

Waveform systematics in the gravitational-wave inference of tidal parameters and equation of state from binary neutron-star signals

Rossella Gamba,¹ Matteo Breschi¹, Sebastiano Bernuzzi¹, Michalis Agathos,¹ and Alessandro Nagar^{2,3}

¹*Theoretisch-Physikalisches Institut, Friedrich-Schiller-Universität Jena, 07743 Jena, Germany*

²*INFN Sezione di Torino, Via P. Giuria 1, 10125 Torino, Italy*

³*Dipartimento di Fisica, Università di Torino, via P. Giuria 1, 10125 Torino, Italy*



(Received 25 September 2020; accepted 4 May 2021; published 7 June 2021)

Gravitational-wave signals from binary neutron-star coalescences carry information about the star's equation of state in their tidal signatures. A major issue in the inference of the tidal parameters (or, directly, of the equation of state) is the systematic error introduced by the waveform approximants. We use a bottom-up approach based on gauge-invariant phase analysis and the Fisher information matrix to investigate waveform systematics and identify biases in parameter estimation. A mock analysis of 15 different binaries indicates that systematics in current waveform models dominate over statistical errors at signal-to-noise ratio (SNR) $\gtrsim 80$ for LIGO and Virgo at design sensitivity. This implies biases in the inference of the reduced tidal parameter that are larger than the statistical 90% credible intervals. For example, while the neutron-star radius could be constrained at the $\sim 5\%$ level at SNR 80, systematics can be at the $\sim 10\%$ level. We apply our approach to GW170817 (SNR ~ 30) and confirm that no significant systematic effects are present. Using frequencies below $\lesssim 1$ kHz for the analysis, we estimate a neutron-star radius of $12.5_{-1.8}^{+1.1}$ km. The latter is consistent with an electromagnetic-informed prior and the recent NICER measurement. Exploring SNR $\gtrsim 100$ in view of third-generation detectors, we find that all the current waveform models lead to differences of at least 1σ in the inference of the reduced tidal parameter. We conclude that current waveform models, including those from numerical relativity, are insufficient to infer the equation of state in the loudest (and potentially most informative) events that will be observed by advanced and third-generation detectors.

DOI: [10.1103/PhysRevD.103.124015](https://doi.org/10.1103/PhysRevD.103.124015)

I. INTRODUCTION

The detection of GW170817 [1], the first coalescing binary neutron-star (BNS) system seen by LIGO-Virgo detectors, demonstrated how gravitational waves (GWs) can be employed as a means to investigate the properties of cold, dense matter [2–5]. Parameter estimation (PE) of GW data gives direct information on the masses, spins, and tidal parameters of the two objects involved in the coalescence. Typical inspiral-merger matched-filtering analyses [6] are commonly performed in the Fourier domain by matching the data to a large number of template waveforms within a Bayesian framework [7,8].

The tools employed during PE are based on Markov-chain Monte Carlo (MCMC) methods or nested sampling algorithms [9]. The template waveforms are obtained from approximate solutions of the two-body problem in general relativity (see, e.g., Refs. [10–13] and references therein). Different approximations and methods give rise to different template families, which—during the process of PE—may in principle lead to different results in the recovery of the source parameters. Errors and biases due to waveform modeling choices are commonly labeled as *waveform systematics*, and are the main topic of the present paper.

Significant waveform systematics, larger than statistical uncertainties, have yet to be observed for binary neutron-star systems: looking at the results coming from the recent observations of BNS mergers, the parameters of both GW190425 [14] and GW170817 [6] have been demonstrated to be largely consistent between different waveform families. However, recent studies [15–21] have pointed out that the measured tidal parameters can be strongly biased depending on the employed tidal and point-mass descriptions of the waveform approximant. The agreement between the different GW models employed in the PE of the observed BNS signals is then mainly due to the relatively low signal-to-noise ratio (SNR). With the increasing sensitivities of next-generation detectors [22–27], waveform systematics will affect the measurements, thus leading to discordant (or inconclusive) results.

In this context, the necessity of understanding the errors introduced by waveform systematics arises. In this paper, we aim at tackling the issue for SNRs relevant for advanced and third-generation (3G) detectors, and provide a bottom-up approach to guide future BNS analyses. In particular, in Sec. II, we summarize the current knowledge of the theoretical tools which are employed to *a priori* predict the presence of waveform systematics. We expand on the

argument of Ref. [18] and propose a way, inspired by Refs. [28,29], to estimate the bias that may affect tidal parameters. In Sec. III, we summarize the key features of the GW models used in our analysis and compare them by computing their gauge-invariant phasing. In Sec. IV, we perform mock PE experiments (*injections*) with 15 binaries having a signal-to-noise ratio of ~ 80 , to study the posterior distributions of the typical parameters of interest of a BNS merger, such as tidal deformabilities, mass ratio, and spins. Differently from previous studies, we focus on injections of different masses and equations of state (cf. Refs. [15–17], where fewer binaries have been considered) and nonspinning waveforms (See Ref. [17] for spin effects.) In particular, we discuss the impact of waveform systematics on the inference of the tidal parameters and the estimation of the radii of the single NSs. In Sec. V, we apply the methods developed during the previous sections to GW170817. We reanalyze the event’s data and find that analyses which consider frequencies up to 1 kHz are less affected by waveform systematics. Finally, in Sec. VI, we estimate the impact of waveform systematics for BNS events detected with third-generation detectors and find that statistical errors will be comparable to waveform systematics from SNRs > 100 for values of the reduced tidal parameter $\tilde{\Lambda} \simeq 400\text{--}1000$ [see Eq. (2)].

Throughout the whole paper, we label the two bodies as A and B . We denote the component masses as m_A, m_B , the dimensionless spins of the bodies as $\chi_{A,B}$, the total mass as $M = m_A + m_B$, and we define the chirp mass of the binary as $\mathcal{M}_c = (m_A m_B)^{3/5} / (M)^{1/5}$. We define the quadrupolar tidal parameters as

$$\Lambda_A \equiv \frac{2}{3} C_A^{-5} k_A^{(2)}, \quad (1)$$

where $k_A^{(2)}$ is the dimensionless gravitoelectric Love number [30,31], and $C_A \equiv Gm_A / (c^2 R_A)$ is the compactness parameter. Λ_A is also denoted by $\tilde{\lambda}_2$ [32]. The quadrupole tidal parameters enter at the leading order in the phase of the waveform through the reduced tidal parameter [10,29]:

$$\tilde{\Lambda} = \frac{16}{13} \frac{(m_A + 12m_B)m_A^4 \Lambda_A}{M^5} + (A \leftrightarrow B). \quad (2)$$

We often switch between mass-rescaled quantities in geometrical units $c = G = 1$ and physical units. Since $GM_\odot \simeq 4.925490947 \mu\text{s}$ or $\simeq 1.476625038 \text{ km}$, the dimensionless frequency $\hat{\omega} = GM\omega$ relates to the frequency in hertz by

$$f = \frac{\omega}{2\pi} \simeq 32.3125 \hat{\omega} \frac{M_\odot}{M} \text{ kHz}. \quad (3)$$

II. ORIGIN OF SYSTEMATICS

Waveform systematics are intrinsically related to the concept of measurability of the waveform parameters. They become important when the differences due to template choice are larger than those induced by noise fluctuations in the detector (statistical uncertainties)—i.e., when the distributions of the estimated parameters $\bar{\theta}$ have a width σ_θ smaller than the differences $\Delta\theta$ induced by waveform models. In this section, we highlight, with basic analytical arguments, that the systematics on tidal parameters crucially depend on the frequency regime in which the measurement is effectively performed.

Optimal gravitational-wave data analysis of compact binaries are based on matched-filtering techniques in which the data are “best matched” to waveform templates [33]. The accuracy requirements on the waveforms used in the matched filtering depend on whether waveform models are employed for detection or parameter estimation. In the former case, waveforms are only required to be effectual; while in the latter, they are required to be faithful [34]. To quantify these concepts, it is necessary to introduce a metric in the waveform space in order to measure how close two waveforms are. The basic quantity used in GW analysis theory is the Wiener inner product between two waveforms $h(t)$ and $k(t)$, defined by

$$(h|k) = 4\Re \int \frac{\tilde{h}(f)\tilde{k}^*(f)}{S_n(f)} df, \quad (4)$$

where $\tilde{h}(f)$ is the Fourier transform of $h(t)$, and $S_n(f)$ is the power spectral density (PSD) of the detector. The faithfulness (or match) is the normalized and noise-weighted inner product

$$\mathcal{F} = \max_{t_c, \phi_c} \frac{(h|k)}{\sqrt{(h|h)(k|k)}}, \quad (5)$$

where t_c, ϕ_c are the time and phase, respectively, of the waveform at a chosen reference time. Due to the arbitrary nature of these parameters, the match is maximized over them. The match \mathcal{F} defines an “angle” in the waveform space; $\mathcal{F} = 1$ indicates perfect overlap between h and k . The mismatch $\bar{\mathcal{F}} = 1 - \mathcal{F}$ gives the loss in signal-to-noise ratio (squared) when the waveforms are aligned in time and phase. Accuracy requirements for both detection and PE can be expressed in terms of \mathcal{F} .¹ A mismatch of $\bar{\mathcal{F}} = 0.03$ corresponds to $\sim 10\%$ of detection losses [36], which is assumed as the effectualness condition for a template bank.

¹Although this has become a common practice, it would be more appropriate to express these requirements by means of suitable effectualness, faithfulness, and accuracy functionals; see Ref. [35].

Necessary conditions for faithful waveform models can also be expressed in terms of \mathcal{F} [35,36] (see below).

Generally, the parameters of a GW signal are measured using matched-filtering techniques within a Bayesian framework [9]. Defining $\tilde{d}(f) = \tilde{A}_d(f)e^{i\Psi_d(f)}$ as the target (injected or measured) strain, $\tilde{h}(f, \boldsymbol{\theta}) = \tilde{A}_h(f, \boldsymbol{\theta})e^{i\Psi_h(f, \boldsymbol{\theta})}$ as the template waveform, and $\boldsymbol{\theta}$ as the set of parameters on which h depends, the likelihood function is

$$p(d|\boldsymbol{\theta}) \propto e^{-\frac{1}{2}(d-h|d-h)}. \quad (6)$$

Writing $(d-h|d-h) = (d|d) + (h|h) - 2(d|h)$, the maximization of the likelihood can be interpreted as the maximization of the matched-filter signal-to-noise ratio (SNR),

$$\rho = \frac{(d|h)}{\sqrt{(h|h)}} \propto \Re \int \frac{\tilde{A}_d \tilde{A}_h e^{i\Delta\Psi}}{S_n} df, \quad (7)$$

where $\Delta\Psi = \Psi_h - \Psi_d$. The SNR ρ quantifies the amount of signal deposited in the recorded data d that is matching a given template h . The *optimal* SNR, instead, is defined as the matched-filtered SNR computed within the assumption $d = h$:

$$\rho_{\text{opt}} = \sqrt{(h|h)}. \quad (8)$$

This value identifies the SNR we would get if the signal were coincident with the template and the noise realization were identically zero. Further, from expanding Eq. (7) around small noise realizations ($n \simeq 0$), one obtains $\rho = \rho_{\text{opt}} + O(n)$. The optimal SNR coincides with the expected value of the actual matched-filter SNR under different noise realizations, under the assumption of zero-mean noise processes. GW data analysis delivers probability distributions of the sampled parameters (posteriors), which can be characterized by their maximum probability (peak) values and credible intervals. The measurements thus obtained can be affected by statistical uncertainties due to fluctuations of the detector noise and systematic effects due to the waveform models employed. In the high-SNR regime, the former mainly impact the posteriors by widening the credible intervals and scattering the maximum-likelihood estimate, and their importance decreases as the SNR grows, while the latter can influence PE by shifting the posterior distributions with respect to the true values.

A. Statistical errors

Under the assumption of Gaussian noise, the variance $\sigma_{\theta_i}^2$ on the measurement of a generic parameter θ_i due to statistical errors can be computed through the Fisher information matrix F_{ij} (see, e.g., Refs. [35–37]). Given a waveform model $\tilde{h}(f, \boldsymbol{\theta}) = \tilde{A}_h(f)e^{i\Psi_h(f)}$, the element (i, j) of F is defined as

$$F_{ij} = (\partial_i h | \partial_j h) \simeq 4 \int \frac{\tilde{A}_h^2}{S_n} (\partial_i \Psi_h \partial_j \Psi_h) df, \quad (9)$$

where $\partial_i = \frac{\partial}{\partial \theta_i}$, and in the last equation we assume that the amplitude \tilde{A} is not correlated to other parameters [37,38].² From the Cramer-Rao bound, the variance of the distribution of θ_i can then be estimated from Eq. (9) as

$$\sigma_{\theta_i}^2 \geq (F^{-1})_{ii}. \quad (10)$$

The Fisher matrix formalism further allows one to identify the relevant frequency ranges at which different parameters are measured. Focusing on Eq. (9), it is clear that the frequency ranges that contribute to the computation of σ_{θ_i} are those where the integrand

$$I_{ii}(f) = 4 \frac{\tilde{A}_h^2(f)}{S_n(f)} [\partial_i \Psi_h(f) \partial_i \Psi_h(f)] \quad (11)$$

is largest and different from zero. Using post-Newtonian (PN) waveforms, whose amplitude \tilde{A} behaves as $\sim f^{-7/6}$, it is immediate to show that on a logarithmic frequency axis these integrands are of type $I_{ii} \sim f^{-4/3} S_n^{-1}(f) f^{p_i/3}$, where the exponent p_i depends on the particular parameter considered [39]. For example, the chirp mass has $p_{\mathcal{M}} = -10$ and, for a fiducial equal-mass $1.4 + 1.4 M_{\odot}$ BNS detected with the LIGO ZERO_det_high_power noise curve [40], is entirely determined by the signal at low frequencies $\lesssim 30$ Hz. The symmetric mass ratio integrand has $p_{\nu} = -6$, and the SNR integrand has $p_{\text{SNR}} = 0$, which implies that they are given by the useful GW cycles below 50 and 100 Hz, respectively, for the fiducial BNS (see, e.g., Fig. 3 of Ref. [39] and Fig. 2 of Ref. [41]). By contrast, the reduced tidal parameters have $p_{\tilde{\Lambda}} = +10$, i.e., $I_{\tilde{\Lambda}\tilde{\Lambda}} \sim f^2/S_n(f)$. Given that the employed noise curve is an approximately flat function of f between 50 and 800 Hz, the information on tides increases as $\sim f^2$ in this interval, to then reach a finite limit at higher frequencies and decay after merger. This marked difference in the frequency support of $I_{\tilde{\Lambda}\tilde{\Lambda}}$ indicates that the measurement of $\tilde{\Lambda}$ is not strongly correlated to that of the chirp mass and mass ratio [39], and that the magnitude of such a correlation decreases as the upper frequency cutoff is increased, because the tidal contribution becomes easier to distinguish from the rest of the signal [39]. Nonetheless, nontidal parameters can still impact the determination of $\tilde{\Lambda}$ (see Appendix A): the maximum-likelihood values of $\tilde{\Lambda}$ minimize the *overall* high-frequency phase differences $\Delta\Psi$, which can receive a non-negligible contribution from the point-mass sectors of the approximants.

²Defining the amplitude parameter as $A = \mathcal{M}^{5/6} Q(\text{angles})/D_L$, the inverse of the Fisher matrix $\Sigma^{i,j}$ is found to be block diagonal, with $\Sigma^{lnA,j} = 0$ for all parameters j different from A .

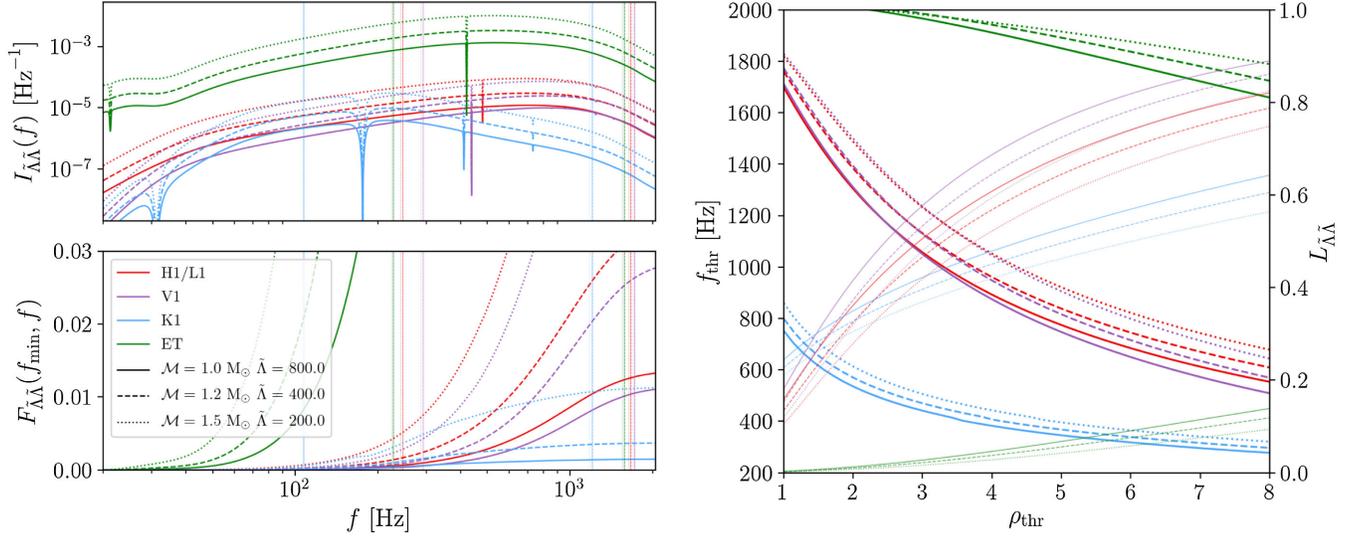


FIG. 1. Left panel: the figure shows $I_{\tilde{\Lambda}}(f)$ (top) and $F_{\tilde{\Lambda}}(f_{\min} = 20 \text{ Hz}, f)$ (bottom) computed for different combinations of $(\mathcal{M}, \tilde{\Lambda})$ employing nonspinning BNS signals with $q = 1$, $D_L = 40 \text{ Mpc}$, $\iota = 0$ and locating the source in the optimal sky location for the involved detector at the GPS time of GW170817 (1187008882.4). We employ `TaylorF2` to compute the derivatives of the phase with respect to $\tilde{\Lambda}$ and `TEOBResumS` to account for corrections in the waveform amplitude. The curves are estimated using design PSD expected for next-generation detectors: red lines refer to LIGO design sensitivity [23], purple lines refer to Virgo design sensitivity [26], blue lines refer to KAGRA design sensitivity [43,44], and green lines refer to Einstein Telescope (configuration D) sensitivity [45,46]. The vertical lines represent the frequencies $f_{5\%}^{\tilde{\Lambda}}$, $f_{95\%}^{\tilde{\Lambda}}$, defined in Eq. (14). Right panel: estimations of f_{thr} (thick lines) and $L_{\tilde{\Lambda}}$ (thin lines) as functions of $\rho_{\text{thr}} \in [1, 8]$ for the cases discussed in the left panel.

The integrands $I_{ii}(f)$ can be employed to quantify the amount of information gathered on the parameter θ_i per frequency bin. We define the cumulative information gathered in an interval $[f_a, f_b]$ of the frequency domain as

$$F_{ii}(f_a, f_b) = \int_{f_a}^{f_b} I_{ii}(f) df. \quad (12)$$

Values of $F_{ii}(f_a, f_b)$ close to zero indicate that the range $[f_a, f_b]$ does not include relevant information on θ_i . From Eq. (10), we obtain [42]

$$\sigma_{\theta_i}^2 \geq \frac{1}{F_{ii}(f_{\min}, f_{\max})}, \quad (13)$$

where f_{\max} and f_{\min} denote the upper and lower bounds of the frequency interval chosen for the analysis.

Through the information distribution $I_{ii}(f)$ and its integral $F_{ii}(f_{\min}, f)$ it is possible to find a most important frequency range where most of the information on θ_i is contained. We define the upper frequency $f_{X\%}^i$ that encloses $X\%$ of information on the i th parameter from

$$F_{ii}(f_{\min}, f_{X\%}^i) = \frac{X}{100} \cdot F_{ii}(f_{\min}, f_{\max}). \quad (14)$$

This definition corresponds to the frequency of the X th percentile of the information distribution $I_{ii}(f)$. It is then

possible to estimate the most important frequency range for the measurement of the i th parameter as the interval that encloses 90% of the total information, identified by the 5% and 95% percentiles $[f_{5\%}^i, f_{95\%}^i]$. Focusing on the tidal parameter $\tilde{\Lambda}$, Fig. 1 (left panel) shows the information distribution $I_{\tilde{\Lambda}}(f)$ and the cumulative information $F_{\tilde{\Lambda}}(f_{\min}, f)$ computed for some exemplary binary configurations using the expected design sensitivity curves for current ground-based detectors and $f_{\min} = 20 \text{ Hz}$. For fiducial BNS mergers with LIGO-Virgo design sensitivities, the most important interval $[f_{5\%}^{\tilde{\Lambda}}, f_{95\%}^{\tilde{\Lambda}}]$ spans a relatively high frequency range: $f_{5\%}^{\tilde{\Lambda}} \approx 300 \text{ Hz}$, and $f_{95\%}^{\tilde{\Lambda}} > 1 \text{ kHz}$. Note that the above intervals are independent of the distance of the source from the detectors—i.e., the same most important frequency interval pertains to a *family* of signals with varying strengths and SNRs.

The key role played by distance and SNR does not lie in the determination of the most important interval, but rather in governing the extent to which the signal can be measured and the parameters extracted. Indeed, the accuracy of the $\tilde{\Lambda}$ estimation depends crucially on the maximum frequency at which we are able to discriminate the signal from noise fluctuations. Within the assumption of $d \approx h$, it is possible to estimate the high-frequency threshold f_{thr} beyond which the additional signal power exceeds the noise contributions as

$$4 \int_{f_{\text{thr}}}^{f_{\text{max}}} \frac{|\tilde{h}(f)|^2}{S_n(f)} df = \rho_{\text{thr}}^2, \quad (15)$$

where ρ_{thr} is an arbitrary threshold value for the SNR. In the case of multiple detectors, the integral on the left-hand side of Eq. (15) has to be replaced with the summation of the integrals evaluated on the different detectors, as it is for a usual summation of SNRs.

With this definition, we guarantee that the signal power enclosed in the frequency range $[f_{\text{thr}}, f_{\text{max}}]$ does not exceed the threshold ρ_{thr} , which should be a negligible value compared to the total SNR. The study of $f_{\text{thr}}(\rho_{\text{thr}})$ for varying ρ_{thr} then allows the characterization of the most informative portion of spectrum in terms of signal power. For a fixed ρ_{thr} , louder signals will have higher f_{thr} s; conversely, extremely noisy signals will have $f_{\text{thr}} \sim f_{\text{min}}$.

Equation (15) is computed using the approximation of optimal SNR, and then the definition of f_{thr} is exact in the limit $d \rightarrow h$. In a realistic scenario, the noise contamination is non-negligible, and consequently f_{thr} can be interpreted as an upper frequency bound beyond which the signal power cannot exceed ρ_{thr} ; this means that, even in the best-case scenario $d \approx h$, the signal power enclosed in the frequency range $[f_{\text{thr}}, f_{\text{max}}]$ will always be lower (or equal, for $d = h$) than the threshold power defined by ρ_{thr} . Once f_{thr} is known, it is possible to evaluate the ratio

$$L_{ii}(f_{\text{thr}}) = \frac{F_{ii}(f_{\text{thr}}, f_{\text{max}})}{F_{ii}(f_{\text{min}}, f_{\text{max}})} \quad (16)$$

that quantifies the fractional information loss on the i th parameter, since f_{thr} represents, by construction, the maximum frequency at which the signal is relevant. Figure 1 (right panel) shows the estimation of f_{thr} and $L_{\tilde{\Lambda}\tilde{\Lambda}}$ as functions of ρ_{thr} . As ρ_{thr} grows, f_{thr} decreases, because a larger power is required to reach the increasing threshold. Conversely, $L_{\tilde{\Lambda}\tilde{\Lambda}}$ increases, since when increasing ρ_{thr} , the support $[f_{\text{thr}}, f_{\text{max}}]$ increases, and this corresponds to a larger loss of information. From the arguments above, it follows, as a rule of thumb, that if $f_{\text{thr}} \ll f_{5\%}^{\tilde{\Lambda}}$, then $L_{\tilde{\Lambda}\tilde{\Lambda}} \approx 1$, and the measurement of the tidal parameter will be strongly affected by noise fluctuation and by sensitivity limits, with the possibility of an uninformative inference for the tidal parameter. On the other hand, if $f_{\text{thr}} \geq f_{5\%}^{\tilde{\Lambda}}$, the estimated value for f_{thr} can qualitatively indicate the range over which modeling differences between waveform models are most relevant.

In Appendix B, we apply the method discussed above to the injections studied in Sec. IV, in order to prove that the injection studies are performed in an informative framework for the tidal parameter $\tilde{\Lambda}$.

Using a GW170817-like template—i.e., a waveform whose intrinsic and extrinsic parameters are fixed to the maximum posterior probability of GW170817—and

setting $\rho_{\text{thr}} = 1$, we find $f_{\text{thr}} \approx 800$ Hz for LIGO design sensitivity and $f_{\text{thr}} \approx 600$ Hz for Virgo design sensitivity, while for a network of three detectors, $f_{\text{thr}} \approx 1$ kHz and $L_{\tilde{\Lambda}\tilde{\Lambda}} \approx 20\%$.³

B. Systematic errors

Let us denote with $|k|$ the norm of a waveform k with respect to the scalar product introduced in Eq. (4). Then, given a waveform model h to approximate the *true* signal s recorded in the data $d = s + n$ (where n denotes the noise contribution), a simple accuracy standard for h can be expressed as [35,36]

$$|d - h|^2 \leq \epsilon^2 |n|^2 \equiv \epsilon^2, \quad (17)$$

with $\epsilon^2 = 1$ (or smaller, for a more strict requirement). A simple geometrical interpretation for Eq. (17) can be obtained by considering that the probability distribution of the data d containing a signal s and noise n is $\propto e^{-1/2(d-s|d-s)}$. Thus, the knowledge of s at the 1σ level is limited to a unit ball in Wiener space,

$$(d - s|d - s) \leq 1. \quad (18)$$

By using the inequality $|h - s| \leq |h - d| + |d - s| = |\delta h| + |n|$ with $\delta h = d - h$, Eq. (17) corresponds to demanding that, in the worst scenario where $|\delta h| = 1$, the errors introduced by systematics biases are equivalent to doubling the noise of the detector [35]. This criterion can be written in terms of the faithfulness as [e.g., Eq. (31) of Ref. [35]]

$$\mathcal{F} > 1 - \frac{\epsilon^2}{2\rho^2}, \quad (19)$$

with $\epsilon^2 \leq 1$. Note that sometimes it is suggested to relax this criterion by taking $\epsilon^2 = N$, the number of intrinsic parameters of the system [47]. The above criteria are necessary conditions that have to be satisfied by faithful waveform models. However, their violation does not guarantee the presence of biases. Indeed in Sec. IV D, we show that all of our simulated signals lie well below the faithfulness thresholds identified by Eq. (19), though not all of them present obvious biases on $\tilde{\Lambda}$. Conversely, if a bias is present, they do not quantify how large the uncertainty on the parameters is.

The biases $\Delta\theta = \bar{\theta} - \theta_{\text{true}}$ between the maximum likelihood (best fit) and the true parameters due to use of a waveform model h instead of the exact waveform can be estimated following Refs. [28,29]. The best-fit (possibly

³However, note that currently known events do not contain as much high-frequency information as the signals displayed here. We will further discuss real GW events in Sec. V.

biased) values $\bar{\theta}$ minimize the function $g(\theta) = (d - h(\theta)|d - h(\theta))$. Therefore, they have to be critical points of g , thus leading to the condition

$$(\partial_j h(\bar{\theta})|d - h(\bar{\theta})) = 0. \quad (20)$$

Linearly expanding $h(\bar{\theta}) \approx h(\theta_{\text{true}}) + \Delta\theta^j \partial_j h(\theta_{\text{true}})$ and inserting it in Eq. (20), one finds that

$$\Delta\theta^i = (F^{-1}(\theta_{\text{true}}))^{ij} (\partial_j h(\theta_{\text{true}})|d - h(\theta_{\text{true}})). \quad (21)$$

This equation can be combined with the accuracy criterion of Eq. (17). Indeed, recalling that $\sigma_{ij}^2 = (F^{-1})_{ij}$, we can write

$$(\sigma_{ij}^2(\theta_{\text{true}}))^{-1} \Delta\theta^i = (\partial_j h(\theta_{\text{true}})|d - h(\theta_{\text{true}})), \quad (22)$$

multiplying both sides by $\Delta\theta^j$, recalling that $h(\bar{\theta}) \approx h(\theta_{\text{true}}) + \Delta\theta^j \partial_j h(\theta_{\text{true}})$ and approximating $d \approx h(\bar{\theta})$ immediately gives

$$(\sigma_{ij}^2(\theta_{\text{true}}))^{-1} \Delta\theta^i \Delta\theta^j \approx (d - h(\theta_{\text{true}})|d - h(\theta_{\text{true}})). \quad (23)$$

Comparing Eq. (17) to Eq. (23), we note that indeed the validity of the former implies that the systematic biases $\Delta\theta^i$ are smaller than the uncertainties due to statistical fluctuations, as expected.

Estimates of the parameters bias using Eq. (21) require knowledge of the derivatives of the waveform model with respect to the parameters. These quantities, however, are nontrivial to evaluate for more sophisticated semianalytical approximants. One might then try to directly minimize the function $g(\theta)$. This in turn requires the minimization of an integral in the multidimensional space of the binary parameters, which can be computationally very expensive. However, we are interested in the bias in the reduced tidal parameter and thus assume that (i) the correlation with the other parameters can be neglected, and (ii) the largest biased parameter is $\tilde{\Lambda}$. The former assumption roughly holds if the signal is sufficiently informative (see above); the latter holds if the point-mass waveforms are sufficiently accurate at low frequencies. In these conditions, minimizing the likelihood over the whole parameter space simply reduces to computing

$$\min_{\tilde{\Lambda}} (d - h(\tilde{\Lambda})|d - h(\tilde{\Lambda})) \quad (24)$$

over a one-dimensional interval of $\tilde{\Lambda}$ values, assuming that all other intrinsic parameters are correctly estimated. While such a minimization has little practical use for GW PE, as the true parameters θ_{true} are unknown, it can nonetheless be used to estimate—given the parameters associated with one particular model—the resulting value of $\tilde{\Lambda}$ that one would get by repeating PE with a different waveform

model. Note that the new model can disagree with the previous one also in the point-mass and spin sector, as assumption (ii) above only requires the two models to agree in the low-frequency limit. In Sec. IV D, we apply this estimate to injection experiments. We find that it is able to correctly capture the behavior of the different approximants studied, and that the estimated values of $\tilde{\Lambda}$ (henceforth denoted as $\tilde{\Lambda}_E$) always fall within the 90% credible intervals of the recovered posterior distributions, with the exception of a few borderline cases where $\tilde{\Lambda}_E$ is nonetheless extremely close to the upper 95th percentile. In Sec. VI, instead, we apply Eq. (23) to two state-of-the-art approximants to estimate the importance of waveform systematics on PE with third-generation detectors.

Note that the arguments presented in this section do not address the impact of prior assumptions in GW PE, but rather focus on the maximum-likelihood estimates, which exactly coincide with the maximum (posterior) probability values only when considering uniform prior distributions. As a general rule of thumb, as long as prior assumptions are more constraining on the source parameters than the actual observational information carried by the waveform, one should expect *a priori* hypotheses to play an important role in PE [29]. Extreme care is then required when dealing with lower-SNR signals. For example, as discussed in Ref. [48], when sampling directly in the component tidal parameters Λ_A, Λ_B , the prior on $\tilde{\Lambda}$ is not independent of the mass ratio of the binary. This, in turn, impacts the computation of credible bounds—and especially of lower bounds, which are used to claim the measurement of tides. In the limit of high SNR, instead, the mean of the posterior distribution can be shown to coincide with the maximum-likelihood estimators [49]. Therefore, it is in this regime that the discussion presented above has to be interpreted.

III. WAVEFORM MODELS

Gravitational waveform models for coalescing compact binaries aim at providing approximate solutions to the GR two-body problem. They map a set of intrinsic parameters θ —for example, the mass ratio q , the chirp mass \mathcal{M}_c , the component dimensionless spins (χ_A, χ_B) , and the dimensionless tidal deformabilities (Λ_A, Λ_B) , into a time or frequency series $h(t; \theta)$ or $\tilde{h}(f; \theta)$. Post-Newtonian (PN) approximants [50,51] construct this mapping by analytically computing the evolution of the orbital phase $\phi(t)$ of a binary system as a perturbative expansion in a small parameter v/c or $x = (v/c)^2$, in which v is the characteristic velocity of the binary. Such models, while cheap from a computational standpoint, are typically unable to reliably describe the waveform at high frequencies [34]—i.e., during the later phases of the evolution of the binary when v becomes a comparable fraction of c . The effective one-body (EOB) approach [13,52–59] resums the PN information (both in the conservative and nonconservative parts of

the dynamics) so as to make it reliable and predictive also in the strong-field, fast-velocity regime. Once improved by numerical relativity (NR) data, this method allows one to compute the complete waveform from the early, quasia-diabatic inspiral up to merger and—when dealing with binary black holes—ringdown. Finally, phenomenological models [60–67] are constructed by first stitching together EOB-based inspirals with numerical relativity simulations, when available, and then devising an accurate, effective, interpolating representation all over the parameter space devised to be computationally efficient.

For our purposes, we choose one representative approximant from each of the three families above. In particular, our analysis will employ the PN `TaylorF2` model, the EOB `TEOBResumS` model, and the phenomenological `IMRPhenomPv2NRtidal` model. In Sec. VI and Appendix E, we will then consider two further approximants: `IMRPhenomPv2NRtidalv2` and `SEOBNRv4Tsurrogate`.

`TaylorF2` is a frequency-domain PN waveform model. The phase of the GW, obtained through a stationary phase approximation (SPA), contains point-mass effects which are fully known up to relative 3.5 PN order [51] and include spin-spin and spin-orbit interactions [68,69]. A higher-order, parametrized, quasi-5.5 PN description of nonspinning point mass effects has also been derived in Ref. [18]. Tidal effects can be included up to relative 7.5 PN order [39,70,71], while quadratic-in-spin effects were included up to 3.5 PN [72]. Throughout the main body of this work, we will employ a 3.5 PN-accurate point-mass baseline, a 7.5 PN description of the tidal phasing, and a 3 PN description of spin-square effects.

`TEOBResumS` is a state-of-the-art EOB waveform model for spin-aligned coalescing compact binaries (either neutron stars or black holes) [72–77]. In this paper, we focus on the tidal sector of `TEOBResumS`, in the form described in Refs. [72,75,78]. In particular, this configuration coincides with the one implemented within `LALInference` [9]. The tidal sector of `TEOBResumS` contains contributions from the multipolar $\ell = 2, 3, 4$ gravitoelectric and $\ell = 2$ gravitomagnetic interactions; the former are included in resummed form stemming from PN and gravitational self-force results [78,79] (see also Refs. [80,81]). Equation-of-state-dependent self-spin effects (also known as quadrupole-monopole terms) are included at next-to-next-to-leading order [75] thanks to a suitable modification of the *centrifugal radius* introduced in Ref. [73], so as to incorporate even-in-spin effects in a way that closely mimics the structure of the Hamiltonian of a point particle on a Kerr black hole. In addition, the models rely on the (iterated) postadiabatic approximation [78,82,83] to compute the full inspiral waveform until about ten orbits before merger, which greatly reduces the computational burden of the waveform generator with negligible losses of accuracy.

These choices, together with rather different treatment of the spin sector and of resummation choices, distinguish `TEOBResumS` from the other state-of-the-art EOB approximant, `SEOBNR` [84,85]. We address the reader to Ref. [86] for a detailed investigation of the differences between the conservative point-mass dynamics of the models. To improve the computational efficiency of the waveform generation, when considering BNS systems, the `SEOBNR` family applies Gaussian process regression to the baseline model `SEOBNRv4T` [87–89]—which includes a description of dynamical tides, but no self-force information—so as to obtain `SEOBNRv4Tsurrogate` [90]. Note that EOB models are the most analytically complete to date, and contain higher-order PN information than that contained in Taylor-expanded PN approximants (e.g., many more test-particle terms at higher PN order as well as resummed tail factor). For this reason, the EOB framework can be Taylor-expanded so as to obtain waveform approximants at (partial) higher PN order than the current fully known 3.5 PN one [18,39,91].

`IMRPhenomPv2NRtidal` is a phenomenological spin precessing model for BNS systems based on the `IMRPhenomPv2` model. In the latter, an effective description of the point-mass waveform is obtained by fitting `SEOBNR-NR` hybrid waveforms⁴ to an analytical representation of the amplitude and phase of the frequency-domain 22 mode h_{22} [62,63]. This representation is further augmented by the `NRtidal` model [92], which provides a description of tidal effects based on a fit of hybrid waveforms composed of PN, `TEOBResumS`, and nonspinning $q \approx 1$ NR simulations.

Recently, Ref. [93] improved this model to `IMRPhenomPv2NRtidalv2` by incorporating a 7.5 PN-accurate low-frequency limit for the tidal sector of the phasing and PN-expanded spin-quadrupole interactions up to 3.5 PN in the waveform phase together with new fits for the amplitude tidal corrections. In this work, we will use both `IMRPhenomPv2NRtidal` and `IMRPhenomPv2NRtidalv2`, imposing that the individual spins are aligned to the orbital angular momentum.

A. Comparing waveform approximants

Let us now turn to discussing in some detail how the differences in the approximants reflect on the GW phase. This is the very first step to take towards the understanding of waveform systematics. In particular, by inspecting the gauge-invariant frequency-domain phase, it is possible to qualitatively assess how the different modeling of point-mass and tidal effects will impact the PE. PN approximants are overall less attractive at high frequencies than EOB models, thus leading to higher recovered tidal parameters.

⁴These waveforms are obtained by stitching together inspiral waveforms for the long inspiral to NR simulations that go through merger and ringdown.

Conversely, phenomenological models are more attractive than EOB, and lower tidal parameters are recovered during PE.

Given the plus and cross polarizations $h_+(t, \theta)$, $h_\times(t, \theta)$ associated with a specific approximant, we define the frequency-domain waveform $\tilde{h}(f, \theta) = \tilde{A}(f, \theta)e^{-i\Psi(f, \theta)}$, where $\tilde{A} = |\tilde{h}_+(f) + i\tilde{h}_\times(f)|$, $\Psi(f) = -\arg(\tilde{h}_+(f) + i\tilde{h}_\times(f))$, and $\tilde{h}_{+, \times}$ are the Fourier transforms of the time domain polarizations. Extracting information directly from the waveform phasing $\Psi(f)$ is complicated by the presence of an affine linear term $\phi_c + 2\pi t_c f$, which can be fixed arbitrarily. A better quantity to discuss waveform phasing is

$$Q_{\hat{\omega}} = \frac{\hat{\omega}^2}{\dot{\hat{\omega}}} = \frac{d\phi(t)}{d \ln \hat{\omega}}, \quad (25)$$

where $\hat{\omega} = 2\pi f M$ is the dimensionless GW frequency. The time-domain GW phase accumulated between two frequencies is given by

$$\phi_{(\omega_1, \omega_2)} = \int_{\hat{\omega}_1}^{\hat{\omega}_2} Q_{\hat{\omega}} d \ln \hat{\omega}. \quad (26)$$

Physically, $Q_{\hat{\omega}}$ is related to the phase acceleration, and the GW phase in the SPA is given by $\Psi''(w) = Q_{\hat{\omega}}(w)/w^2$. The inverse of $Q_{\hat{\omega}}$ is thus the adiabatic parameter whose magnitude controls the validity of the SPA [18,72,94]. Since there is no time/phase shift ambiguity and no necessity of alignment in phase plots with the $Q_{\hat{\omega}}$, the latter quantity is preferable with respect to the phase, because information can be lost in the alignment [82,94,95]. Thus, rather than comparing phase differences, we compute $Q_{\hat{\omega}}$ for the waveform approximants discussed above and extract information from $\Delta Q_{\hat{\omega}} = Q_{\hat{\omega}}^{\text{TEOBResumS}} - Q_{\hat{\omega}}^X$, where X is any other approximant.

Figure 2 shows the quantity $\Delta Q_{\hat{\omega}}$, computed for three reference waveforms with varying $\tilde{\Lambda}$ and zero spins and decomposed into its point-mass $\Delta Q_{\hat{\omega}}^{PM}$ and tidal $\Delta Q_{\hat{\omega}}^T$ contributions. The frequency range is roughly divided at the ‘‘cutoff’’ thresholds of the regimes at which point-mass ($\hat{\omega} < 0.02$) and tidal ($\hat{\omega} > 0.05$) effects are measured according to the Fisher matrix information formalism. During the early inspiral (first column), point-mass contributions dominate over tidal effects, and as expected, the phenomenological description of the inspiral is closer to TEOBResumS than the one offered by TaylorF2. When $0.02 \lesssim \hat{\omega} \lesssim 0.05$ (second column), the importance of tidal effects gradually increases, and the behavior of the two approximants starts differing significantly. Focusing on IMRPhenomPv2NRTidal, we observe that the largest contribution to $\Delta Q_{\hat{\omega}}$ comes from the tidal sector. As $\tilde{\Lambda}$ grows, both $\Delta Q_{\hat{\omega}}^T$ and $\Delta Q_{\hat{\omega}}$ become increasingly positive.

Therefore, matter effects in IMRPhenomPv2NRTidal are stronger than in TEOBResumS. Over the same range ($0.02 \lesssim \hat{\omega} \lesssim 0.05$), tidal terms of TaylorF2 behave in the exact opposite way. Increasing the value of $\tilde{\Lambda}$ leads to more negative $\Delta Q_{\hat{\omega}}^T$. For this approximant, then, matter effects are weaker than TEOBResumS. The trends shown in the intermediate range are maintained by both approximants also for $\hat{\omega} > 0.05$ and up to $\hat{\omega} \approx 0.10$, close to merger frequency (third column). We highlight that the point-mass terms of TaylorF2 grow monotonically, reflecting how the PN approximation breaks down at high frequencies. However, notably, the point mass contribution is positive—i.e., more attractive than TEOBResumS—and larger than or comparable to tidal corrections for moderate values of $\tilde{\Lambda}$. In GW parameter estimation, $\Delta Q_{\hat{\omega}}^{PM}$ then can partially compensate the negative $\Delta Q_{\hat{\omega}}^T$. Globally, IMRPhenomPv2NRTidal is more attractive than TEOBResumS, which implies that when recovering simulated TEOBResumS waveforms with IMRPhenomPv2NRTidal, one may expect to find lower values of $\tilde{\Lambda}$ than the ones injected. Instead, when recovering simulated TEOBResumS waveforms with TaylorF2, one may expect to find higher values of $\tilde{\Lambda}$ than the ones injected.

Spin effects are studied with a similar approach in Fig. 3, which shows $\Delta Q_{\hat{\omega}}$ computed for three waveforms with fixed $\tilde{\Lambda} = 400$ and varying magnitude of the dimensionless spins (χ_A, χ_B). We consider configurations with spins aligned to the orbital angular momentum and such that $\chi_A = \chi_B = \chi$. Focusing on the point-mass contribution, we observe that increasing χ does not impact significantly the magnitude of $\Delta Q_{\hat{\omega}}^{PM}$ for IMRPhenomPv2NRTidal. On the other hand, spin-induced effects are noticeably more repulsive in TaylorF2 than in TEOBResumS over the whole frequency range considered. Concerning $\Delta Q_{\hat{\omega}}^T$, we observe that the differences at $\hat{\omega} < 0.02$ are no longer negligible with respect to the point-mass contributions, and in general are larger than those found for nonspinning binaries. These differences can be attributed to the model of the spin-quadrupole terms. We recall that a spinning NS acquires a quadrupole moment due to its own rotation, which in turn causes a distortion of the gravitational field outside the body. The magnitude of such a quadrupole moment is an equation-of-state-dependent quantity, which can be parametrized through a coefficient C_Q [75,96]. The importance of this term in parameter estimation was pointed out in, e.g., Ref. [41], which showed how neglecting it can lead to biases on the recovery of the mass ratio and the total mass. Both TaylorF2 and IMRPhenomPv2NRTidal include these corrections only up to 3 PN, or next-to-leading order (NLO), whereas TEOBResumS also incorporates tail-dependent corrections in resummed form, as well as next-to-next-to-leading-order (NNLO) effects. The resummation weakens the effect

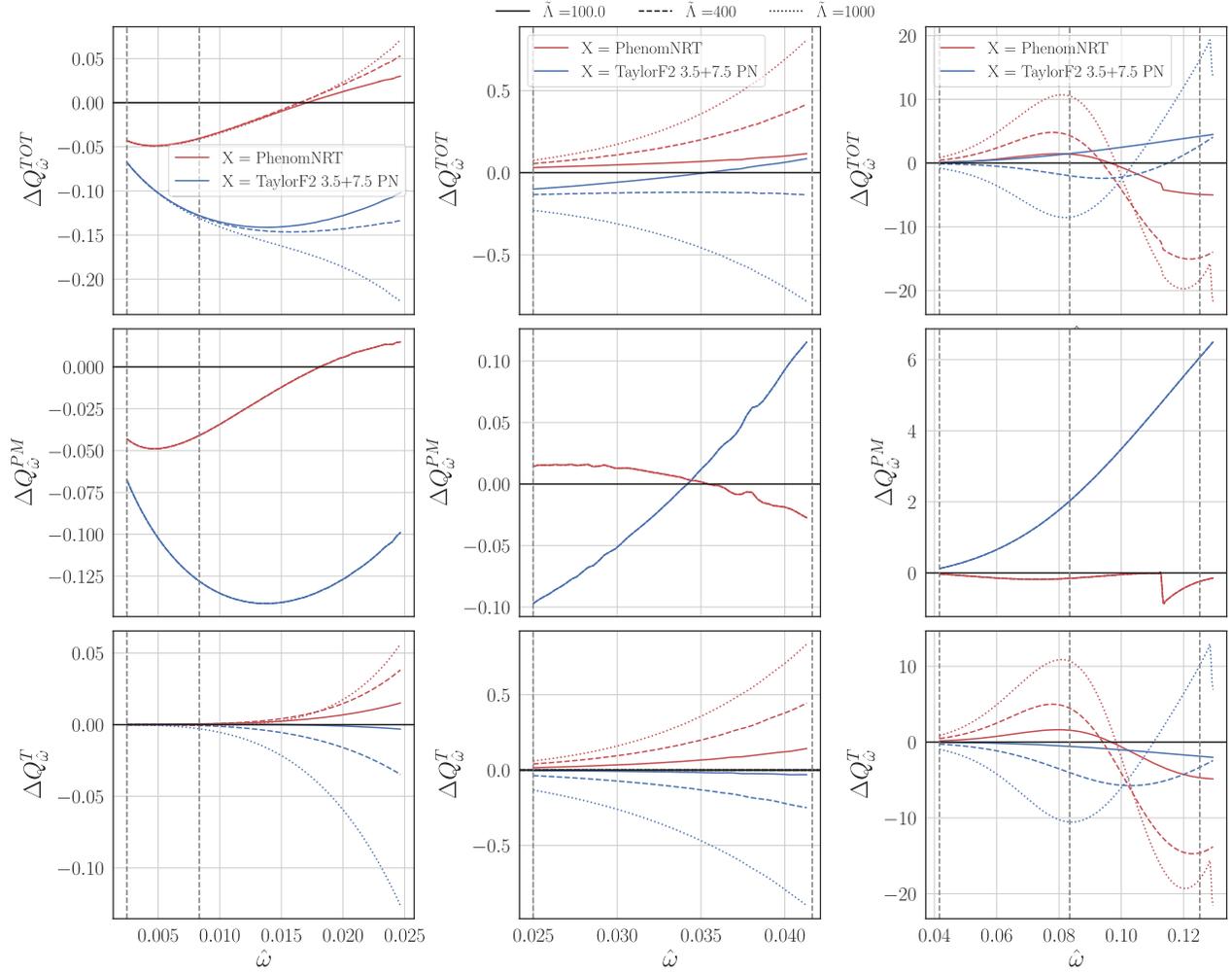


FIG. 2. The $\Delta Q_\omega = Q_\omega^{\text{TEOB}} - Q_\omega^X$ function computed for three waveforms with fixed spins $\chi_1 = \chi_2 = 0$ and varying $\tilde{\Lambda} = \{100, 400, 1000\}$, represented by continuous, dashed, and dotted lines, respectively. ΔQ_ω is then further decomposed into its point-mass $\Delta Q_\omega^{\text{PM}}$ (second row) and tidal $\Delta Q_\omega^{\text{T}}$ (third row) contributions, so that $\Delta Q_\omega^{\text{TOT}} = \Delta Q_\omega^{\text{PM}} + \Delta Q_\omega^{\text{T}}$. We display the same curves over three different frequency ranges and scales, roughly corresponding to the regimes in which point-mass effects are dominant, comparable or negligible with respect to tidal effects. We observe that $\Delta Q_\omega^{\text{T}}$ for TaylorF2 and IMRPhenomPv2NRTidal have opposite behaviors, with TaylorF2 being more repulsive and IMRPhenomPv2NRTidal more attractive than TEObResumS.

of quadrupole-monopole terms above $\hat{\omega} \approx 0.06$ [72,75]—i.e., above the frequency at which the NSs touch and hydrodynamical effects become relevant [97]. Note that the weaker effect of the (effective) EOS-dependent self-spin terms with respect to the PN expressions at high frequencies is also suggested by NR simulations [98], with the latter also suggesting stronger (effective) spin-orbit effects than PN.⁵

Overall, when considering injections of TEObResumS highly spinning waveforms, we expect IMRPhenomPv2NRTidal to underestimate tidal parameters, and TaylorF2 to overestimate them.

⁵But note that in the hydrodynamical regime it is, strictly speaking, not possible to interpret these as spin interactions and to compare them to PN.

IV. INJECTION STUDY

We present a full Bayesian PE study on 15 signals injected with TEObResumS and recovered with TaylorF2 and IMRPhenomPv2NRTidal. We interpret our results in light of the Q_ω analysis of Sec. III and find that it correctly indicates the behaviors of the studied waveform approximants.

A. Method

In order to study waveform systematics in a controlled environment, we generate artificial data (*injections*) using the TEObResumS model (with all higher modes up to $\ell = 8$) for 15 different nonspinning binary configurations, described by the intrinsic parameters $(m_A, m_B, \Lambda_A, \Lambda_B)$ and reported in Table I with the alternative representation

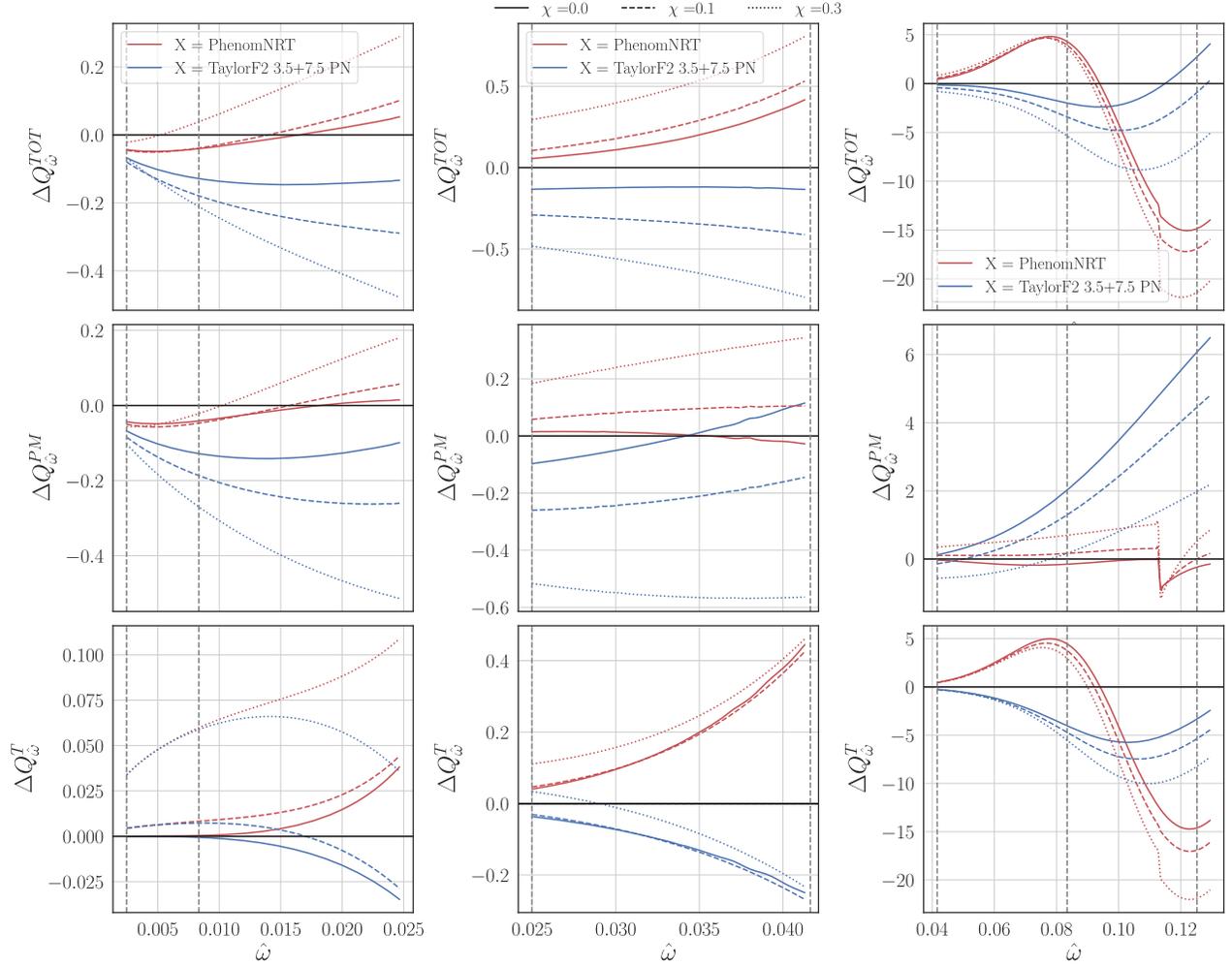


FIG. 3. The $\Delta Q_{\hat{\omega}}^{\text{TOT}} = Q_{\hat{\omega}}^{\text{TEOB}} - Q_{\hat{\omega}}^{\text{X}}$ function computed for three waveforms with fixed $\tilde{\Lambda} = 400$ and varying spins $\chi_1 = \chi_2 = \{0, 0.1, 0.3\}$, represented by continuous, dashed, and dotted lines, respectively. $\Delta Q_{\hat{\omega}}$ is then further decomposed into its point-mass $\Delta Q_{\hat{\omega}}^{\text{PM}}$ (second row) and tidal $\Delta Q_{\hat{\omega}}^{\text{T}}$ (third row) contributions, so that $\Delta Q_{\hat{\omega}}^{\text{TOT}} = \Delta Q_{\hat{\omega}}^{\text{PM}} + \Delta Q_{\hat{\omega}}^{\text{T}}$. All quantities are plotted over three different frequency ranges and scales. Note that $\Delta Q_{\hat{\omega}}^{\text{T}}$ is comparable to $\Delta Q_{\hat{\omega}}^{\text{PM}}$ at $\hat{\omega} < 0.02$. This effect can be attributed to the spin-spin interactions.

$(M, q, \tilde{\Lambda})$. The waveform polarizations are then projected on the three LIGO-Virgo detectors, locating the source at the sky position of GW170817.⁶ The injections are 64 s long with a sampling rate of 4096 Hz, and they are performed with zero noise configuration—i.e., no additional noise is included in the analyzed strains, in order to minimize the statistical fluctuations and to work in a framework as close as possible to the one described in Sec. II. We use Advanced LIGO and Advanced Virgo design amplitude spectral densities (ASD) [23–26]. The SNRs of the injected signals span a range from 82 to 94 (depending on the specific combination of masses and tidal parameters) that result in

⁶We use the maximum posterior values for sky location and distance from LVC analysis [1] combined with the information coming from Ref. [99].

louder signals than the current BNS observations [6,14]. For the estimation of the posterior distributions, we adopt the Bayesian framework offered by the `lalinferen` sampler as implemented in the software LSC Algorithm Library Suite (LALSuite) [9,100,101]. The waveform models used in the matched filtering analysis are the already described TaylorF2 and IMRPhenomPv2NRTidal. We perform two sets of injections, in part already discussed in Ref. [19]. In the first set, matter effects are modeled using two independent quadrupolar tidal parameters Λ_A, Λ_B . In the second set, we use the spectral parametrization of the EOS [3,102,103]. Within this framework, the EOS of cold dense NS matter is represented as a smooth function, parametrized in a four-dimensional space by the coefficients $(\gamma_0, \gamma_1, \gamma_2, \gamma_3)$. Each combination of these values specifies an adiabatic index $\Gamma(P)$:

TABLE I. Comparison between the properties of the injected signals and the recovered marginalized one-dimensional posteriors. For each simulation, we report medians and 90% credible regions. For each approximant and frequency range we additionally display the values of $\tilde{\Lambda}_E$ obtained as described in Sec. IV D.

| Injection | | IMRPhenomPv2_NRT | | | | | | | | | | | | | | | | | |
|------------------|------------------|------------------|--------------------------------|--|--|--------------------------------------|---------------------|--|--|--------------------------------------|---------------------|--|--|--------------------------------------|---------------------|--|--|--------------------------------------|---------------------|
| | | 1 kHz | | | | 2 kHz | | | | 1 kHz | | | | 2 kHz | | | | | |
| f_{cut} | M_{inj} | q_{inj} | $\tilde{\Lambda}_{\text{inj}}$ | M | q | $\tilde{\Lambda}$ | $\tilde{\Lambda}_E$ |
| 2B | 2.70 | 1.00 | 127 | 2.69 ^{+0.05} _{-0.01} | 0.86 ^{+0.12} _{-0.18} | 115 ⁺¹⁰⁵ ₋₇₆ | 116 | 2.69 ^{+0.05} _{-0.01} | 0.86 ^{+0.13} _{-0.18} | 98 ⁺⁵² ₋₅₂ | 117 | 2.69 ^{+0.06} _{-0.01} | 0.84 ^{+0.15} _{-0.18} | 150 ⁺¹⁴² ₋₁₀₆ | 136 | 2.69 ^{+0.06} _{-0.02} | 0.83 ^{+0.15} _{-0.18} | 106 ⁺⁹² ₋₇₂ | 130 |
| SLy | 3.00 | 1.00 | 191 | 2.99 ^{+0.06} _{-0.01} | 0.85 ^{+0.13} _{-0.18} | 132 ⁺⁸¹ ₋₇₄ | 168 | 2.99 ^{+0.06} _{-0.02} | 0.85 ^{+0.13} _{-0.18} | 143 ⁺⁴⁹ ₋₅₃ | 169 | 2.99 ^{+0.07} _{-0.02} | 0.83 ^{+0.15} _{-0.18} | 169 ⁺¹²⁴ ₋₁₀₅ | 307 | 3.00 ^{+0.07} _{-0.02} | 0.81 ^{+0.17} _{-0.18} | 158 ⁺⁸⁷ ₋₈₄ | 162 |
| LS220 | 3.20 | 1.00 | 202 | 3.19 ^{+0.06} _{-0.02} | 0.85 ^{+0.13} _{-0.18} | 132 ⁺⁶⁶ ₋₆₇ | 158 | 3.19 ^{+0.07} _{-0.02} | 0.85 ^{+0.14} _{-0.18} | 150 ⁺⁴⁷ ₋₅₄ | 158 | 3.19 ^{+0.08} _{-0.02} | 0.85 ^{+0.15} _{-0.18} | 168 ⁺¹⁰⁸ ₋₉₆ | 288 | 3.20 ^{+0.08} _{-0.03} | 0.81 ^{+0.17} _{-0.18} | 171 ⁺⁸⁶ ₋₈₃ | 274 |
| SFH0 | 2.92 | 1.00 | 252 | 2.91 ^{+0.06} _{-0.02} | 0.85 ^{+0.14} _{-0.18} | 172 ⁺⁹⁰ ₋₈₄ | 222 | 2.91 ^{+0.06} _{-0.02} | 0.84 ^{+0.14} _{-0.18} | 185 ⁺⁵⁶ ₋₅₉ | 224 | 2.92 ^{+0.07} _{-0.02} | 0.83 ^{+0.15} _{-0.18} | 233 ⁺¹³⁵ ₋₁₂₇ | 257 | 2.91 ^{+0.07} _{-0.02} | 0.81 ^{+0.17} _{-0.18} | 233 ⁺⁹⁶ ₋₉₇ | 253 |
| DD2 | 3.18 | 1.00 | 332 | 3.17 ^{+0.07} _{-0.02} | 0.84 ^{+0.14} _{-0.18} | 218 ⁺⁷³ ₋₇₆ | 257 | 3.18 ^{+0.07} _{-0.02} | 0.81 ^{+0.17} _{-0.16} | 248 ⁺⁵⁵ ₋₆₁ | 297 | 3.17 ^{+0.08} _{-0.02} | 0.82 ^{+0.16} _{-0.18} | 311 ⁺¹¹⁸ ₋₁₁₅ | 345 | 3.18 ^{+0.08} _{-0.03} | 0.81 ^{+0.17} _{-0.18} | 335 ⁺⁹⁶ ₋₉₉ | 343 |
| SFH0 | 2.80 | 1.00 | 334 | 2.79 ^{+0.06} _{-0.01} | 0.84 ^{+0.14} _{-0.18} | 228 ⁺¹⁰⁶ ₋₁₀₃ | 285 | 2.79 ^{+0.06} _{-0.01} | 0.84 ^{+0.14} _{-0.18} | 248 ⁺⁷⁵ ₋₇₅ | 286 | 2.79 ^{+0.07} _{-0.02} | 0.83 ^{+0.16} _{-0.19} | 319 ⁺¹⁵⁶ ₋₁₅₃ | 462 | 2.79 ^{+0.07} _{-0.02} | 0.81 ^{+0.17} _{-0.18} | 342 ⁺¹¹³ ₋₁₁₈ | 444 |
| ALF2 | 3.00 | 1.00 | 382 | 2.99 ^{+0.07} _{-0.02} | 0.84 ^{+0.14} _{-0.18} | 258 ⁺⁸⁸ ₋₈₈ | 330 | 3.00 ^{+0.07} _{-0.02} | 0.81 ^{+0.17} _{-0.17} | 280 ⁺⁷² ₋₇₂ | 330 | 2.99 ^{+0.07} _{-0.02} | 0.83 ^{+0.16} _{-0.19} | 370 ⁺¹³⁸ ₋₁₃₃ | 467 | 3.00 ^{+0.07} _{-0.03} | 0.81 ^{+0.17} _{-0.17} | 400 ⁺¹⁰⁹ ₋₁₁₀ | 457 |
| SLy | 2.68 | 1.00 | 401 | 2.67 ^{+0.06} _{-0.01} | 0.84 ^{+0.14} _{-0.18} | 286 ⁺¹²⁰ ₋₁₂₄ | 349 | 2.67 ^{+0.06} _{-0.01} | 0.84 ^{+0.14} _{-0.18} | 295 ⁺⁷² ₋₈₀ | 350 | 2.68 ^{+0.07} _{-0.02} | 0.85 ^{+0.15} _{-0.19} | 397 ⁺¹⁷³ ₋₁₆₉ | 434 | 2.68 ^{+0.07} _{-0.02} | 0.82 ^{+0.16} _{-0.19} | 423 ⁺¹²² ₋₁₂₄ | 430 |
| SLy | 2.69 | 0.88 | 401 | 2.68 ^{+0.06} _{-0.02} | 0.83 ^{+0.15} _{-0.18} | 290 ⁺¹²³ ₋₁₂₄ | 358 | 2.68 ^{+0.06} _{-0.02} | 0.83 ^{+0.16} _{-0.18} | 294 ⁺⁷⁷ ₋₈₅ | 313 | 2.68 ^{+0.07} _{-0.02} | 0.81 ^{+0.17} _{-0.18} | 404 ⁺¹⁷⁶ ₋₁₆₉ | 557 | 2.68 ^{+0.07} _{-0.02} | 0.81 ^{+0.17} _{-0.18} | 426 ⁺¹²⁷ ₋₁₂₅ | 386 |
| SFH0 | 2.72 | 0.88 | 412 | 2.70 ^{+0.06} _{-0.02} | 0.83 ^{+0.15} _{-0.19} | 299 ⁺¹¹⁹ ₋₁₁₉ | 341 | 2.70 ^{+0.06} _{-0.02} | 0.82 ^{+0.16} _{-0.18} | 304 ⁺⁷⁶ ₋₈₂ | 350 | 2.71 ^{+0.07} _{-0.02} | 0.82 ^{+0.17} _{-0.19} | 416 ⁺¹⁷² ₋₁₆₆ | 440 | 2.71 ^{+0.07} _{-0.02} | 0.81 ^{+0.17} _{-0.18} | 439 ⁺¹³² ₋₁₃₁ | 435 |
| SFH0 | 2.71 | 1.00 | 413 | 2.70 ^{+0.06} _{-0.01} | 0.84 ^{+0.15} _{-0.18} | 293 ⁺¹¹⁷ ₋₁₁₇ | 388 | 2.70 ^{+0.06} _{-0.02} | 0.83 ^{+0.15} _{-0.18} | 306 ⁺⁷⁵ ₋₈₄ | 352 | 2.71 ^{+0.07} _{-0.02} | 0.82 ^{+0.16} _{-0.18} | 409 ⁺¹⁷⁴ ₋₁₆₉ | 453 | 2.71 ^{+0.07} _{-0.02} | 0.82 ^{+0.16} _{-0.18} | 443 ⁺¹²⁴ ₋₁₂₄ | 449 |
| LS220 | 2.69 | 0.86 | 714 | 2.68 ^{+0.07} _{-0.02} | 0.81 ^{+0.16} _{-0.18} | 532 ⁺¹³⁷ ₋₁₄₇ | 633 | 2.69 ^{+0.07} _{-0.03} | 0.78 ^{+0.19} _{-0.16} | 532 ⁺⁹⁹ ₋₁₀₈ | 589 | 2.68 ^{+0.07} _{-0.02} | 0.81 ^{+0.17} _{-0.19} | 764 ⁺¹⁹¹ ₋₁₉₅ | 756 | 2.68 ^{+0.06} _{-0.02} | 0.82 ^{+0.16} _{-0.18} | 856 ⁺¹⁴⁹ ₋₁₅₈ | 756 |
| LS220 | 2.68 | 1.00 | 715 | 2.68 ^{+0.06} _{-0.02} | 0.83 ^{+0.15} _{-0.19} | 525 ⁺¹³⁵ ₋₁₄₀ | 660 | 2.68 ^{+0.06} _{-0.02} | 0.79 ^{+0.19} _{-0.16} | 528 ⁺⁸⁷ ₋₁₀₄ | 613 | 2.68 ^{+0.06} _{-0.02} | 0.82 ^{+0.16} _{-0.18} | 758 ⁺¹⁸⁸ ₋₁₉₅ | 754 | 2.67 ^{+0.06} _{-0.02} | 0.84 ^{+0.15} _{-0.18} | 844 ⁺¹⁴⁵ ₋₁₄₄ | 758 |
| DD2 | 2.71 | 1.00 | 840 | 2.70 ^{+0.06} _{-0.02} | 0.83 ^{+0.15} _{-0.18} | 628 ⁺¹³⁰ ₋₁₄₄ | 758 | 2.71 ^{+0.06} _{-0.03} | 0.78 ^{+0.19} _{-0.16} | 622 ⁺⁹³ ₋₁₁₂ | 711 | 2.70 ^{+0.06} _{-0.02} | 0.83 ^{+0.15} _{-0.18} | 916 ⁺¹⁸⁸ ₋₁₉₂ | 918 | 2.70 ^{+0.06} _{-0.02} | 0.84 ^{+0.15} _{-0.18} | 1011 ⁺¹⁴⁴ ₋₁₅₂ | 920 |
| DD2 | 2.48 | 1.00 | 1366 | 2.47 ^{+0.05} _{-0.02} | 0.83 ^{+0.15} _{-0.16} | 1104 ⁺¹⁸⁶ ₋₁₈₉ | 1269 | 2.48 ^{+0.04} _{-0.02} | 0.78 ^{+0.19} _{-0.12} | 1057 ⁺¹⁴¹ ₋₁₃₈ | 1170 | 2.47 ^{+0.04} _{-0.01} | 0.84 ^{+0.14} _{-0.17} | 1542 ⁺²⁴⁶ ₋₂₄₅ | 1477 | 2.47 ^{+0.04} _{-0.01} | 0.85 ^{+0.13} _{-0.17} | 1668 ⁺²⁰⁰ ₋₂₀₇ | 1484 |

$$\Gamma(P) = \exp \left[\sum_{k=0}^3 \gamma_k \log(P/P_0)^k \right], \quad (27)$$

where P_0 is some reference pressure. The adiabatic index is by definition related to the pressure-density function $P(\rho)$ through $\Gamma = \rho \frac{d \ln P}{d \rho}$. The complete EOS is then built by fixing the low-density sector ($P < P_0$) to the SLy description, and integrating the differential equation for $\rho(P)$ implied by the definition of Γ in the core of the NS ($P > P_0$). Once the EOS is fixed, it is possible to calculate the tidal polarizability parameters $\Lambda_{A,B}$, which are then used to model the tidal effects in the waveforms. These analyses give a posterior distribution for the coefficients γ_i , which can be mapped into the EOSs and radii of the merging NSs. However, this method assumes implicitly that both NSs are described by the same EOS and that no strong first-order phase transitions happen in the core of the NS.

For both of the previous methods, the analyses are performed with two different maximal frequencies, $f_{\max} = 1$ kHz and $f_{\max} = 2$ kHz—i.e., in frequency ranges $f \in [23 \text{ Hz}, 1 \text{ kHz}]$ and $f \in [23 \text{ Hz}, 2 \text{ kHz}]$ —in order to study whether the extension to the higher-frequency cutoff increases the differences between approximants. The priors distributions are flat in mass components, in a range corresponding to $\mathcal{M}_c \in [1.0, 2.2] M_\odot$ and $q \in [1, 8]$. We use an aligned-spin configuration with isotropic priors on the spin components and $a_{i,z} \in [-0.05, +0.05]$, $i = A, B$. Regarding the tidal parameters, the prior distributions are uniform in the free parameters involved in the analysis: when we adopt the EOS-insensitive description, $p(\Lambda_i) \propto 1$ in the range $\Lambda_i \in [0, 5000]$ for $i = A, B$; while for the spectral parametrization cases, the prior distribution is uniform in the spectral parameters in the ranges $\gamma_0 \in [0.2, 2]$, $\gamma_1 \in [-1.6, 1.7]$, $\gamma_2 \in [-0.6, 0.6]$, $\gamma_3 \in [-0.02, 0.02]$, and additionally $\Gamma(P)$ is constrained to be in the range $[0.5, 4.5]$. This setup is identical to the one proposed in Ref. [3]. In comparison to previous studies, we employ a larger set of simulated signals, in order to better understand the behavior of the studied approximants when different sources are considered [15,16].

In the remainder of this section, we (i) examine the measurement of mass and spin parameters; (ii) discuss the systematic effects that different approximants induce in the recovered tidal parameters, NS radii, and EOS reconstruction; and (iii) apply the faithfulness criteria previously described to our data.

B. Masses, mass ratio, and spins

We first discuss the determination of the nontidal parameters. Figure 4 shows the recovered posterior distributions of the total mass M and mass ratio q parameters obtained with low-spin priors. The estimates obtained are consistent between the different approximants and frequency cutoffs, and with the injected signals, with the real

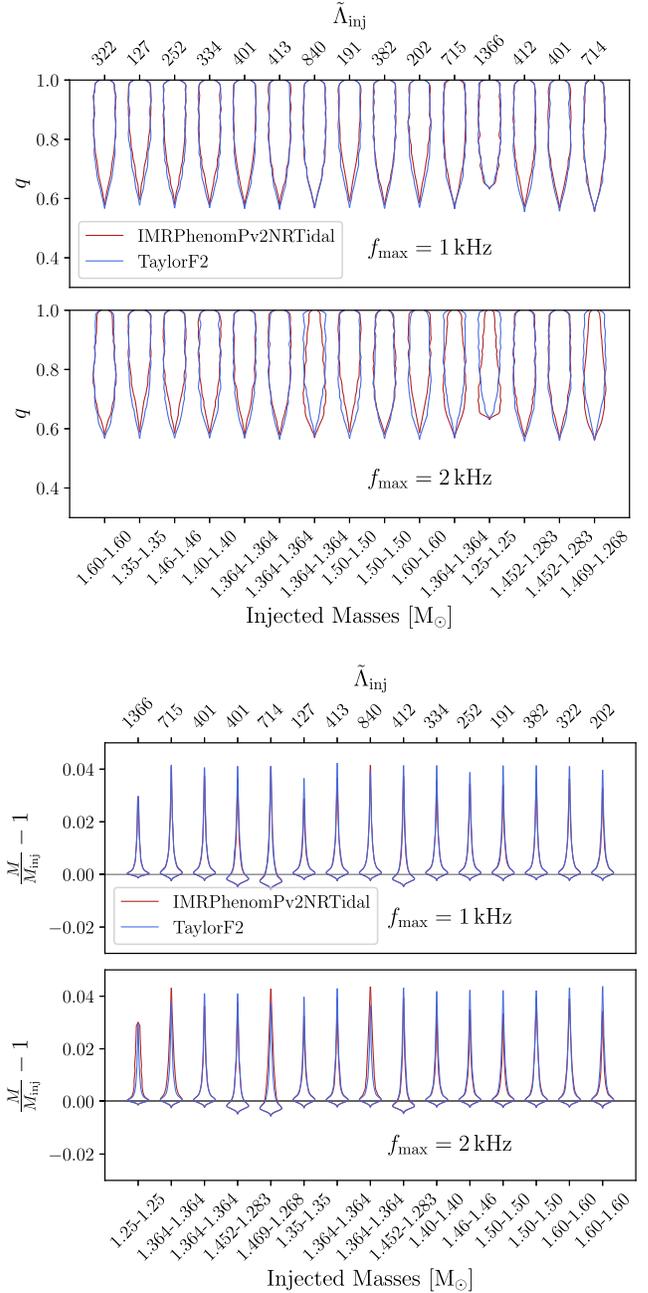


FIG. 4. Distributions of the mass ratio q (top) and the deviation from the injected total mass M_{inj} (bottom), displayed for all the simulated signals of Table I. The distributions recovered are consistent between different approximants and frequency ranges. The unequal mass signals cannot be distinguished from their equal-mass counterpart.

values always falling inside the 90% credible intervals. This indicates that the systematic differences in phasing observed at low frequencies (see the first column of Fig. 2) are smaller than statistical uncertainties. We find that the injected unequal-mass signals (with $q = 0.86$ and $q = 0.88$) cannot be distinguished from the equal-mass ones. This can partly be attributed to the known existing

correlation between mass ratio and spin parameters [37]. In PN waveforms, the leading-order spin interactions are described by the parameter β , given by [37,68]

$$\beta = \chi_{\text{eff}} - \frac{38\nu}{113}(\chi_1 + \chi_2), \quad (28)$$

where

$$\chi_{\text{eff}} = \frac{m_1\chi_1 + m_2\chi_2}{M} \cdot \hat{\mathbf{L}} \quad (29)$$

is the mass-weighted sum of the component spin parameters, and is often used during PE as a measure of the collective spin of the binary, as it is a conserved quantity of the orbit-averaged precession equations over precession timescales [104]. A Fisher matrix analysis reveals that spin parameters, which at leading order have $p = -4$ in the notation of Sec. II, are measured over a very similar range of frequencies to the (symmetric) mass ratio ν [39,41], and so they are strongly correlated. In more detail, positive-aligned spins have a repulsive effect on the binary

dynamics. By contrast, decreasing the symmetric mass ratios (i.e., more unequal-mass systems) accelerates the coalescence. The two effects are thus in direct competition, and spin effects can be reproduced by varying ν [105]. As a consequence, widening the spin priors leads to larger mass ratio distributions. Hence, different prior assumptions on mass ratio and component spins can lead to very different posterior distributions, and are of key importance when interpreting the data. In Appendix A, from Eq. (A5), we see that this correlation may also reflect on the estimate of $\tilde{\Lambda}$ even in the case of high-SNR signals, leading to an overall broadening of the $\tilde{\Lambda}$ posteriors.

C. Tidal parameter and NS radius

We now discuss systematics in the inference on tidal parameters and the effect on constraints on the NS radius. Figure 5 shows the posterior distributions of the tidal parameters $\tilde{\Lambda}$ recovered through PN and phenomenological approximants. The values coming from the posterior samples are rescaled by the *true* injected value, adopting the auxiliary parameter

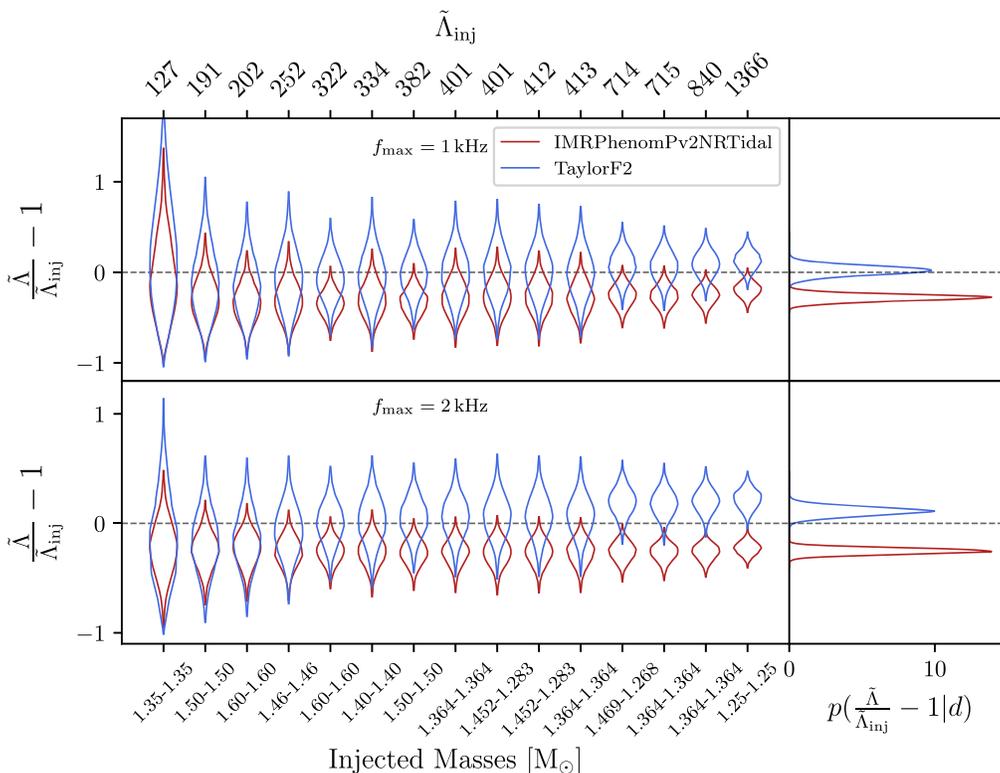


FIG. 5. Main panel: violin plots of the fractional deviation between the injected values of $\tilde{\Lambda}$ and the recovered posteriors. The color code depends on the approximant employed for the PE (red for IMRPhenomPv2NRTidal, blue for TaylorF2), and the results are displayed for two different frequency cutoffs f_{max} , 1 kHz (top) and 2 kHz (bottom). As matter effects grow, the deviation between the two approximants and the TEOBResumS baseline increases, reaching approximately $\pm 20\%$ when $\tilde{\Lambda}^{\text{inj}} = 1366$. In the right panel, we display the combined posterior distribution of the fractional deviation, reweighted as described in the text. An increase in f_{max} makes the differences larger in the recovery of $\tilde{\Lambda}$ between models. This is especially true for TaylorF2: PN waveform models are known to become less accurate close to merger frequencies.

$$\varepsilon_{\tilde{\Lambda}}^{(i)} = \frac{\tilde{\Lambda}^{(i)}}{\tilde{\Lambda}_{\text{inj}}^{(i)}} - 1, \quad (30)$$

which encodes the fractional deviation from the injected value for each simulated signal i . We observe that, as the injected values of $\tilde{\Lambda}$ increase, the relative uncertainties of the recovered posterior distributions decrease and modeling differences become more relevant (the median of the distributions are shifted with respect to zero). The combination of these two effects leads to evident biases in the recovered values. The overall bias due to waveform effects is quantified by the combined posterior distribution $p(\varepsilon_{\tilde{\Lambda}}|d)$ shown in the right panel of Fig. 5. This quantity is estimated by weighting each posterior distribution $p(\varepsilon_{\tilde{\Lambda}}^{(i)}|d^{(i)})$ by the respective prior distribution $p(\varepsilon_{\tilde{\Lambda}}^{(i)})$, computed from the prior distributions for $\tilde{\Lambda}^{(i)}$. The result is multiplied by the prior distribution $p(\varepsilon_{\tilde{\Lambda}})$ for the combined parameter $\varepsilon_{\tilde{\Lambda}}$, taken as uniform in the range $[-2, +2]$, i.e.,

$$p(\varepsilon_{\tilde{\Lambda}}|d) = p(\varepsilon_{\tilde{\Lambda}}) \prod_i \frac{p(\varepsilon_{\tilde{\Lambda}}^{(i)}|d^{(i)})}{p(\varepsilon_{\tilde{\Lambda}}^{(i)})}, \quad (31)$$

where the index i runs over all the injected binaries. We find that IMRPhenomPv2NRtidal systematically recovers lower values than those injected with TeOBResumS, while TaylorF2 tends to systematically overestimate tidal parameters as matter effects increase, although it is able to capture the injected values for $\tilde{\Lambda} \leq 400$. These results can be understood in terms of the Q_ω analysis of Sec. III, coupled to the relevant frequency ranges computed and discussed in Appendix B. To summarize, the analyzed signals contain useful information up to approximately 1 kHz, depending on the source parameters. We are then consistently in the situation where f_{thr} is larger than $f_{5\%}^{\tilde{\Lambda}}$, whose values lie around 240–300 Hz. Then, as shown in the third column of Fig. 2, for IMRPhenomPv2NRtidal systematic differences in $\tilde{\Lambda}$ are dominated by the tidal sector, which is more attractive than TeOBResumS and leads to lower estimates of $\tilde{\Lambda}$. The attractive point-mass contribution of TaylorF2, instead, leads to slight underestimates of the tidal parameters for values of $\tilde{\Lambda} \approx 100$, while for $\tilde{\Lambda} \approx 400$ it compensates the tidal sector. The latter dominates for larger values of $\tilde{\Lambda}$, and—being too repulsive—causes overestimates of matter effects.

Translating information on the tidal parameters of a NS into information on the NS EOS and radius is not straightforward. Given that waveform models do not explicitly depend on the NS radius, it is not possible to directly extract R from GW data. It is necessary, instead, to rely on either some representation of the EOS [102, 106–110], or on quasi-universal (EOS-insensitive) relations

TABLE II. Summary of the quasiuniversal relations used in Sec. IV C. While the De *et al.* (De) relation immediately links the radius of a $1.4 M_\odot$ NS to the mass-weighted tidal deformability and chirp mass of a BNS system, the Yagi-Yunes (YY) and Raithel *et al.* (R) relations require the numerical inversion of $\tilde{\Lambda}(R, q)$. For more detail, see Appendix D.

| Shorthands | References | Expressions |
|----------------------------|------------|-----------------|
| De (De <i>et al.</i>) | [111] | Eq. (D1) |
| YY (Yagi and Yunes) | [113,114] | Eqs. (D2), (D4) |
| R (Raithel <i>et al.</i>) | [112] | Eq. (D5) |

(URs), which phenomenologically link macroscopic quantities of the binary between each other. In particular, we employ the spectral parametrization of Refs. [102,108] and the EOS universal relations of De and Lattimer [111], of Raithel *et al.* [112], and of Yagi and Yunes [113,114]. The EOS-insensitive relations used here are summarized in Appendix D and Table II.

In the reminder of this subsection, we focus on the implication of waveform systematics on the recovery of the NS radii and EOS reconstruction. We additionally gauge the further biases that can be introduced by employing quasiuniversal relations for the recovery of R . To do so, we apply the above URs to the analyses performed by sampling the component tidal parameters Λ_i independently of each other, *as well as* (spectral) parametrized EOS runs. Indeed, the parametrized posterior EOSs obtained are usually employed in conjunction with the component mass posteriors m_i to solve the TOV structure equations and obtain a direct estimate of R . At the same time, however, given an EOS and the component masses, it is possible to compute $\tilde{\Lambda}$, apply some URs, and obtain another—in principle equivalent—estimate of R . This allows for a direct comparison of the effects of using universal relations in place of parametrized EOS runs, independently of the choice of the sampling parameters (and, therefore, of the implied priors on $\tilde{\Lambda}$). Figure 6 shows the distributions of the deviation in the estimates of $R_{1.4M_\odot}$ (top panel) and of the radius of the lighter star R_2 (bottom panel) with respect to the real radii values corresponding to the parameters and EOSs listed in Table I. For a fixed approximant and distribution of $\tilde{\Lambda}$, URs tend to recover smaller values of R_2 than those obtained by solving the TOV equations with the spectral EOSs. The same observation holds for $R_{1.4M_\odot}$, although the estimates are overall more consistent between the two methods. The errors due purely to waveform modeling can instead be gauged by comparing distributions of R_2 and $R_{1.4M_\odot}$ obtained with different approximants and the same UR. We find that IMRPhenomPv2NRtidal tends to underestimate R , while TaylorF2 behaves in the opposite way. The difference between the estimates offered by TaylorF2 and IMRPhenomPv2NRtidal becomes more relevant as tidal effects grow, mirroring the behavior

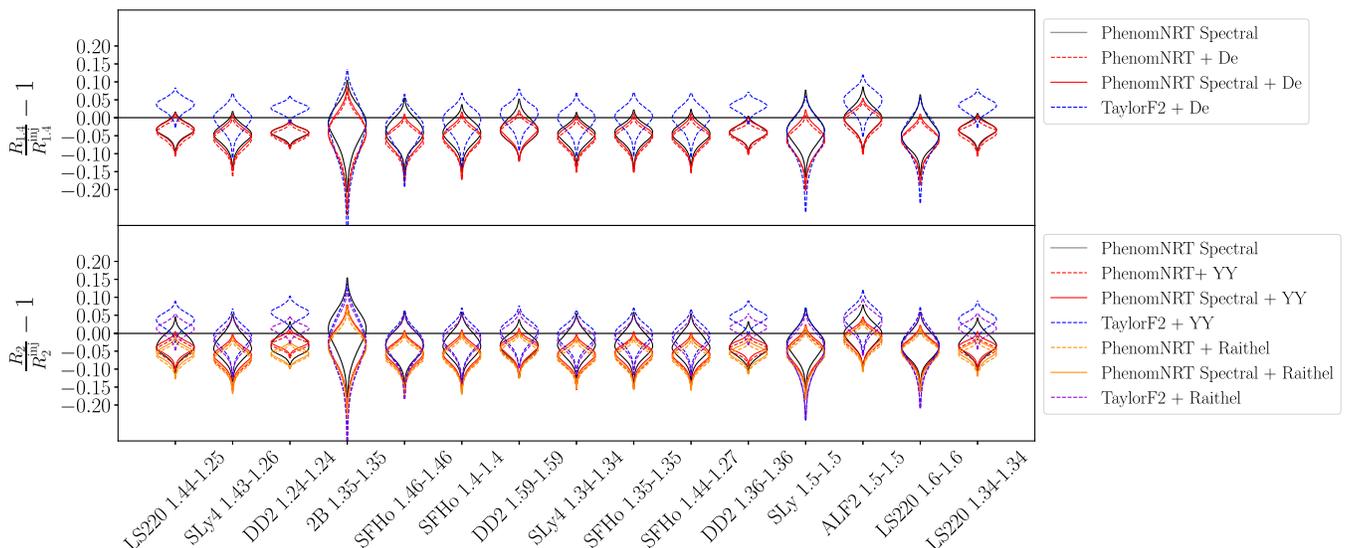


FIG. 6. Fractional deviation between the real “injected” values R_{inj} , computed for each signal listed in Table I, and the distributions of R computed either by using the universal relations of Table II (colored lines) or by considering the parametrized EOS posteriors and solving the NS structure equations (black lines). Whenever posteriors from the parametrized EOS (“spectral”) runs are involved, we employ solid lines. Conversely, whenever we use posteriors from analyses performed by sampling Λ_A and Λ_B independently, we use dashed lines. The under/over estimates displayed by IMRPhenomPv2NRTidal and TaylorF2 in Fig. 5 translate into similar biases on R , amounting up to $\pm 5\%$.

of the $\tilde{\Lambda}$ distributions. The overall differences in radii are then given by the combination of waveform systematics and systematics coming from the method chosen to map the intrinsic parameter of the source into the radii. The former are dominant for large values of $\tilde{\Lambda}$, whereas for smaller values of $\tilde{\Lambda}$ the two are comparable. In particular, this means that some combinations of approximant and mapping method may attenuate (or exacerbate) the cumulative error, depending on whether they tend to over- or underestimate R . In our study, with an injected EOB waveform, we find that while using URs negatively impacts IMRPhenomPv2NRTidal analyses, TaylorF2 runs would gain from employing URs rather than a parametrized spectral analysis. The overall bias can amount to approximately $\pm 5\%$ between TEOBResumS and PN/phenomenological waveforms, and 10% between IMRPhenomPv2NRTidal and TaylorF2.

D. Faithfulness thresholds and PE biases

Finally, we apply the accuracy criteria of Sec. II to our data and show that, while criteria based on faithfulness alone are of little use to predicting the presence of biases, an estimate of the parameter bias can be obtained using Eq. (24). We begin by computing the unfaithfulness \mathcal{F} between waveform models evaluated with the same set of true parameters θ_{inj} through Eq. (5). We place all sources in GW170817’s sky location, and employ the analytical aLIGODesignSensitivityP1200087 PSD [23], provided by pycbc [115]. The results are summarized in Fig. 7. We find that both TaylorF2 and

IMRPhenomPv2NRTidal give values largely above the nominal threshold of $\bar{\mathcal{F}} = 0.03$ (not shown in the figure), which corresponds to $\sim 10\%$ of detection losses [36,37]. When considering the thresholds provided by Eq. (19) in its weaker formulation ($\epsilon^2 = 6$), we find that most signals fall above the value corresponding to

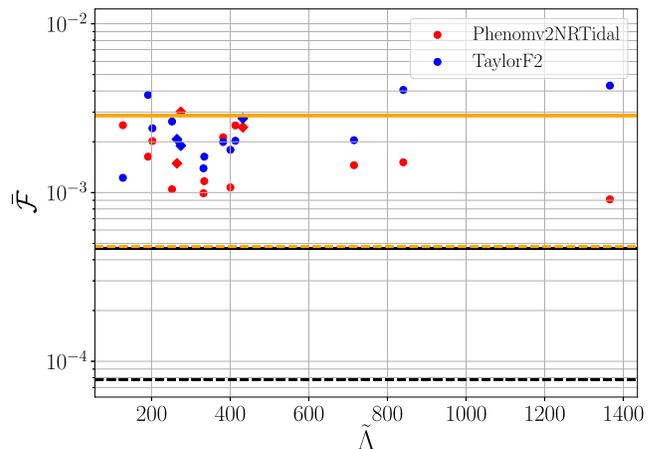


FIG. 7. Unfaithfulness values between the IMRPhenomPv2-NRTidal (red) or TaylorF2 (blue) waveforms and the injected TEOBResumS waveforms. The plus and cross polarizations are projected onto the Livingston detector, and the source is fixed at the GW170817 sky location. Horizontal straight lines correspond to the threshold values of $\bar{\mathcal{F}}$ obtained with SNRs of 80 (black) and 32.4 (orange), computed through Eq. (19) with $\epsilon = 6$ (straight) and $\epsilon = 1$ (dashed). Dots correspond to waveforms with $q = 1$, while diamonds correspond to signals with $q \neq 1$.

GW170817's network SNR (straight orange line), and largely below the threshold corresponding to a network SNR or 80 (straight black line)—i.e., the SNR above which all of our injections are performed. By tightening the constraints and enforcing $\epsilon^2 = 1$ (dashed lines), we find that already at the SNR of GW170817, none of the considered waveforms are faithful enough to ensure that no waveform systematics will be observed. However, not all our injections give largely biased or incompatible results. These facts indicate that these criteria give necessary but not sufficient rules to identify biases and highlight the strong dependence of the criteria themselves on the arbitrarily chosen value of ϵ^2 .

To test the procedure outlined in Sec. II B, we obtain an estimate of the biased values $\tilde{\Lambda}_E$ by applying Eq. (24), and minimizing the quantity $\sum_i (h_i^{\text{TEOBResumS}} - h_i^{\text{IMRPhenomPv2NRTidal, TaylorF2}} | h_i^{\text{TEOBResumS}} - h_i^{\text{IMRPhenomPv2NRTidal, TaylorF2}})$, where the sum is performed over the network interferometers considered (Livingston, Hanford and Virgo, in our case). In particular, for each waveform $h^{\text{IMRPhenomPv2NRTidal, TaylorF2}}$, we fix the intrinsic parameters (m_A, m_B, χ_A, χ_B) to their real injected values, and vary $\Lambda_A = \Lambda_B$ over the one-dimensional interval $\tilde{\Lambda} \in [\min(0, \tilde{\Lambda}_{\text{inj}} - 500), \tilde{\Lambda}_{\text{inj}} + 500]$. The simplifying choice of imposing $\Lambda_A = \Lambda_B$ can be justified by considering that in our injection study we were unable to distinguish $q = 1$ from $q \neq 1$ systems. While this might not be true for more asymmetric systems than those studied in the present paper, the issue can be easily circumvented by employing Binary-Love universal relations [113]. The straightforward procedure described leads to the values displayed in Table I. We find that the $\tilde{\Lambda}_E$ values computed, while often slightly overestimated with respect to the medians of the distributions of the tidal parameters recovered through PE, fall into the 90% $\tilde{\Lambda}$ credible limits in the large majority of cases, thus providing a good approximation of the overall behavior of the approximants employed. Due to the overestimate of $\tilde{\Lambda}_E$, the bias $\Delta\tilde{\Lambda}_B = |\tilde{\Lambda}_E - \tilde{\Lambda}_{\text{inj}}|$ is larger than the real bias $\Delta\tilde{\Lambda}_{\text{true}} = |\tilde{\Lambda}^{\text{median}} - \tilde{\Lambda}_{\text{inj}}|$ for the TaylorF2 approximant, and smaller for IMRPhenomPv2NRTidal. Estimates of waveform systematics based on the above method might then be slightly optimistic (pessimistic) when comparing TEOBResumS to IMRPhenomPv2NRTidal (TaylorF2).

V. GW170817

We now apply the approach developed and tested in the previous sections to the analysis of GW170817.

We perform a Bayesian analysis of GW170817 using the IMRPhenomPv2NRTidal, TaylorF2, and TEOBResumS approximants, involving `pbilby` [116]. We adopt an almost identical configuration to the one presented in Ref. [117] (see also Ref. [6]). In more detail,

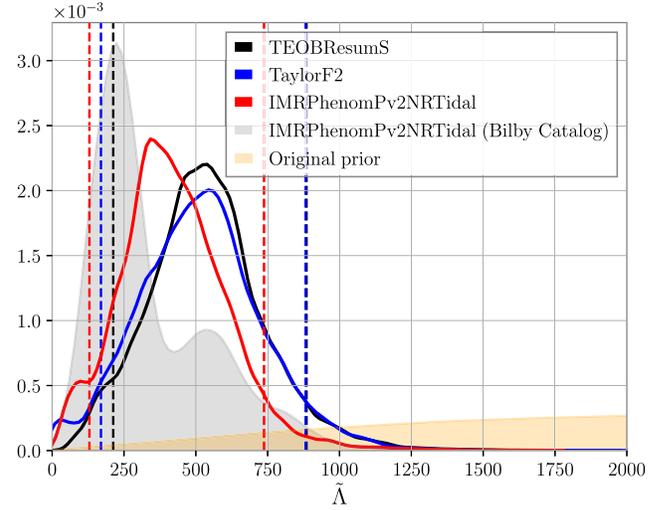


FIG. 8. Analysis of GW170817 data. Marginalized one-dimensional $\tilde{\Lambda}$ posteriors, obtained by analyzing the data up to $f_{\text{max}} = 1024$ Hz with three approximants: TaylorF2 (blue), IMRPhenomPv2NRTidal (red), and TEOBResumS (black). The posteriors shown are reweighted to flat in $\tilde{\Lambda}$ prior, as is done in, e.g., Ref. [6]. The public `bilby` posteriors from Ref. [117] (gray) are also displayed. Note that the `bilby` analysis uses $f_{\text{max}} = 2048$ Hz.

we consider a strain of 128 s around the GPS time 1187008882.43 s. Data are downloaded directly from the GWOSC [118], in its cleaned and deglitched version (v2). We employ the PSDs provided by Ref. [6], and we fix the sky location to the one provided by EM constraints. Further, as we are mainly interested in estimating the intrinsic parameters of the source, we marginalize over distance, time, and phase. The sampling is performed with uniform priors in chirp mass $\mathcal{M} \in [1.18, 1.21] M_{\odot}$ and mass ratio $q \in [0.125, 1]$, with the additional constraints $m_A, m_B \in [1.001, 4.314] M_{\odot}$. The quadrupolar tidal coefficients Λ_A, Λ_B are uniformly sampled in the interval $[0, 5000]$. The main differences with respect to the analysis of Ref. [117] lie in (i) the different spin priors employed, which are taken to be aligned to the orbital angular momentum and such that $(\chi_A, \chi_B) \in [-0.05, 0.05]$, and (ii) the high-frequency cutoff of 1024 Hz that we impose (instead of the 2048 Hz of Ref. [117]).

Using the formalism of the Fisher matrix outlined in Sec. II A, we investigate in which frequency region the tidal information is effectively extracted, in accordance with the extracted posterior samples: the Fisher matrix element $I_{\tilde{\Lambda}\tilde{\Lambda}}$ has its main support in the frequency band from 200 Hz to 1.5 kHz. Subsequently, we compute f_{thr} according to Eq. (15) and, in order to achieve a more realistic result, we neglect the contributions above merger frequency f_{mrg} , where this quantity is estimated using numerical relativity fits introduced in Ref. [119]. As shown in Fig. 9, we find that the SNR of GW170817 is located at frequencies lower

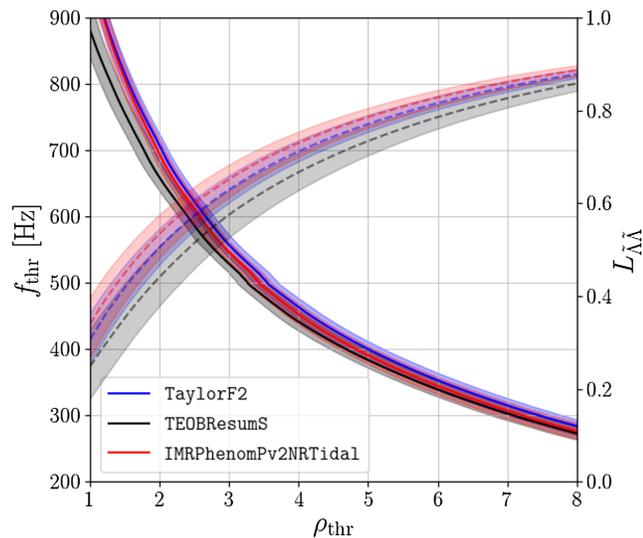


FIG. 9. Estimation of f_{thr} (solid lines) and $L_{\tilde{\Lambda}\tilde{\Lambda}}$ (dashed lines) from GW170817 posterior samples extracted with TaylorF2 (blue), IMRPhenomPv2NRTidal (red), and TEOBResumS (black). The shadowed bands represent the 90% credible regions. In order to evaluate f_{thr} , we use the identical waveform models involved in the extraction of the posterior samples, and we set a cutoff frequency at merger f_{mrg} , estimated with the fits introduced in Ref. [119]. For the estimation of $L_{\tilde{\Lambda}\tilde{\Lambda}}$, we limit ourselves to 6.5PN phase corrections. The plotted values are computed using a network of three detectors (two LIGOs and Virgo) and setting $f_{\text{max}} = 2$ kHz. For a single interferometer at $\rho_{\text{thr}} = 1$, we get $f_{\text{thr}} \approx 700$ Hz for LIGO’s detectors and $f_{\text{thr}} \approx 250$ Hz for Virgo.

than ~ 700 Hz (depending on the chosen value of ρ_{thr}). More precisely, we can say that the signal power enclosed above ~ 900 – 1000 Hz does not exceed an SNR of 1 (which roughly corresponds to 3% of the total SNR), while the power above ~ 550 Hz cannot contribute more than an SNR of 3 (10% of the total SNR). The large variability of f_{thr} with the chosen value of ρ_{thr} indicates that a relatively small fraction of the SNR is accumulated over a rather large frequency interval.

From the estimation of f_{thr} and the discussion of Sec. II, one expects the masses of the binary to be measured rather accurately. The reduced tidal parameter, instead, will be affected by significant statistical uncertainties: from the posterior samples, we estimate a loss of tidal information of $L_{\tilde{\Lambda}\tilde{\Lambda}} \sim 20\%$ – 50% , for $\rho_{\text{thr}} \sim 1$ – 3 , and for $f_{\text{thr}} \sim 1000$ – 550 Hz.

The marginalized tidal parameter posteriors reweighted to flat in $\tilde{\Lambda}$ prior are shown in Fig. 8. All the measurements agree within the 95% confidence region, thus indicating that waveform systematics are not the main source of uncertainty. However, the distributions for the different approximants do suggest the presence of some systematic effects. These posteriors should be interpreted in terms of the phasing plots in Fig. 2 for $f \lesssim 700$ kHz (the low-frequency part of the right panel) and $\tilde{\Lambda} \lesssim 1000$.

TABLE III. Analysis of GW170817 data: log evidences and corresponding standard deviations computed using different waveform approximants: TEOBResumS, TaylorF2, and IMRPhenomPv2NRTidal. The values obtained, according to standard Bayesian statistics, indicate that it is not possible to identify a preferred waveform model exclusively by relying on the GW170817 data.

| Approximant | $\ln p(d \text{Approx.})$ |
|---------------------|---------------------------|
| TaylorF2 | 523.078 ± 0.102 |
| TEOBResumS | 522.585 ± 0.102 |
| IMRPhenomPv2NRTidal | 522.261 ± 0.103 |

The phasing analysis of Sec. III shows that IMRPhenomPv2NRTidal is more attractive than TEOBResumS and TaylorF2; the systematic differences in the relevant frequency regime are dominated by the tidal part (IMRPhenomPv2NRTidal vs others) or by a mixture of the point mass and tides (TaylorF2 vs TEOBResumS). This is consistent with the slightly smaller $\tilde{\Lambda}$ measured with the IMRPhenomPv2NRTidal with respect to the other approximants and attributable to the particular design of IMRPhenomPv2NRTidal (PN tides at LO in the low-frequency regime, TEOBResumS in the middle-frequency regime, and NR data at higher frequencies; with LO tides stronger than PN NLO, NNLO, and EOB tides at low frequencies, and NR tides typically stronger than EOB tides [97,120]). TEOBResumS measurement is instead compatible with TaylorF2. This is again understandable from the phasing plots discussed in Sec. III: for $\hat{\omega} \sim 0.035$ – 0.06 , the differences in the point mass and tidal sector between the approximants have opposite signs and partially compensate each other. Nonetheless, it is not possible from this analysis to identify whether a model is preferred by the data available, which is consistent with the conclusion of Refs. [1,121]. We report in Table III the evidences given by the different approximants. We conclude that systematics effects are observable in GW170817, but they do not dominate the measurement of $\tilde{\Lambda}$. These effects are nonetheless expected *a priori* from the phasing analysis of Sec. III.

Note that our IMRPhenomPv2NRTidal posteriors do not present the double peak in $\tilde{\Lambda}$ that is instead found in Refs. [1,121]. The reason for this difference lies in the high-frequency cutoff imposed. This same effect had already been noticed in Ref. [122]. The authors, using the spin-aligned IMRPhenomDNRTidal model [62,63] to analyze the data, together with the relative binning technique [123], found a double-peak structure in the posterior of $\tilde{\Lambda}$ with $f_{\text{max}} = 1.5$ kHz; that, however, disappeared when f_{max} was lowered to 1 kHz. Further investigation by Narikawa *et al.* [124] showed how the bimodality in $\tilde{\Lambda}$ reflects the distributions found when analyzing the Hanford and Livingston data separately. The authors found that the

distribution of $\tilde{\Lambda}$ derived by the Livingston detector changes irregularly when they vary the maximum frequency above 1100 Hz, and that the Handford detector does not display such a behavior. The authors suggest that an in-depth study of noise properties and glitch removal techniques might be necessary to improve our understanding of the August binary. To this end, Pankow *et al.* [125] performed injection-recovery studies to test the goodness of glitch removal techniques, and found that for the studied signals, the mitigation methods applied are able to remove the glitch adequately. Conversely, Wade *et al.* [126] performed injection-recovery studies with different Gaussian noise realizations, and found that bimodalities in $\tilde{\Lambda}$ similar to those observed for GW170817 may arise in some cases. Repeating our analysis with `IMRPhenomPv2NRtidal` and $f_{\max} = 2$ kHz, we too reobtain the double peak in $\tilde{\Lambda}$. The evidence of the newer analysis is, however, compatible to the one reported in Table III: $\ln p(d|\text{IMRPhenomPv2NRtidal}, 2 \text{ kHz}) = 521.860 \pm 0.103$. Similarly, the maximum values of the log-likelihood are comparable between the two analyses: $\max(\ln p(d|\theta)^{1 \text{ kHz}}) \sim 534.26$ and $\max(\ln p(d|\theta)^{2 \text{ kHz}}) \sim 534.12$. This is not unexpected, and had already been observed in Ref. [122]. This implies that negligible SNR is accumulated above 1 kHz, and—given the yet unknown nature of the bimodality—that caution should be exercised when interpreting posteriors for $\tilde{\Lambda}$ obtained when considering data above 1 kHz.

Overall, we find consistent values for intrinsic parameters such as masses and spins with Refs. [1,121] and higher $\tilde{\Lambda}$ values. To translate the information on $\tilde{\Lambda}$ to constraints on the NS radius R , we apply the URs of Ref. [111] to the reweighted `TEOBResumS` $\tilde{\Lambda}$ posteriors and estimate the radius of a $1.4 M_{\odot}$ NS. We find $R_{1.4} = 12.5^{+1.1}_{-1.8}$ km. This value is slightly larger than—though still compatible with—the one obtained in Ref. [1]. The effect of the key choices of our analysis—i.e., the high-frequency cutoff employed, the use of `TEOBResumS` and the low-spin priors imposed—is then that of pushing towards higher R values and stiffer EOSs. In the literature, additional radius estimates have been computed by including further astrophysical information. We find our result, which focuses on the implications of GW data alone, to be in good agreement with the radii obtained when additionally accounting for electromagnetic priors [127] and the measurement given by NICER [128,129], which both also favor $\tilde{\Lambda}$ values larger than ≈ 200 .

VI. TIDAL INFERENCE WITH 3G DETECTORS

Third-generation detectors such as the Einstein Telescope [22,27] and Cosmic Explorer [130] are expected to start taking data in the late 2020s. Their increased sensitivity at high frequencies will significantly improve the detection of tidal signatures in the inspiral, and even

allow the detection of GWs from the remnant. Typical SNRs expected for GW170817-like events detected by ET are of the order of 1700. As a consequence, the importance of waveform systematics is expected to further increase with respect to second-generation detectors.

To summarize some the arguments of Sec. II, the SNR enters the determination of $\tilde{\Lambda}$ through two main channels. First, it determines the maximum useful frequency f_{thr} [see Eq. (15)], above which variations of ρ can be fully attributed to statistical fluctuations and which determines the regimes in which tidal measurements are performed. Second, it is related to the width of the distribution of the tidal parameter $\sigma_{\tilde{\Lambda}} = \tilde{\Lambda}^{95\%ile} - \tilde{\Lambda}^{5\%ile}$. If the signal is loud enough—as is expected with ET and CE— f_{thr} will be above the merger frequency for a large fraction of events. Therefore, when studying the signal with inspiral-merger-only waveform models, the effect of varying the SNR will mainly affect $\sigma_{\tilde{\Lambda}}$. To obtain a quantitative estimate of $\sigma_{\tilde{\Lambda}}$ for 3G detectors, we fit the values found in our injection study and extrapolate them to higher SNRs. We find that a good approximation of the behavior of $\sigma_{\tilde{\Lambda}}$ over the SNR range we considered is obtained by assuming that

$$\sigma_{\tilde{\Lambda}}(\rho) = \frac{c}{\rho - \rho_0}. \quad (32)$$

This functional form is valid only for $\rho > \rho_0$, in which case the denominator can be expanded as a geometrical series, and incorporates the corrections to the leading-order $1/\rho$ asymptotical behavior expected from the Fisher matrix analysis. Fitting Eq. (32) to the data, we find $(c, \rho_0) = (7497.97, 63.09)$ for `TaylorF2` and $(4372.21, 66.78)$ for `IMRPhenomPv2`. As could already be observed from Fig. 5, `IMRPhenomPv2NRtidal` constrains the tidal parameter better than its PN counterpart: $\sigma_{\tilde{\Lambda}}^{\text{Phenom}}(\rho)$ is (almost) parallel to $\sigma_{\tilde{\Lambda}}^{\text{TaylorF2}}(\rho)$ but shifted to lower values. To obtain a unique estimate of $\sigma_{\tilde{\Lambda}}$, we compute the mean value: $\bar{\sigma}_{\tilde{\Lambda}} = (\sigma_{\tilde{\Lambda}}^{\text{Phenom}} + \sigma_{\tilde{\Lambda}}^{\text{TaylorF2}})/2$.

The expression of $\bar{\sigma}_{\tilde{\Lambda}}$ can then be used to compute the SNR at which two independent measurements $\tilde{\Lambda}_1$ and $\tilde{\Lambda}_2$, whose difference we denote as $\Delta\tilde{\Lambda}$, become statistically inconsistent. Figure 10 shows the quantity $\Delta\tilde{\Lambda}/\bar{\sigma}_{\tilde{\Lambda}}$ as a function of the optimal SNR ρ for values of $\Delta\tilde{\Lambda} \in [-100, 100]$. When $\Delta\tilde{\Lambda}/\bar{\sigma}_{\tilde{\Lambda}} \approx 1$, statistical fluctuations are of the same order of magnitude as systematical effects. For $|\Delta\tilde{\Lambda}| \approx 100$, we see that this condition is satisfied already at the threshold $\rho \approx 125$. As $|\Delta\tilde{\Lambda}|$ decreases, the threshold SNR increases, reaching $\rho \approx 300$ in correspondence of a $\Delta\tilde{\Lambda} \approx 20$.

The above considerations are independent of the exact waveform models employed and do not tackle the issue of estimating the $\Delta\tilde{\Lambda}$ associated with two specific chosen approximants. While it is clear from the injection study of

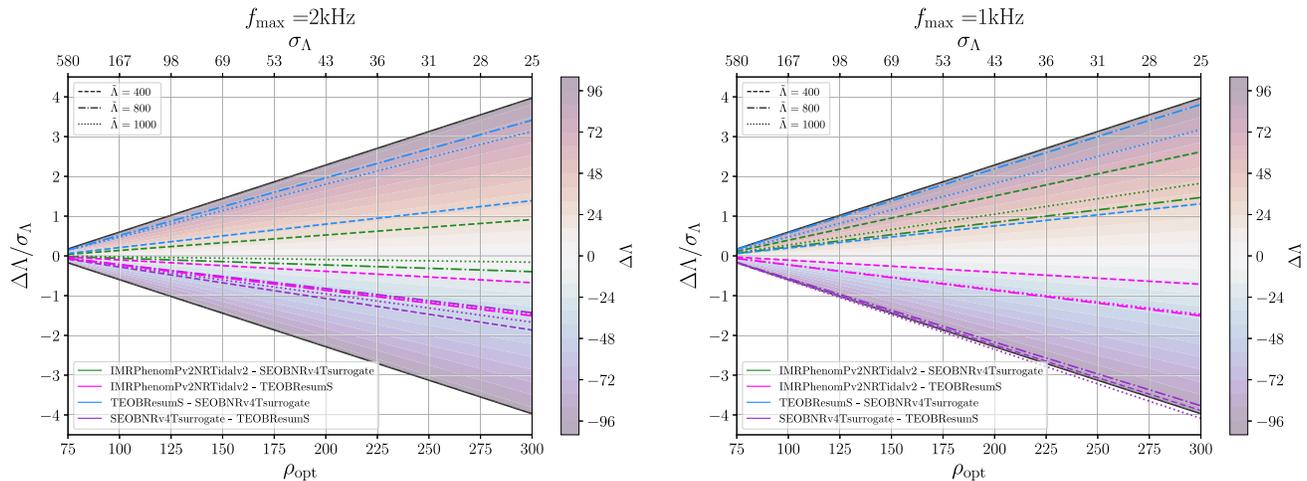


FIG. 10. The ratio between systematic effects $\Delta\tilde{\Lambda}$ and statistical uncertainties $\sigma_{\tilde{\Lambda}}$, shown as a function of the SNR ρ for a range of different $\Delta\tilde{\Lambda} \in [-100, +100]$. Colored lines refer to values of $\Delta\tilde{\Lambda}/\sigma_{\tilde{\Lambda}}$ estimated between a baseline approximant Y and a recovery approximant X , labelled as “ X - Y ” in the bottom-left legend, and computed for two different frequency cutoffs (left and right panels) and three different values of $\tilde{\Lambda}$ (dashed, dotted, and dash-dotted lines). We find that $\Delta\tilde{\Lambda}/\sigma_{\tilde{\Lambda}} \approx 1$ at SNRs ranging from 175 to 200 for all $\tilde{\Lambda}$ values. Therefore, with 3G detectors, all the current approximants will appear to be statistically inconsistent.

Sec. IV that large $\Delta\tilde{\Lambda}$ are to be expected when employing TaylorF2 and IMRPhenomPv2NRTidal, we take a step further and qualitatively estimate the bias $\Delta\tilde{\Lambda} = \tilde{\Lambda}_E - \tilde{\Lambda}$ through Eq. (24) for two additional state-of-the-art approximants, IMRPhenomPv2NRTidalv2 [93] and SEOBNRv4Tsurrogate [90]. We thus compare the latter and TEOBResumS in pairs and report the differences with respect to two baselines (TEOBResumS and SEOBNRv4Tsurrogate). Following the procedure described in Sec. IV D, we consider values of $\tilde{\Lambda}$ equal to 400, 800, and 1000, place the sources in GW170817’s location, and employ the EinsteinTelescope-P1600143 PSD [131]. We compute waveforms from 30 to 2048 Hz (left panel) or 1024 Hz (right panel). Results are again displayed in Fig. 10.

We find that both SEOBNRv4Tsurrogate and IMRPhenomPv2NRTidalv2 “underestimate” the values of $\tilde{\Lambda}$ at the TEOBResumS baseline (right panel), and that the $|\Delta\Lambda|$ values found are always below ≈ 100 . This indicates that tides are stronger in the SEOBNRv4Tsurrogate and IMRPhenomPv2NRTidalv2 models than in TEOBResumS. When restricting below 1 kHz (large $\tilde{\Lambda}$), the systematic bias in $\tilde{\Lambda}$ due to the differences between IMRPhenomPv2NRTidalv2 and TEOBResumS is $\lesssim 2\sigma$, corresponding to $\Delta\tilde{\Lambda} \pm 50$, while it varies $\sim 2\sigma$ – 4σ when considering differences with respect to SEOBNRv4Tsurrogate. This indicates that the differences between IMRPhenomPv2NRTidalv2 and TEOBResumS are mostly related to the modeling of tides at high frequencies, while the tides in the EOB models differ from each other already at lower frequencies.

Some caution is needed when interpreting the results obtained for the different waveform approximants: in

Sec. IV D, we have seen that at times the estimated $\tilde{\Lambda}_E$ would overestimate $\tilde{\Lambda}^{\text{median}}$ by up to 100. This difference was acceptable at the injected SNRs, but it indicates that our estimate might not be precise enough at the SNRs which characterize 3G detectors. Nonetheless, we expect the behavior of the approximants (i.e., their being more/less attractive) to be correctly captured.

Overall, our findings indicate that above SNR ≈ 100 – 200 , $\sigma_{\tilde{\Lambda}}$ will be small enough that the models will appear to be fully inconsistent with each other. The estimated systematic biases reflect differences in the tidal modeling at frequencies corresponding to the very last orbits and thus accessible to NR. We stress that at frequencies $\hat{\omega} \gtrsim 0.06$, the NSs are in contact, and the waveform modeling based on tidal interactions can only be considered an effective description, since the dynamics are dominated by hydrodynamics [97]. We demonstrate in Appendix C that current NR simulations are not sufficiently accurate to produce faithful waveforms. New, more precise NR simulations appear crucial to further develop tidal waveform models for future detectors.

VII. CONCLUSIONS

In this paper, we discussed a possible approach for the analysis of waveform systematics in the estimation of tidal effects in BNS. We demonstrated the effectiveness of our method in a mock experiment using a large set of injected signals and applied the method to GW170817. We recommend using this method for future analysis and point out that the approximants used for the main analysis of GW170817 should be significantly improved for future robust analysis at SNR ~ 80 and beyond. We expand on these conclusions here below.

The bottom-up method employed in this work is composed of three steps: First, the waveform approximants should be compared using the Q_ω analysis in order to understand the effect of the modeling choices (and the physics implemented in the models) on the GW phase. The Q_ω diagnostic is key to determining the waveforms’ differences, and it is free from the ambiguities introduced in the phase comparisons by the time/phase shift. Second, it is important to identify the frequency regime at which the tidal information is effectively extracted. This can be accomplished by computing fractional losses L_{ij} defined in Sec. II. Third, the PE results should be interpreted in terms of the ΔQ_ω analysis on the relevant frequency interval.

Our mock experiments show that this procedure is effective in identifying the main biases introduced by the waveform models. Note in this respect that the “target” model used for the computation of the ΔQ_ω should be chosen among those that are considered sufficiently faithful on the relevant frequency regime. For example, for analyzing biases at low frequencies, the target model should contain maximal analytical information (vs minimal fitting), and high-order Taylor or EOB models represent the best choice in this respect. At very high frequencies, numerical relativity data would be the best choice, although the accuracy of the data is not yet sufficient for robust statements (See Appendix C).

The analysis of GW170817 shows that the measurement of the tides shows minor systematic effects if performed up to 1 kHz [122,127]. Extending the analysis to higher frequencies introduces some waveform effects, albeit still compatible with others in the 90% confidence region. In particular, comparing our results to Fig. 9 of Ref. [6], we observe a shift in the posterior of $\tilde{\Lambda}$ computed with `IMRPhenomPv2NRTidal` that can be fully understood from the Q_ω phasing analysis presented here. When applying `IMRPhenomPv2NRTidal` to 2 kHz, the posteriors have a double peak that is not present with the 1 kHz cutoff. The similar inferences using the `TaylorF2` and `TEOBResumS` approximants are instead related to the fact that, in the relevant frequency regime, the differences between the `TaylorF2` point mass and the tidal effects have opposite signs and partly compensate each other (see Fig. 2).

By applying the URs of Ref. [111] to the $\tilde{\Lambda}$ values obtained in our GW170817 reanalysis, we obtain a new measurement of $R_{1.4 M_\odot}$ which—based exclusively on the information gathered from GW data—is in good agreement with results coming from independent astrophysical observations; i.e., the NICER radius measurement and the information coming from EM observations [127,128].

Significant waveform systematics are to be expected for GW170817-like signals already for the current advanced detectors at design sensitivity. Further, systematic errors increase for large values of the tidal parameter $\tilde{\Lambda}$. These

events, with large tidal deformabilities or high SNRs, are the best candidates for an actual *measure* (vs the upper limit) of the tidal parameters and EOS constraints. At design sensitivity, the expected bias in the reduced tidal parameter using `TaylorF2` and `IMRPhenomPv2-NRTidal` is about 2σ (for average BNS parameters as quantified in Fig. 5). This would be reflected in systematics on the NS radius of about 1 km (10%), that are comparable to or well above the current best estimates of the NS radius, also including electromagnetic constraints [3,111,127,132,133].

Moving to higher sensitivities and 3G detectors, we estimate that the systematics between the approximants that currently have the smallest differences among themselves become dominant over statistical errors at a SNR of 200 and for $\tilde{\Lambda} \gtrsim 400$ (Fig. 10). This implies that EOS constraints from the potentially most informative (and rare) events will be harmed by tidal waveform systematics.

ACKNOWLEDGMENTS

We thank Jocelyn Read, Derek Davis, and Katerina Chatziioannou for useful discussions and comments on the manuscript. R. G. acknowledges support from the Deutsche Forschungsgemeinschaft (DFG) under Grant No. 406116891 within the Research Training Group No. RTG 2522/1. M. B. and S. B. acknowledge support by the EU H2020 under ERC Starting Grant No. BinGraSp-714626. M. B. acknowledges support from the Deutsche Forschungsgemeinschaft (DFG) under Grant No. 406116891 within the Research Training Group No. RTG 2522/1. Data analysis was performed on the supercomputers ARA in Jena and ARCCA in Cardiff. We acknowledge the computational resources provided by Friedrich Schiller University Jena, supported in part by DFG Grants No. INST 275/334-1 FUGG and No. INST 275/363-1 FUGG, and Cardiff University, funded by STFC Grant No. ST/I006285/1. Data postprocessing was performed on the Virgo “Tullio” server in Torino, supported by INFN. This research has made use of data obtained from the Gravitational Wave Open Science Center [134], a service of LIGO Laboratory, the LIGO Scientific Collaboration, and the Virgo Collaboration. LIGO is funded by the U.S. National Science Foundation. Virgo is funded by the French Centre National de Recherche Scientifique (CNRS), the Italian Istituto Nazionale della Fisica Nucleare (INFN), and the Dutch Nikhef, with contributions by Polish and Hungarian institutes.

APPENDIX A: EFFECT OF THE POINT-MASS SECTOR ON $\tilde{\Lambda}$

In this Appendix, we explicitly show how uncertainties in the point-mass phase (of both statistic and systematic nature) can affect the determination of the tidal parameter $\tilde{\Lambda}$. Starting from Eq. (7), writing

$\exp(i\Delta\Psi) = \cos(\Delta\Psi) + i \sin(\Delta\Psi)$, and expanding the cosine around $\Delta\Psi \approx 0$, the SNR becomes

$$\begin{aligned} \rho &= \frac{4}{\sqrt{\langle h|h \rangle}} \int \frac{\tilde{A}_d \tilde{A}_h}{S_n} \left[1 - \frac{(\Delta\Psi)^2}{2} + O(\Delta\Psi^3) \right] df \\ &\simeq \rho_{\text{opt}} - \frac{2}{\sqrt{\langle h|h \rangle}} \int \frac{\tilde{A}^2 (\Delta\Psi)^2}{S_n} df, \end{aligned} \quad (\text{A1})$$

where the last step assumes $\tilde{A}_h \approx \tilde{A}_d = \tilde{A}$. By defining $\bar{\theta}$ as the set of parameters such that $\Delta\Psi(\bar{\theta}; f) \approx 0$ over the whole frequency range considered $[f_{\text{min}}, f_{\text{max}}]$, and expanding $\Delta\Psi$ in Eq. (A2) around $\bar{\theta}$, the second integral in Eq. (A2) can be connected to the Fisher matrix:

$$\int \frac{\tilde{A}^2 (\Delta\Psi)^2}{S_n} df \approx \int \frac{\tilde{A}^2 \partial_i \Psi_h \partial_j \Psi_h}{S_n} \Delta\theta^i \Delta\theta^j df, \quad (\text{A2})$$

with $\partial_i = \partial/\partial\theta^i$ and $\Delta\theta^i = \theta^i - \bar{\theta}^i$ (repeated indices imply a summation). Under the assumption of high SNR, the integrals over f in Eq. (A2) can be split as

$$-\frac{\Delta\theta^l \Delta\theta^m}{2} \int_{f_{\text{min}}}^{f_c} I_{l,m} df - \frac{\Delta\theta^{\tilde{\Lambda}} \Delta\theta^{\tilde{\Lambda}}}{2} \int_{f_c}^{f_{\text{max}}} I_{\tilde{\Lambda},\tilde{\Lambda}} df, \quad (\text{A3})$$

where f_c is a ‘‘cutoff frequency’’ that identifies the beginning of the relevant frequency support of $I_{\tilde{\Lambda},\tilde{\Lambda}}$ [see Eq. (14)] and has the value of ≈ 300 Hz for fiducial BNS. Equation (A3) clearly shows the different frequency regimes at which the parameters are measured during PE. $\bar{\mathcal{M}}$, \bar{q} , and $\bar{\chi}$ are determined during the early inspiral ($f \leq f_c$); $\tilde{\Lambda}$ at higher frequencies ($f \geq f_c$). Sampling methods will tend to recover the parameters $\theta \rightarrow \bar{\theta}$. However, due to the varying sensitivity of the detector over different frequency ranges, the parameters measured during the early inspiral $\theta_{\text{insp}} = (\mathcal{M}, q, \chi)$ converge faster than tidal parameters. Let us then go back to Eq. (A2) and express its left-hand side as

$$-\frac{\Delta\theta^l \Delta\theta^m}{2} \int_{f_{\text{min}}}^{f_c} I_{l,m} df - \frac{1}{2} \int_{f_c}^{f_{\text{max}}} \frac{\tilde{A}^2 (\Delta\Psi)^2}{S_n} df. \quad (\text{A4})$$

The first integral has, again, been expanded about the set of parameters $\bar{\theta}_{\text{insp}}$. Taking the limit $\theta_{\text{insp}} \rightarrow \bar{\theta}_{\text{insp}}$, its contribution tends to zero by definition. The remaining second integral can be explicitly written as

$$-\frac{1}{2} \int_{f_c}^{f_{\text{max}}} \frac{\tilde{A}^2}{S_n} [\Delta\Psi^{\text{PM}}(\bar{\theta}_{\text{insp}}, \tilde{\Lambda} = 0) + \Delta\Psi^T(\bar{\theta}_{\text{insp}}, \tilde{\Lambda})]^2 df, \quad (\text{A5})$$

where we have separated $\Delta\Psi$ into its point-mass ($\Delta\Psi^{\text{PM}}$) and tidal ($\Delta\Psi^T$) contributions. Critically,

$\Delta\Psi^{\text{PM}}(\bar{\theta}_{\text{insp}}, \tilde{\Lambda} = 0)$ is not necessarily close to zero above f_c , as the parameters $\bar{\theta}_{\text{insp}}$ are determined over a different regime and chosen to minimize $\Delta\Psi^{\text{PM}}(\bar{\theta}_{\text{insp}}, \tilde{\Lambda} = 0)$ below f_c . The value $\tilde{\Lambda}$ therefore will have to minimize not only $\Delta\Psi^T$ over $[f_c, f_{\text{max}}]$, but rather the sum of $\Delta\Psi^T$ and $\Delta\Psi^{\text{PM}}$. This means that both the tidal and the point-mass sectors of a waveform model can introduce biases in the recovery of tidal parameters, and that overall phase differences accumulated over f_c are absorbed mainly by $\tilde{\Lambda}$.

APPENDIX B: TIDAL INFORMATION

In this Appendix, we apply the method presented in Sec. II A to the signals involved in the PE studies of Sec. IV, proving that the injections are actually performed in an informative framework for the tidal parameter, in which statistical fluctuations cannot be considered as the dominant source of the differences observed in the tidal parameter (see Fig. 5).

Table IV shows the values of the frequency support $[f_{5\%}^{\tilde{\Lambda}}, f_{95\%}^{\tilde{\Lambda}}]$ defined in Eq. (14) computed for the injected signals, including all the detectors involved in the analysis. For all the cases, $f_{\text{mrg}} > 1$ kHz, indicating the presence of a signal in the high-frequency regime, and $f_{95\%}^{\tilde{\Lambda}} > 1$ kHz, meaning that the tidal contributions are relevant above this value. Furthermore, Table IV reports the values of f_{thr} and $L_{\tilde{\Lambda},\tilde{\Lambda}}$, defined in Eqs. (15) and (16), respectively, computed for the same signals for $\rho_{\text{thr}} = 1, 3$. We observe that these

TABLE IV. Values of $f_{5\%}^{\tilde{\Lambda}}$, $f_{95\%}^{\tilde{\Lambda}}$, f_{thr} , and $L_{\tilde{\Lambda},\tilde{\Lambda}}$ computed for the signals involved in the injection studies in Sec. IV. We recall that the injected signals have extrinsic properties identical to the maximum posterior parameters of GW170817 [1]. The reported values are estimated with a three-detector network (two LIGOs and Virgo) at a design sensitivity using the TEOBResumS waveform model.

| EOS | M | q | f_{mrg} | $f_{5\%}^{\tilde{\Lambda}}$ | $f_{95\%}^{\tilde{\Lambda}}$ | $\rho_{\text{thr}} = 1$ | | $\rho_{\text{thr}} = 3$ | |
|-------|------|------|------------------|-----------------------------|------------------------------|-------------------------|---------------------------------------|-------------------------|---------------------------------------|
| | | | | | | f_{thr} | $L_{\tilde{\Lambda},\tilde{\Lambda}}$ | f_{thr} | $L_{\tilde{\Lambda},\tilde{\Lambda}}$ |
| DD2 | 2.71 | 1.00 | 1287 | 245 | 1460 | 1085 | 0.18 | 731 | 0.52 |
| LS220 | 2.68 | 1.00 | 1366 | 259 | 1800 | 1152 | 0.23 | 740 | 0.57 |
| LS220 | 2.69 | 0.86 | 1241 | 242 | 1332 | 1055 | 0.15 | 731 | 0.51 |
| SFHo | 2.71 | 1.00 | 1426 | 271 | 1825 | 1207 | 0.23 | 766 | 0.59 |
| SFHo | 2.72 | 0.88 | 1416 | 278 | 1862 | 1252 | 0.25 | 772 | 0.61 |
| SLy | 2.68 | 1.00 | 1588 | 273 | 1746 | 1211 | 0.22 | 772 | 0.60 |
| SLy | 2.69 | 0.88 | 1480 | 272 | 1816 | 1208 | 0.23 | 766 | 0.59 |
| DD2 | 2.48 | 1.00 | 1206 | 240 | 1666 | 1033 | 0.21 | 693 | 0.55 |
| DD2 | 3.18 | 1.00 | 1192 | 249 | 1715 | 1125 | 0.19 | 782 | 0.49 |
| 2B | 2.70 | 1.00 | 1646 | 293 | 1834 | 1311 | 0.24 | 804 | 0.63 |
| SLy | 3.00 | 1.00 | 1540 | 278 | 1744 | 1254 | 0.20 | 828 | 0.56 |
| LS220 | 3.20 | 1.00 | 1288 | 255 | 1443 | 1332 | 0.30 | 826 | 0.63 |
| SFHo | 2.92 | 1.00 | 1449 | 281 | 1874 | 1285 | 0.24 | 802 | 0.59 |
| SFHo | 2.80 | 1.00 | 1519 | 273 | 1698 | 1222 | 0.20 | 788 | 0.58 |
| ALF2 | 3.00 | 1.00 | 1299 | 250 | 1395 | 1121 | 0.15 | 787 | 0.50 |

ρ_{thr} values correspond to power losses of $\sim 1.5\%$ and $\sim 4\%$ of the total power, respectively. For $\rho_{\text{thr}} = 1$, we have $f_{\text{thr}} > 1$ kHz, showing that the signal power is relevant above this threshold. For this value, $L_{\tilde{\lambda}\tilde{\lambda}} \leq 30\%$. These facts are reflected in a lower variance on the posterior distribution for $\tilde{\Lambda}$ coming from the PE analyses with $f_{\text{max}} = 2048$ Hz with respect to those with $f_{\text{max}} = 1024$ Hz. Finally, for all the injected signals, we have $f_{\text{thr}} > f_{5\%}^{\tilde{\Lambda}}$, which proves that these data contain information on the tidal parameter in an accessible frequency range.

APPENDIX C: FAITHFULNESS OF NUMERICAL RELATIVITY WAVEFORMS

Numerical relativity (NR) simulations are fundamental for understanding the merger physics and the waveform morphology in the high-frequencies regime. They incorporate hydrodynamical effects and can model not only the late-inspiral–merger parts of the coalescence, but also the postmerger phase. While NR waveforms are often regarded as exact with respect to those provided by waveform approximants in the same regime, the complex 3D simulations can introduce significant uncertainties, e.g., Refs. [97,135–138]. The latter are due both to systematics (finite radius extraction of the GWs, numerical dissipation, etc.) and to finite grid resolution. Systematics are difficult to control, but finite resolution errors can be studied by simulating at different resolutions and performing convergence tests.

In this Appendix, we apply the method of Sec. IV D to a set of NR waveforms taken from the CoRe database [139], with the aim of testing the accuracy of current state-of-the-art NR simulations and guiding future efforts. In particular, we consider multiorbit and eccentricity-reduced simulations performed with the BAM code, and we focus on the late-inspiral–merger phase, where waveforms are shown to be convergent. To the best of our knowledge, accuracy standards for BNS NR waveforms at multiple grid resolutions have been computed only in Ref. [135] for data that are currently superseded by the those produced with simulations employing high-order numerical fluxes [136,138] and higher resolutions than we consider here. We use here a sample of CoRe waveforms computed at multiple resolutions and produced in Refs. [92,140–143].

Table V displays the faithfulness values computed for a set of BAM waveforms. Each value is obtained by comparing the two highest-resolution simulations available for each considered set of intrinsic parameters. For each resolution R , the simulations compute the multipoles $h_{\ell m}(t)$; the waveform polarizations h_+^R, h_\times^R are reconstructed from

$$h_+ - ih_\times = D_L^{-1} \sum_{\ell=2}^{\infty} \sum_{m=-\ell}^{\ell} h_{\ell m}(t) {}_{-2}Y_{\ell m}(t, \psi), \quad (\text{C1})$$

TABLE V. Faithfulness values \mathcal{F} computed considering frequencies from f_{low} to f_{mrg} between simulations with the same intrinsic parameters and two different resolutions, extracted at $r/M = 1000$. The source is situated in the same sky location as GW170817, and the waveform polarizations h_+ and h_\times are computed and projected on the Livingston detector. We employ the aLIGO Design Sensitivity P1200087 [23] PSD from pycbc [115] to compute the matches and compare the values obtained to the thresholds \mathcal{F}_{thr} calculated with Eq. (19) with $\epsilon^2 = 1$ or $\epsilon^2 = 6$. A tick \checkmark indicates that $\mathcal{F} > \mathcal{F}_{\text{thr}}$. Conversely, a cross \times indicates that $\mathcal{F} < \mathcal{F}_{\text{thr}}$.

| Sim | n ^a | \mathcal{F} | SNR | | | | | | |
|----------|----------------|---------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | | | 14 | | 30 | | 80 | | |
| | | | 6 | 1 | 6 | 1 | 6 | 1 | |
| BAM:0011 | [96, 64] | 0.991298 | \checkmark | \times | \times | \times | \times | \times | \times |
| BAM:0017 | [96, 64] | 0.985917 | \checkmark | \times | \times | \times | \times | \times | \times |
| BAM:0021 | [96, 64] | 0.957098 | \times |
| BAM:0037 | [216, 144] | 0.998790 | \checkmark | \checkmark | \checkmark | \times | \times | \times | \times |
| BAM:0048 | [108, 72] | 0.983724 | \times |
| BAM:0058 | [64, 64] | 0.999127 | \checkmark | \checkmark | \checkmark | \times | \times | \times | \times |
| BAM:0064 | [240, 160] | 0.997427 | \checkmark | \times | \checkmark | \times | \times | \times | \times |
| BAM:0091 | [144, 108] | 0.997810 | \checkmark | \checkmark | \checkmark | \times | \times | \times | \times |
| BAM:0094 | [144, 108] | 0.996804 | \checkmark | \checkmark | \checkmark | \times | \times | \times | \times |
| BAM:0095 | [256, 192] | 0.999550 | \checkmark |
| BAM:0107 | [128, 96] | 0.995219 | \checkmark | \times | \times | \times | \times | \times | \times |
| BAM:0127 | [128, 96] | 0.999011 | \checkmark | \checkmark | \checkmark | \times | \times | \times | \times |

^aNumber of grid point (linear resolution) of the finest grid refinement, roughly covering the diameter of one NS.

where ${}_{-2}Y_{\ell m}(t, \psi)$ are the spin-weighted spherical harmonics of spin $s = -2$ and D_L is the luminosity distance. Assuming for simplicity that the radiation is emitted along the z axis, perpendicular to the orbital plane, one has that $\iota = \psi = 0$, and ${}_{-2}Y_{2\pm 2}(0, 0) = \sqrt{5/(64\pi)}(1 \pm 1)^2$. Fixing the source in GW170817’s sky location and projecting the polarizations on the Livingston detector, matches are finally computed over a frequency range $f \in [f_{\text{low}}, f_{\text{mrg}}]$, where f_{low} is defined as the frequency at which the amplitude of the Fourier transform $\text{FT}[\text{Re}(h_{22})]$ is highest and f_{mrg} is the merger frequency—i.e., the instantaneous frequency corresponding to the peak of the amplitude $|h_{22}(t)|$. Such values are then contrasted to the threshold faithfulness \mathcal{F}_{thr} of Eq. (19). Similarly to Sec. IV D, we choose ϵ^2 to be equal to 1, for a stricter requirement, or to the number of intrinsic parameters of a BNS system ($\epsilon^2 = 6$). Note that while $\mathcal{F} < \mathcal{F}_{\text{thr}}$ is a necessary but not sufficient condition for biases to appear, $\mathcal{F} > \mathcal{F}_{\text{thr}}$ is a sufficiently strong requirement to ensure that two waveforms are faithful. While for low-SNR signals most of the waveforms considered are accurate enough, we find that—out of the twelve simulations examined—none passes the accuracy test when (SNR 80, $\epsilon^2 = 1$), and only one (BAM:0095) manages to pass it when (SNR 80, $\epsilon^2 = 6$)

and (SNR 30, $\epsilon^2 = 1$). Note that the stars are resolved in this case with $\gtrsim 200$ grid points.

Our findings indicate that the largest portion of the NR simulations available to date may not yet be sufficiently accurate for GW data-analysis purposes. High-order methods for hydrodynamics and resolutions > 200 grid points per star appear necessary for GW modeling.

APPENDIX D: UNIVERSAL RELATIONS

In this Appendix, we collect the quasi-universal relations employed in the main text.

- (1) *De et al.*: This phenomenological relation was first introduced in Ref. [111], and it links the chirp mass of a BNS system \mathcal{M}_c and its mass-weighted tidal parameter $\tilde{\Lambda}$ to the radius of a $1.4 M_\odot$ star, $R_{1.4}$. Explicitly,

$$R_{1.4} \simeq (11.2 \pm 0.2) \frac{\mathcal{M}_c}{M_\odot} \left(\frac{\tilde{\Lambda}}{800} \right)^{1/6} \text{ km} \quad (\text{D1})$$

Note that this expression is valid for GW170817-like systems, and it is expected to fail for stars lighter than $1 M_\odot$ or heavy systems, with a chirp mass \mathcal{M}_c larger than 1.5.

- (2) *Binary-Love and C-Love*: These relations were obtained in Refs. [113,114]. The Binary-Love relation links the asymmetric combination of the tidal parameters, $\Lambda_a = (\Lambda_1 - \Lambda_2)/2$, to the symmetric one, $\Lambda_s = (\Lambda_1 + \Lambda_2)/2$, in a mass-ratio (q) dependent way:

$$\Lambda_a = F_n(q) \Lambda_s \frac{a + \sum_{i=1}^3 \sum_{j=1}^2 b_{ij} q^j \Lambda_s^{-i/5}}{a + \sum_{i=1}^3 \sum_{j=1}^2 c_{ij} q^j \Lambda_s^{-i/5}}, \quad (\text{D2})$$

$$F_n(q) \equiv \frac{1 - q^{10/(3-n)}}{1 + q^{10/(3-n)}}, \quad (\text{D3})$$

where the coefficients n , a , b_{ij} , c_{ij} can be found in, e.g., Ref. [113]. It can be used in PE to reduce the dimensionality of the parameter space by linking Λ_2

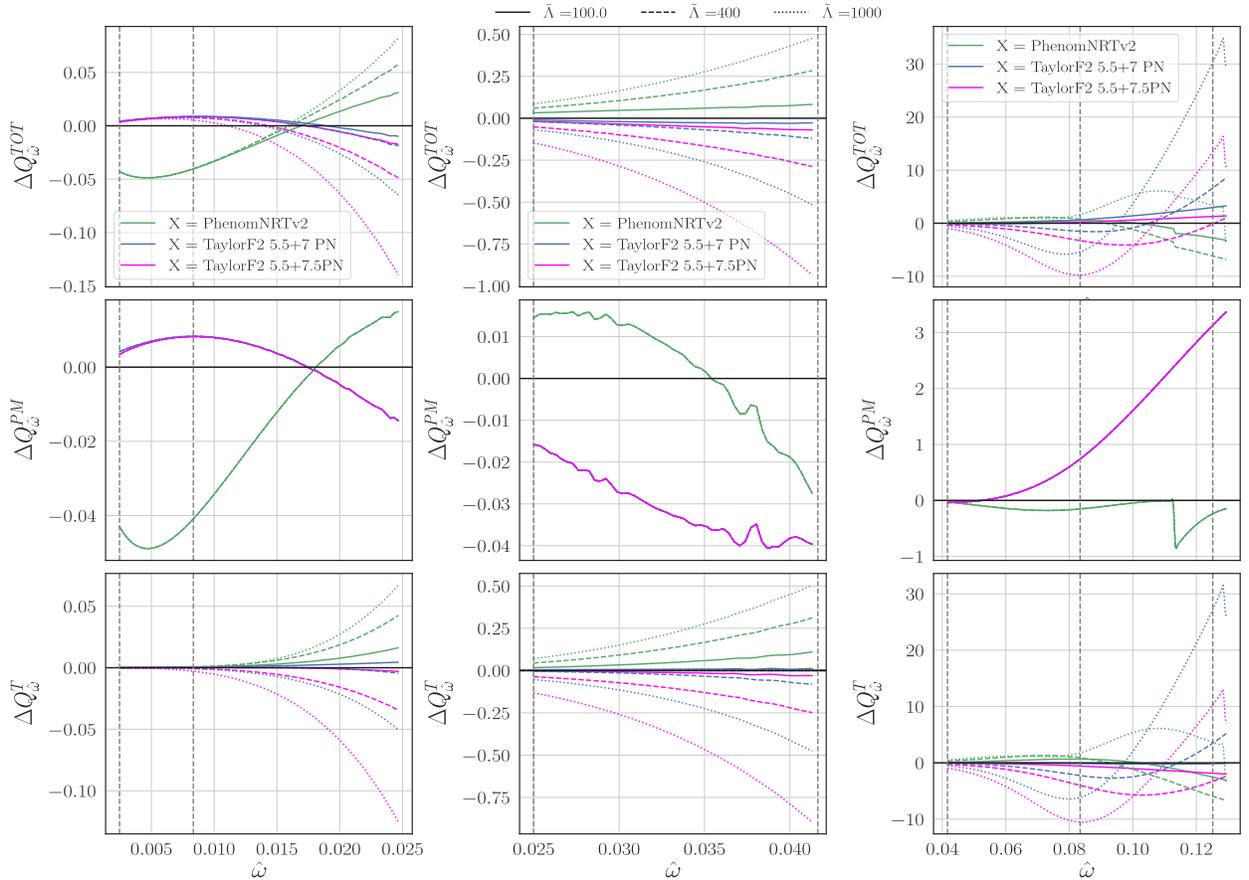


FIG. 11. This figure is the equivalent of Fig. 2 with TaylorF2 at 5.5 PN PM and 7 PN tides (blue), 5.5 PN PM and 7.5 PN tides (magenta), and IMRPhenomPv2NRTidalv2 (green). Note that the 5.5 PN PM and NRTidalv2 descriptions improve the ΔQ_ω^{PM} and ΔQ_ω^T in the low- and high-frequency regimes, respectively. When coupled to a 5.5 PN PM, 7.5 PN tides are closer to TEOBResumS for $\tilde{\Lambda} = 100$ than the 7 PN description, which in turn performs better for $\tilde{\Lambda} = 400$ over a large portion of the frequency range.

and Λ_1 [144]. The C-Love relation, instead, links the compactness of a NS to its tidal deformability:

$$C_i(\Lambda_i) = \sum_{k=0}^2 a_k (\ln \Lambda_i)^k, \quad (\text{D4})$$

and $a_0 = 0.3616998$, $a_1 = -0.0354818$, $a_2 = 0.0006193849$. To obtain an estimate of the radius of one of the NSs, we combine them. Indeed, rather than directly applying Eq. (D4) to the posterior samples of Λ_2 , we wish to map $\tilde{\Lambda}$ onto R , as $\tilde{\Lambda}$ is the better measured quantity from GW analysis. To do so, we obtain the relation $\Lambda_1 = \Lambda_1(\Lambda_2, q)$ from the inversion of Eq. (D2), and we compute $\Lambda_2 = \Lambda_2(\tilde{\Lambda}, q)$ from the definition of $\tilde{\Lambda}$. Finally, we apply Eq. (D4).

- (3) *Raithel et al.*: The relation found in Ref. [112] is based on a quasi-Newtonian approximation of the full relativistic expression for the tidal deformability of a star, given by Eq. (96) of Ref. [31] with $\beta \approx 1$. Explicitly,

$$\tilde{\Lambda} = \tilde{\Lambda}_0(1 + \delta_0(1 - q)^2), \quad (\text{D5})$$

$$\tilde{\Lambda}_0 = \frac{15 - \pi^2}{3\pi^2} \xi^{-5} (1 - 2\xi)^{5/2}, \quad (\text{D6})$$

$$\delta_0 = \frac{3}{104} (1 - 2\xi)^{-2} (-10 + 94\xi - 83\xi^2), \quad (\text{D7})$$

$$\xi = \frac{2^{1/5} G M_c}{R c^2}. \quad (\text{D8})$$

The above equations can be inverted numerically to obtain $R(\tilde{\Lambda}, q, M_c)$.

APPENDIX E: Q_ω ANALYSIS WITH OTHER APPROXIMANTS

In the present Appendix, we repeat the discussion of the second part of Sec. III, and we compute $\Delta Q_\omega^{\text{TOT}}$, ΔQ_ω^T , and ΔQ_ω^{PM} for two additional state-of-the-art approximants: (i) IMRPhenomPv2NRTidalv2, which differs from the IMRPhenomPv2NRTidal model exclusively in its tidal

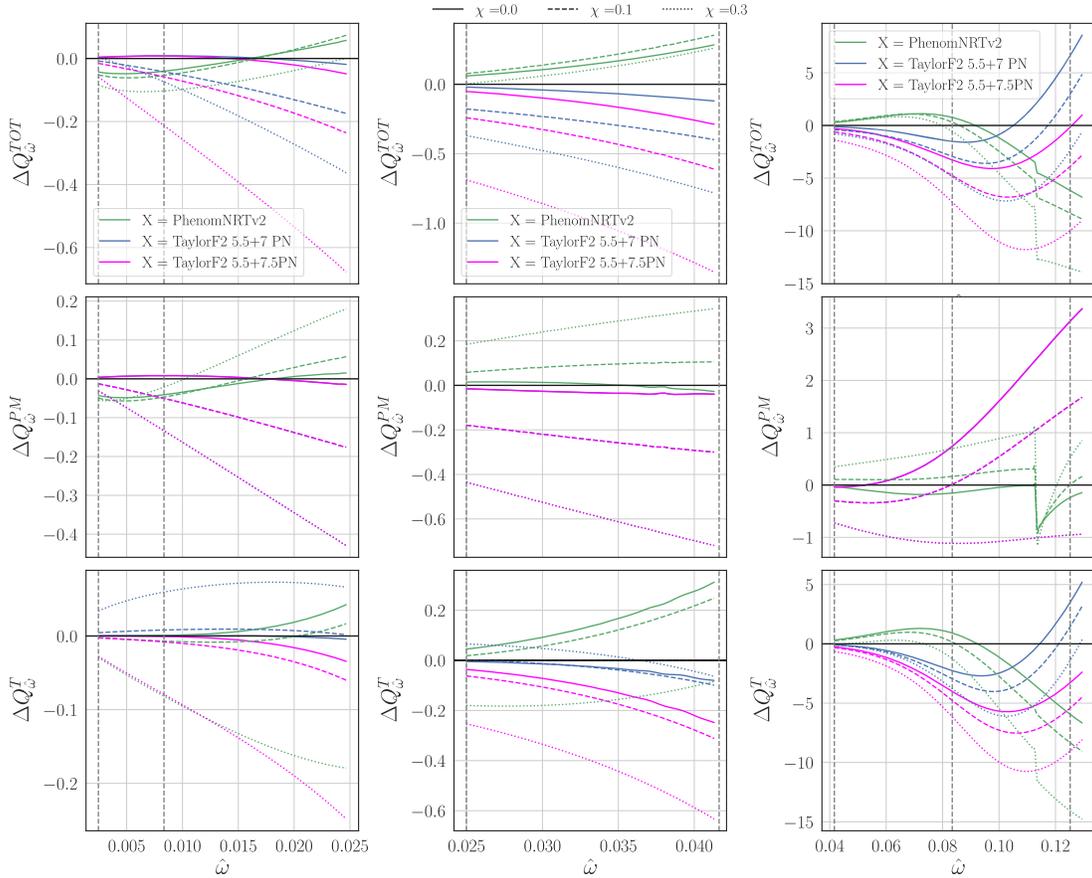


FIG. 12. This figure is the equivalent of Fig. 2 with TaylorF2 at 5.5 PN PM and 7 PN tides (blue), 5.5 PN PM and 7.5 PN tides (magenta), and IMRPhenomPv2NRTidalv2 (green). We note that in the low-frequency regime, the 5.5 PN PM, which gives the best approximation of the TEOBResumS PM between the approximants considered for nonspinning equal-mass binaries, becomes increasingly more negative as spins grow. On the other hand, the Phenom description over the same range consistently has $|\Delta Q_\omega^{PM}| < 0.2$. When considering the high-frequency contributions, instead, the hierarchy displayed in Fig. 3 is maintained.

sector, which now incorporates a 7.5 PN low-frequency limit and PN-expanded spin-quadrupole interactions up to 3.5 PN in the waveform phase; and (ii) TaylorF2, endowed with quasi-5.5 PN point-mass terms, 7.5 PN tides, and spin-spin terms up to 3.5 PN. We note that it is not possible to obtain meaningful information from Q_ω for SEOBNRv4Tsurrogate, as it is not continuous and derivable over the whole range of frequencies considered.

Figure 11 shows $Q_{\hat{\omega}}$ for the aforementioned approximants, computed for three reference signals with varying $\tilde{\Lambda}$ and zero spins, once again divided into three $\hat{\omega}$ intervals, corresponding to the regimes in which tidal contributions are roughly smaller than, comparable to, or dominant with respect to $\Delta Q_{\hat{\omega}}^{PM}$. Inspecting the first column (which corresponds to the early frequencies interval), we notice that the $\Delta Q_{\hat{\omega}}^{PM}$ of TaylorF2 is comparable to that of IMRPhenomPv2NRTidalv2, and is overall closer to TeOBResumS's description than the one provided by considering a 3.5 PN point-mass baseline. When considering the tidal sector, instead, we note that the behaviors of NRTidalv2 and 7.5 PN tides are opposites from the start. Moving to higher frequencies, tidal effects dominate both the late inspiral and merger regimes for both approximants. IMRPhenomPv2NRTidalv2 is less attractive than

IMRPhenomPv2NRTidal, while the 7.5 PN tidal description is more repulsive than the 6 PN one. Close to merger, we find that—as expected—the point-mass contribution of the 5.5 PN approximant becomes large and positive, and partially compensates the negative $\Delta Q_{\hat{\omega}}^T$. Overall, we find that IMRPhenomPv2NRTidalv2 provides a description that is much closer to TeOBResumS than the one offered by IMRPhenomPv2NRTidal, albeit being still slightly more attractive.

Spinning configurations are studied in Fig. 12, which shows $\Delta Q_{\hat{\omega}}$ for three target signals with $\tilde{\Lambda} = 400$ and increasing spins $\chi_1 = \chi_2$. Focusing on the low-frequency contribution to $\Delta Q_{\hat{\omega}}$, we observe that IMRPhenomPv2NRTidalv2 is now overall closer to TeOBResumS than IMRPhenomPv2NRTidal and TaylorF2 were. Additionally, the 5.5 PN point mass, which in the non-spinning case follows closely the behavior of TeOBResumS, becomes more negative as spins grow. Moving to the high-frequency regime, the improvements of NRTidalv2 have a positive effect on $\Delta Q_{\hat{\omega}}^T$ of the phenomenological approximant. Indeed, while the PhenomEOB-PN hierarchy displayed in Fig. 3 for spinning binaries is maintained, the differences decrease and NRTidalv2 is closer to the TeOBResumS description than NRTidal.

-
- [1] B. P. Abbott *et al.* (Virgo and LIGO Scientific Collaborations), *Phys. Rev. Lett.* **119**, 161101 (2017).
 - [2] D. Radice, A. Perego, F. Zappa, and S. Bernuzzi, *Astrophys. J.* **852**, L29 (2018).
 - [3] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. Lett.* **121**, 161101 (2018).
 - [4] R. Essick, P. Landry, and D. E. Holz, *Phys. Rev. D* **101**, 063007 (2020).
 - [5] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Classical Quantum Gravity* **37**, 045006 (2020).
 - [6] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. X* **9**, 031040 (2019).
 - [7] A. Ghosh, W. Del Pozzo, and P. Ajith, *Phys. Rev. D* **94**, 104070 (2016).
 - [8] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Classical Quantum Gravity* **37**, 055002 (2020).
 - [9] J. Veitch *et al.*, *Phys. Rev. D* **91**, 042003 (2015).
 - [10] E. E. Flanagan and T. Hinderer, *Phys. Rev. D* **77**, 021502 (2008).
 - [11] T. Damour and A. Nagar, *Phys. Rev. D* **81**, 084016 (2010).
 - [12] T. Damour and A. Nagar, *Fundam. Theor. Phys.* **162**, 211 (2011).
 - [13] T. Damour, P. Jaranowski, and G. Schfer, *Phys. Rev. D* **91**, 084024 (2015).
 - [14] B. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Astrophys. J. Lett.* **892**, L3 (2020).
 - [15] R. Dudi, F. Pannarale, T. Dietrich, M. Hannam, S. Bernuzzi, F. Ohme, and B. Bruegmann, *Phys. Rev. D* **98**, 084061 (2018).
 - [16] A. Samajdar and T. Dietrich, *Phys. Rev. D* **98**, 124030 (2018).
 - [17] A. Samajdar and T. Dietrich, *Phys. Rev. D* **100**, 024046 (2019).
 - [18] F. Messina, R. Dudi, A. Nagar, and S. Bernuzzi, *Phys. Rev. D* **99**, 124051 (2019).
 - [19] M. Agathos, F. Zappa, S. Bernuzzi, A. Perego, M. Breschi, and D. Radice, *Phys. Rev. D* **101**, 044006 (2020).
 - [20] T. Narikawa, N. Uchikata, K. Kawaguchi, K. Kiuchi, K. Kyutoku, M. Shibata, and H. Tagoshi, *Phys. Rev. Research* **2**, 043039 (2020).
 - [21] A. Chen, N. K. Johnson-McDaniel, T. Dietrich, and R. Dudi, *Phys. Rev. D* **101**, 103008 (2020).
 - [22] B. Sathyaprakash *et al.*, in *46th Rencontres de Moriond on Gravitational Waves and Experimental Gravity* (2011), pp. 127–136 [arXiv:1108.1423].
 - [23] J. Aasi *et al.* (LIGO Scientific Collaboration), *Classical Quantum Gravity* **32**, 115012 (2015).
 - [24] B. P. Abbott *et al.* (VIRGO, KAGRA, and LIGO Scientific Collaborations), *Living Rev. Relativity* **21**, 3 (2018); **19**, 1(E) (2016).
 - [25] G. M. Harry (LIGO Scientific Collaboration), *Classical Quantum Gravity* **27**, 084006 (2010).

- [26] F. Acernese *et al.* (VIRGO Collaboration), *Classical Quantum Gravity* **32**, 024001 (2015).
- [27] M. Maggiore *et al.*, *J. Cosmol. Astropart. Phys.* **03** (2020) 050.
- [28] C. Cutler and M. Vallisneri, *Phys. Rev. D* **76**, 104018 (2007).
- [29] M. Favata, *Phys. Rev. Lett.* **112**, 101101 (2014).
- [30] T. Hinderer, *Astrophys. J.* **677**, 1216 (2008).
- [31] T. Damour and A. Nagar, *Phys. Rev. D* **80**, 084035 (2009).
- [32] K. Yagi, *Phys. Rev. D* **89**, 043011 (2014).
- [33] B. J. Owen and B. Sathyaprakash, *Phys. Rev. D* **60**, 022002 (1999).
- [34] T. Damour, B. R. Iyer, and B. S. Sathyaprakash, *Phys. Rev. D* **57**, 885 (1998).
- [35] T. Damour, A. Nagar, and M. Trias, *Phys. Rev. D* **83**, 024006 (2011).
- [36] L. Lindblom, B. J. Owen, and D. A. Brown, *Phys. Rev. D* **78**, 124020 (2008).
- [37] C. Cutler and E. E. Flanagan, *Phys. Rev. D* **49**, 2658 (1994).
- [38] E. Poisson and C. M. Will, *Phys. Rev. D* **52**, 848 (1995).
- [39] T. Damour, A. Nagar, and L. Villain, *Phys. Rev. D* **85**, 123007 (2012).
- [40] D. Shoemaker *et al.*, LIGO Document No. LIGO-T0900288-v3, 2010.
- [41] I. Harry and T. Hinderer, *Classical Quantum Gravity* **35**, 145010 (2018).
- [42] B. Z. Bobrovsky, E. Mayer-Wolf, and M. Zakai, *Ann. Stat.* **15**, 1421 (1987).
- [43] Y. Aso, Y. Michimura, K. Somiya, M. Ando, O. Miyakawa, T. Sekiguchi, D. Tatsumi, and H. Yamamoto (KAGRA Collaboration), *Phys. Rev. D* **88**, 043007 (2013).
- [44] T. Akutsu *et al.* (KAGRA Collaboration), *Nat. Astron.* **3**, 35 (2019).
- [45] M. Punturo, M. Abernathy, F. Acernese, B. Allen, N. Andersson *et al.*, *Classical Quantum Gravity* **27**, 194002 (2010).
- [46] S. Hild *et al.*, *Classical Quantum Gravity* **28**, 094013 (2011).
- [47] K. Chatzioannou, A. Klein, N. Yunes, and N. Cornish, *Phys. Rev. D* **95**, 104004 (2017).
- [48] W. Kastaun and F. Ohme, *Phys. Rev. D* **100**, 103023 (2019).
- [49] M. Vallisneri, *Phys. Rev. D* **77**, 042001 (2008).
- [50] L. Blanchet, *Living Rev. Relativity* **17**, 2 (2014).
- [51] A. Buonanno, B. Iyer, E. Ochsner, Y. Pan, and B. Sathyaprakash, *Phys. Rev. D* **80**, 084043 (2009).
- [52] A. Buonanno and T. Damour, *Phys. Rev. D* **59**, 084006 (1999).
- [53] A. Buonanno and T. Damour, *Phys. Rev. D* **62**, 064015 (2000).
- [54] T. Damour, P. Jaranowski, and G. Schaefer, *Phys. Rev. D* **62**, 084011 (2000).
- [55] T. Damour, *Phys. Rev. D* **64**, 124013 (2001).
- [56] T. Damour, P. Jaranowski, and G. Schäfer, *Phys. Rev. D* **78**, 024009 (2008).
- [57] D. Bini, T. Damour, and A. Geralico, *Phys. Rev. Lett.* **123**, 231104 (2019).
- [58] D. Bini, T. Damour, and A. Geralico, *Phys. Rev. D* **102**, 024062 (2020).
- [59] D. Bini, T. Damour, and A. Geralico, *Phys. Rev. D* **102**, 024061 (2020).
- [60] L. Santamaria, F. Ohme, P. Ajith, B. Brügmann, N. Dorband *et al.*, *Phys. Rev. D* **82**, 064016 (2010).
- [61] M. Hannam, P. Schmidt, A. Bohé, L. Haegel, S. Husa, F. Ohme, G. Pratten, and M. Pürrer, *Phys. Rev. Lett.* **113**, 151101 (2014).
- [62] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. J. Forteza, and A. Bohé, *Phys. Rev. D* **93**, 044007 (2016).
- [63] S. Husa, S. Khan, M. Hannam, M. Pürrer, F. Ohme, X. J. Forteza, and A. Bohé, *Phys. Rev. D* **93**, 044006 (2016).
- [64] L. London, S. Khan, E. Fauchon-Jones, X. J. Forteza, M. Hannam, S. Husa, C. Kalaghatgi, F. Ohme, and F. Pannarale, *Phys. Rev. Lett.* **120**, 161102 (2018).
- [65] 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).
- [66] G. Pratten, S. Husa, C. Garcia-Quiros, M. Colleoni, A. Ramos-Buades, H. Estelles, and R. Jaume, *Phys. Rev. D* **102**, 064001 (2020).
- [67] G. Pratten *et al.*, arXiv:2004.06503.
- [68] K. G. Arun, A. Buonanno, G. Faye, and E. Ochsner, *Phys. Rev. D* **79**, 104023 (2009); **84**, 049901(E) (2011).
- [69] B. Mikoczi, M. Vasuth, and L. A. Gergely, *Phys. Rev. D* **71**, 124043 (2005).
- [70] J. Vines, E. E. Flanagan, and T. Hinderer, *Phys. Rev. D* **83**, 084051 (2011).
- [71] Q. Henry, G. Faye, and L. Blanchet, *Phys. Rev. D* **102**, 044033 (2020).
- [72] A. Nagar *et al.*, *Phys. Rev. D* **98**, 104052 (2018).
- [73] T. Damour and A. Nagar, *Phys. Rev. D* **90**, 044018 (2014).
- [74] A. Nagar, G. Riemenschneider, and G. Pratten, *Phys. Rev. D* **96**, 084045 (2017).
- [75] A. Nagar, F. Messina, P. Rettegno, D. Bini, T. Damour, A. Geralico, S. Akcay, and S. Bernuzzi, *Phys. Rev. D* **99**, 044007 (2019).
- [76] A. Nagar, G. Pratten, G. Riemenschneider, and R. Gamba, *Phys. Rev. D* **101**, 024041 (2020).
- [77] A. Nagar, G. Riemenschneider, G. Pratten, P. Rettegno, and F. Messina, *Phys. Rev. D* **102**, 024077 (2020).
- [78] S. Akcay, S. Bernuzzi, F. Messina, A. Nagar, N. Ortiz, and P. Rettegno, *Phys. Rev. D* **99**, 044051 (2019).
- [79] S. Bernuzzi, T. Dietrich, and A. Nagar, *Phys. Rev. Lett.* **115**, 091101 (2015).
- [80] D. Bini, T. Damour, and G. Faye, *Phys. Rev. D* **85**, 124034 (2012).
- [81] D. Bini and T. Damour, *Phys. Rev. D* **90**, 124037 (2014).
- [82] T. Damour, A. Nagar, and S. Bernuzzi, *Phys. Rev. D* **87**, 084035 (2013).
- [83] A. Nagar and P. Rettegno, *Phys. Rev. D* **99**, 021501 (2019).
- [84] Y. Pan, A. Buonanno, A. Taracchini, M. Boyle, L. E. Kidder, A. H. Mroué, H. P. Pfeiffer, M. A. Scheel, B. Szilágyi, and A. Zenginoglu, *Phys. Rev. D* **89**, 061501 (2014).
- [85] S. Babak, A. Taracchini, and A. Buonanno, *Phys. Rev. D* **95**, 024010 (2017).
- [86] P. Rettegno, F. Martinetti, A. Nagar, D. Bini, G. Riemenschneider, and T. Damour, *Phys. Rev. D* **101**, 104027 (2020).
- [87] T. Hinderer *et al.*, *Phys. Rev. Lett.* **116**, 181101 (2016).

- [88] J. Steinhoff, T. Hinderer, A. Buonanno, and A. Taracchini, *Phys. Rev. D* **94**, 104028 (2016).
- [89] A. Bohé *et al.*, *Phys. Rev. D* **95**, 044028 (2017).
- [90] B. D. Lackey, M. Prer, A. Taracchini, and S. Marsat, *Phys. Rev. D* **100**, 024002 (2019).
- [91] F. Messina and A. Nagar, *Phys. Rev. D* **95**, 124001 (2017); **96**, 049907(E) (2017).
- [92] T. Dietrich, S. Bernuzzi, and W. Tichy, *Phys. Rev. D* **96**, 121501 (2017).
- [93] T. Dietrich, A. Samajdar, S. Khan, N. K. Johnson-McDaniel, R. Dudi, and W. Tichy, *Phys. Rev. D* **100**, 044003 (2019).
- [94] L. Baiotti, T. Damour, B. Giacomazzo, A. Nagar, and L. Rezzolla, *Phys. Rev. D* **84**, 024017 (2011).
- [95] L. Baiotti, T. Damour, B. Giacomazzo, A. Nagar, and L. Rezzolla, *Phys. Rev. Lett.* **105**, 261101 (2010).
- [96] E. Poisson, *Phys. Rev. D* **57**, 5287 (1998).
- [97] S. Bernuzzi, A. Nagar, M. Thierfelder, and B. Brügmann, *Phys. Rev. D* **86**, 044030 (2012).
- [98] T. Dietrich, S. Bernuzzi, M. Ujevic, and W. Tichy, *Phys. Rev. D* **95**, 044045 (2017).
- [99] B. P. Abbott *et al.* (GROND, SALT Group, OzGrav, DFN, INTEGRAL, Virgo, Insight-Hxmt, MAXI Team, Fermi-LAT, J-GEM, RATIR, IceCube, CAASTRO, LWA, ePESSTO, GRAWITA, RIMAS, SKA South Africa/MeerKAT, H.E.S.S., 1M2H Team, IKI-GW Follow-up, Fermi GBM, Pi of Sky, DWF (Deeper Wider Faster Program), Dark Energy Survey, MASTER, AstroSat Cadmium Zinc Telluride Imager Team, Swift, Pierre Auger, ASKAP, VINROUGE, JAGWAR, Chandra Team at McGill University, TTU-NRAO, GROWTH, AGILE Team, MWA, ATCA, AST3, TOROS, Pan-STARRS, NuSTAR, ATLAS Telescopes, BOOTES, CaltechNRAO, LIGO Scientific, High Time Resolution Universe Survey, Nordic Optical Telescope, Las Cumbres Observatory Group, TZAC Consortium, LOFAR, IPN, DLT40, Texas Tech University, HAWC, ANTARES, KU, Dark Energy Camera GW-EM, CALET, Euro VLBI Team, and ALMA Collaborations), *Astrophys. J.* **848**, L12 (2017).
- [100] LIGO Scientific Collaboration, LIGO Algorithm Library—LALSuite, free software (GPL) (2018).
- [101] J. Veitch and A. Vecchio, *Phys. Rev. D* **81**, 062003 (2010).
- [102] L. Lindblom, *Phys. Rev. D* **82**, 103011 (2010).
- [103] M. F. Carney, L. E. Wade, and B. S. Irwin, *Phys. Rev. D* **98**, 063004 (2018).
- [104] E. Racine, *Phys. Rev. D* **78**, 044021 (2008).
- [105] E. Baird, S. Fairhurst, M. Hannam, and P. Murphy, *Phys. Rev. D* **87**, 024035 (2013).
- [106] J. S. Read, B. D. Lackey, B. J. Owen, and J. L. Friedman, *Phys. Rev. D* **79**, 124032 (2009).
- [107] G. Raaijmakers, T. Riley, and A. Watts, *Mon. Not. R. Astron. Soc.* **478**, 2177 (2018).
- [108] L. Lindblom and N. M. Indik, *Phys. Rev. D* **89**, 064003 (2014); **93**, 129903(E) (2016).
- [109] P. Landry, R. Essick, and K. Chatziioannou, *Phys. Rev. D* **101**, 123007 (2020).
- [110] R. Essick, I. Tews, P. Landry, S. Reddy, and D. E. Holz, *Phys. Rev. C* **102**, 055803 (2020).
- [111] S. De, D. Finstad, J. M. Lattimer, D. A. Brown, E. Berger, and C. M. Biwer, *Phys. Rev. Lett.* **121**, 091102 (2018); **121**, 259902(E) (2018).
- [112] C. A. Raithel, *Eur. Phys. J. A* **55**, 80 (2019).
- [113] K. Yagi and N. Yunes, *Classical Quantum Gravity* **33**, 13LT01 (2016).
- [114] K. Yagi and N. Yunes, *Phys. Rep.* **681**, 1 (2017).
- [115] A. Nitz *et al.*, gwastro/pycbc: Pycbc release v1.16.9, (2020).
- [116] R. Smith and G. Ashton, *Mon. Not. R. Astron. Soc.* **498**, 4492 (2020).
- [117] I. Romero-Shaw *et al.*, *Mon. Not. R. Astron. Soc.* **499**, 3295 (2020).
- [118] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *SoftwareX* **13**, 100658 (2021).
- [119] M. Breschi, S. Bernuzzi, F. Zappa, M. Agathos, A. Perego, D. Radice, and A. Nagar, *Phys. Rev. D* **100**, 104029 (2019).
- [120] S. Bernuzzi, A. Nagar, T. Dietrich, and T. Damour, *Phys. Rev. Lett.* **114**, 161103 (2015).
- [121] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. X* **9**, 011001 (2019).
- [122] L. Dai, T. Venumadhav, and B. Zackay, [arXiv:1806.08793](https://arxiv.org/abs/1806.08793).
- [123] B. Zackay, L. Dai, and T. Venumadhav, [arXiv:1806.08792](https://arxiv.org/abs/1806.08792).
- [124] T. Narikawa, N. Uchikata, K. Kawaguchi, K. Kiuchi, K. Kyutoku, M. Shibata, and H. Tagoshi, *Phys. Rev. Research* **1**, 033055 (2019).
- [125] C. Pankow *et al.*, *Phys. Rev. D* **98**, 084016 (2018).
- [126] L. Wade, J. D. E. Creighton, E. Ochsner, B. D. Lackey, B. F. Farr, T. B. Littenberg, and V. Raymond, *Phys. Rev. D* **89**, 103012 (2014).
- [127] D. Radice and L. Dai, *Eur. Phys. J. A* **55**, 50 (2019).
- [128] G. Raaijmakers *et al.*, *Astrophys. J. Lett.* **887**, L22 (2019).
- [129] G. Raaijmakers *et al.*, *Astrophys. J. Lett.* **893**, L21 (2020).
- [130] D. Reitze *et al.*, *Bull. Am. Astron. Soc.* **51**, 035 (2019).
- [131] B. P. Abbott *et al.* (LIGO Scientific Collaboration), *Classical Quantum Gravity* **34**, 044001 (2017).
- [132] E. Annala, T. Gorda, A. Kurkela, and A. Vuorinen, *Phys. Rev. Lett.* **120**, 172703 (2018).
- [133] C. D. Capano, I. Tews, S. M. Brown, B. Margalit, S. De, S. Kumar, D. A. Brown, B. Krishnan, and S. Reddy, *Nat. Astron.* **4**, 625 (2020).
- [134] <https://www.gw-openscience.org>
- [135] S. Bernuzzi, M. Thierfelder, and B. Brügmann, *Phys. Rev. D* **85**, 104030 (2012).
- [136] D. Radice, L. Rezzolla, and F. Galeazzi, *Mon. Not. R. Astron. Soc.* **437**, L46 (2014).
- [137] D. Radice, S. Bernuzzi, and C. D. Ott, *Phys. Rev. D* **94**, 064011 (2016).
- [138] S. Bernuzzi and T. Dietrich, *Phys. Rev. D* **94**, 064062 (2016).
- [139] T. Dietrich, D. Radice, S. Bernuzzi, F. Zappa, A. Perego, B. Brügmann, S. V. Chaurasia, R. Dudi, W. Tichy, and M. Ujevic, *Classical Quantum Gravity* **35**, 24LT01 (2018).
- [140] S. Bernuzzi, A. Nagar, S. Balmelli, T. Dietrich, and M. Ujevic, *Phys. Rev. Lett.* **112**, 201101 (2014).
- [141] T. Dietrich, M. Ujevic, W. Tichy, S. Bernuzzi, and B. Brügmann, *Phys. Rev. D* **95**, 024029 (2017).
- [142] T. Dietrich and T. Hinderer, *Phys. Rev. D* **95**, 124006 (2017).
- [143] T. Dietrich, S. Bernuzzi, B. Brügmann, M. Ujevic, and W. Tichy, *Phys. Rev. D* **97**, 064002 (2018).
- [144] K. Chatziioannou, C.-J. Haster, and A. Zimmerman, *Phys. Rev. D* **97**, 104036 (2018).