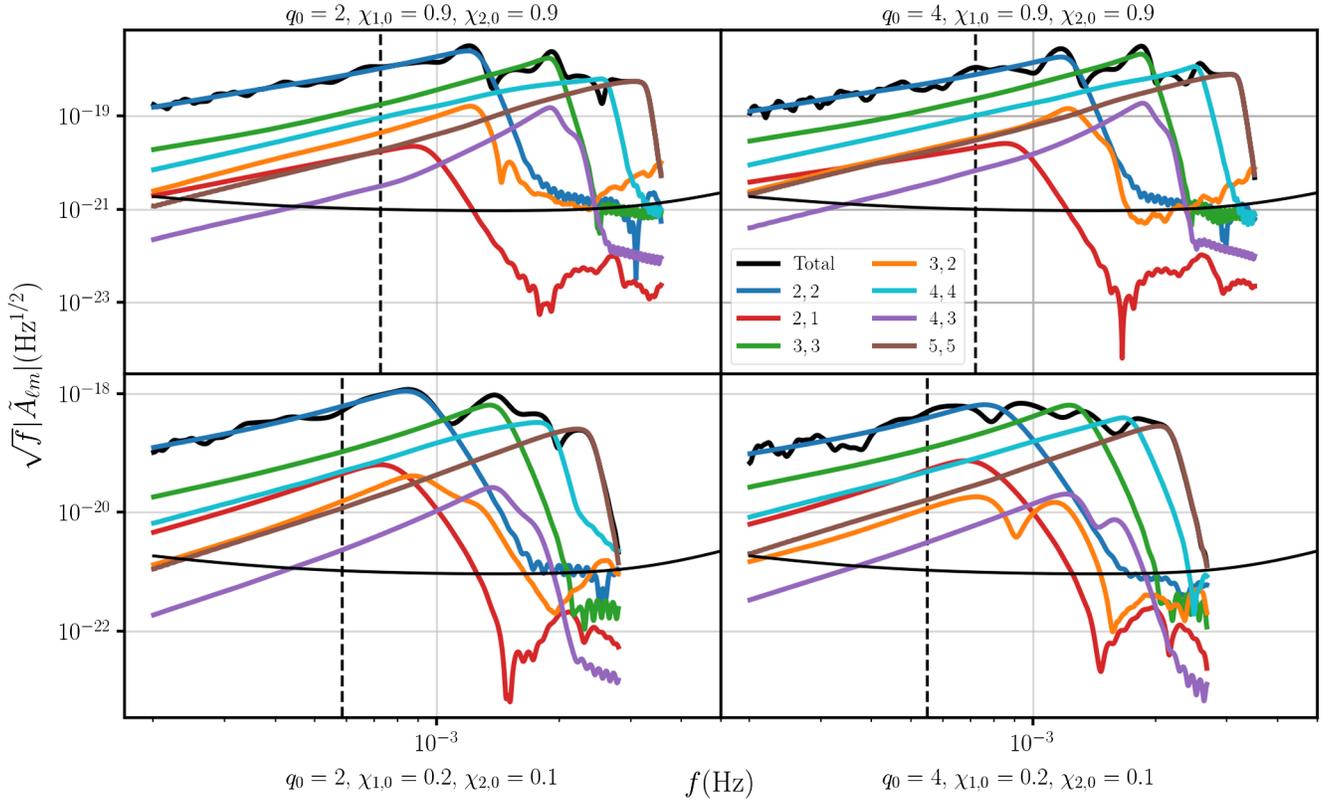


FIG. 2. Amplitude in frequency domain in the TDI channel  $A$  broken into the contributions of each of the harmonics included in the pSEOBNRv5HM model,  $\tilde{A}_{\ell m}$ . The latter is computed using the harmonic decomposition of Eq. (2.1) in the TDI equations (2.10a). The thin black line shows 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 polarization, 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  $z_0 = 2.2$ . For systems at  $z_0 = 3.7$ , one should simply rescale the amplitudes by  $18\,331/33\,691 \simeq 0.54$ , 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-spin systems (upper row), the (3, 2) and (4, 3) harmonics are louder than the (2, 1) harmonic, whereas the opposite is true for low-spin systems (lower row).

#### 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 [131], 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 centered around the median for the intrinsic parameters (i.e., the masses and spins) as a function of the SNR of the system. The color, shape, and filling of the point indicate, respectively, the total mass, spin, and mass

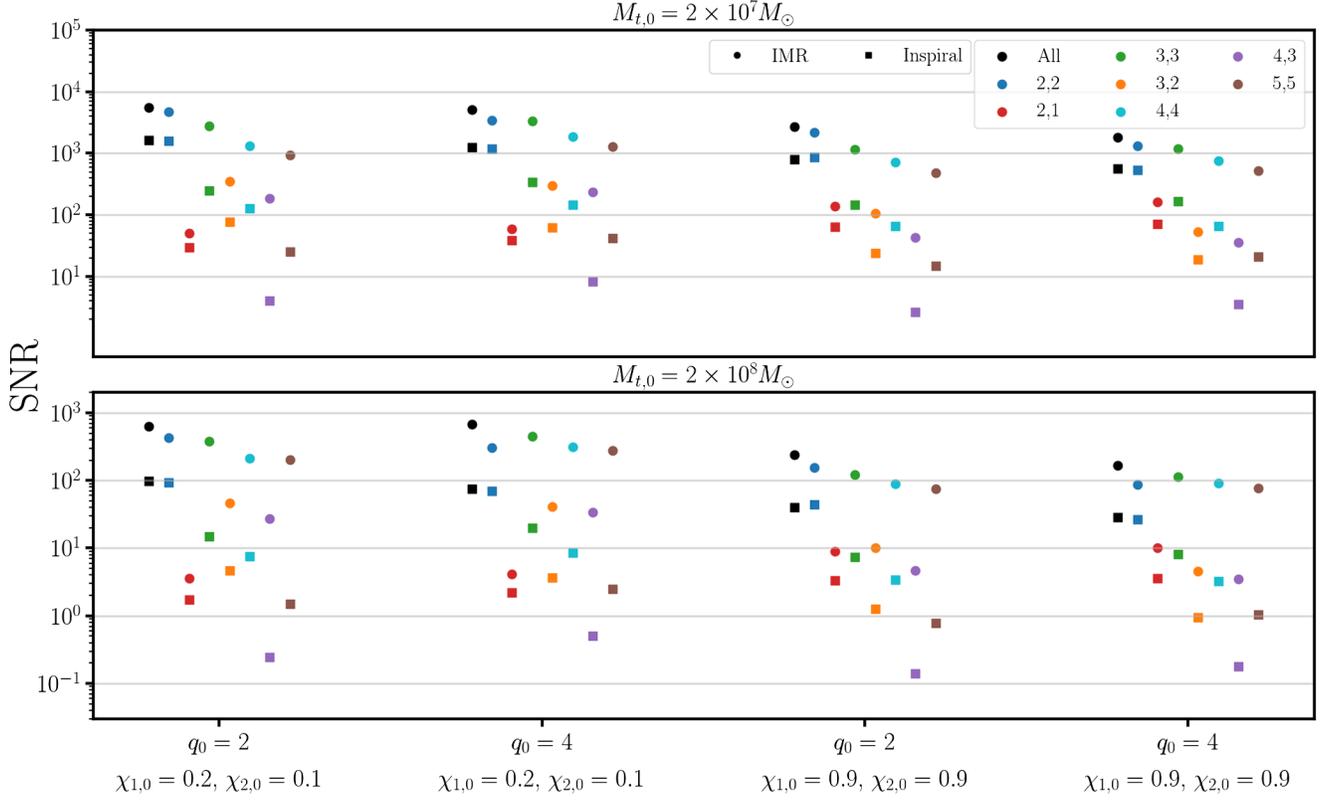


FIG. 3. Total SNR (black) and SNR in the individual harmonics (colors) for all the systems at  $z_0 = 2.2$ . The values at  $z_0 = 3.7$  can be obtained by rescaling by  $18\,331/33\,691 \simeq 0.54$ . Circles show the full IMR SNR and squares the SNR in the inspiral stage.

ratio of the system, indicated in the legends. Each point is doubled because of the two redshifts used:  $z_0 = 2.2$  ( $z_0 = 3.7$ ) 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  $1/\text{SNR}$ , as expected in the high SNR regime [77,78,132]. This relation is more scattered for the spin parameters, especially for  $\chi_1$ : Systems with  $q_0 = 4$  have better spin measurement, in agreement with [75]. This is because, for such systems (nonfilled 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 [133] found. This is the benefit of using a fully consistent modeling 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.

For very massive systems with low SNR, we find a multimodality 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

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  $z_0 = 2.2$ . Values at  $z_0 = 3.7$  can be obtained by rescaling by  $18\,331/33\,691 \simeq 0.54$ . 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.

		$M_{t,0} = 2 \times 10^7 M_\odot$			$M_{t,0} = 2 \times 10^8 M_\odot$		
		IMR	Merger-ringdown	Inspiral	IMR	Merger-ringdown	Inspiral
$\chi_{1,0} = \chi_{2,0} = 0.9$	$q_0 = 2$	5505	5253	1645	623	616	98
	$q_0 = 4$	5174	5017	1269	672	668	75
$\chi_{1,0} = 0.2, \chi_{2,0} = 0.1$	$q_0 = 2$	2678	2558	795	241	238	40
	$q_0 = 4$	1832	1741	571	167	165	29

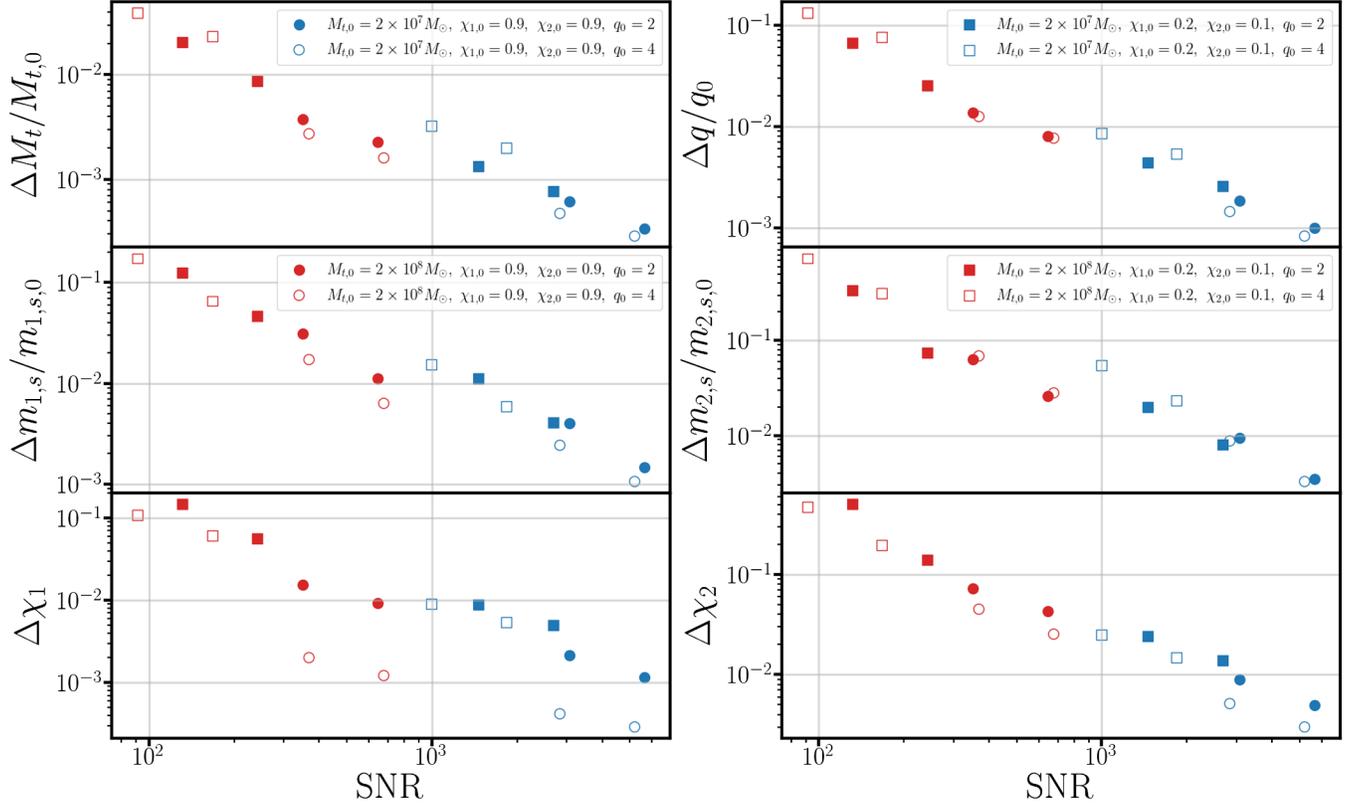


FIG. 4. Measurement error (90% confidence interval centered around the median) of intrinsic parameters. For mass parameters, we show the errors relative to the injection value. We inject signals within GR and recover them with GR templates (i.e., we employ SEOBNRv5HM). All points are doubled because of the two redshift possibilities;  $z_0 = 2.2$  ( $z_0 = 3.7$ ) 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  $1/\text{SNR}$  trend.

in the upper-right part in 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 small compared to the total SNR ( $\sim 30$ ; 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 the ones obtained with the IMR phenomenological model IMRPhenomTHM [91], 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 nonzero deviations when running the Bayesian analysis. In Figs. 7 and 8, we show the width of the 90% confidence interval centered 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 subdominant (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  $M_{t,0} = 2 \times 10^7 M_\odot$  (blue points) and follow the  $1/\text{SNR}$  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  $q_0 = 4$  (nonfilled 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 subdominant harmonics are poorly constrained for very massive systems (red points),  $\Delta\delta f_{\ell m}$  and  $\Delta\delta\tau_{\ell m} \sim 1$ ,

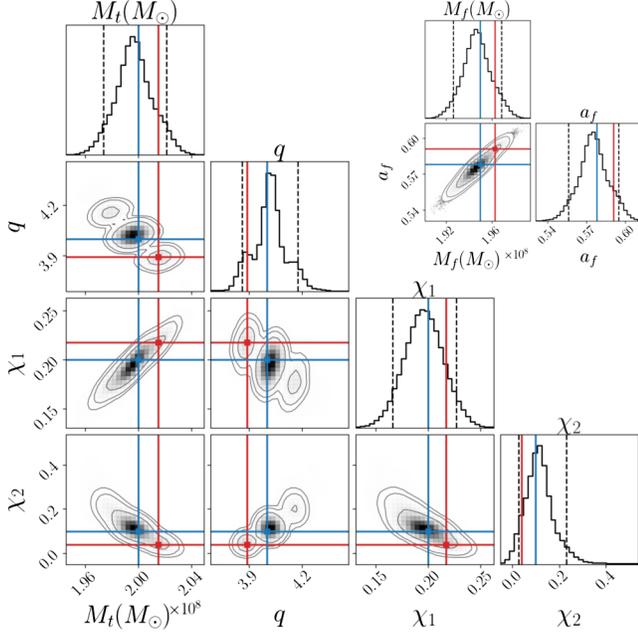


FIG. 5. Corner plot for the system with  $M_{t,0} = 2 \times 10^8 M_{\odot}$ ,  $\chi_{1,0} = 0.2$ ,  $\chi_{2,0} = 0.1$ ,  $q_0 = 4$ ,  $z_0 = 2.2$ , and  $\text{SNR} = 167$ . The blue lines show the injection and the red ones a secondary mode. Contours show the 68%, 90%, and 95% confidence intervals, and dashed lines the 0.05 and 0.95 quantiles. On the upper-right part, we show the inferred properties of the remnant. The multimodality 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 multimodality in  $M_f$  and  $a_f$ .

which, given our prior range, translates into uninformative measurements. This is due to the low SNR of these harmonics. However, we get rather good constraints, within 0.1, for systems with  $M_{t,0} = 2 \times 10^7 M_{\odot}$  (blue

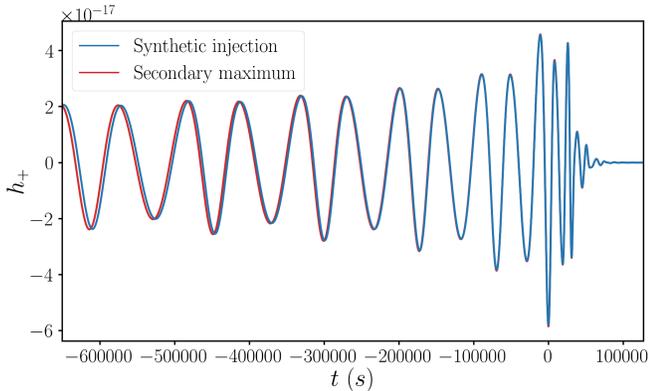
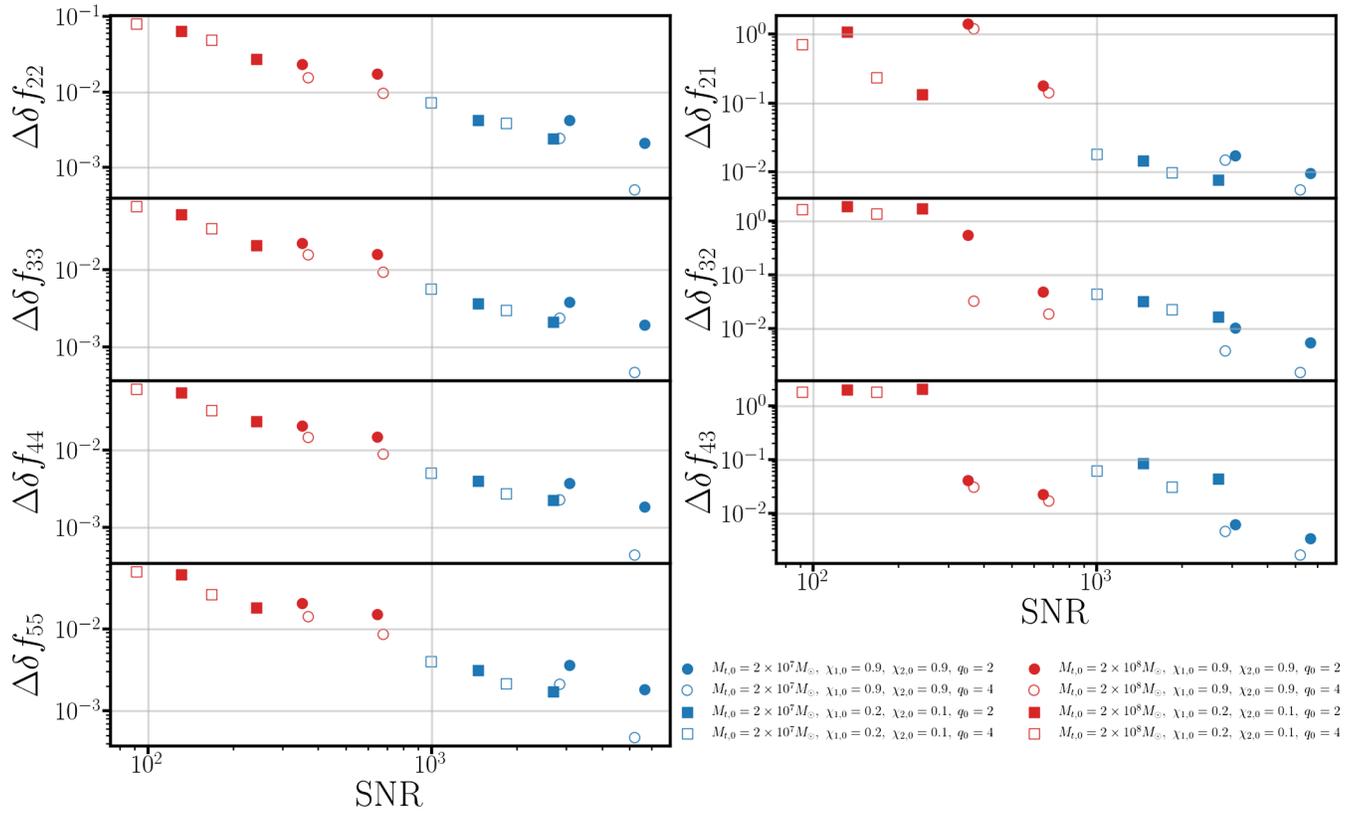
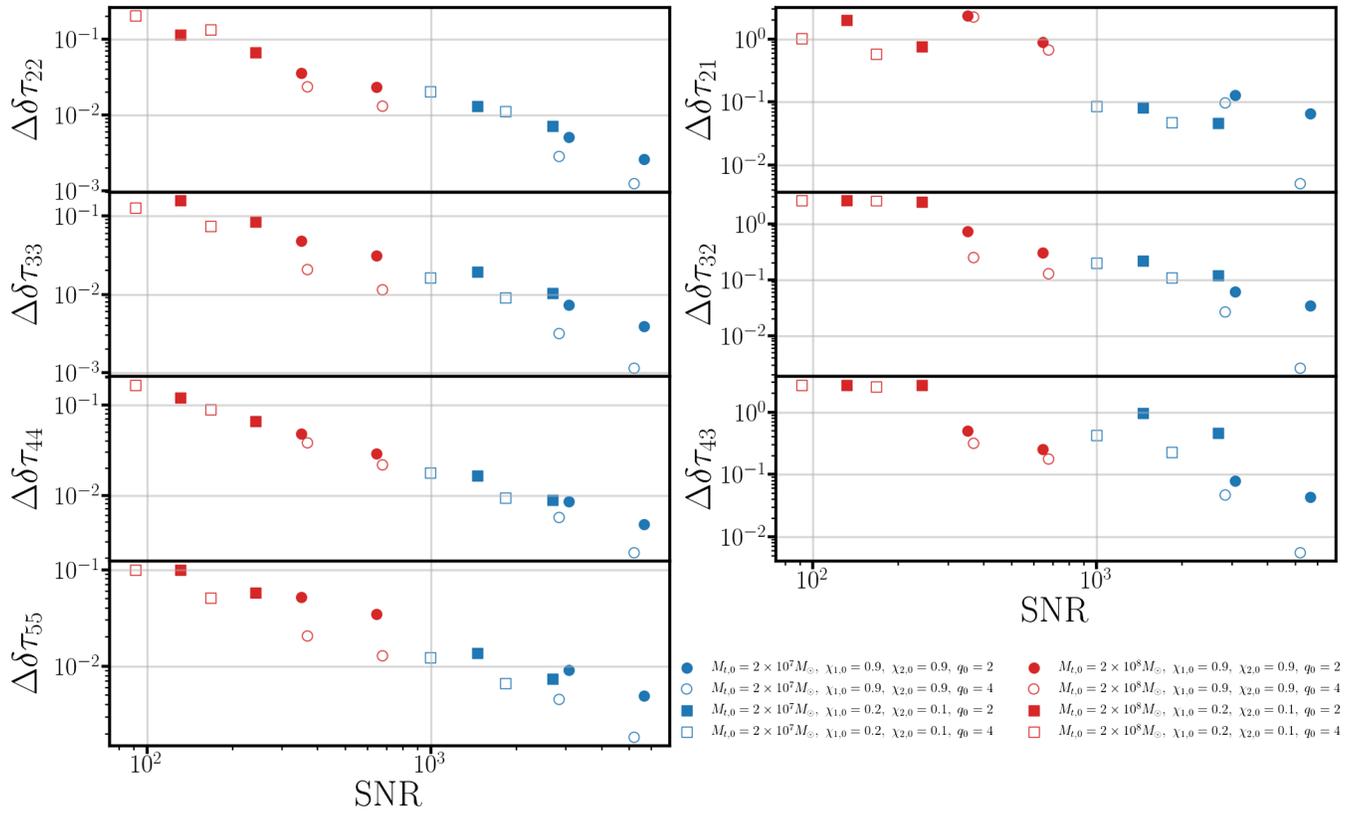


FIG. 6. Comparison between synthetic-injection waveform and the one at a secondary maximum (same color code as in Fig. 5). The late inspiral and merger-ringdown signals match almost perfectly. The early inspiral signals are dephased, but the SNR in the inspiral is so low that it makes little difference.

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 applying the Rayleigh criterion [58] to both the damping time and the frequency to decide on the distinguishability of QNMs is too stringent. Indeed, as can be seen from the upper-left panel in 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]. However, given the little correlation between  $\tau_{\ell m}$  and  $f_{\ell m}$ , it is often enough to apply the Rayleigh criterion only to the damping time or to the frequency, as suggested in previous works [58, 59, 134, 135]. In the cases shown here, the frequency is enough to decide on the distinguishability. The upper-right and lower-left panels 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 subdominant harmonics. Note that the system in the lower-left panel has a total mass of  $2 \times 10^7 M_{\odot}$ , illustrating that the confidence regions of QNMs are not always all well separated for systems with  $M_{t,0} = 2 \times 10^7 M_{\odot}$ , although they do tend to yield better results than for systems with  $M_{t,0} = 2 \times 10^8 M_{\odot}$ , as illustrated in Figs. 7 and 8.

We find some cases of multimodality 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 multimodality can be understood as the subdominant harmonics trying to “match” each other.


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

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

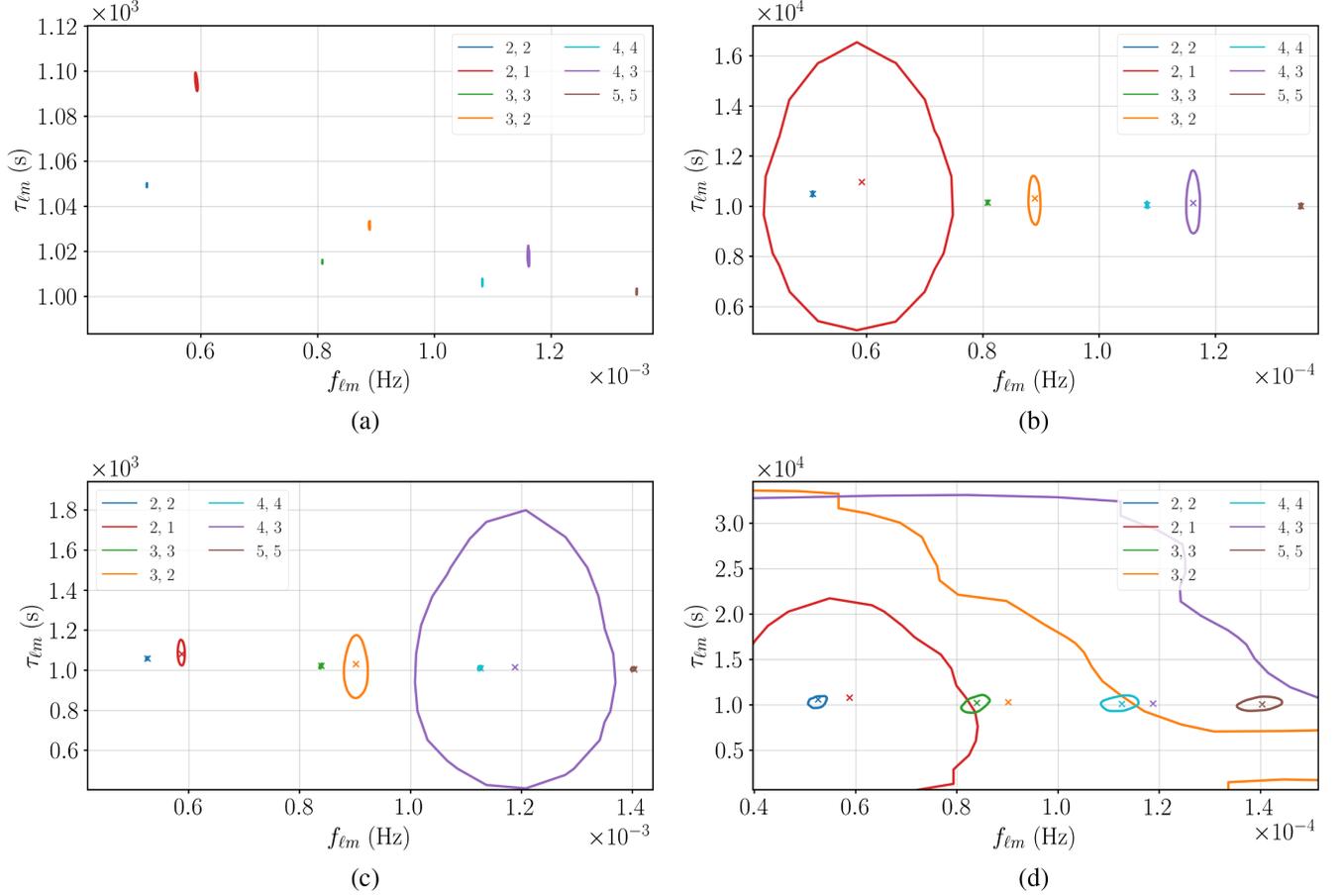


FIG. 9. 90% confidence interval on the QNMs for four different systems. Crosses indicate the true values. Between columns only the total mass varies, and between rows mass ratio, spins, and redshift vary, keeping the same total mass. In every case, the synthetic injection is GR. The top-left panel illustrates why applying the Rayleigh criterion to both the damping time and the frequency to decide on the resolvability of QNMs 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 seven 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). (a)  $M_{t,0} = 2 \times 10^7 M_{\odot}$ ,  $\chi_{1,0} = \chi_{2,0} = 0.9$ ,  $q_0 = 4$ ,  $z_0 = 2.2$ , SNR = 5174. (b)  $M_{t,0} = 2 \times 10^8 M_{\odot}$ ,  $\chi_{1,0} = \chi_{2,0} = 0.9$ ,  $q_0 = 4$ ,  $z_0 = 2.2$ , SNR = 672. (c)  $M_{t,0} = 2 \times 10^7 M_{\odot}$ ,  $\chi_{1,0} = 0.2$ ,  $\chi_{2,0} = 0.1$ ,  $q_0 = 2$ ,  $z_0 = 3.7$ , SNR = 1457. (d)  $M_{t,0} = 2 \times 10^8 M_{\odot}$ ,  $\chi_{1,0} = 0.2$ ,  $\chi_{2,0} = 0.1$ ,  $q_0 = 2$ ,  $z_0 = 3.7$ , SNR = 132.

## 2. Non-GR injections

We now consider non-GR injections, and we generate synthetic signals with nonzero 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., nonspinning BHs), they were computed in theories such as Einstein-Maxwell-dilaton [136], dynamical Chern-Simons gravity [42], Einstein-dilaton-Gauss-Bonnet gravity [40,41,137], higher-curvature gravity theories [53], and for some solutions in massive (bi)gravity [138–140]. 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-Simons theory [54]. Estimates for QNMs of spinning BHs in non-GR theories have also used the connection between the light ring and QNMs [41,141–143], which is formally valid only in the eikonal  $\ell \rightarrow \infty$  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, for both frequencies and damping times. In Figs. 11 and 12, we show the width of the 90% confidence interval centered 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 in Figs. 11 and 12. For this value of the

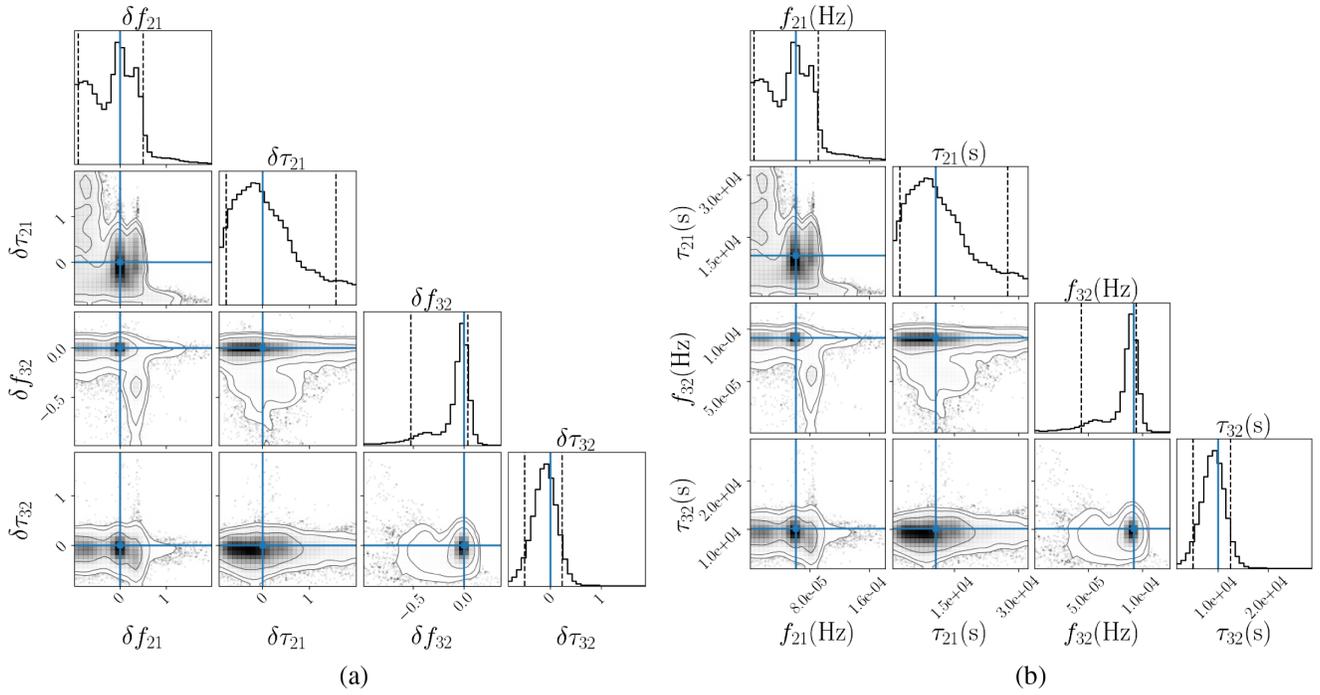


FIG. 10. Reduced corner plot of deviation parameters and QNMs for a GR injection of the system  $M_{t,0} = 2 \times 10^8$ ,  $\chi_{1,0} = \chi_{2,0} = 0.9$ ,  $q_0 = 2$ ,  $z_0 = 3.7$ , and  $\text{SNR} = 339$ . We find a multimodality in the  $\delta f_{21}$  posterior, which is the less-well-measured 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. (a) Deviation parameters. (b) QNMs.

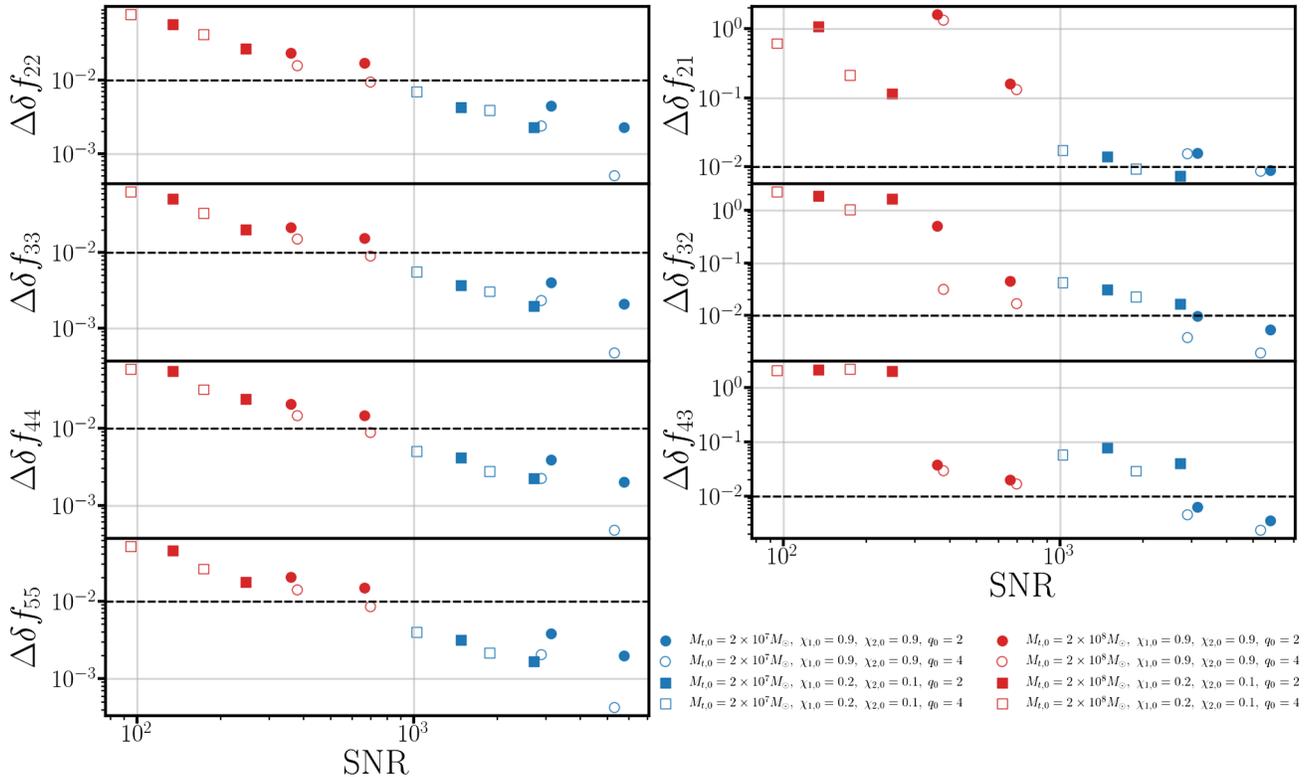


FIG. 11. Width of the 90% confidence interval centered around the median for  $\delta f_{\ell m}$  for injections with  $\delta f_{\ell m} = \delta \tau_{\ell m} = 0.01$ , as indicated by the black dashed lines.

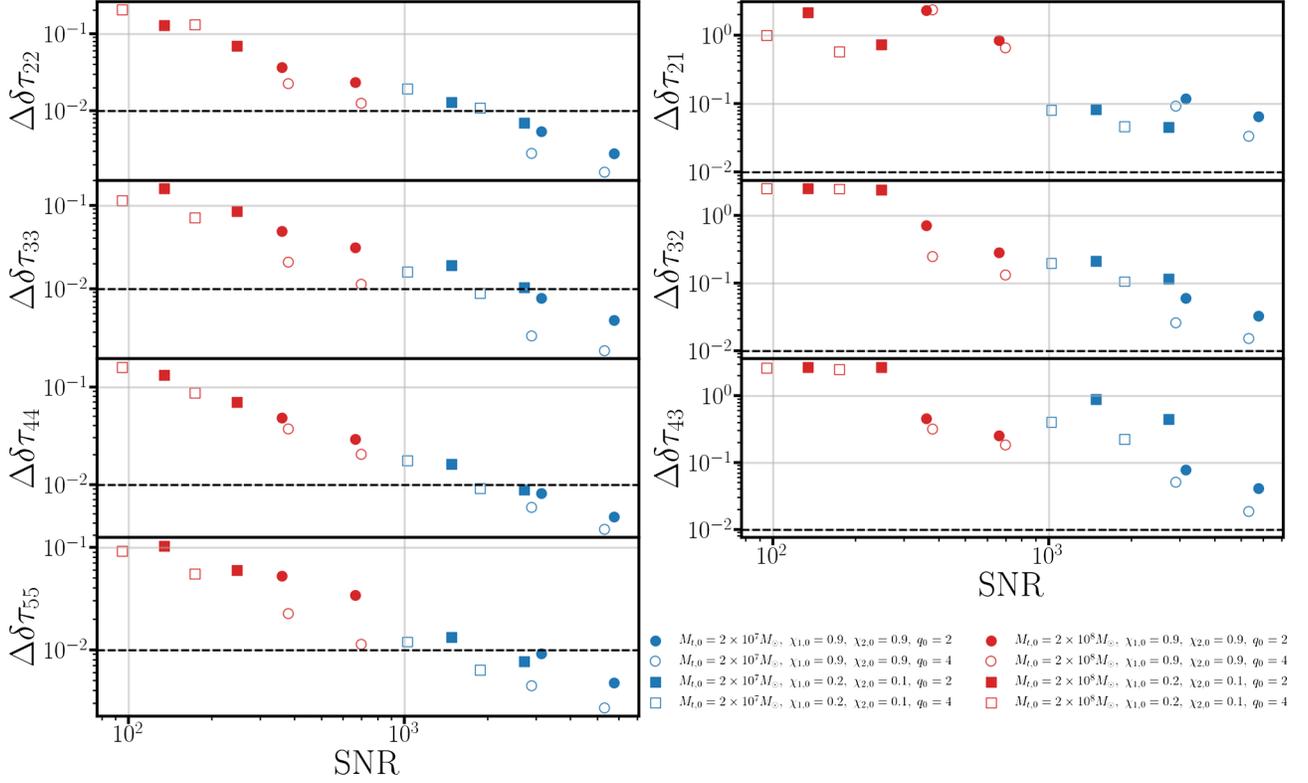


FIG. 12. Width of the 90% confidence interval centred around the median for  $\delta\tau_{\ell m}$  for injections with  $\delta f_{\ell m} = \delta\tau_{\ell m} = 0.01$ , as indicated by the black dashed lines.

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, in both the frequency and the damping time of the dominant harmonics (left column), and for the higher SNR systems, even of the subdominant harmonics (right column). Detecting a deviation in several harmonics, preferably in both the frequency and the damping time, would reinforce our confidence that we are truly observing effects in gravity theories alternative to GR.

## V. IMPACT OF SYSTEMATICS

In order to assess if the lack of accuracy in waveform modeling 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 [144] at the highest available resolution. It provides the signal of a BBH with mass ratio 2 and aligned spins of magnitude 0.3. First, we perform injections for the four total mass and redshift combinations detailed in Sec. III, allowing for deviations from GR in the QNMs when running our Bayesian analysis. Next, we investigate more methodically how systematic effects come into play and propose a novel method to assess their impact.

### A. Results on astrophysical systems

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, giving  $\text{SNR} = 175$  and  $1312$ , 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. We recall that our model does not explicitly include overtones beyond  $n = 0$ , as discussed in Sec. II. Including them in the merger-ringdown signal could, in principle, reduce this systematic bias but would

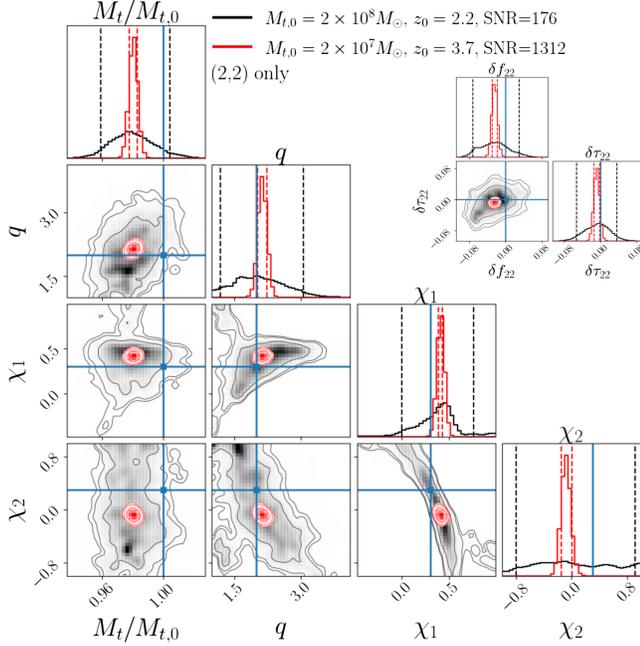


FIG. 13. Corner plot of intrinsic and deviation parameters for NR injections recovered with pSEOBNRv5HM templates. Contours show the 68%, 90%, and 95% confidence intervals, and dashed lines the 0.05 and 0.95 quantiles. We compare the results for  $M_{t,0} = 2 \times 10^8 M_\odot$ ,  $z_0 = 2.2$  (black) and  $M_{t,0} = 2 \times 10^7 M_\odot$ ,  $z_0 = 3.7$  (red); both have  $q_0 = 2$  and  $\chi_{1,0} = \chi_{2,0} = 0.3$ . 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 mismodeling 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.

require fitting for the amplitude, phase, and starting time of each overtone, which has proven very difficult so far [145], and could end up having the opposite effect on the accuracy of the waveform. 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 incompatible with the true parameters at more than 95% confidence, in particular, for the GR deviation ones. 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 demonstrate in Sec. V B. This is not surprising, since the accuracy of the SEOBNRv5 model degrades when including higher harmonics, as comprehensive comparisons to  $\sim 440$  NR waveforms and NR surrogate models have

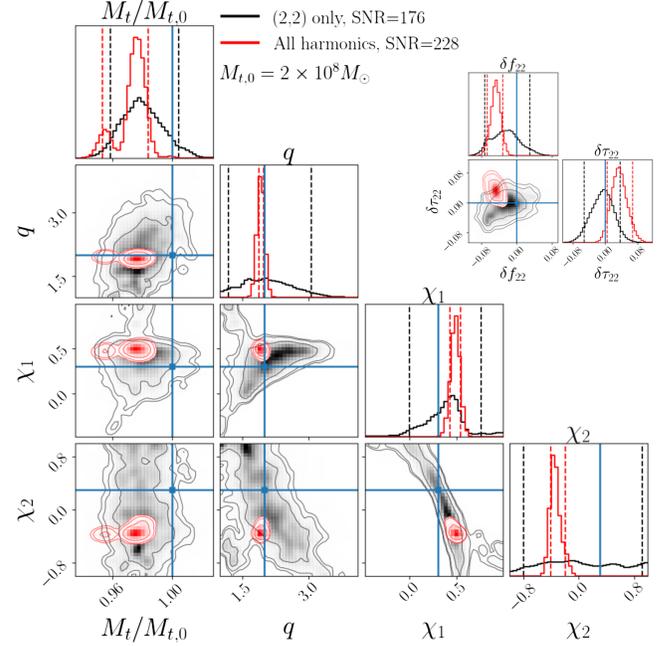


FIG. 14. The same as Fig. 13, but now comparing results when including only the (2, 2) harmonic (black) versus when including all harmonics of pSEOBNRv5HM (red), for  $M_{t,0} = 2 \times 10^8 M_\odot$  and  $z_0 = 2.2$ . The inclusion of higher harmonics worsens the match between NR and pSEOBNRv5HM waveforms, 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.

shown [81]. One of the difficulties lies in the relative alignment between harmonics. When considering harmonics individually, we have freedom in the alignment (in phase and time) of the waveforms. When including several harmonics, we have a single phase shift and a single time shift that can be varied for all harmonics simultaneously. The relative alignments between them are fixed and might not agree with the ones of NR waveforms. In the next section, we further illustrate how such tests of GR become less reliable when including higher harmonics.

## B. Exploring systematic effects

Our goal here is to understand how systematic effects come into play and, in particular, how they depend on which portion of the signal (inspiral or merger-ringdown) dominates. In addition to total masses of  $2 \times 10^7 M_\odot$  and  $2 \times 10^8 M_\odot$ , we also consider  $M_t = 2 \times 10^6 M_\odot$ . The minimum frequency used to analyze the signal remains  $f_{\min} = 2 \times 10^{-4} [M_{t,0} / (2 \times 10^7 M_\odot)]$  Hz. This is a pessimistic choice for signals with  $M_t = 2 \times 10^6 M_\odot$ , as they can accumulate significant SNR in the early inspiral, but with this choice, the length of the signal that is analyzed (in geometric units) is kept fixed for different total masses.

TABLE II. Full IMR SNR and its decomposition into the contribution of the merger-ringdown and inspiral stages for the NR waveform SXS:BBH:2125 [144], assuming a total SNR of 175. We show results both for when only the (2, 2) mode is considered and when all harmonics modeled in SEOBNRv5HM are included.

IMR	$M_{t,0} = 2 \times 10^6 M_\odot$		$M_{t,0} = 2 \times 10^7 M_\odot$		$M_{t,0} = 2 \times 10^8 M_\odot$	
	Merger-ringdown	Inspiral	Merger-ringdown	Inspiral	Merger-ringdown	Inspiral
(2, 2) only	175	102	161	68	168	49
All harmonics	175	106	165	58	171	35

Moreover, we fix the SNR of the sources rather than their distance, to allow for a fair comparison between systems. Unless specified, we take the SNR to be 175 [as for the system with  $M_t = 2 \times 10^8 M_\odot$  at  $z = 2.2$  when including only the (2, 2) harmonic shown in Sec. V]. We stress that the long-wavelength approximation for the LISA response is expected to break for  $M_t = 2 \times 10^6 M_\odot$ . Overall, the signals considered in this section might not be astrophysically realistic (e.g., we expect systems with  $M_t = 2 \times 10^7 M_\odot$  to have much larger SNR than 175), but we seek to have a methodic understanding of how systematic effects appear for different signal morphologies. We start by considering injections with the (2, 2) harmonic only and then with all other harmonics included in SEOBNRv5HM. In both cases, the injected signal and templates have the same harmonic content. In Table II, we give the merger-ringdown and inspiral SNR of the systems we consider, assuming an IMR SNR of 175, for both harmonic content options.

### 1. (2, 2) only

In Fig. 15, we compare the posterior distributions obtained for different total masses. Note that, unlike in previous plots, we show the chirp mass, defined as  $\mathcal{M}_c = (m_1^3 m_2^3 / (m_1 + m_2))^{1/5}$ , rather than the total mass. This is the mass combination that is best measured during the inspiral. We find that, as the total mass decreases, the chirp mass and the mass ratio are better measured and in better agreement with the true value. This is because the fraction of the SNR in the inspiral is larger for lighter systems, and this is the portion of the signal where our templates are in better agreement with NR waveforms. In Fig. 16, we compare the adimensionalized TDI strain  $A$  for the highest likelihood point in each case. The upper panel focuses on the inspiral, and we can see that the agreement is better for lighter mass systems. The lower panel focuses on the merger-ringdown portion [we recall that  $t/M_t = 0$  is the peak of the (2, 2) amplitude], and we can see that the system with  $M_t = 2 \times 10^8 M_\odot$  is the one in the best agreement with the injection, as expected from the fact that it is the system for which the fraction of SNR in the merger-ringdown is the largest.

Looking at the caption in Fig. 15, it might appear surprising that the system with  $M_t = 2 \times 10^6$  is the one

for which the GR deviation parameters are the largest (while still being very much consistent with 0). The reason behind it is that the determination of intrinsic parameters for this system really comes from the inspiral, and the GR deviation parameters are then estimated to maximize the match in the merger-ringdown portion of the signal. In contrast, the intrinsic parameters of the  $M_t = 2 \times 10^8 M_\odot$  system are chosen to maximize the match in the merger-ringdown. This becomes clearer if we increase the SNR, as in Fig. 17, where we take the IMR SNR to be 707. The posteriors become thinner and the intrinsic parameters in the  $M_t = 2 \times 10^8 M_\odot$  case are no longer compatible with the true value. Moreover, for  $M_t = 2 \times 10^6 M_\odot$ , we definitely favor nonzero deviations from GR. The fact that this is needed to maximize the

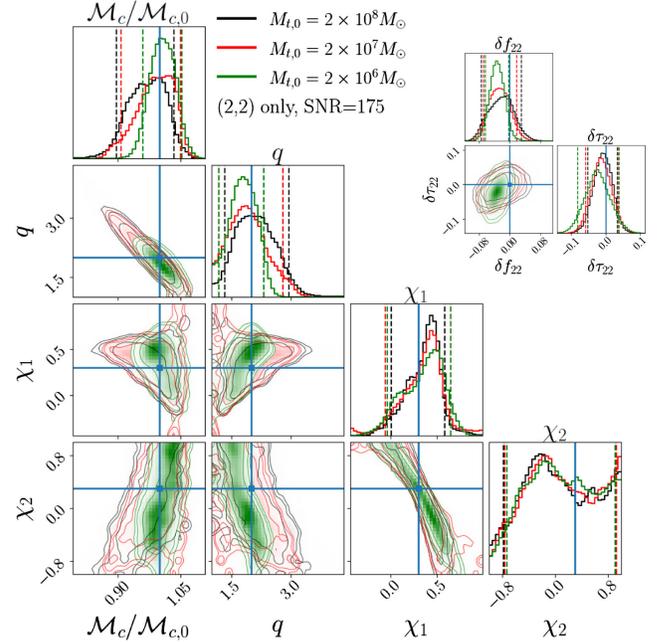


FIG. 15. Corner plot for the parameters estimated using pSEOBNRv5HM to analyze a mock NR injection for different total masses, considering only the (2, 2) harmonic and fixing the SNR of the injection to be 175. Contours show the 68%, 90%, and 95% confidence intervals, and dashed lines the 0.05 and 0.95 quantiles. In the upper-right part, we show the posterior on the fractional deviations to GR QNMs.

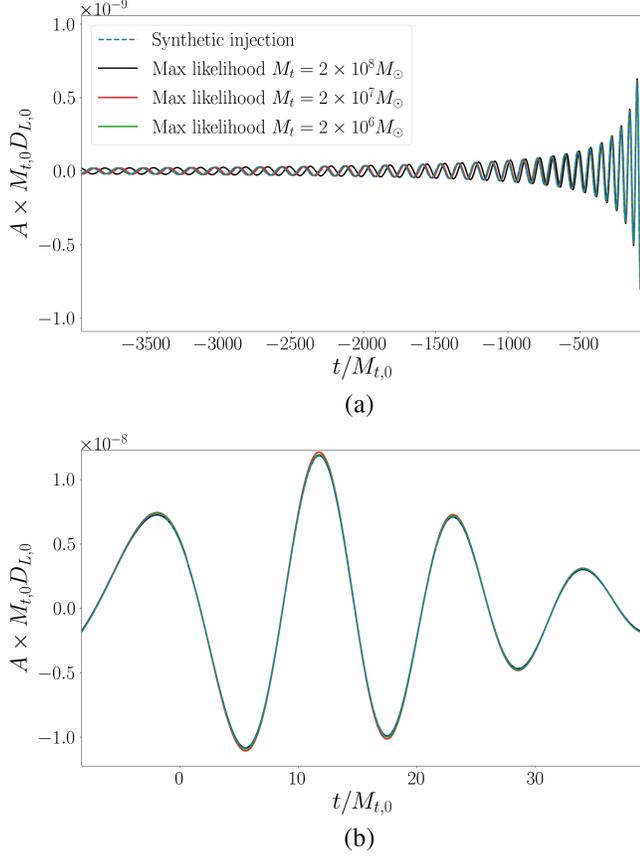


FIG. 16. Comparison between the waveforms of the maximum likelihood points (from the analyses shown in Fig. 15) and the injection waveform. We plot the adimensionalized TDI strain  $A$  as a function of time in geometric units to allow for the comparison between different total masses. (a) Inspiral. (b) Merger-ringdown.

match, even for an SNR of 175, is likely a consequence of the model used for the merger-ringdown [see Eq. (2.3) and below]. When changing the QNMs, the amplitude and phase coefficients in Eqs. (2.6a) and (2.6b) change consistently, possibly providing a better match, in particular, around the merger, where we use a purely phenomenological description, and discrepancies between our templates and NR waveforms are larger. Moreover, our model might not be accounting for higher overtones accurately enough. Therefore, favoring a nonzero value for the GR deviation parameters does not necessarily reflect that the physical QNMs are different from the Kerr ones but rather that, given the functional form assumed, such QNM values provide overall a better fit to data. It would be interesting to explore if we still find such deviations from GR when allowing for the additional modifications around the merger proposed in [67].

Finally, in Fig. 18, we show the distribution of log-likelihood values obtained from running parameter estimation allowing or not for deviations from GR

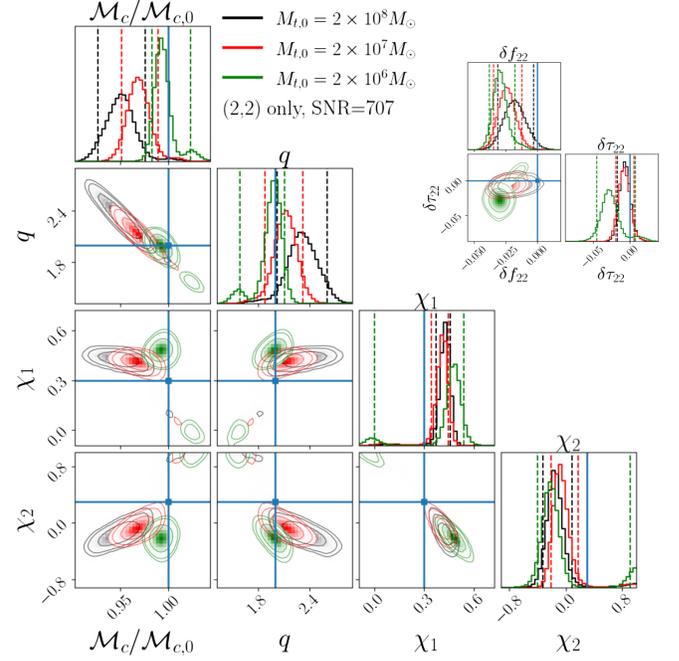


FIG. 17. The same as Fig. 15 for an IMR SNR of 707.

(pSEOBNRv5HM versus SEOBNRv5HM) for the three values of the total mass (fixing the SNR to 175). We find that, as the total mass decreases, the likelihood values increase, and allowing for deviations from GR improves more the fit, in agreement with Fig. 15. We can use the results shown in this plot to estimate the SNR from which it will be favored to allow for deviations from GR. For this purpose, we introduce the Akaike information criterion (AIC) [146], defined as

$$\text{AIC} = 2n_p - 2 \ln \hat{\mathcal{L}}, \quad (5.1)$$

where  $\hat{\mathcal{L}}$  is the maximum likelihood and  $n_p$  is the number of free parameters. The latter accounts for the

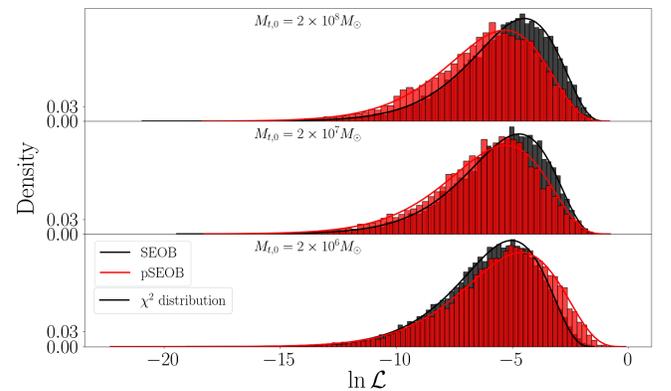


FIG. 18. Distribution of log-likelihood values when analyzing with SEOBNRv5HM and pSEOBNRv5HM for different total masses.

dimensionality penalty. When choosing between different models to describe observed data, the one with minimum AIC is favored. Moreover, we can estimate the log-Bayes factor between model 1 and model 2 as

$$\ln \mathcal{B} = -\frac{1}{2}(\text{AIC}_1 - \text{AIC}_2). \quad (5.2)$$

$$\ln \mathcal{L} = \sum_{c \in [A,E]} (d_c | s_c(\theta)) - \frac{1}{2}(d_c | d_c) - \frac{1}{2}(s_c(\theta) | s_c(\theta)) \quad (5.3)$$

$$= \sum_{c \in [A,E]} (d_c | d_c) \sqrt{\frac{(s_c(\theta) | s_c(\theta))}{(d_c | d_c)}} \left( \frac{(d_c | s_c(\theta))}{\sqrt{(d_c | d_c)(s_c(\theta) | s_c(\theta))}} - \frac{1}{2} \sqrt{\frac{(d_c | d_c)}{(s_c(\theta) | s_c(\theta))}} - \frac{1}{2} \sqrt{\frac{(s_c(\theta) | s_c(\theta))}{(d_c | d_c)}} \right). \quad (5.4)$$

For a zero-noise injection,  $d_c = s_{0,c}$ , and  $(d_c | d_c) = \text{SNR}_c^2$  is the SNR of the signal in the TDI channel  $c$ . If we vary the loudness of the injection by changing the distance (as we do here) and rescale the template by the same factor, which corresponds to rescaling the distance, then  $(d_c | d_c)$  is the only term that depends on the SNR. Thus, we get the simple scaling for the log-likelihood:

$$\ln \mathcal{L} = \ln \mathcal{L}_{175} \left( \frac{\text{SNR}}{175} \right)^2. \quad (5.5)$$

Note that this scaling does not rely on approximating the likelihood as a Gaussian on the event parameters. Combining Eqs. (5.1), (5.2), and (5.5), we find that the log-Bayes factor for pSEOBNRv5HM versus SEOBNRv5HM is

$$\ln \mathcal{B} = (\ln \hat{\mathcal{L}}_{\text{pSEOBNRv5HM},175} - \ln \hat{\mathcal{L}}_{\text{SEOBNRv5HM},175}) \left( \frac{\text{SNR}}{175} \right)^2 - (n_{p,\text{pSEOBNRv5HM}} - n_{p,\text{SEOBNRv5HM}}). \quad (5.6)$$

In general, it can be difficult to accurately estimate  $\ln \hat{\mathcal{L}}$  from the MCMC alone, because it lies in the higher-end tail of the log-likelihood distribution (see Fig. 18). It becomes harder as the dimensionality of the parameter space increases. To get a better estimate of the maximum likelihood, we start from the Gaussian approximation. Under this hypothesis,

$$\ln \mathcal{L} = \ln \hat{\mathcal{L}} - \frac{1}{2} \theta' C^{-1} \theta \quad (5.7)$$

$$= \ln \hat{\mathcal{L}} - \frac{1}{2} \sum_i^{n_p} \frac{\theta_i^2}{\sigma_i^2}, \quad (5.8)$$

In [147], we justify the use of the expression above in the case of Gaussian posteriors and discuss how it relates to the standard indistinguishability criterion [148–152]. Here,  $n_p$  is 9 when using SEOBNRv5HM and 11 when using pSEOBNRv5HM. From the definition of the likelihood [Eq. (2.14)], we can write the log-likelihood (up to an additional constant that depends only on the noise properties of the detector) as

where  $\theta'$  are the coordinates of  $\theta$  in the basis of eigenvectors of the covariance matrix  $C$  and  $\sigma_i^2$  are the eigenvalues of  $C$ . From Eq. (5.8), we can see that  $2(\ln \hat{\mathcal{L}} - \ln \mathcal{L})$  follows a  $\chi^2$  distribution with  $n_p$  degrees of freedom. Since the mean of a  $\chi^2$  distribution with  $n_p$  degrees of freedom is  $n_p$ , denoting the mean by  $\langle \cdot \rangle$ , we have

$$\ln \hat{\mathcal{L}} = \langle \ln \mathcal{L} \rangle + \frac{n_p}{2}. \quad (5.9)$$

The mean log-likelihood is less dependent on sampling the tails of the distribution than the maximum log-likelihood. Thus, we can use the log-likelihood samples we get from MCMC to estimate it and then use Eq. (5.9) to estimate  $\hat{\mathcal{L}}$ . We recall that our prior on  $\theta$  is flat, so the likelihood values obtained with MCMC are fair draws of the distribution followed by the log-likelihood. In Fig. 18, we overplot in full lines the probability density function of the theoretical distribution of log-likelihood values assuming that  $\ln \mathcal{L} \sim \ln \hat{\mathcal{L}} - \frac{1}{2} \chi^2(n_p)$ , and  $\hat{\mathcal{L}}$  was estimated with the procedure described above. We can see that the agreement between this prediction and the distribution we obtain with MCMC is remarkable, even though the likelihood is not actually Gaussian in the event parameters. Moreover, we have verified that the integrated weight of the theoretical probability density function above the maximum likelihood we find with our MCMC is typically below  $1/N_s$ , where  $N_s$  is the number of MCMC samples. This indicates that our sampling of the log-likelihood function is compatible with its theoretical estimate.

Using this method to estimate  $\hat{\mathcal{L}}$ , we find that, for  $\text{SNR} = 175$ , we have  $\ln \mathcal{B} < 0$  for all three total masses. This is in agreement with Fig. 15, where the posteriors are compatible with GR at least at  $\sim 68\%$  confidence. We now estimate the SNR for which pSEOBNRv5HM would

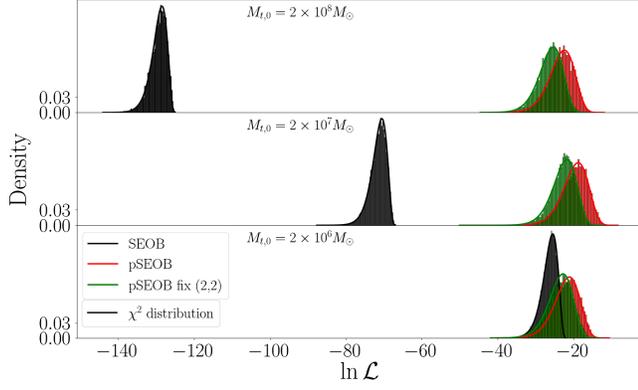


FIG. 19. The same as Fig. 18, now including all harmonics and considering an additional case where  $\delta f_{22}$  and  $\delta \tau_{22}$  are fixed to 0.

be definitely favored with respect to SEOBv5HM. Following the Kass-Raftery scale [153], we adopt the criterion  $\ln \mathcal{B} > 3$  to estimate that a model is favored with respect to another. For  $M_t = 2 \times 10^6 M_\odot$ , we estimate  $\text{SNR} \gtrsim 330$ , and, for  $M_t = 2 \times 10^7 M_\odot$  and  $M_t = 2 \times 10^8 M_\odot$ ,  $\text{SNR} \gtrsim 598$  and  $\text{SNR} \gtrsim 977$ , respectively. These values are in agreement with Fig. 17, where having a zero value for the GR modifications is supported for  $M_t = 2 \times 10^8 M_\odot$ , but it is not in the other cases.

## 2. All harmonics

In Fig. 19, we show the equivalent of Fig. 18 when including all harmonics, fixing the total IMR SNR to 175. We consider one extra case: using for the template the pSEOBv5HM model with the (2, 2) QNM fixed to its Kerr value (i.e.,  $\delta f_{22} = \delta \tau_{22} = 0$ ). The motivation for this is the following: The rationale for this test of GR is that the parameters of the binary are measured from the inspiral; from there, we can estimate the final mass and spin and, therefore, the QNM spectrum. We then seek to measure deviations from this spectrum. However, for heavier systems, most of the information comes from the merger-ringdown and the measurements are no longer “independent.” Thus, we want to investigate how the parameter estimation changes when not allowing for deviations in the dominant mode.

First, we note that the likelihood values we get when including all harmonics are lower than when including only the (2, 2) mode, despite the significant increase in the number of free parameters in the pSEOBv5HM case (23 versus 11), showing that the agreement with the injected NR waveform worsens. Focusing on the SEOBv5HM case first (in black), we find that the hierarchy of likelihood values between the different total masses is compatible with the values reported in Table II: The best fit is for the  $M_t = 2 \times 10^6 M_\odot$  system, because most of its SNR comes from the inspiral, where our templates are more reliable. In the pSEOBv5HM case, we find that the quality of the fit improves

significantly, with a larger improvement for more massive systems. Surprisingly, we find higher likelihood values for the  $M_t = 2 \times 10^7 M_\odot$  system than for the  $M_t = 2 \times 10^6 M_\odot$  one. This happens because, in the pSEOBv5HM case, the additional parameters allow a substantial improvement in the match in the merger-ringdown portion of the signal but not in the inspiral. To better understand this, we start from Eq. (5.4) and make the assumption that  $(d_c | d_c) = (s_{0,c} | s_{0,c}) \simeq (s_c(\hat{\theta}) | s_c(\hat{\theta}))$ , where  $\hat{\theta}$  is the maximum likelihood point. In other words, we assume that the loudness of the recovered signal is virtually the same as the one of the injection. If we further assume that the contribution from each channel is roughly the same, we can write

$$\ln \hat{\mathcal{L}} = -2\text{SNR}^2(1 - \text{FF}(s_0, h)), \quad (5.10)$$

where FF is the fitting factor between the true signal and our templates defined as the maximized overlap:

$$\text{FF}(s_0, h) = \max_\theta \mathcal{O}(s_0 | h(\theta)), \quad (5.11)$$

$$\mathcal{O}(h_1, h_2) = \frac{(h_1 | h_2)}{\sqrt{(h_1 | h_1)(h_2 | h_2)}}. \quad (5.12)$$

In the previous equations, we can split between the contribution coming from the inspiral (I) and the one coming from the merger-ringdown (MRD):

$$\ln \hat{\mathcal{L}} = -2((\text{SNR}^2(1 - \text{FF}))_I + (\text{SNR}^2(1 - \text{FF}))_{\text{MRD}}). \quad (5.13)$$

From Table II, we see that the fraction of SNR coming from the merger-ringdown for the  $M_t = 2 \times 10^7 M_\odot$  system is larger than the one coming from the inspiral for the  $M_t = 2 \times 10^6 M_\odot$  one. Thus, if introducing the QNM deviations allows a sufficient improvement in the match in the merger-ringdown for the  $M_t = 2 \times 10^7 M_\odot$  system, we can obtain higher likelihood values for it than for the  $M_t = 2 \times 10^6 M_\odot$  system, unlike in the SEOBv5HM case. As in the case of the (2, 2) harmonic only, we can estimate the SNRs for which pSEOBv5HM is favored with respect to SEOBv5HM, now including all harmonics. Note that now  $n_p = 23$  for pSEOBv5HM. We find  $\text{SNR} \gtrsim 68$ ,  $\text{SNR} \gtrsim 93$ , and  $\text{SNR} \gtrsim 214$  for  $M_t = 2 \times 10^8 M_\odot$ ,  $M_t = 2 \times 10^7 M_\odot$ , and  $M_t = 2 \times 10^6 M_\odot$ , respectively. We stress that these values are below the typical SNRs we expect for MBHBs, in particular, for sources with  $M_t = 2 \times 10^6 M_\odot$  and  $M_t = 2 \times 10^7 M_\odot$ . Systems with  $M_t = 2 \times 10^8 M_\odot$  could also have SNR above the respective limit, as is the case for the system shown in Fig. 14, for which GR is indeed excluded at more than 95% confidence. Moreover, our estimates are in

agreement with the corner plots shown in Figs. 25–27 in Appendix D, where we show the posterior for the different total masses in all three scenarios. For an IMR SNR of 175, only the lighter system is compatible with GR in the pSEOBNRv5HM case. 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. This result is in qualitative agreement with our estimate of  $\text{SNR} \gtrsim 214$  for pSEOBNRv5HM to be favored with respect to SEOBNRv5HM for the  $M_t = 2 \times 10^6 M_\odot$  system, which is the closest one among our systems (in terms of relative contribution of the inspiral and merger-ringdown) to the system considered in Ref. [62].

From Figs. 25–27, it is possible to see that the posteriors in the pSEOBNRv5HM with fixed (2,2) QNM case are slices of the pSEOBNRv5HM posteriors on the  $\delta f_{22} = \delta \tau_{22} = 0$  hypersurfaces, as they should be. It is harder to see that the posteriors in the SEOBNRv5HM correspond indeed to slices of the pSEOBNRv5HM posteriors on the  $\delta f_{\ell m} = \delta \tau_{\ell m} = 0$  hypersurfaces, because this slice of the posterior yields much lower likelihood values than explored by the sampler in the pSEOBNRv5HM case, as can be seen from Fig. 19, except in the  $M_t = 2 \times 10^6 M_\odot$  case (Fig. 27). We find that the chirp mass and the mass ratio tend to be in “better” agreement with their true values in the pSEOBNRv5HM fixed (2,2) case than in the full pSEOBNRv5HM one, but we have no conclusive evidence of this. The spins are usually wrongly measured, with only a secondary mode in the SEOBNRv5HM case for  $M_t = 2 \times 10^6 M_\odot$  containing the true values. This suggests that the inclusion of spins in our waveforms is one of the bottlenecks for performing accurate parameter estimation.

It might seem surprising that the likelihood values in the pSEOBNRv5HM with fixed (2, 2) case are only a little lower than in the full case, as illustrated in Fig. 19, even more given that (2, 2) is the dominant mode. A likely explanation for this is as follows. Deviations in the QNMs can partially compensate for the misalignment in harmonics between NR and pSEOBNRv5HM waveforms, as they change the amplitude and phase of the different harmonics, and improve the match, reinforcing our claim in the discussion around Fig. 15: Measuring nonzero GR deviations does not necessarily mean that the physical QNMs are different from their predicted Kerr values. However, the phase shift and the time shift that we fit when doing parameter estimation are defined with respect to the peak of the (2, 2) amplitude. Therefore, it is less crucial to allow for deviations in this harmonic. It would be interesting to investigate how the match worsens when fixing other QNMs, in particular, the next-to-dominant (3, 3), or when fixing several but not all of them. We leave this for future investigation.

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 paid to the modeling of higher harmonics, as their inclusion increases biases, and of spins, as their measurement tends to be very biased. Let us stress that even the NR waveforms we currently use are not accurate enough for SNRs of thousands [81,152] 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 modeling 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 analyzed them with templates in GR (SEOBNRv5HM) and with parametrized deviations from GR (pSEOBNRv5HM), using the newly released SEOBNRv5HM waveform family [79–82,154]. 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 for both the synthetic injection and the Bayesian analysis. We find that having a consistent modeling 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 of  $10^7 M_\odot$ . 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 seven included in the pSEOBNRv5HM model in the most favorable cases (i.e., total mass  $10^7 M_\odot$ , mass ratio  $\sim 2$ –4, highly spinning, and at  $z \sim 2$ ).

Then, we have assessed the impact of systematics on ringdown tests by using NR waveforms to perform synthetic injections and analyzing them with pSEOBNRv5HM.

In order to estimate at which SNRs we would expect to erroneously measure deviations from GR, we have developed a novel approach based on the Akaike information criterion. We have found that, in particular, when higher harmonics are included, parameter estimation is significantly biased already for SNRs of  $\mathcal{O}(100)$ , 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 jeopardize 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 [90]. 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 [86]. 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 [155–157], typically by modifying the post-Newtonian expansion of the GW phase [158], 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 [160] codes and getting access to NR waveforms, as well as Danny Laghi and Vasco Gennari for comments on the Bayes’ factor scale. We express our gratitude to the anonymous referee for their valuable comments on our manuscript. 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 [159]. 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 neighboring points, but for high SNRs these oscillations become important. This is illustrated in Fig. 20, where we plot the log-likelihood as a function of  $\chi_2$  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  $\sim -15$ . 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  $\sim 5$  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

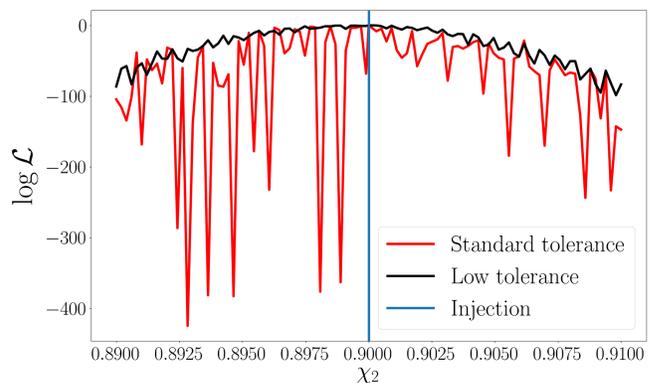


FIG. 20. Log-likelihood as a function of  $\chi_2$  when using two different tolerances, when integrating the EOB equations of motion, for a system with  $M_{r,0} = 2 \times 10^8 M_\odot$ ,  $\chi_{1,0} = \chi_{2,0} = 0.9$ ,  $q_0 = 2$ ,  $z_0 = 2.2$ , and  $\text{SNR} = 623$ . When using waveforms computed with standard tolerance, the likelihood can be a very nonsmooth function of the binary parameters, inducing secondary maxima and yielding very small values even close to the injected value.

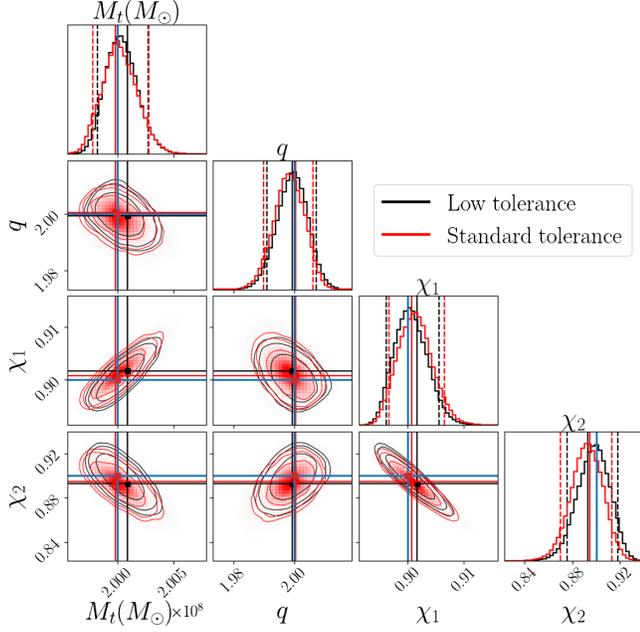


FIG. 21. Comparison between parameter estimation results when using low-tolerance waveforms (black) and standard ones (red) for the same system as in Fig. 20. Blue lines indicate the injection point; black and red symbols indicate the maximum-posterior point of the low-tolerance and standard-tolerance runs, respectively. The nonsmoothness of the likelihood function illustrated in Fig. 20 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.

pronounced. However, as we show in Fig. 21, 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. VC in Ref. [113] in the context of sky localization with LISA. For the system shown in Fig. 21, 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.

For some systems, this apparent bias persists even when using low-tolerance waveforms, as shown in Fig. 22. In order to further validate the argument that this is caused by the nonsmoothness of the waveform across parameter space, we compare to the results obtained using the IMRPhenomTHM waveform model [91]. This is an IMR time-domain approximant, built from a

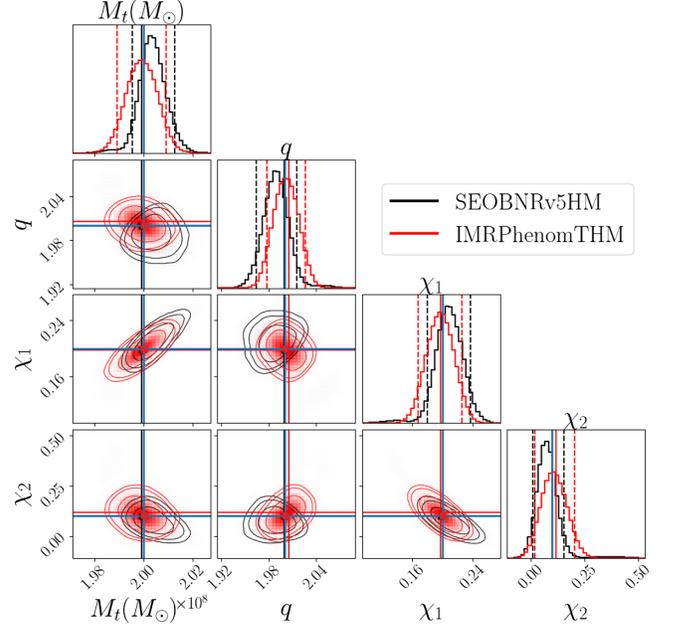


FIG. 22. Comparison of the parameter estimation results obtained using low-tolerance SEOBNRv5HM waveforms (black) and IMRPhenomTHM (red) waveforms for a system with  $M_{t,0} = 2 \times 10^8 M_\odot$ ,  $\chi_{1,0} = 0.2$ ,  $\chi_{2,0} = 0.1$ ,  $q_0 = 2$ ,  $z_0 = 2.2$ , and  $\text{SNR} = 241$ . The latter shows no apparent bias when looking at one-dimensional marginalized 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.

phenomenological approach, in the spirit of the frequency-domain approximants of the phenomenological family of templates [92–94,161–163]. It is based on post-Newtonian expressions [158] 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 subdominant (3,2) and (4,3) harmonics. For nonprecessing 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. 22. 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.

## 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, for both the synthetic injection and parameter estimation. We restrict ourselves to the GR case. In Fig. 23, we show how

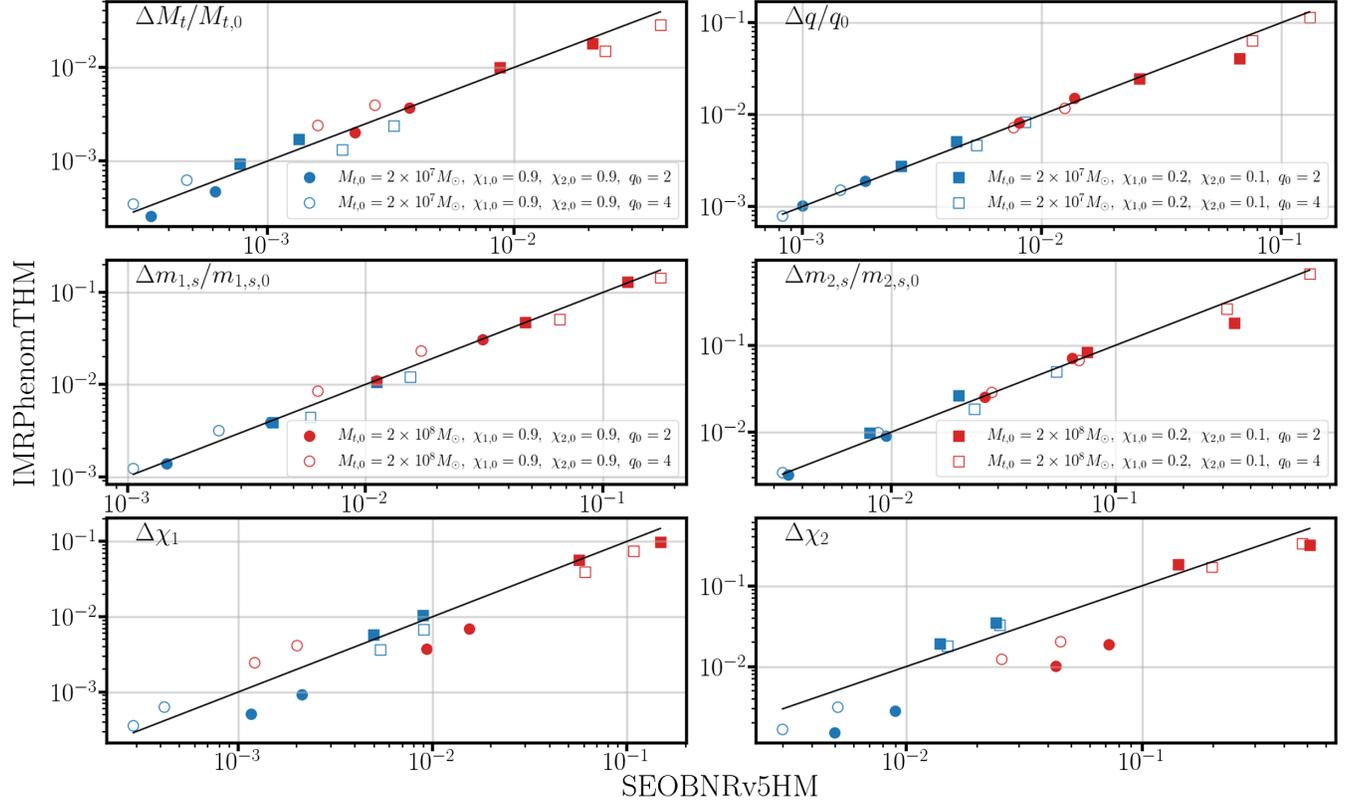


FIG. 23. Comparison between the measurement errors with SEOBNRv5HM ( $x$  axes) and with IMRPhenomTHM ( $y$  axes). Black lines show  $y = x$ . Errors are similar, with a slightly higher discrepancy for the spins, in particular, for high-spin systems (circles).

the measurement errors on intrinsic parameters compare when using the IMRPhenomTHM and the SEOBNRv5HM waveforms. Black lines correspond 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. 24 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 in 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.

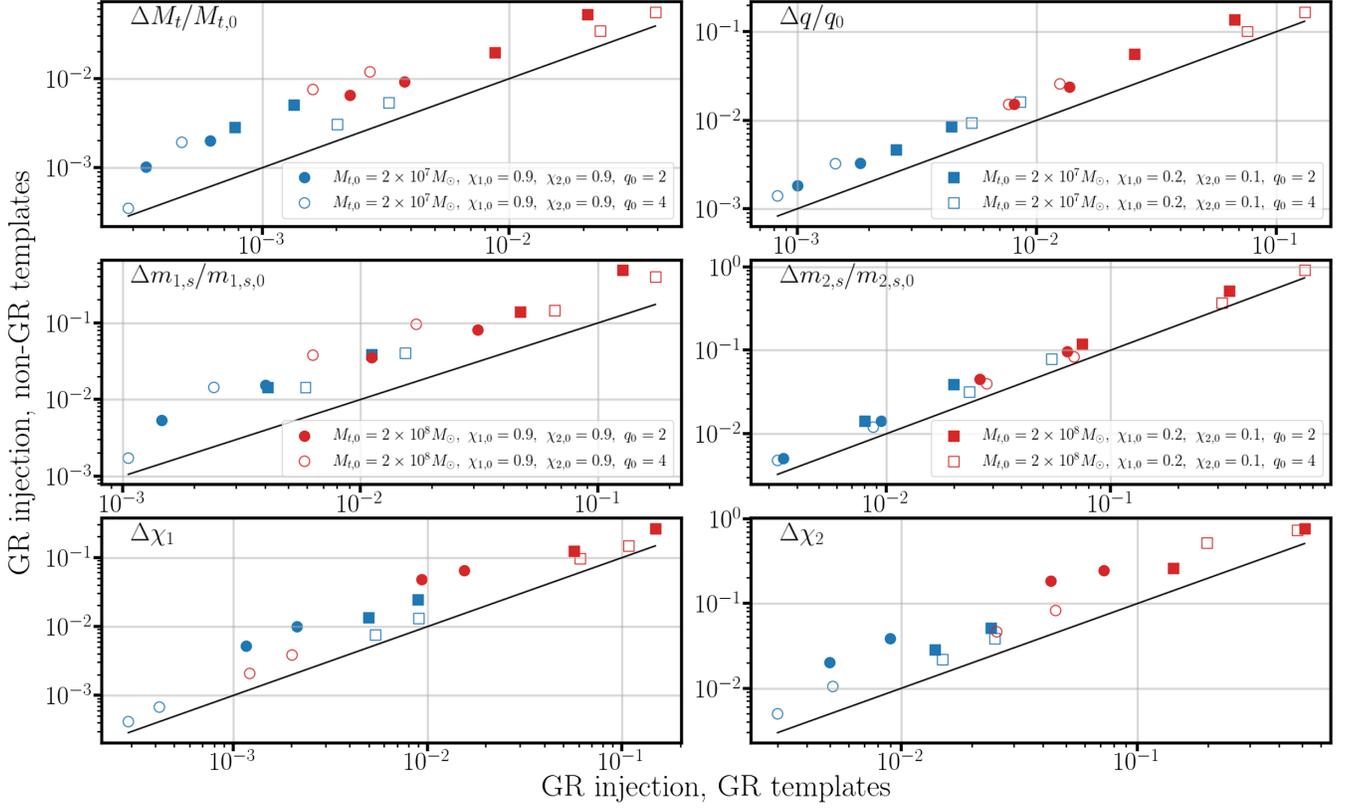


FIG. 24. Comparison of the measurement error on intrinsic parameters when setting deviation parameters to 0 ( $x$  axes) and when letting them vary ( $y$  axes) for a GR injection. Black lines show  $y = x$ . Measurements worsen when allowing for deviations from GR to vary but remain comparable to the GR case.

#### APPENDIX D: IMPACT OF SYSTEMATICS: CORNER PLOTS

We show here the corner plots comparing the posteriors in the three scenarios considered in Sec. V B for the three choices of total mass.

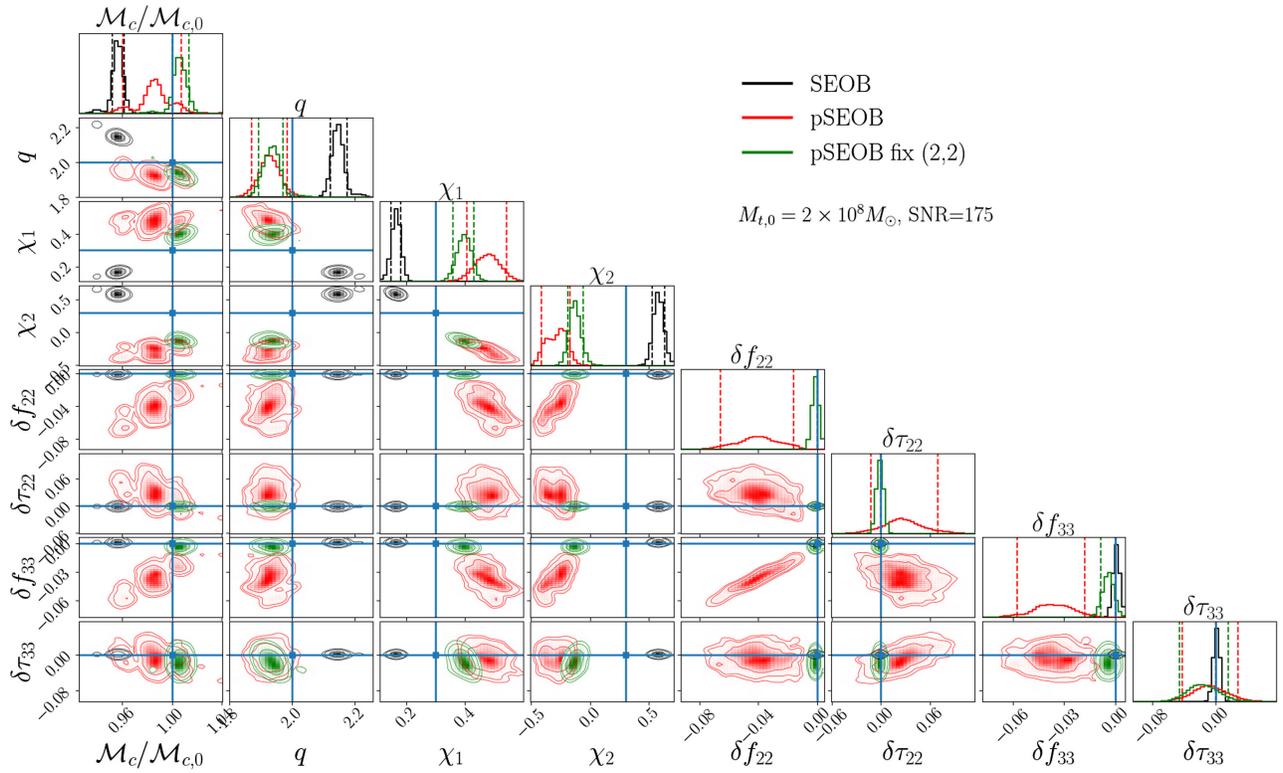


FIG. 25. Corner plot for the parameters estimated in three different scenarios for a total mass of  $2 \times 10^8 M_\odot$  including all harmonics of SEOBNRv5HM and fixing the SNR of the injection to be 175. Contours show the 68%, 90%, and 95% confidence intervals, and dashed lines the 0.05 and 0.95 quantiles.

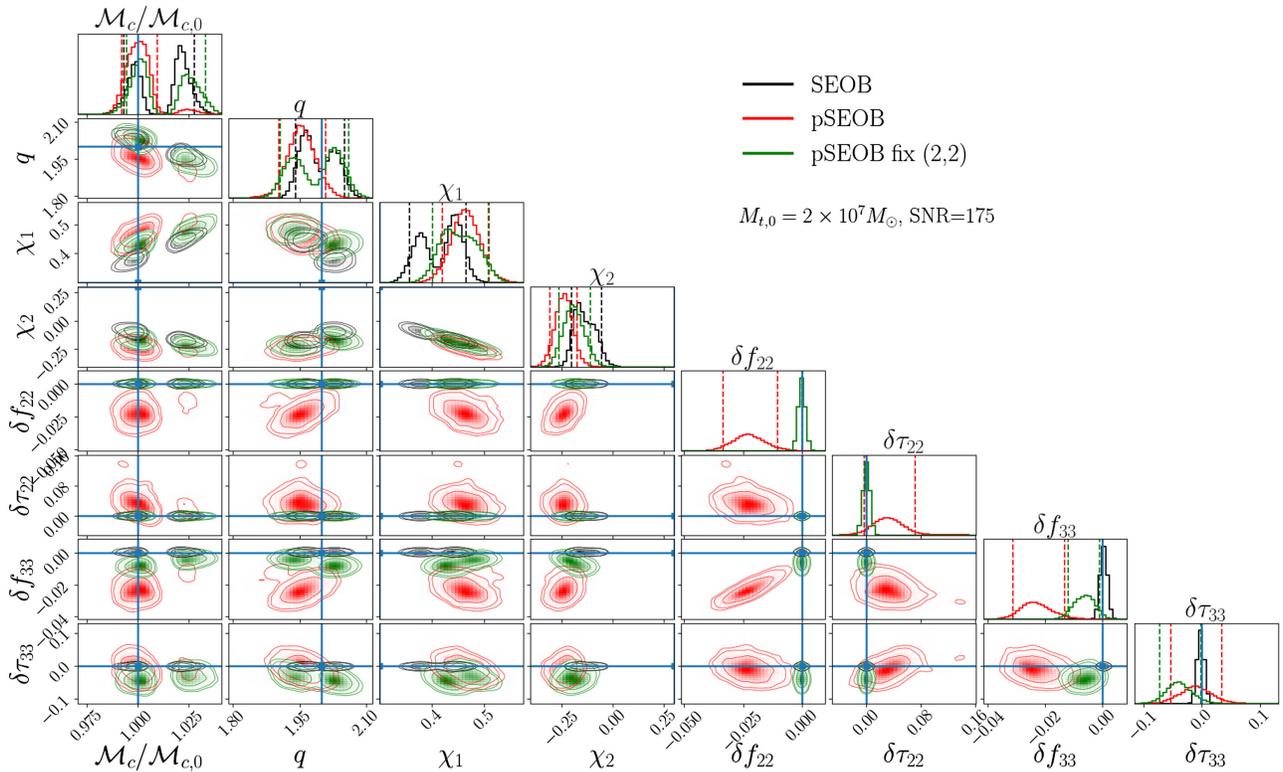
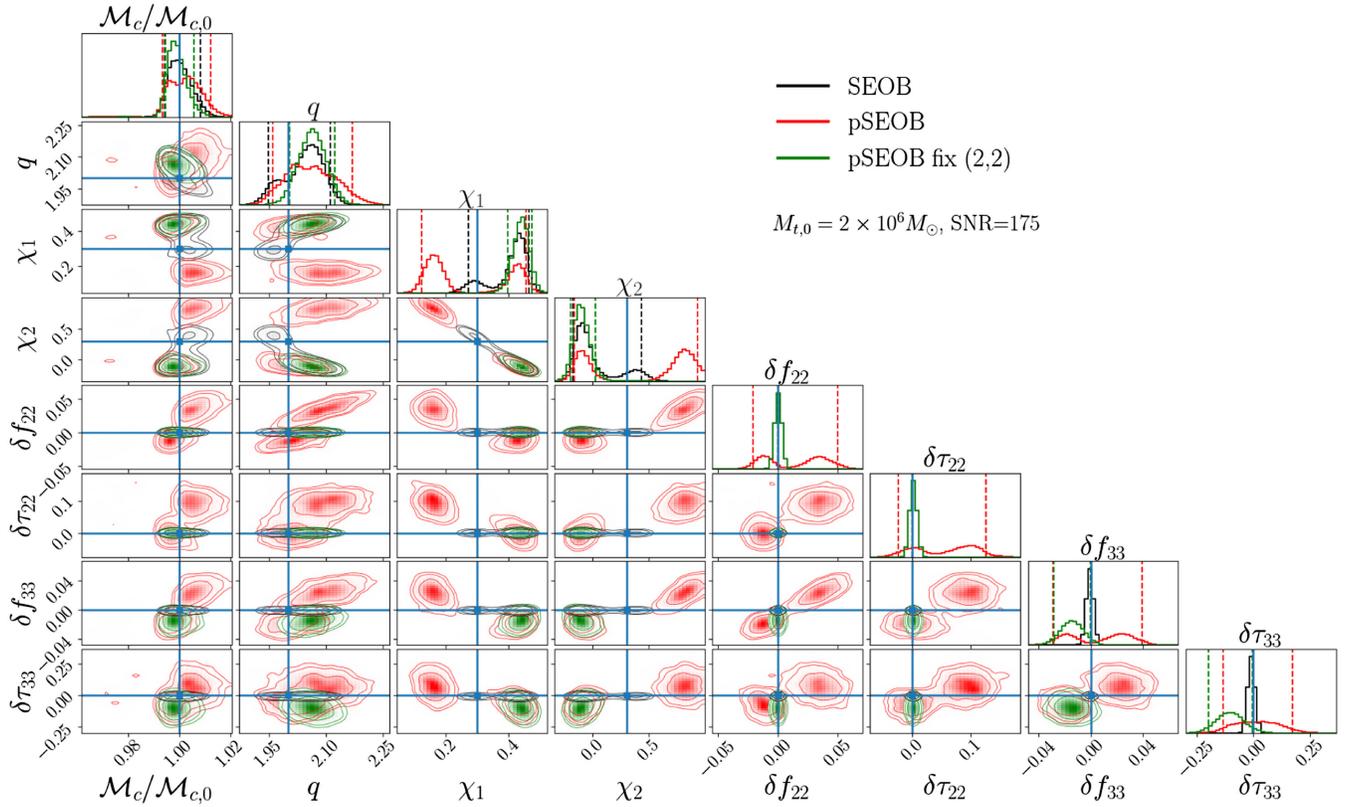


FIG. 26. The same as Fig. 25 for  $M_t = 2 \times 10^7 M_\odot$ .

FIG. 27. The same as Fig. 25 for  $M_t = 2 \times 10^6 M_\odot$ .

- [1] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. D* **109**, 022001 (2024).
- [2] R. Abbott *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), *Phys. Rev. X* **13**, 041039 (2023).
- [3] J. Aasi *et al.* (LIGO Scientific Collaboration), *Classical Quantum Gravity* **32**, 074001 (2015).
- [4] F. Acernese *et al.* (Virgo Collaboration), *Classical Quantum Gravity* **32**, 024001 (2015).
- [5] T. Akutsu *et al.* (KAGRA Collaboration), *Prog. Theor. Exp. Phys.* **2021**, 05A101 (2021).
- [6] A. H. Nitz, S. Kumar, Y.-F. Wang, S. Kasta, S. Wu, M. Schäfer, R. Dhurkunde, and C. D. Capano, *Astrophys. J.* **946**, 59 (2023).
- [7] S. Olsen, T. Venumadhav, J. Mushkin, J. Roulet, B. Zackay, and M. Zaldarriaga, *Phys. Rev. D* **106**, 043009 (2022).
- [8] B. P. Abbott *et al.* (KAGRA, LIGO Scientific, and Virgo Collaborations), *Living Rev. Relativity* **21**, 3 (2018).
- [9] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Astrophys. J. Lett.* **818**, L22 (2016).
- [10] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Astrophys. J. Lett.* **882**, L24 (2019).
- [11] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Astrophys. J. Lett.* **913**, L7 (2021).
- [12] R. Abbott *et al.* (KAGRA, Virgo, and LIGO Scientific Collaborations), *Phys. Rev. X* **13**, 011048 (2023).
- [13] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Astrophys. J. Lett.* **900**, L13 (2020).
- [14] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. Lett.* **119**, 161101 (2017).
- [15] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. Lett.* **121**, 161101 (2018).
- [16] B. P. Abbott *et al.* (LIGO Scientific, Virgo, 1M2H, Dark Energy Camera GW-E, DES, DLT40, Las Cumbres Observatory, VINROUGE, MASTER Collaborations), *Nature (London)* **551**, 85 (2017).
- [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 Collaborations), *Astrophys. J. Lett.* **848**, L12 (2017).
- [18] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. Lett.* **116**, 221101 (2016); **121**, 129902(E) (2018).
- [19] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. D* **100**, 104036 (2019).
- [20] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. D* **103**, 122002 (2021).
- [21] R. Abbott *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), *Phys. Rev. D* (to be published).
- [22] C. V. Vishveshwara, *Nature (London)* **227**, 936 (1970).
- [23] W. H. Press and S. A. Teukolsky, *Astrophys. J.* **185**, 649 (1973).
- [24] S. Chandrasekhar and S. L. Detweiler, *Proc. R. Soc. A* **344**, 441 (1975).
- [25] K. D. Kokkotas and B. G. Schmidt, *Living Rev. Relativity* **2**, 2 (1999).
- [26] V. Ferrari and L. Gualtieri, *Gen. Relativ. Gravit.* **40**, 945 (2008).
- [27] E. Berti, V. Cardoso, and A. O. Starinets, *Classical Quantum Gravity* **26**, 163001 (2009).
- [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, *Classical Quantum Gravity* **29**, 232002 (2012).
- [30] M. W. Horbatsch and C. P. Burgess, *J. Cosmol. Astropart. Phys.* **05** (2012) 010.
- [31] E. Berti, V. Cardoso, L. Gualtieri, M. Horbatsch, and U. Sperhake, *Phys. Rev. D* **87**, 124020 (2013).
- [32] T. P. Sotiriou and S.-Y. Zhou, *Phys. Rev. Lett.* **112**, 251102 (2014).
- [33] K. Yagi, L. C. Stein, and N. Yunes, *Phys. Rev. D* **93**, 024010 (2016).
- [34] E. Barausse and K. Yagi, *Phys. Rev. Lett.* **115**, 211105 (2015).
- [35] H. O. Silva, J. Sakstein, L. Gualtieri, T. P. Sotiriou, and E. Berti, *Phys. Rev. Lett.* **120**, 131104 (2018).
- [36] C. A. R. Herdeiro, E. Radu, N. Sanchis-Gual, and J. A. Font, *Phys. Rev. Lett.* **121**, 101102 (2018).
- [37] F.-L. Julié, *J. Cosmol. Astropart. Phys.* **10** (2018) 033.
- [38] F.-L. Julié and E. Berti, *Phys. Rev. D* **100**, 104061 (2019).
- [39] F.-L. Julié, H. O. Silva, E. Berti, and N. Yunes, *Phys. Rev. D* **105**, 124031 (2022).
- [40] J. L. Blázquez-Salcedo, F. S. Khoo, and J. Kunz, *Phys. Rev. D* **96**, 064008 (2017).
- [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).
- [42] C. Molina, P. Pani, V. Cardoso, and L. Gualtieri, *Phys. Rev. D* **81**, 124021 (2010).
- [43] V. Cardoso and L. Gualtieri, *Phys. Rev. D* **80**, 064008 (2009); **81**, 089903(E) (2010).
- [44] P. Pani, E. Berti, and L. Gualtieri, *Phys. Rev. Lett.* **110**, 241103 (2013).
- [45] P. Pani, E. Berti, and L. Gualtieri, *Phys. Rev. D* **88**, 064048 (2013).
- [46] Z. Mark, H. Yang, A. Zimmerman, and Y. Chen, *Phys. Rev. D* **91**, 044025 (2015).
- [47] O. J. C. Dias, M. Godazgar, and J. E. Santos, *Phys. Rev. Lett.* **114**, 151101 (2015).
- [48] V. Cardoso, M. Kimura, A. Maselli, E. Berti, C. F. B. Macedo, and R. McManus, *Phys. Rev. D* **99**, 104077 (2019).
- [49] R. McManus, E. Berti, C. F. B. Macedo, M. Kimura, A. Maselli, and V. Cardoso, *Phys. Rev. D* **100**, 044061 (2019).
- [50] L. Pierini and L. Gualtieri, *Phys. Rev. D* **103**, 124017 (2021).
- [51] L. Pierini and L. Gualtieri, *Phys. Rev. D* **106**, 104009 (2022).
- [52] P. A. Cano, K. Fransen, T. Hertog, and S. Maenaut, *Phys. Rev. D* **108**, 024040 (2023).
- [53] P. A. Cano, K. Fransen, T. Hertog, and S. Maenaut, *Phys. Rev. D* **108**, 124032 (2023).
- [54] P. Wagle, N. Yunes, and H. O. Silva, *Phys. Rev. D* **105**, 124003 (2022).
- [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, *Classical Quantum Gravity* **21**, 787 (2004).
- [57] E. E. Flanagan and S. A. Hughes, *Phys. Rev. D* **57**, 4535 (1998).
- [58] E. Berti, V. Cardoso, and C. M. Will, *Phys. Rev. D* **73**, 064030 (2006).
- [59] X. Jiménez Forteza, S. Bhagwat, P. Pani, and V. Ferrari, *Phys. Rev. D* **102**, 044053 (2020).
- [60] X. J. Forteza, S. Bhagwat, S. Kumar, and P. Pani, *Phys. Rev. Lett.* **130**, 021001 (2023).
- [61] R. Brito, A. Buonanno, and V. Raymond, *Phys. Rev. D* **98**, 084038 (2018).
- [62] A. Ghosh, R. Brito, and A. Buonanno, *Phys. Rev. D* **103**, 124041 (2021).
- [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).
- [64] R. Cotesta, A. Buonanno, A. Bohé, A. Taracchini, I. Hinder, and S. Ossokine, *Phys. Rev. D* **98**, 084028 (2018).
- [65] A. Buonanno and T. Damour, *Phys. Rev. D* **59**, 084006 (1999).
- [66] A. Buonanno and T. Damour, *Phys. Rev. D* **62**, 064015 (2000).
- [67] E. Maggio, H. O. Silva, A. Buonanno, and A. Ghosh, *Phys. Rev. D* **108**, 024043 (2023).
- [68] H. O. Silva, A. Ghosh, and A. Buonanno, *Phys. Rev. D* **107**, 044030 (2023).
- [69] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. Lett.* **116**, 061102 (2016).
- [70] G. Carullo, W. Del Pozzo, and J. Veitch, *Phys. Rev. D* **99**, 123029 (2019); **100**, 089903(E) (2019).
- [71] M. Isi and W. M. Farr, arXiv:2107.05609.
- [72] P. Amaro-Seoane *et al.* (LISA Collaboration), arXiv:1702.00786.
- [73] A. Klein *et al.*, *Phys. Rev. D* **93**, 024003 (2016).
- [74] S. Bhagwat, C. Pacilio, E. Barausse, and P. Pani, *Phys. Rev. D* **105**, 124063 (2022).

- [75] R. Cotesta, S. Marsat, A. Toubiana, and B. Emanuele (to be published).
- [76] E. Berti, A. Sesana, E. Barausse, V. Cardoso, and K. Belczynski, *Phys. Rev. Lett.* **117**, 101102 (2016).
- [77] L. S. Finn and D. F. Chernoff, *Phys. Rev. D* **47**, 2198 (1993).
- [78] M. Vallisneri, *Phys. Rev. D* **77**, 042001 (2008).
- [79] D. P. Mihaylov, S. Ossokine, A. Buonanno, H. Estelles, L. Pompili, M. Pürrer, and A. Ramos-Buades, [arXiv:2303.18203](https://arxiv.org/abs/2303.18203).
- [80] M. Khalil, A. Buonanno, H. Estelles, D. P. Mihaylov, S. Ossokine, L. Pompili, and A. Ramos-Buades, *Phys. Rev. D* **108**, 124036 (2023).
- [81] L. Pompili *et al.*, *Phys. Rev. D* **108**, 124035 (2023).
- [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, *Phys. Rev. D* **108**, 124038 (2023).
- [83] C. Pitte, Q. Baghi, S. Marsat, M. Besançon, and A. Petiteau, *Phys. Rev. D* **108**, 044053 (2023).
- [84] A. Sesana, M. Volonteri, and F. Haardt, *Mon. Not. R. Astron. Soc.* **377**, 1711 (2007).
- [85] A. Sesana, J. Gair, E. Berti, and M. Volonteri, *Phys. Rev. D* **83**, 044036 (2011).
- [86] M. Bonetti, A. Sesana, F. Haardt, E. Barausse, and M. Colpi, *Mon. Not. R. Astron. Soc.* **486**, 4044 (2019).
- [87] P. Dayal, E. M. Rossi, B. Shiralilou, O. Piana, T. R. Choudhury, and M. Volonteri, *Mon. Not. R. Astron. Soc.* **486**, 2336 (2019).
- [88] E. Barausse, I. Dvorkin, M. Tremmel, M. Volonteri, and M. Bonetti, *Astrophys. J.* **904**, 16 (2020).
- [89] E. Barausse and A. Lapi, [arXiv:2011.01994](https://arxiv.org/abs/2011.01994).
- [90] M. A. Latif and A. Ferrara, *Pub. Astron. Soc. Aust.* **33**, e051 (2016).
- [91] H. Estellés, S. Husa, M. Colleoni, D. Keitel, M. Mateur-Lucena, C. García-Quirós, A. Ramos-Buades, and A. Borchers, *Phys. Rev. D* **105**, 084039 (2022).
- [92] P. Ajith *et al.*, *Phys. Rev. D* **77**, 104017 (2008); **79**, 129901(E) (2009).
- [93] P. Ajith *et al.*, *Phys. Rev. Lett.* **106**, 241101 (2011).
- [94] L. Santamaria *et al.*, *Phys. Rev. D* **82**, 064016 (2010).
- [95] N. Aghanim *et al.* (Planck Collaboration), *Astron. Astrophys.* **641**, A6 (2020); **652**, C4(E) (2021).
- [96] A. Buonanno, G. B. Cook, and F. Pretorius, *Phys. Rev. D* **75**, 124018 (2007).
- [97] B. J. Kelly and J. G. Baker, *Phys. Rev. D* **87**, 084004 (2013).
- [98] A. Bohé *et al.*, *Phys. Rev. D* **95**, 044028 (2017).
- [99] 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).
- [100] T. Damour and A. Nagar, *Phys. Rev. D* **90**, 044018 (2014).
- [101] E. Berti, V. Cardoso, and M. Casals, *Phys. Rev. D* **73**, 024013 (2006); **73**, 109902(E) (2006).
- [102] A. Kumar Mehta, P. Tiwari, N. K. Johnson-McDaniel, C. K. Mishra, V. Varma, and P. Ajith, *Phys. Rev. D* **100**, 024032 (2019).
- [103] E. Berti and A. Klein, *Phys. Rev. D* **90**, 064012 (2014).
- [104] Y. Pan, A. Buonanno, A. Taracchini, L. E. Kidder, A. H. Mroué, H. P. Pfeiffer, M. A. Scheel, and B. Szilágyi, *Phys. Rev. D* **89**, 084006 (2014).
- [105] M. H.-Y. Cheung *et al.*, *Phys. Rev. Lett.* **130**, 081401 (2023).
- [106] L. C. Stein, *J. Open Source Software* **4**, 1683 (2019).
- [107] X. Jiménez-Forteza, D. Keitel, S. Husa, M. Hannam, S. Khan, and M. Pürrer, *Phys. Rev. D* **95**, 064024 (2017).
- [108] F. Hofmann, E. Barausse, and L. Rezzolla, *Astrophys. J. Lett.* **825**, L19 (2016).
- [109] S. Gossan, J. Veitch, and B. S. Sathyaprakash, *Phys. Rev. D* **85**, 124056 (2012).
- [110] J. Meidam, M. Agathos, C. Van Den Broeck, J. Veitch, and B. S. Sathyaprakash, *Phys. Rev. D* **90**, 064009 (2014).
- [111] M. Isi, M. Giesler, W. M. Farr, M. A. Scheel, and S. A. Teukolsky, *Phys. Rev. Lett.* **123**, 111102 (2019).
- [112] C. Cutler, *Phys. Rev. D* **57**, 7089 (1998).
- [113] S. Marsat, J. G. Baker, and T. Dal Canton, *Phys. Rev. D* **103**, 083011 (2021).
- [114] M. Tinto and S. V. Dhurandhar, *Living Rev. Relativity* **8**, 4 (2005).
- [115] 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).
- [116] N. Karnesis, M. L. Katz, N. Korsakova, J. R. Gair, and N. Stergioulas, *Mon. Not. R. Astron. Soc.* **526**, 4814 (2023).
- [117] N. Karnesis, M. L. Katz, N. Korsakova, J. R. Gair, and N. Stergioulas, Eryn, <https://github.com/mikekatz04/Eryn> (2023).
- [118] M. Volonteri, G. Lodato, and P. Natarajan, *Mon. Not. R. Astron. Soc.* **383**, 1079 (2008).
- [119] J. Binney and S. Tremaine, *Galactic Dynamics* (Princeton University Press, Princeton, 1987).
- [120] M. J. Rees, *Annu. Rev. Astron. Astrophys.* **22**, 471 (1984).
- [121] J. Antoniadis *et al.*, *Astron. Astrophys.* **678**, A50 (2023).
- [122] G. Agazie *et al.* (NANOGrav Collaboration), *Astrophys. J. Lett.* **951**, L8 (2023).
- [123] D. J. Reardon *et al.*, *Astrophys. J. Lett.* **951**, L6 (2023).
- [124] H. Xu *et al.*, *Res. Astron. Astrophys.* **23**, 075024 (2023).
- [125] A. Sesana, E. Barausse, M. Dotti, and E. M. Rossi, *Astrophys. J.* **794**, 104 (2014).
- [126] J. M. Bardeen and J. A. Petterson, *Astrophys. J. Lett.* **195**, L65 (1975).
- [127] T. Bogdanovic, C. S. Reynolds, and M. C. Miller, *Astrophys. J. Lett.* **661**, L147 (2007).
- [128] D. Merritt and M. Milosavljevic, *Living Rev. Relativity* **8**, 8 (2005).
- [129] M. Colpi, *Space Sci. Rev.* **183**, 189 (2014).
- [130] M. Tremmel, F. Governato, M. Volonteri, T. R. Quinn, and A. Pontzen, *Mon. Not. R. Astron. Soc.* **475**, 4967 (2018).
- [131] C. L. Rodriguez, B. Farr, V. Raymond, W. M. Farr, T. B. Littenberg, D. Fazi, and V. Kalogera, *Astrophys. J.* **784**, 119 (2014).
- [132] C. Cutler and E. E. Flanagan, *Phys. Rev. D* **49**, 2658 (1994).
- [133] V. Baibhav, E. Berti, and V. Cardoso, *Phys. Rev. D* **101**, 084053 (2020).
- [134] S. Bhagwat, X. J. Forteza, P. Pani, and V. Ferrari, *Phys. Rev. D* **101**, 044033 (2020).
- [135] I. Ota and C. Chirenti, *Phys. Rev. D* **101**, 104005 (2020).
- [136] V. Ferrari, M. Pauri, and F. Piazza, *Phys. Rev. D* **63**, 064009 (2001).
- [137] P. Pani and V. Cardoso, *Phys. Rev. D* **79**, 084031 (2009).

- [138] R. Brito, V. Cardoso, and P. Pani, *Phys. Rev. D* **88**, 023514 (2013).
- [139] R. Brito, V. Cardoso, and P. Pani, *Phys. Rev. D* **87**, 124024 (2013).
- [140] E. Babichev, R. Brito, and P. Pani, *Phys. Rev. D* **93**, 044041 (2016).
- [141] K. Glampedakis, G. Pappas, H. O. Silva, and E. Berti, *Phys. Rev. D* **96**, 064054 (2017).
- [142] P. Jai-akson, A. Chatrabhuti, O. Evnin, and L. Lehner, *Phys. Rev. D* **96**, 044031 (2017).
- [143] K. Glampedakis and G. Pappas, *Phys. Rev. D* **97**, 041502 (2018).
- [144] M. Boyle *et al.*, *Classical Quantum Gravity* **36**, 195006 (2019).
- [145] V. Baibhav, M. H.-Y. Cheung, E. Berti, V. Cardoso, G. Carullo, R. Cotesta, W. Del Pozzo, and F. Duque, *Phys. Rev. D* **108**, 104020 (2023).
- [146] H. Akaike, *IEEE Trans. Autom. Control* **19**, 716 (1974).
- [147] A. Toubiana and J. R. Gair, [arXiv:2401.06845](https://arxiv.org/abs/2401.06845).
- [148] E. E. Flanagan and S. A. Hughes, *Phys. Rev. D* **57**, 4566 (1998).
- [149] L. Lindblom, B. J. Owen, and D. A. Brown, *Phys. Rev. D* **78**, 124020 (2008).
- [150] S. T. McWilliams, B. J. Kelly, and J. G. Baker, *Phys. Rev. D* **82**, 024014 (2010).
- [151] K. Chatziioannou, A. Klein, N. Yunes, and N. Cornish, *Phys. Rev. D* **95**, 104004 (2017).
- [152] M. Pürrer and C.-J. Haster, *Phys. Rev. Res.* **2**, 023151 (2020).
- [153] R. E. Kass and A. E. Raftery, *J. Am. Stat. Assoc.* **90**, 773 (1995).
- [154] A. Ramos-Buades, A. Buonanno, H. Estellés, M. Khalil, D. P. Mihaylov, S. Ossokine, L. Pompili, and M. Shiferaw, *Phys. Rev. D* **108**, 124037 (2023).
- [155] N. Yunes and F. Pretorius, *Phys. Rev. D* **80**, 122003 (2009).
- [156] M. Agathos, W. Del Pozzo, T. G. F. Li, C. Van Den Broeck, J. Veitch, and S. Vitale, *Phys. Rev. D* **89**, 082001 (2014).
- [157] A. K. Mehta, A. Buonanno, R. Cotesta, A. Ghosh, N. Sennett, and J. Steinhoff, *Phys. Rev. D* **107**, 044020 (2023).
- [158] L. Blanchet, *Living Rev. Relativity* **17**, 2 (2014).
- [159] <http://git.ligo.org/waveforms/software/pyseobnr>.
- [160] LIGO Scientific Collaboration, LIGO Algorithm Library—LALSuite, free software (GPL) (2018).
- [161] S. Husa, S. Khan, M. Hannam, M. Pürrer, F. Ohme, X. Jiménez Forteza, and A. Bohé, *Phys. Rev. D* **93**, 044006 (2016).
- [162] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. Jiménez Forteza, and A. Bohé, *Phys. Rev. D* **93**, 044007 (2016).
- [163] 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).