A novel ringdown amplitude-phase consistency test

Xisco Jiménez Forteza[1,2], Swetha Bhagwat[3,4], Sumit Kumar[1,2], Paolo Pani[3]
[1] *Max Planck Institute for Gravitational Physics (Albert Einstein Institute), Callinstraße 38, 30167 Hannover, Germany*
[2] *Leibniz Universität Hannover, 30167 Hannover, Germany*
[3] *Dipartimento di Fisica, "Sapienza" Università di Roma & Sezione INFN Roma1, Piazzale Aldo Moro 5, 00185, Roma, Italy and*
[4] *Institute for Gravitational Wave Astronomy & School of Physics and Astronomy, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK*

The ringdown signal emitted during a binary black hole coalescence can be modeled as a linear superposition of the characteristic damped modes of the remnant black hole that get excited during the merger phase. While checking the consistency of the measured frequencies and damping times against the Kerr BH spectrum predicted by General Relativity (GR) is a cornerstone of strong-field tests of gravity, the consistency of measured excitation amplitudes and phases have been largely left unexplored. For a nonprecessing, quasi-circular binary black hole merger, we find that GR predicts a narrow region in the space of mode amplitude ratio and phase difference, independently of the spin of the binary components. Using this unexpected result, we develop a new null test of strong-field gravity which demands that the measured amplitudes and phases of different ringdown modes should lie within this narrow region predicted by GR. We call this the *amplitude-phase consistency test* and introduce a procedure for performing it using information from the ringdown signal. Lastly, we apply this test to the GW190521 event, using the multimodal ringdown parameters inferred by Capano et al. (2021) [1]. While ringdown measurements errors for this event are large, we show that GW190521 is consistent with the amplitude-phase consistency test. Our test is particularly well suited for accommodating multiple loud ringdown detections as those expected in the near future, and can be used complementarily to standard black-hole spectroscopy as a proxy for modified gravity, compact objects other than black holes, and binary precession.

Introduction. A binary black hole (BBH) ringdown is the gravitational-wave (GW) signal emitted as the remnant black hole (BH) formed during a BBH coalescence relaxes towards a stationary configuration [2–5]. According to Einstein's general theory of relativity (GR), the final state is uniquely described by the Kerr metric [6–9]. After a transient regime, the dynamics of this relaxation phase can be described as the evolution of perturbations (set up by the binary's initial conditions) on the metric of the remnant BH. The GW signal emitted during this phase can be well-approximated as an infinite linear superposition of quasi-normal modes (QNMs), i.e. damped sinusoids modes with discrete characteristic complex frequency spectra [10–13]. Each mode is identified by three integers (l, m, n) and defined in terms of its frequency f_{lmn}, damping time τ_{lmn}, excitation amplitude A_{lmn}, and phase ϕ_{lmn} (see Eq. (2) below).

While the frequencies and damping times depend only on the remnant's mass and spin, each ringdown mode can be independently excited with amplitude and phase set by the features and symmetries of the initial perturbation. For ringdown signals produced by quasi-circular BBH coalescences, A_{lmn} and ϕ_{lmn} are solely determined by the BBH parameters – specifically, their masses and spins – and can be calculated using numerical relativity (NR) simulations [14–16]. Thus, ringdown allows us to check two key predictions of GR in the strong-field regime: a) if the QNM spectrum is consistent with that of a Kerr BH with the expected mass and spin, and b) if the mode excitation is consistent with the dynamics of the (pre-)merger as predicted for fully nonlinear BBH simulations within GR.

Here we focus on the latter possibility and propose a novel test of GR that we dub *ringdown amplitude-phase consistency (APC) test*. After a suitable normalization, we find that only a narrow region of the mode amplitude-phase plane is allowed for a BBH ringdown in GR. On the other hand, different values of the amplitudes and phases are generically expected not only for BBHs in GR extensions [17–21], but also within GR if the binary components are not BHs [22] (e.g. in neutron-stars or even more exotic boson-star [23–26] coalescences). In both cases the remnant can still be a Kerr BH (therefore passing standard QNM tests), but the QNM amplitudes and phases are model dependent and can provide a way to distinguish the nature of the merger.

One virtue of this test is that it does not require information from the inspiral phase other than the binary extrinsic parameters. It is therefore particularly well suited for massive BBHs, where the inspiral is short and the parameter estimation of the binary intrinsic parameters (e.g. the mass ratio) is affected by large uncertainties, which jeopardize the accuracy of inspiral-merger-ringdown (IMR) consistency tests [27]. This is the case of GW190521 [28], which is also the only event detected thus far for which a measurement of subdominant angular-mode parameters has been reported [1]. GW190521 is an outstanding event also because at least its primary component mass lies in the BH mass gap predicted by the pair-instability supernova theory, thus suggesting also exotic explanations [29–31] that would impact on the QNM amplitudes and phases.

QNM amplitude and phase fits. We focus on nonprecessing, quasi-circular BBHs; we comment on the role

of spin misalignment and eccentricity in the final discussion. It is convenient to use the dominant ($l = m = 2$, $n = 0$) mode as a baseline and work with the intrinsic amplitude ratio and phase difference defined as

$$A^R_{lmn} \equiv \frac{A_{lmn}}{A_{220}}\,, \qquad \delta\phi_{lmn} \equiv \frac{m}{2}\phi_{220} - \phi_{lmn}\,, \tag{1}$$

respectively [32]. These quantities are partially independent of the extrinsic parameters like sky position, distance, polarization ψ, and inclination ι. In particular, the definition of $\delta\phi_{lmn}$ removes the degeneracy of the mode phase ϕ_{lmn} and the coalescence phase φ (see Appendix A 6 and Ref. [33] for a related discussion). In this parametrization, the ringdown waveform can be analytically modeled as

$$h_+ + i\,h_\times = A_{220}\sum_{lmn}\Big(e^{-i\frac{m}{2}\phi_{220}}A^R_{lmn}e^{i\delta\phi_{lmn}}S_{lmn}(\iota,\varphi) \times e^{i2\pi f_{lmn}t}e^{-t/\tau_{lmn}}\Big)\,, \tag{2}$$

where $S_{lmn}(\iota,\varphi)$ are the spin 2-weighted spheroidal harmonic basis functions which for simplicity we approximate to spherical harmonic functions [34]. This assumption can be easily relaxed and, as later discussed, introduces a systematic error no larger than 1% for spinning remnants with $a_f \lesssim 0.9$ and for the modes considered in this work.

For a quasi-circular BBH, A^R_{lmn} and $\delta\phi_{lmn}$ are functions of the binary mass ratio $q = m_1/m_2 \geq 1$ and spins $\{\chi_1, \chi_2\}$. We fit for A^R_{lmn} and $\delta\phi_{lmn}$ mode-wise for a set of 142 nonprecessing and noneccentric NR simulations from the SXS catalog [35]. Our simulation set is consistent with that used to calibrate the `SEOBNRv4HM` model [36] and spans $q \in [1, 10]$ and $\chi_{1,2} \in [-0.9, 0.9]$. We fit for the $(lmn) \in \{(330), (210), (440)\}$ modes and provide ready-to-use analytical fits in terms of q and a post-Newtonian motivated effective spin $\chi_{\rm pheno}(q, \chi_1, \chi_2)$ whose explicit form depends on the mode under consideration (see Appendix A). In the nonspinning limit, these fits are consistent with the those obtained in [36–39], while they correct the results given in [40] for the phase difference. Here, we will highlight the main aspects of the fit significant to the APC test and refer to Appendix A for details.

In Figure 1, each shaded area delimits the regions on the $A^R_{lmn} - \delta\phi_{lmn}$ plane estimated by fitting the amplitudes and phases of a given mode using Eq. (2) to our dataset of ringdown NR waveforms. They therefore bracket the admissible range for BBH ringdowns within GR. The dashed curves correspond to the nonspinning limit ($\chi_{1,2} = 0$), wherein A^R_{lmn} and $\delta\phi_{lmn}$ parametrically depends on the mass ratio q only. On the other hand, the spread of each shaded region around the dashed curve quantifies the effects of the progenitor spins $\chi_{1,2}$. Given the parity and polarization conventions of the odd/even m modes used for the SXS waveforms, we plot $\mathrm{mod}(\delta\phi_{lmn}, \pi)$ for the (210) and (330) modes and $\mathrm{mod}(\delta\phi_{lmn}, 2\pi)$ for the (440) mode (see Appendix A).

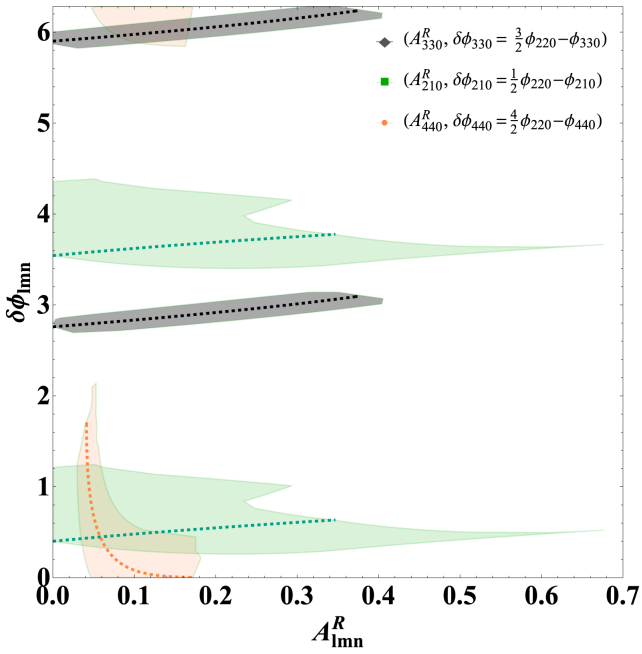

FIG. 1. The regions of the $A^R_{lmn} - \delta\phi_{lmn}$ plane allowed for BBHs in GR for various modes. They are estimated by fitting the parameters $A^R_{lmn} - \delta\phi_{lmn}$ of Eq. (2) to 142 nonprecessing waveforms, covering the parameter space $q \in [1, 10]$ and $\chi_{1,2} \in [-0.9, 0.9]$. Given the parity of the odd/even m modes, we plot $\mathrm{mod}(\delta\phi_{lmn}, \pi)$ for the (210) and (330) modes and $\mathrm{mod}(\delta\phi_{lmn}, 2\pi)$ for the 440 mode. The dashed curve within each region corresponds to the nonspinning limit.

We find that, for the (330) mode, this GR admissible region of the parameter space is remarkably narrow, since the effects of the progenitor spins are small. In particular, $\delta\phi_{330} \in [2.68, \pi]$ and $A_{330} \in [0, 0.42]$ for our entire dataset. On the other hand, spin effects on both A^R_{lmn} and $\delta\phi_{lmn}$ are larger for the (440) mode and, especially, for the (210) mode, which translates into wider allowed regions.

The ringdown APC test. *Any* nonprecessing, quasi-circular BBH ringdown in GR must be characterized by amplitude and phases lying in the narrow GR admissible region on the $A^R_{lmn} - \delta\phi_{lmn}$ space presented in Fig. 1. We can use this feature to devise a null test of strong gravity wherein we demand that the measured amplitudes and phases of the modes in a ringdown signal must lie in this narrow region. Conceptually, the ringdown APC test involves ensuring that the posterior distribution of the estimated mode amplitude ratios and phase differences have significant support in the allowed region. Since this should be the case for all quasi-circular BBH mergers regardless of the binary parameters, this test can be easily extended to incorporate a population of observations.

However, while performing this test it is important to take into account the uncertainty about the ringdown start time compared to the global peak time $t^{\rm p}$ at which the strain amplitude $|h(t) = \Sigma_{lmn} h_{lmn}|$ of the signal maximizes [40, 41]. Given the time $t^{\rm p}_{lmn}$ at which

each $|h_{lmn}|$ mode peaks (as computed from NR fits), we arbitrarily shift each mode by $\Delta t = t^{\rm p} - t^{\rm p}_{220}$, with $t^{\rm p}_{220} < t^{\rm p}_{lm0}$ [40, 42]. This induces a correction to the amplitude ratio,

$$A^R_{lmn}(t^{\rm p}) = A^R_{lmn}(t^{\rm p}_{220}) e^{\frac{\Delta t}{\tau_{220}} - \frac{\Delta t}{\tau_{lmn}}} \,. \tag{3}$$

Note that the τ_{lmn} are similar for the modes considered here. For example, for a BH with final spin $a_f \approx 0.85$ consistent with GW190521 [1, 28], $\tau_{330}/\tau_{220} \sim 0.986$, $\tau_{440}/\tau_{220} \sim 0.974$, and $\tau_{210}/\tau_{220} \sim 0.956$. Therefore, even a conservative choice $\Delta t \approx 10M$ (which is equivalent to $\Delta t \approx 2(t^{\rm p}_{330} - t^{\rm p}_{220}) \approx (t^{\rm p}_{210} - t^{\rm p}_{220})$ [40] and therefore sets a conservative upper bound on Δt) translates into a $\approx (1, 2, 3)\%$ correction for the amplitude of the (330), (440), (210) modes, respectively, well below the current statistical uncertainties [1].

A similar correction needs to be accounted for the intrinsic phase, $\phi_{lmn}(t^{\rm p}) = \phi_{lmn}(t^{\rm p}_{220}) + \omega_{lmn}\Delta t$. Using Eq. (1), this translates into the phase difference

$$\delta\phi_{lmn}(t^{\rm p}) = \delta\phi_{lmn}(t^{\rm p}_{220}) + \left(\frac{m}{2}\omega_{220} - \omega_{lmn}\right)\Delta t. \tag{4}$$

Since $\omega_{lmn} \approx \frac{l}{2}\omega_{220}$ (as can be exactly proven in the eikonal, $l = m \gg 1$, limit [12, 43]), for $l = m$ modes such as (330) and (440) the phase correction induced by Δt is $\approx 10\%$ and $\approx 20\%$ for the (330) and (440) mode, respectively, again assuming $\Delta t = 10M$ (see Appendix A 6). This is a rather conservative choice for the (330) mode, since in this case $\Delta t \lesssim 5M$ [40], leading to a systematic uncertainty no larger than 4%. In other words, for the $l = m$ mode the phase fits can be directly compared to the inference results as,

$$\delta\phi_{lmn}(t^{\rm p}) \approx \delta\phi_{lmn}(t^{\rm p}_{220}) \quad \forall \quad l = m. \tag{5}$$

On the other hand, for the (210) mode the uncertainty coming from the last term in Eq. (4) is much larger.

Application on GW190521. GW190521 is the only event observed by the LIGO-Virgo-KAGRA (LVK) Collaboration thus far that has allowed for a measurement of the subdominant angular mode in the ringdown [1]. The total signal-to-noise ratio (SNR) of this event is $\rho \approx 14$, of which $\rho \approx 12$ from the ringdown phase only – with $\sqrt{14^2 - 12^2} \sim 7$ from the inspiral –, due to the large total source mass of the binary ($M_{\rm tot} = 151^{+29}_{-17} M_\odot$ [28]). This makes GW190521 ideal for ringdown-only tests and, as a proof of principle, here we apply our ringdown APC test to this event.

In Fig. 2, we plot the 67% and 95% credible regions in the $A^R_{330} - \delta\phi_{330}$ plane obtained from the posterior distributions provided in [1]. The orange contour refers to the parameter estimation obtained by assuming the Kerr QNM spectrum as predicted in GR, whereas the blue contour generalizes this result to the case in which the QNM frequencies and damping times are allowed to vary. Interestingly, this more general and agnostic assumption does not deteriorate the confidence region significantly.

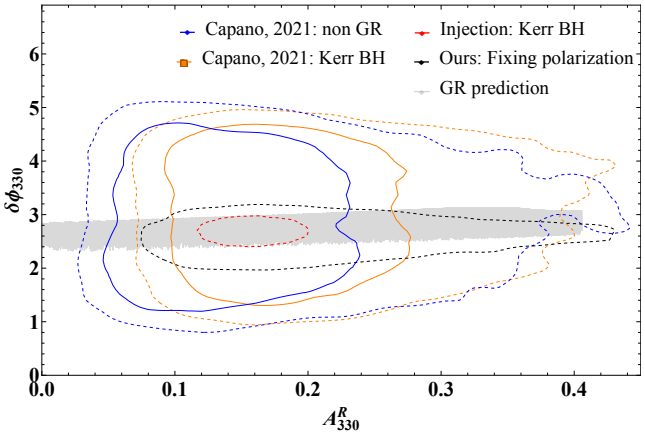

FIG. 2. Ringdown APC test applied to GW190521. The amplitude ratio A^R_{330} and the phase difference $\delta\phi_{330}$ contours obtained from [1] are compared to the GR prediction (shadowed gray area). The solid and dashed contours represent the 67% and 95% credible levels, respectively. The orange and blue contours correspond to the posteriors obtained by fixing the GR QNM spectrum or allowing deviations from the Kerr QNMs, respectively. The black dot-dashed contour provides the 95% credible region obtained as in [1] but fixing the polarization to the maximum likelihood value ($\psi = 0$) given by the IMR analysis [44]. The red dot-dashed contour provides the 95% credible region for the NR injection SXS:0258 consistent with GW190521 and with SNR $\rho = 30$.

The gray shadowed area marks the GR-permissible region on the $A^R_{330} - \delta\phi_{330}$ plane, obtained by taking already into account the following uncertainties – the 1σ deviations on the best fit results, the $\sim 1\%$ and $\sim 4\%$ uncertainties on A^R_{330} and $\delta\phi_{330}$, respectively, sourced by the ringdown start-time ambiguity discussed above, and the spherical-vs-spheroidal harmonic approximation. The black dot-dashed contour provides the 95% credible region obtained as in [1] but fixing the polarization angle ψ to the maximum likelihood value estimated from the full IMR analysis in [44], as already done in [1] for the other two antenna pattern parameters, namely the right ascension and declination (see also [45–47] for similar approaches in standard BH spectroscopy tests). Estimating the polarization angle independently is required to break the degeneracy between ψ and $\delta\phi_{lmn}$. Indeed, ψ may be estimated solely from the inspiral-merger regime, while the intrinsic dependence of $\delta\phi_{lmn}$ on q and $\chi_{\rm pheno}$ (see Appendix A 6) only arises in the ringdown phase. In this case we observe a small bias on $\delta\phi_{330}$ that may be sourced by the weak constraints obtained on ψ for GW190521. Finally, the red dot-dashed contour in Fig. 2 denotes the 95% credible region obtained from injecting the SXS NR waveform SXS:0258 (i.e. with parameters similar to GW190521 [48]) into Gaussian noise at SNR $\rho = 30$. For this injection we considered a three-detector configuration (Livingston-Hanford-Virgo) with the polarization angle fixed to the IMR maximum likelihood value and performed the parameter estimation us-

ing the PyCBC Inference library [49]. This injection is representative for the quality of the test achievable with a GW190521-like event at SNR twice as much the actual value for GW190521. As expected in this case the confidence region shrinks and the test is significantly more accurate.

As shown in Fig. 2, the 1σ credible intervals obtained in [1] contain the GR permissible range. This establishes that the posteriors are compatible with GR and that GW19021 passes the ringdown APC test.

Discussion. The APC test provides a novel strategy to test GR using the ringdown mode excitation. NR waveforms of BBH mergers can be used to obtain accurate empirical relations for the mode amplitude ratio A^R_{lmn} and phase difference $\delta\phi_{lmn}$ as functions of the binary mass ratio and spins. We found a tight relation between A^R_{lmn} and $\delta\phi_{lmn}$ which can be tested with GW multimode ringdown measurements, as in GW190521. Indeed, we showed that for GW190521 the $A^R_{330}-\delta\phi_{330}$ relation is consistent with the 1σ posterior distributions presented in [1].

We have tested the accuracy of our fits to SXS waveforms against both the RIT [50] and Maya [51] NR catalogs (see Appendix A 4). The fitting uncertainties obtained on the (330) mode are well below the statistical uncertainties of the current GW observations. Likewise, approximating the spheroidal harmonics in Eq. (2) with the spherical harmonics basis used in NR simulations introduces mode mixing [34, 52, 53] which is typically small. For GW190521 this affects A^R_{330} and $\delta\phi_{330}$ at the percent level. The uncertainty on the ringdown start time also affects both A^R_{330} and $\delta\phi_{330}$ at the percent level. This holds true for any $l=m$ mode, while the phase of $l\neq m$ modes is more affected by a shift of the peak time, making $l\neq m$ modes not optimal for our test.

Overall, for the (330) subleading mode the total systematic errors accumulated from the fit, ringdown start time, and spherical-harmonic approximation may reach a value of about $\sim 13\%$ for both A^R_{lmn} and $\delta\phi_{lmn}$. For GW190521, the marginalized 1σ statistical uncertainty on these quantity is about 100%, therefore much larger than the systematic deviations accumulated mainly from the current fit uncertainties. However, the situation might change for louder detections as those routinely expected in the third-generation era [54, 55], in which case systematic errors of the fit might limit the accuracy of the APC test, unless the quality of NR waveforms improves. On the other hand, the measurements of the polarization angle ψ (which is degenerate with the phase difference $\delta\phi_{lmn}$) and of other binary's intrinsic parameters are expected to improve as more interferometers are added to the network, or through an electromagnetic counterpart, and will anyway improve with third-generation detectors, therefore allowing for a more accurate test.

We have focused on quasi-circular binaries with aligned spins, although the same test can be applied to more generic cases. In Appendix A 7, we estimate that unless the eccentricity $e\gtrsim 0.3$ ($e\gtrsim 0.6$) its effect on the fits of $\delta\phi_{330}$ (A^R_{330}) is negligible. Likewise, we have found that the binary spins have a subleading effects on $A^R_{lmn}-\delta\phi_{lmn}$ compared to the mass ratio, especially for the (330) mode which is typically the most relevant one for the test. This suggests that – at least for this mode – the effect of spin precession should also be subdominant. On the other hand, for the (210) and (440) modes the effect of the binary spins is more relevant, so for these modes spin precession should be accurately modeled.

Indeed, as any null-hypothesis consistency test, its potential failure would signal a departure from the adopted baseline model, and such a departure could have multiple origins. Most conservatively, a failure of the test proposed here can be seen as evidence for spin precession or large eccentricity in the BBH. Another possibility is that the observed ringdown was not originated from a BBH, even if the remnant is a standard Kerr BH and GR is correct. Indeed, the coalescence of sufficiently massive neutron stars, boson stars [23–26], or many other exotic compact objects [22] might still produce a ringdown with QNMs consistent with those of a Kerr BH in GR, but with different QNM amplitudes and phases, thus failing our test while passing standard tests based on BH spectroscopy. In this context, although measurement errors are large, it is relevant that GW190521 passes the APC test. It would be interesting to assess whether this is in tension with alternative explainations for this event, e.g. a Proca star merger [29]. Finally, a departure from the APC test can signal a violation of GR, even in those cases in which stationary BHs are described by the same Kerr solution and would therefore pass standard BH spectroscopy tests. Disentangling these three possibilities might require a generalization of our fits and multiple loud ringdown detections – since, for example, the effects of precession might appear only in a subclass of events, whereas GR deviations should be ubiquitous.

While we focused on applying the test using only the ringdown signal and prior knowledge of the binary extrinsic parameters (coming either from the inspiral or from an independent sky localization), a variant of the test is to check for IMR consistency using the mode excitation amplitudes. As discussed in Appendix B, in principle one could invert the $A^R_{lmn}(q,\chi_{\rm pheno})$ and $\delta\phi_{lmn}(q,\chi_{\rm pheno})$ relations to infer an estimate of the mass ratio (and spins) from the QNM excitations and then compare these estimates with those inferred from the inspiral. However, owing to the mild dependence of $\delta\phi_{lmn}$ on the binary parameters the quality of this test is expected to be rather poor. A more promising avenue is to neglect the phases and use only the amplitude ratios of several subleading QNMs. This might be particularly relevant for systems like GW190521 where, due to the short duration and relatively low SNR at low frequencies, the inference on the binary parameters is controversial and model dependent [44, 56, 57]. Also in this case the better accuracy (especially at low frequency) of third-generation detectors might significantly improve this test (see also [58] for a conceptual framework in this direction).

Overall, given the amount of observations expected in the near future and the prospect in the third-generation era, we argue that the APC test provides an excellent arena to complement standard BH spectroscopy tests in the strong-gravity regime.

Acknowledgments. We acknowledge the Max Planck Gesellschaft for support, and we are grateful to the Atlas cluster computing team at AEI Hannover for their help. The authors are specially thankful to Lionel London, Cecilio García-Quiros, and Juan Calderon-Bustillo for the invaluable discussions and further clarifications about the NR phase alignment and phase conventions. X. Jimenez is also thankful to P. Mourier for the useful discussions about the correspondence of the fit and parameter-estimation results. S.B. is supported by the UKRI Stephen Hawking Fellowship, grant ref. EP/W005727. P.P. acknowledges financial support provided under the European Union's H2020 ERC, Starting Grant agreement no. DarkGRA–757480. We also acknowledge support under the MIUR PRIN (Grant 2020KR4KN2 "String Theory as a bridge between Gauge Theories and Quantum Gravity") and FARE (GW-NEXT, CUP: B84I20000100001, 2020KR4KN2) programmes, and from the Amaldi Research Center funded by the MIUR program "Dipartimento di Eccellenza" (CUP: B81I18001170001).

[1] C. D. Capano, M. Cabero, J. Westerweck, J. Abedi, S. Kastha, A. H. Nitz, A. B. Nielsen, and B. Krishnan, (2021), arXiv:2105.05238 [gr-qc].
[2] S. Chandrasekhar and S. L. Detweiler, Proc. Roy. Soc. Lond. **A344**, 441 (1975).
[3] S. A. Teukolsky, Astrophys. J. **185**, 635 (1973).
[4] W. H. Press and S. A. Teukolsky, Astrophys. J. **185**, 649 (1973).
[5] S. A. Teukolsky and W. H. Press, Astrophys. J. **193**, 443 (1974).
[6] J. D. Bekenstein, Phys. Rev. **D7**, 2333 (1973).
[7] B. Carter, Phys. Rev. Lett. **26**, 331 (1971).
[8] S. W. Hawking, Phys. Rev. Lett. **26**, 1344 (1971).
[9] S. W. Hawking and J. B. Hartle, Commun. Math. Phys. **27**, 283 (1972).
[10] E. Berti, V. Cardoso, and A. O. Starinets, Class. Quant. Grav. **26**, 163001 (2009), arXiv:0905.2975 [gr-qc].
[11] S. Chandrasekhar, *The mathematical theory of black holes* (Oxford Classic Texts in the Physical Sciences, 1985).
[12] V. Ferrari and B. Mashhoon, Phys. Rev. D **30**, 295 (1984).
[13] K. D. Kokkotas and B. G. Schmidt, Living Rev. Rel. **2**, 2 (1999), arXiv:gr-qc/9909058 [gr-qc].
[14] X. Jiménez-Forteza, D. Keitel, S. Husa, M. Hannam, S. Khan, and M. Pürrer, Phys. Rev. **D95**, 064024 (2017), arXiv:1611.00332 [gr-qc].
[15] C. O. Lousto and J. Healy, Phys. Rev. **D93**, 124074 (2016), arXiv:1601.05086 [gr-qc].
[16] F. Hofmann, E. Barausse, and L. Rezzolla, Astrophys. J. **825**, L19 (2016), arXiv:1605.01938 [gr-qc].
[17] M. Okounkova, Phys. Rev. D **102**, 084046 (2020), arXiv:2001.03571 [gr-qc].
[18] M. Okounkova, L. C. Stein, J. Moxon, M. A. Scheel, and S. A. Teukolsky, Phys. Rev. D **101**, 104016 (2020), arXiv:1911.02588 [gr-qc].
[19] W. E. East and J. L. Ripley, Phys. Rev. Lett. **127**, 101102 (2021), arXiv:2105.08571 [gr-qc].
[20] M. Elley, H. O. Silva, H. Witek, and N. Yunes, (2022), arXiv:2205.06240 [gr-qc].
[21] H. Lim, G. Khanna, and S. A. Hughes, (2022), arXiv:2204.06007 [gr-qc].
[22] V. Cardoso and P. Pani, Living Rev. Rel. **22**, 4 (2019), arXiv:1904.05363 [gr-qc].
[23] C. Palenzuela, P. Pani, M. Bezares, V. Cardoso, L. Lehner, and S. Liebling, Phys. Rev. D **96**, 104058 (2017), arXiv:1710.09432 [gr-qc].
[24] T. Helfer, E. A. Lim, M. A. G. Garcia, and M. A. Amin, Phys. Rev. D **99**, 044046 (2019), arXiv:1802.06733 [gr-qc].
[25] M. Bezares and C. Palenzuela, Class. Quant. Grav. **35**, 234002 (2018), arXiv:1808.10732 [gr-qc].
[26] M. Bezares, M. Bošković, S. Liebling, C. Palenzuela, P. Pani, and E. Barausse, Phys. Rev. D **105**, 064067 (2022), arXiv:2201.06113 [gr-qc].
[27] R. Abbott *et al.* (LIGO Scientific, VIRGO, KAGRA), (2021), arXiv:2112.06861 [gr-qc].
[28] R. Abbott *et al.* (LIGO Scientific, Virgo), Phys. Rev. Lett. **125**, 101102 (2020), arXiv:2009.01075 [gr-qc].
[29] J. C. Bustillo, N. Sanchis-Gual, A. Torres-Forné, J. A. Font, A. Vajpeyi, R. Smith, C. Herdeiro, E. Radu, and S. H. W. Leong, Phys. Rev. Lett. **126**, 081101 (2021), arXiv:2009.05376 [gr-qc].
[30] M. Fishbach and D. E. Holz, Astrophys. J. Lett. **904**, L26 (2020), arXiv:2009.05472 [astro-ph.HE].
[31] J. C. Bustillo, S. H. W. Leong, K. Chandra, B. McKernan, and K. E. S. Ford, (2021), arXiv:2112.12481 [gr-qc].
[32] J. Calderón Bustillo, A. Bohé, S. Husa, A. M. Sintes, M. Hannam, and M. Pürrer, (2015), arXiv:1501.00918 [gr-qc].
[33] V. Baibhav, E. Berti, and V. Cardoso, Phys. Rev. D **101**, 084053 (2020), arXiv:2001.10011 [gr-qc].
[34] E. Berti and A. Klein, Phys. Rev. **D90**, 064012 (2014), arXiv:1408.1860 [gr-qc].
[35] The SXS Collaboration, "SXS Gravitational Waveform Database," (2019).
[36] R. Cotesta, A. Buonanno, A. Bohé, A. Taracchini, I. Hinder, and S. Ossokine, (2018), arXiv:1803.10701 [gr-qc].
[37] S. Borhanian, K. G. Arun, H. P. Pfeiffer, and B. S. Sathyaprakash, (2019), arXiv:1901.08516 [gr-qc].
[38] L. London and E. Fauchon-Jones, Class. Quant. Grav. **36**, 235015 (2019), arXiv:1810.03550 [gr-qc].
[39] I. Ota and C. Chirenti, Phys. Rev. D **101**, 104005 (2020), arXiv:1911.00440 [gr-qc].
[40] X. Jiménez Forteza, S. Bhagwat, P. Pani, and V. Ferrari, Phys. Rev. D **102**, 044053 (2020), arXiv:2005.03260 [gr-qc].
[41] S. Bhagwat, X. J. Forteza, P. Pani, and V. Ferrari, Phys. Rev. **D101**, 044033 (2020), arXiv:1910.08708 [gr-qc].
[42] H. Estellés, S. Husa, M. Colleoni, D. Keitel, M. Mateu-

Lucena, C. García-Quirós, A. Ramos-Buades, and A. Borchers, (2020), arXiv:2012.11923 [gr-qc].
[43] V. Cardoso, A. S. Miranda, E. Berti, H. Witek, and V. T. Zanchin, Phys. Rev. D **79**, 064016 (2009), arXiv:0812.1806 [hep-th].
[44] A. H. Nitz and C. D. Capano, Astrophys. J. Lett. **907**, L9 (2021), arXiv:2010.12558 [astro-ph.HE].
[45] R. Cotesta, G. Carullo, E. Berti, and V. Cardoso, (2022), arXiv:2201.00822 [gr-qc].
[46] M. Isi and W. M. Farr, (2022), arXiv:2202.02941 [gr-qc].
[47] E. Finch and C. J. Moore, (2022), arXiv:2205.07809 [gr-qc].
[48] The progenitor parameters of the simulation are $q = 2$ and $\chi_{\rm pheno} = 0.48$, with dimensionless final spin $a_f = 0.84$ and we have scaled the waveform to the total detector-frame mass consistent with the ringdown. The injected waveform includes the following modes $(2,\pm2),(2,\pm1),(3,\pm3),(3,\pm2),(4,\pm4),(4,\pm3)$. We fix the polarization to the maximum likelihood value $\psi = 0$, consistent with [44].
[49] C. M. Biwer, C. D. Capano, S. De, M. Cabero, D. A. Brown, A. H. Nitz, and V. Raymond, Publ. Astron. Soc. Pac. **131**, 024503 (2019), arXiv:1807.10312 [astro-ph.IM].
[50] M. Campanelli, J. Healy, C. Lousto, and Y. Zlochower, "CCRG@RIT Catalog of Numerical Simulations," (2022).
[51] K. Jani, J. Healy, J. A. Clark, L. London, P. Laguna, and D. Shoemaker, "Georgia Tech catalog of binary black hole simulations," (2016).
[52] G. B. Cook and M. Zalutskiy, Phys. Rev. D **90**, 124021 (2014), arXiv:1410.7698 [gr-qc].
[53] C. García-Quirós, M. Colleoni, S. Husa, H. Estellés, G. Pratten, A. Ramos-Buades, M. Mateu-Lucena, and R. Jaume, Phys. Rev. D **102**, 064002 (2020), arXiv:2001.10914 [gr-qc].
[54] M. Maggiore *et al.*, JCAP **03**, 050 (2020), arXiv:1912.02622 [astro-ph.CO].
[55] V. Kalogera *et al.*, (2021), arXiv:2111.06990 [gr-qc].
[56] S. Kastha, C. D. Capano, J. Westerweck, M. Cabero, B. Krishnan, and A. B. Nielsen, (2021), arXiv:2111.13664 [gr-qc].
[57] R. Abbott *et al.* (LIGO Scientific, VIRGO, KAGRA), (2021), arXiv:2111.03606 [gr-qc].
[58] S. Bhagwat, C. Pacilio, E. Barausse, and P. Pani, (2021), arXiv:2201.00023 [gr-qc].
[59] X. Jiménez Forteza and S. Bhagwat, (2022), in preparation.
[60] D. Keitel, X. Jimenez-Forteza, *et al.*, Phys. Rev. **D96**, 024006 (2017), arXiv:1612.09566 [gr-qc].
[61] E. Barausse, A. Buonanno, S. A. Hughes, G. Khanna, S. O'Sullivan, and Y. Pan, Phys. Rev. **D85**, 024046 (2012), arXiv:1110.3081 [gr-qc].
[62] L. London, D. Shoemaker, and J. Healy, Phys. Rev. **D90**, 124032 (2014), [Erratum: Phys. Rev.D94,no.6,069902(2016)], arXiv:1404.3197 [gr-qc].
[63] X. J. Forteza and P. Mourier, (2021), arXiv:2107.11829 [gr-qc].
[64] B. P. Abbott *et al.* (LIGO Scientific, Virgo), Astrophys. J. **883**, 149 (2019), arXiv:1907.09384 [astro-ph.HE].

Appendix A: Fits for A^R_{lmn} and $\delta\phi_{lmn}$

For each of the 142 NR SXS waveforms in our dataset, we fit for A^R_{lmn} and $\delta\phi_{lmn}$ for $(lmn) = (330)$, (210), and (440) modes. The waveform dataset spans $q \in [1, 10]$ and $\chi_{1,2} \in [-0.9, 0.9]$. We fit for the amplitudes A_{lmn} and phases ϕ_{lmn} of each h_{lmn} mode using the following ansatz,

$$h_{lmn} = A_{lmn} e^{-i\phi_{lmn}} e^{i2\pi f_{lmn} t} e^{-t/\tau_{lmn}}\,, \tag{A1}$$

where the frequencies and damping times are fixed to the values predicted by GR, whereas A_{lmn} and ϕ_{lmn} refers to the same starting time $t = 0$. On the other hand, the ringdown modes h_{lmn} are extracted at a reference time $t^{\rm p}_{220}$ corresponding to the peak of the dominant (220) mode. To refer all modes to the same starting time we account for a time shift $\Delta t = t^{\rm p} - t^{\rm p}_{220}$, where $t^{\rm p} \lesssim 10M$ is the global peak time of the signal. As discussed in the main text, this ansatz does not fully match Eq. (2) due to the peak time ambiguity, which adds some small uncertainty on the phase values $\delta\phi_{lmn}$ for the $l = m$ modes.

1. Error estimate for NR waveforms

NR waveforms contain two sources of uncertainties that are of interest to our study; i) resolution uncertainties which are produced by the finiteness of the numerical grid; and ii) extrapolation errors which are produced from computing the data at a finite radii and extrapolating them to future null infinity. To quantify resolution errors, for each simulation we compute the mismatch (as defined, e.g., in Eq. (1) of [41]) between waveforms at the two highest resolutions. Then, to estimate the extrapolation errors, we use the highest-resolution waveform and compute the mismatch between the waveform extrapolated with either second or third polynomial order. We finally compute the distribution of the mismatch across our dataset. In Table I we give the order of magnitude of the mismatch at the median value for the distribution for each mode. Extrapolation errors are negligible relative to resolution errors for each mode, and the maximum mismatch is at most of $\mathcal{O}(10^{-3})$. A more detailed study on NR systematics in the context of ringdown amplitude and phase fits are presented in a companion paper [59]. In Appendix A 4 we will also compare the results of our fits obtained using different catalogs of NR BBH waveforms.

2. Fits for A^R_{lmn}

For a quasi-circular nonprecessing binary, the mode amplitude ratios and phase differences generically depend on all intrinsic binary parameters such as mass ratio q and spin amplitudes $\chi_{1,2}$. However, as in the case of an

Mode	Error	Mismatch
h_{22}	Resolution	$\mathcal{O}(10^{-3})$
	Extrapolation	$\mathcal{O}(10^{-6})$
h_{33}	Resolution	$\mathcal{O}(10^{-4})$
	Extrapolation	$\mathcal{O}(10^{-6})$
h_{21}	Resolution	$\mathcal{O}(10^{-3})$
	Extrapolation	$\mathcal{O}(10^{-6})$
h_{44}	Resolution	$\mathcal{O}(10^{-3})$
	Extrapolation	$\mathcal{O}(10^{-6})$

TABLE I. Resolution and extrapolation error estimates for NR waveforms in the SXS catalog used for our fits.

effective PN waveform modeling [37], it turns out that also A^R_{lmn} depends most strongly on certain combinations of the spins. We therefore obtain analytical (approximate) relations for A^R_{lmn} as a function of the BBH parameters using the ansatz [37]

$$A^R_{lm0} = a_0\,\delta + a_1\,\delta^2 + a_2\,\chi \quad \forall \text{ odd modes} \tag{A2}$$

$$A^R_{lm0} = a_0\,(1-3\eta) + a_1\,(1-3\eta)^2 + a_2\,(1-3\eta)^3 + a_3 + \chi_s\,, \ \forall \text{even modes} \tag{A3}$$

where the $\{(3,3,0),(2,1,0)\} \in$ odd modes and $(4,4,0) \in$ even modes, $\eta = q/(1+q)^2$, $\delta = \sqrt{1-4\eta}$,

$$\chi = \frac{\chi_a + \chi_s\sqrt{1-4\eta}}{2} \tag{A4}$$

is a particular combination of the spin parameters and, for (anti)aligned spins, $\chi_{s,a} = (m_1\chi_1 \pm m_2\chi_2)/(m_1+m_2)$ with $m_{1,2}$ being the progenitor BH masses. This ansatz automatically enforces $A^R_{lmn}(q) \to 0$ for $q \to 1$ for all odd modes in the $\chi_1 = \chi_2$ limit, which arises from the binary's symmetry under $m_1 \leftrightarrow m_2$. Note that the above ansatz differs from the ones we have used for the nonspinning fits presented in Ref. [40].

We fit the data in two hierarchical steps following [14]: we first fit the nonspinning waveforms using Eqs. (A2) with $\chi_{1,2} = 0$. Then, we fit for the spinning BBH waveforms, keeping the values of the coefficients obtained from the nonspinning fit to constrain the final result in the nonspinning limit. This improves the accuracy of the fit in the region of the parameter space where the NR solutions are known to be more accurate [14, 60]. We get the following analytical ready-to-use fits:

$$A^R_{330} = 0.572\sqrt{1-4\eta} - 0.144(1-4\eta) + 0.035\chi\,, \tag{A5}$$

$$A^R_{210} = \left|0.328\sqrt{1-4\eta} + 0.115(1-4\eta) - 0.414\chi\right|\,, \tag{A6}$$

$$A^R_{440} = 0.251\left(1 + 59.773\eta^3 + -16.307\eta^2 - 3\eta\right) - 0.011\chi_s\,. \tag{A7}$$

We set the amplitudes to be positive, shifting by a factor π the phase, $\delta\phi_{lmn} \to \delta\phi_{lmn} + \pi$, for those cases in which the fit provides a negative amplitude. For the (210) mode, we observe that A^R_{210} tends to negative values at low mass ratio $q \lesssim 2$ and high spin $\chi \sim 0.4$. In this case we added the absolute value to the ansatz in order to keep the $A^R_{lmn} > 0$ convention . Note that the fits recover the test particle limit [40, 61] as $q \to \infty$, and the nonspinning regime as $\chi_{1,2} \to 0$. Indeed, in the nonspinning limit we verified that our fit agrees reasonably well with the fits in [39, 40, 62].

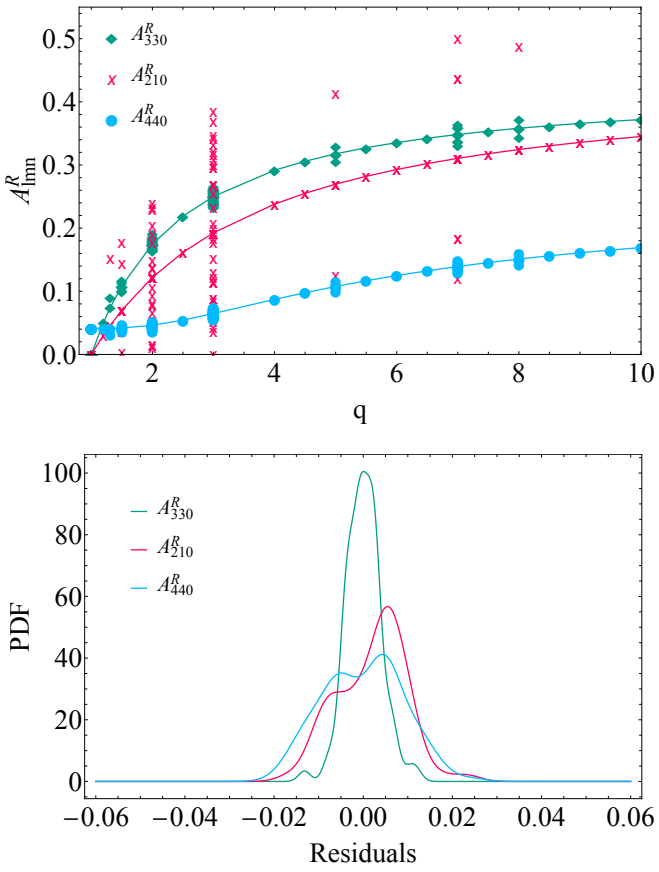

FIG. 3. Top panel: amplitude ratio A^R_{lmn} in terms of the mass ratio q obtained from fitting 142 waveforms for the (330), (210), and (440) modes. The spread of the points on the vertical direction quantifies the effects of the binary spin parameter, $\chi_{\rm pheno} = \chi, \chi_s$ for the odd/even modes, respectively. The solid line shows the fit for the nonspinning case. Bottom panel: Normalised residual distributions for the three modes. Notice that effects are relatively small for the (330) and (210) compared to their typical amplitudes. Further analysis about the errors is provided in a followup paper [59].

In the top panel of Fig. 3, we present the amplitude ratio A^R_{lmn} as a function of q for all 142 simulations. The green diamonds, red crosses, and blue dots correspond to A^R_{330}, A^R_{210}, and A^R_{440}, respectively. The solid lines denotes A^R_{lmn} for the nonspinning BBHs, i.e., $A^R_{lmn} = A^R_{lmn}(q, \chi_{1,2} = 0)$. For all modes considere here, A^R_{lmn} increases with the mass ratio, i.e. for more asymmetric binaries. Spins effects are small for A^R_{330} and A^R_{440}, leading to a small scatter around the solid lines. This also

suggests that spin effects are generically small for these modes, even when accounting for spin misalignment.

In order to show the spin and mass-ratio dependence more clearly, in the top panel of Fig 4 we present a contour plot of A^R_{lmn} as a function of q and χ, χ_s. We observe that, at variance with the other modes, A^R_{210} depends significantly on the binary spins. Finally, we quantify the goodness of the fit using residuals. In the bottom panel of Fig. 3, we present the normalized fit residual distributions, where residuals are the difference between the amplitude/phase obtained by the fit and that obtained from the raw NR data,i.e., Residuals $= (A^R_{lmn,\rm NR-data} - A^R_{lmn,\rm fit})$ or $(\delta\phi_{lmn,\rm NR-data} - \delta\phi_{lmn,\rm fit})$ for each simulation used. We see that the residues are centered around zero with a small spread. Compared to their absolute value, these errors are small for the (330) and (210) mode and modest for the (440) mode.

3. Fits for $\delta\phi_{lmn}$

a. *Disentangling NR conventions from physical phase*

The NR waveforms from different catalogs adopt different conventions for phases, so one must appropriately account for this to combine/compare phases across several NR waveform catalogs. These conventions arise from the choice of – a) the tetrad adopted to extract the NR waveform which adds an overall polarization angle ψ_0; and b) from rotations of the BH orbital plane by an angle φ_0. Two waveforms (say A and B) from different catalogs with the same physical intrinsic parameters and aligned in time are related by [32, 53],

$$h^A_{lmn}(t) = e^{\iota(\psi_0 + m\varphi_0)} h^B_{lmn}(t). \tag{A8}$$

The polarization angle across the NR codes is either $\psi_0 = 0$ or $\psi_0 = \pi$ to preserve the rotating-counterrotating mode symmetry, $h_{lm} = (-1)^l h^*_{l-m}$, for circularly-polarized, nonprecessing waveforms[1] [53]. We are interested only in the physical contribution to the phase, $\delta_{lmn}(\vec{\lambda})$, which depends only on the binary intrinsic parameters $\vec{\lambda}$. A generic ringdown phase ϕ_{lmn} results from the sum of the three contributions [32, 62]

$$\phi_{lmn} = \delta_{lmn}(\vec{\lambda}) + m\,\varphi_0 + \{0, \pi\}\ . \tag{A9}$$

Unlike the physical phase $\delta_{lmn}(\vec{\lambda})$ that depends on the BBH parameters, the extrinsic phase terms ψ_0 and φ_0 may vary across sets of NR simulations and codes. However, the dependency on φ_0 is eliminated out if we fit for the following quantity

$$\delta\phi_{lmn} := \frac{m}{2}\phi_{22n} - \phi_{lmn} = \frac{m}{2}\delta_{220}(\vec{\lambda}) - \delta_{lmn}(\vec{\lambda}) + \left\{0, \left(\frac{m}{2} - 1\right)\pi\right\}\ . \tag{A10}$$

The phase difference $\delta\phi_{lmn}$ depends only on the intrinsic binary parameters $\vec{\lambda}$ and on a global phase factor which is either zero or $(\frac{m}{2} - 1)\pi$ depending on the simulation. For instance, we can identify the convention used in a NR waveform by knowing that, in the low-frequency inspiral regime, the phase difference between the dominant (220) mode and a higher mode (lmn) satisfies $\mathrm{mod}(m\,\phi_{220} - 2\phi_{lmn}, 2\,\pi) = 0, 2\pi$, for both the even and the odd modes (see Appendix D of [53] and [42]). For the SXS data, we have checked that this value is consistent with $\psi_0 = 0$. On the other hand, the RIT and Maya [50, 51] waveform catalogs adopt the $\psi_0 = \pi$ convention [35].

b. *Phase fits*

Similar to the case of the amplitude ratio, we produce ready-to-use fits for $\delta\phi_{lmn} = \frac{m}{2}\phi_{22n} - \phi_{lmn}$ as a function of the BBH parameters. We use the following ansatz informed by the leading order PN expressions on $\delta\phi_{lmn}$ [36]

$$\delta\phi_{330} = b_0\,\delta + b_1\,\chi_\phi + c_0\,, \tag{A11}$$
$$\delta\phi_{210} = b_0\,\delta + b_1\,\chi_\phi + b_2\,\chi^2_\phi + c_0\,, \tag{A12}$$
$$\delta\phi_{440} = a_0\,\eta^{d_0} + b_1\,\chi_s\,, \tag{A13}$$

where $\chi_\phi \equiv \frac{1}{2}(\chi_a + \frac{\sqrt{1-4\eta}}{1-2\eta}\chi_s)$ is another phenomenological fit parameter.

We follow a similar hierarchical fitting procedure as the one previously described for the amplitude-ratio fits. The fits obtained for the $\delta\phi_{lmn}$ read

$$\delta\phi_{330} = 2.759 + 0.406\sqrt{1-4\eta} - 0.055\chi_\phi \tag{A14}$$
$$\delta\phi_{210} = 0.401 + 0.286\sqrt{1-4\eta} + 0.402\chi_\phi + 0.652\chi^2_\phi$$
$$\delta\phi_{440} = 4245.459\eta^{5.646} - 0.365\chi_s\,.$$

In Fig. 5 we present $\delta\phi_{lmn}$ (as defined in Eq. (A10)) obtained at $t_r - t^{\rm p}_{22} = 10M$. Again the solid lines correspond to the nonspinning BBH case, while the spread of points around the lines quantify the spin dependence of the result. As q increases, we observe that $\delta\phi_{330}$ and $\delta\phi_{440}$ mildly increase while $\delta\phi_{210}$ rapidly vanishes. We stress that for the SXS catalog the convention is such that $\mathrm{mod}(\delta\phi_{330}, \pi) = \pi$ in the low-frequency PN regime. For $q \approx 1$, we see that $\delta\phi_{lmn}$ develops a dependence on the BBH spin parameters, which is reduced as q increases. This is consistent with [36], where $\mathrm{mod}(\delta\phi_{330}, \pi)$ is evaluated at $t_r = t^{\rm p}_{22}$.

We again evaluate the fit residuals for each NR waveform; the distribution of the residuals is shown in the

[1] The rotating-counterrotating mode symmetry implies $h_{lm} = (-1)^l h^*_{l,-m}$. We can define a new waveform h^p_{lm} up to a polarization angle ψ_0 as $h^p_{lm} = e^{i\psi_0}(-1)^l h_{lm}$ and $h^p_{l-m} = e^{i\psi_0}(-1)^l h_{l-m}$. Then, $h^p_{l-m} = e^{2i\psi_0}(-1)^l h^{p,*}_{lm}$ and $h^p_{l-m} = (-1)^l h^{p,*}_{lm}$ if and only if $\psi_0 = 0, \pi$.

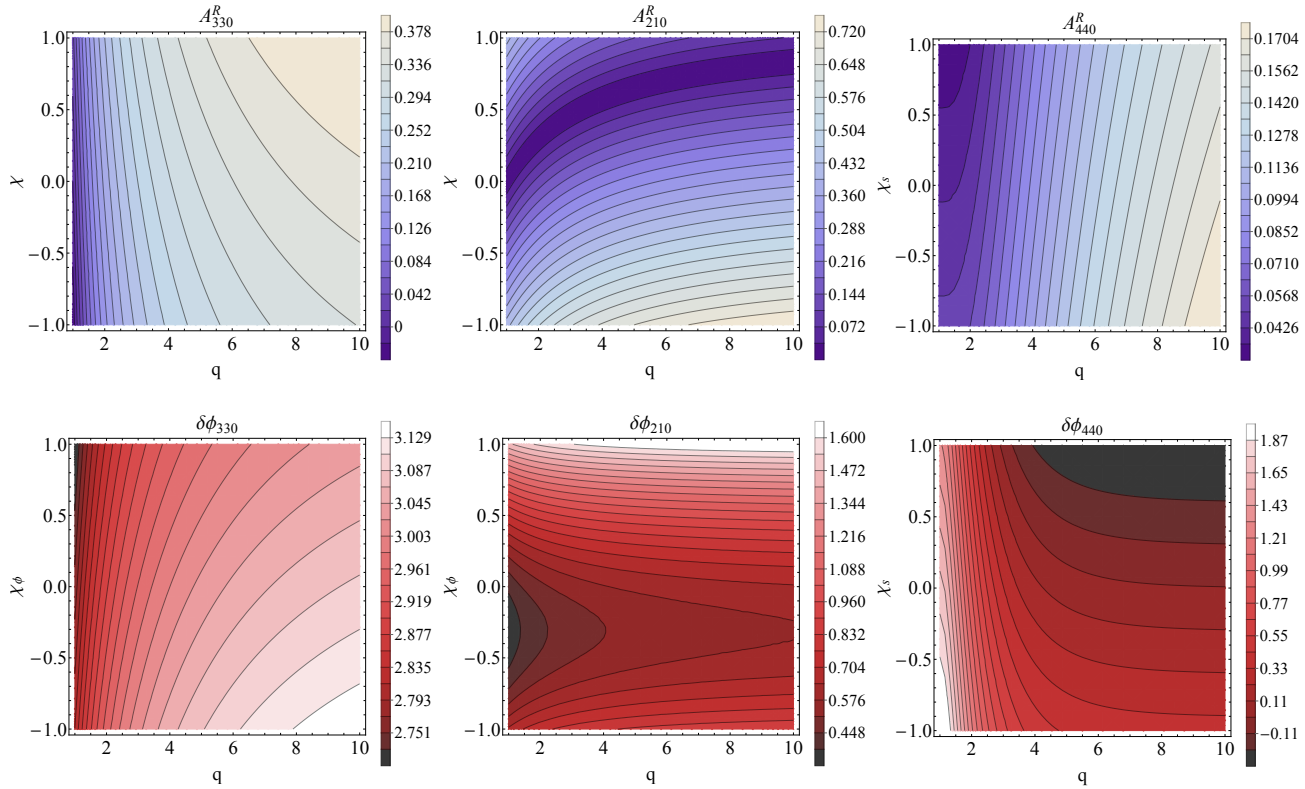

FIG. 4. Contour plots showing the amplitude ratio A^R_{lmn} (top) and phase difference $\delta\phi_{lmn}$ (bottom) as functions of q and $\chi_{\rm pheno} = \chi, \chi_s, \chi_\phi$ (depending on the panels) for the $(lmn) = (330)$ (left), (210) (middle), and (440) (right) mode. As $\chi_{\rm pheno}$ increases, the magnitude of A^R_{330} increases as well, whereas A^R_{210} decreases and A^R_{440} mildly decreases. The small region around $q \sim 1-2$ and $\chi > 0$ in which A^R_{210} increases with χ is forced by fitting $|A^R_{210}|$, which induces a sign flip on the trend of $A^R_{210}(\chi)$. In the bottom panels, $\delta\phi_{lmn}$ decreases linearly with χ_ϕ for the (330) mode while it increases as χ^2_ϕ for the (210) mode. The value of the phase difference $\delta\phi_{lmn}$ decreases with $\chi_{\rm pheno}$ for the (330) and (440) modes while it increases for the (210) mode.

bottom panel of Fig. 5. Similar to Fig. 3, the largest errors are obtained for $\delta\phi_{440}$. Finally, in the bottom panel of Fig 4 we show $\delta\phi_{lmn}$ as a function of q and χ, χ_s, χ_ϕ for the (330) (left), (210) (middle) and (440) (right) modes. As in the case of the amplitude ratio, the phase $\delta\phi_{330}$ is only mildly affected by the spin, whereas a stronger spin dependence occurs for the 221 and (440) modes.

4. Comparison of the fits using other NR catalogs

The RIT and Maya public catalogs [50, 51] provide a large set of NR simulations that can be used for fitting and testing. In particular, the public data provided for the RIT catalog is tested to be in the convergent regime – resolution errors shall dominate – and shows good global IMR agreement with the SXS data for all the modes up to $l = 5$. We have calibrated the amplitudes and phases using data from the SXS catalog since the latter is the only code that provides data at different resolutions and extrapolation levels, which we used for the error estimates previously presented. On the other hand, we can use NR data from the RIT and Maya catalogs to benchmark our results. This test is particularly useful since the two families of codes use significantly different numerical schemes to solve Einstein's equations for a BBH [2]. In particular, both A^R_{lmn} and $\delta\phi_{lmn}$ computed from different codes may be affected dominantly by the finite extraction and extrapolation effects which may add differences up to a few percent [60]. We observe that the systematic errors (i.e., a shift of the median value of the distribution with respect to zero), are below the fitting errors which are instead characterized by the width of the distributions. This observation holds true for all the modes considered here.

To compare the data from the different catalogs, we have to revisit again the various conventions used in each of them, for example the $\psi_0 = 0, \pi$ rotational factors arising from different tetrad choices in the simulations. For the RIT and Maya data we need to replace $\delta\phi_{lmn} \to -\delta\phi_{lmn} - \frac{m}{2}\pi$, where the latter factor results from the

[2] For instance, while the SXS waveforms are solved in the generalized harmonic gauge, the RIT and Maya are solved using the BSSNOK formulation.

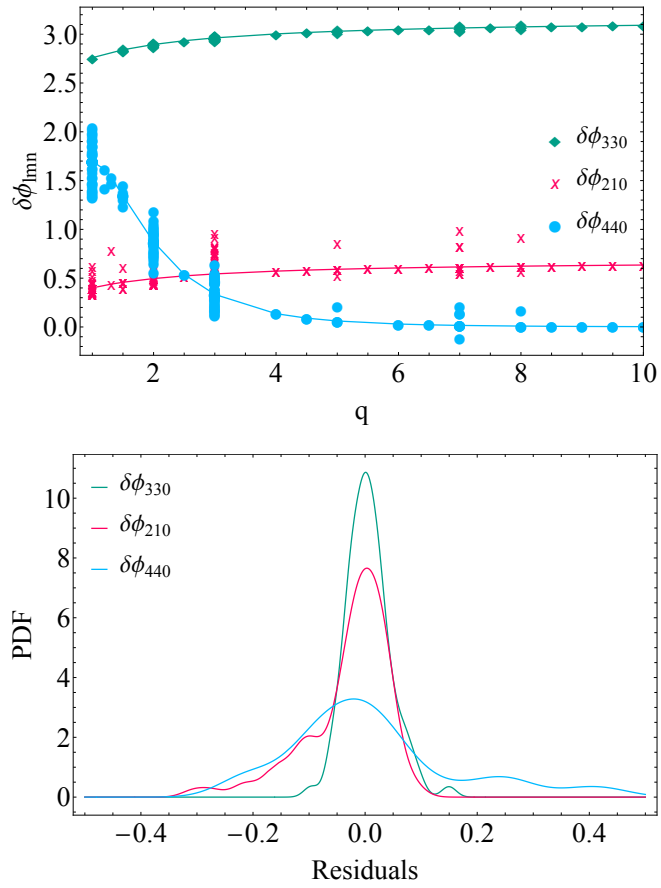

FIG. 5. Same as Fig. 3 but for the phase difference $\delta\phi_{lmn}$.

different tetrad conventions used in these codes. The additional minus sign comes from the reversed definition of the imaginary component of the h_{lm} modes between the SXS dataset and the RIT and Maya datasets, which just implies that $\phi^{\rm SXS}_{lmn} \leftrightarrow -\phi^{\rm RIT,Maya}_{lmn}$[35].

In Fig. 6 we compare the SXS amplitude ratio and phase difference of the (330) mode with the RIT and Maya data. For the amplitude ratio (left panels of Fig. 6), we find a good match between the Maya and SXS data while the results show an offset of about $A^R_{330} \sim 0.05$ when using the waveforms from the RIT catalog. This systematic offset is still smaller than the value of the uncertainty observed in GW190521, $\delta A^R_{330} \sim 0.1$, but it may become important for louder future events. In the case of the phase difference (right panels of Fig. 6), despite we find a better agreement between all the three codes, the RIT code shows slightly larger tails. The standard deviation obtained from the normalized residuals distributions is about ~ 0.1 rad for the three codes, and it is therefore irrelevant for current or near-future GW observations.

5. Identifying fit outliers in the data

We have observed that the BH spins have a minor impact on the values of A^R_{lmn} and $\delta\phi_{lmn}$ compared to the effect of the mass ratio q. This allow us to use an ansatz linear in $\chi_{\rm pheno}$ for all the modes and magnitudes except for $\delta\phi_{210}$, in which the effects of $\chi_{\rm pheno}$ are found to be larger. While this property makes it easier to model A^R_{lmn} and $\delta\phi_{lmn}$ in terms of the physical parameters, we have found that some outliers may still decrease the quality of the spin-dependent part. We believe that the outliers are dominated by numerical noise and we remove them from our calibration dataset as explained in the following. We first look for highly significant outliers on both A^R_{lmn} and $\delta\phi_{lmn}$ by testing the spin fit $\chi_{\rm pheno}$ at each mass ratio q. More specifically, we select the data at mass ratio $q = 1, 2, 3, 6, 7, 8$ to perform a bootstrapping analysis on the $\chi_{\rm pheno}$ axis. This is achieved by computing the spinning fit at each q, that in general will contain N points, for a data set of $N-1$ points. We iterate the fit for the N points at each q, and we compute the value of the standard deviation $\sigma(q^i, \chi^i_{\rm pheno})$. Then, we select the median value of $\bar{\sigma}$ and we discard all points beyond a conservative deviation of $4\bar{\sigma}$. In Fig. 7 we show an example of this procedure applied to the amplitude ratio A^R_{330}. We see that the blue cross placed at the low-right corner is easily identified by this algorithm. However, it is important to keep a conservative criterion to avoid rejecting systematically false outliers.

6. On the Δt dependence for the phase difference

As discussed in the main text, in order to refer the NR fits to the parameter estimation to the same initial time we need to shift each mode by $\Delta t = t^{\rm p} - t^{\rm p}_{220}$, where $t^{\rm p}$ is the global peak time of the signal and $t^{\rm p}_{220}$ is an arbitrary reference time taken to be the peak time of $|h_{220}|$. This time shift introduces a dephasing $\Delta\phi_{lmn} = \omega_{lmn}\Delta t$ and an extra term in the phase difference $\delta\phi_{lmn}(t^{\rm p})$ (see Eq. (4))

$$\left(\frac{m}{2}\omega_{220} - \omega_{lmn}\right)\Delta t\,. \tag{A15}$$

Fortunately, as can be shown analytically in the geodesics approximation valid in the eikonal limit $l = m \gg 1$ [12, 43], for the fundamental ($n = 0$) modes the following approximation holds: $\omega_{lm0} \sim l/2\,\omega_{220}$. Thus, for the $l = m$ modes the extra terms in Eq. (A15) is small. In Fig. 8, we show $\omega_{lmn} - l/2\,\omega_{220}$ for the (330) and (440) modes as a function of the remnant spin, showing that the difference is in the range $\approx [0.04, 0.08]$ for any spin. This yields an ambiguity in $\delta\phi_{lmn}$ approximately of $(\omega_{330} \sim 3/2\,\omega_{220})\Delta t \sim 0.04\Delta t/M$ and $(\omega_{440} \sim 4/2\,\omega_{220})\Delta t \sim 0.07\Delta t/M$. Choosing a very conservative error estimate on Δt, namely $\Delta t = 2(t^{\rm p}_{33} - t^{\rm p}_{22}) = 10M$, we conclude that ignoring the dephasing introduced by Δt will at most introduce an overall uncertainty $\delta\phi_{330} \sim 0.4$ rad

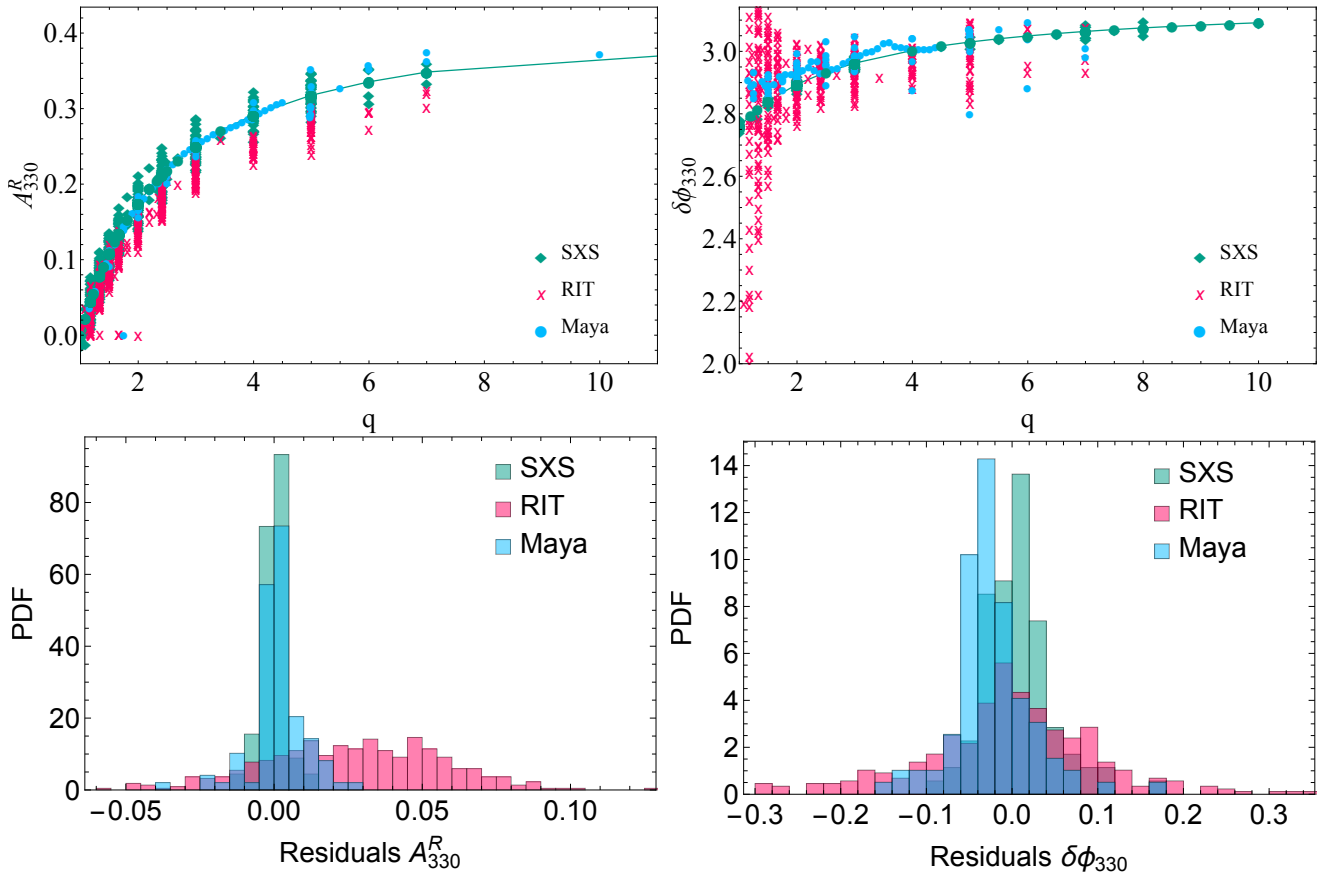

FIG. 6. Top panel: amplitude ratio (left) and phase difference (right) in terms of the mass ratio q and the phenomenological spin parameter $\chi_{\rm pheno}$ obtained from fitting 142 waveforms for the (330) modes and from data from the SXS, RIT, and Maya catalogs. The solid line joining the points represents nonspinning solution. Bottom panel: Normalized residual distributions obtained for the data from the three datasets for the amplitude ratio (left) and phase difference (right).

and $\delta\phi_{330} \sim 0.8\,$rad. Note that, since $(t^{\rm p}_{33} - t^{\rm p}_{22}) \sim 5M$ for all the NR simulations explored [40, 42], in practice we expect the uncertainty on $\delta\phi_{330}$ to be typically half of this conservative estimate.

7. The effect of the eccentricity

The coalescence of eccentric binaries can in principle modify the initial perturbation conditions of the final BH, and hence the QNM amplitudes and phases. In general, the amplitude ratio A^R_{lmn} and the phase difference $\delta\phi_{lmn}$ should also depend on the eccentricity e. In Fig. 9 we examine the impact of the eccentricity on the values of A^R_{330} and $\delta\phi_{330}$ for a set of mass ratios $q = 2, 3, 4$, i.e., where both the standard BH spectroscopy test and our APC test are most promising [1, 63]. To obtain these estimates, we have used the data from the RIT catalog [50]. Notice that the values on $\delta\phi_{330}$ are significantly modified only at relatively large values of the eccentricity with $e \gtrsim 0.3$, while this value raises up to $e \gtrsim 0.6$ for A^R_{330}. These values are still above the upper limit threshold of $e \sim 0.1$ obtained from the search of eccentric objects during the first and second LIGO observation runs [64], thus not relevant for the bulk of the events observed by current ground based GW observatories.

Appendix B: Consistency between mode-excitation and BBH mass ratio in GW190521

In this work we have proposed a new APC test of GR that can be done for GW190521-like systems using the ringdown alone (and possibly prior knowledge of the binary extrinsic parameters [33]). In this section, we highlight another possible null test of GR that makes use of the amplitude ratio and its relation to the BBH mass ratio, and therefore requires the entire inspiral-merger-ringdown (IMR) signal. The basic concept here is that one could estimate the mass ratio q from the ringdown by inverting the $A^R_{lmn} = A^R_{lmn}(q)$ relation and then check whether the inferred value is consistent with q measured

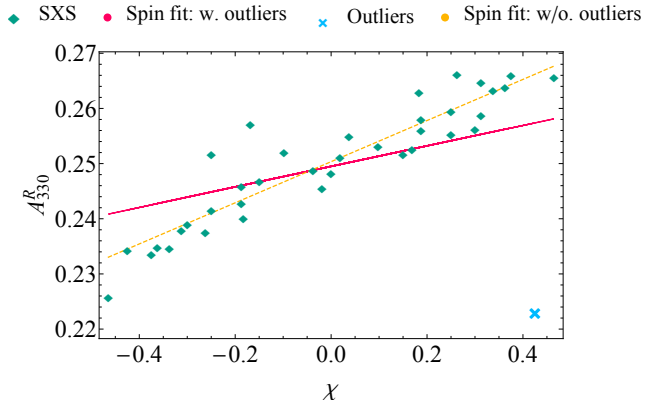

FIG. 7. Amplitude ratio fits A^R_{330} for $q = 3$ in terms of the effective spin χ. SXS data are denoted by green diamonds while we use red dots to denote the results obtained from fitting the data with the PN motivated ansatz (A2), which is linear in χ. One can easily identify an outlier point at $A^R_{330} = 0.223$ and $\chi = 0.850$ (blue cross). The solid orange curve provides the results of the linear fit after removing the outlier point. In particular, this corresponds to the SXS simulation SXS:0293.

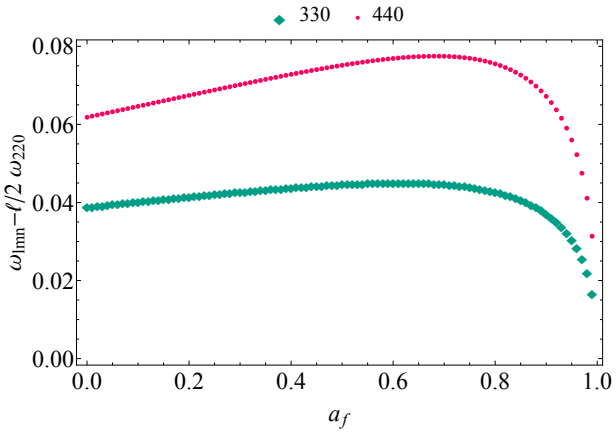

FIG. 8. Difference $\omega_{lmn} - l/2\,\omega_{220}$ as a function of the remnant spin. The $\omega_{lmn} \sim l/2\,\omega_{220}$ approximation is used in the calibration of $\delta\phi_{lmn}$ for the (330) and (440) modes. For the (330) mode and a conservative choice $\Delta t \sim 10M$ this adds an uncertainty of about 0.4 rad. On the other hand, such approximation worsens for the (440) mode.

independently from the full IMR signal[3]. This is complementary to the standard IMR consistency tests performed by the LVK Collaboration [27] and is based on an idea similar to that used to design the merger-ringdown test for the BBH population presented in Ref. [58].

Our fits provide empirically 2-to-2 maps $(A^R_{lmn}, \delta\phi_{lmn}) \to (q, \chi_{\rm pheno})$ which can be inverted to obtain q. Note that the sensitivity of this test toward the measurement of the spin is limited since A^R_{330} and A^R_{440} have a rather mild dependence on the spins (see Fig. 3). However, even neglecting the spin dependence, one can simply map A^R to q to a good approximation. For instance, for $q = 2$ we get $A^R_{330} = 0.14, 0.19$ for $\chi_{1,2} = \pm 0.85$, so the spin dependence is minor. In principle this kind of test can also be designed with $\delta\phi_{lmn}$. However, the dependence of $\delta\phi_{lmn}$ on the BBH parameters is weak, and so one expects larger errors on the inferred BBH parameters.

We scrutinize GW190521 for consistency between mode-excitation and BBH mass ratio. Owing to its large total mass and short inspiral signal in the LIGO-Virgo band, the inspiral parameter estimation of GW190521 is particularly sensitive to model systematics and there is some tension among the binary parameters (including the mass ratio) inferred with different waveforms [44, 56, 57]. Thus, one cannot perform a reliable IMR-like test on this signal. In Fig. 10 we compare the different estimates for q using parameter estimation posteriors provided in these IMR studies with the expected value of q inferred from ringdown. For the ringdown estimate of q, we compare the estimate of A^R_{330} obtained in [1] along with our fit results. Specifically, we show

(i) the parameter-estimation results on A^R_{lmn} from [1] (red) and translated to q by inverting our A^R_{330} mode fit. The red dot (Ours) provides its best likelihood value obtained from the marginalized distribution $q - A^R_{330}$.

(ii) the parameter-estimation results on q and $\chi_{1,2}$ obtained by the LVK collaboration [57] (blue) and translated to A^R_{lmn} by using our (330) mode fit. The blue-square (LVK) provides its best likelihood value obtained from the marginalised distribution $q - A^R_{330}$. Since the LVK does not provide an independent distribution on A^R_{330}, we get the elongated blue contours using our fit.

We observe that Ours, Capano+, and LVK are all consistent with the parameter-estimation posterior distribution at the 1σ and 2σ credible level. Note that the q distribution from Capano+ is obtained from an independent set of fits [37]. We also obtained a rather flat distribution on the phenomenological spin parameter with $\chi \in [0, 1]$, which is expected given the mild dependence of both $A^R_{330}(q, \chi_{\rm pheno})$ and $\delta\phi_{330}(q, \chi_{\rm pheno})$ on χ. However, this paradigm may change shortly with louder detections such as those expected from third-generation detectors [54, 55] and LISA [58].

[3] While in principle an IMR consistency test can be done by directly checking the consistency of the fundamental mode amplitude A_{220} as a function of the binary parameters, this quantity depends on several (both intrinsic and extrinsic) parameters so its constraining power is limited, e.g., by correlations. On the other hand the amplitude ratio A^R_{lmn} depends mainly on the binary mass ratio and spins.

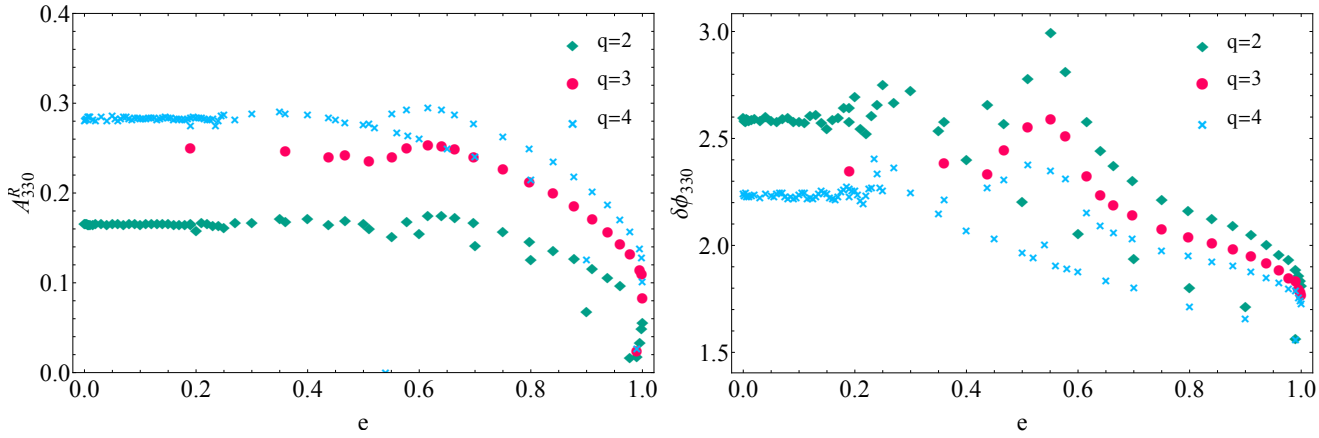

FIG. 9. The impact of the eccentricity e on the estimates of the amplitude ratio A^R_{330} (left) and phase difference $\delta\phi_{330}$ (right) for a set of mass ratios $q = 2, 3, 4$ and $\chi_{\rm pheno} = 0$ using the RIT catalog [50]. Notice that the effects of the eccentricity become relevant only when $e \gtrsim 0.3$ for the $\delta\phi_{330}$ and when $e \gtrsim 0.6$ for A^R_{330}.

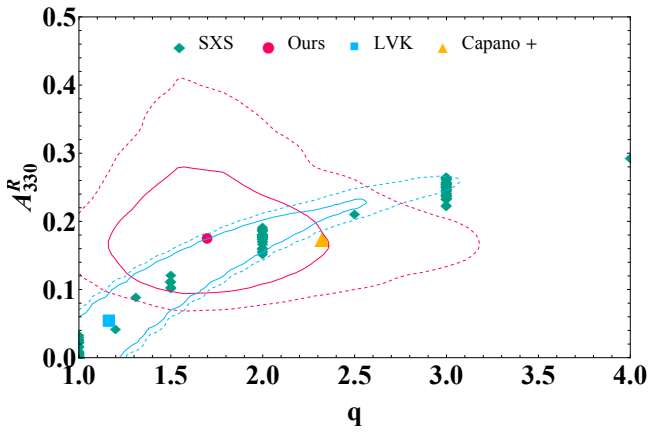

FIG. 10. Amplitude ratio A^R_{330} estimated from the set of NR waveforms used in this work and restricted to $q < 3$. The mild dependence on the spin parameters is reflected by the relatively small spread of the points at fixed q. We add on top the different estimates of the $A^R_{330} - q$ relation obtained from different mass ratio estimates: Capano+ (orange triangle), Ours (red dot) and LVK (blue square). The values of A^R_{330} and q predicted from NR are consistent at the 2σ level with the distributions obtained on q and A^R_{lmn} obtained from [57] and [1].