

## Toward a self-consistent framework for measuring black hole ringdowns

Teagan A. Clarke<sup>1,2,\*</sup> Maximiliano Isi<sup>3</sup> Paul D. Lasky<sup>1,2</sup> Eric Thrane<sup>1,2</sup> Michael Boyle<sup>4</sup> Nils Deppe<sup>4,5</sup>  
Lawrence E. Kidder<sup>4</sup> Keefe Mitman<sup>6</sup> Jordan Moxon<sup>6</sup> Kyle C. Nelli<sup>6</sup> William Throwe<sup>4</sup> and Nils L. Vu<sup>6</sup>

<sup>1</sup>*School of Physics and Astronomy, Monash University, Clayton, Victoria 3800, Australia*

<sup>2</sup>*OzGrav: The ARC Centre of Excellence for Gravitational Wave Discovery,  
Clayton, Victoria 3800, Australia*

<sup>3</sup>*Center for Computational Astrophysics, Flatiron Institute, New York, New York 10010, USA*

<sup>4</sup>*Cornell Center for Astrophysics and Planetary Science, Cornell University, Ithaca, New York 14853, USA*

<sup>5</sup>*Department of Physics, Cornell University, Ithaca, New York 14853, USA*

<sup>6</sup>*Theoretical Astrophysics, Walter Burke Institute for Theoretical Physics,  
California Institute of Technology, Pasadena, California 91125, USA*



(Received 5 February 2024; accepted 7 May 2024; published 11 June 2024)

The ringdown portion of a binary black hole merger consists of a sum of modes, each containing an infinite number of tones that are exponentially damped sinusoids. In principle, these can be measured as gravitational-waves with observatories like LIGO/Virgo/KAGRA, however in practice it is unclear how many tones can be meaningfully resolved. We investigate the consistency and resolvability of the overtones of the quadrupolar  $\ell = m = 2$  mode by starting at late times when the gravitational waveform is expected to be well approximated by the  $\ell mn = 220$  tone alone. We present a Bayesian inference framework to measure the tones in numerical relativity data. We measure tones at different start times, checking for consistency: we classify a tone as stably recovered if and only if the 95% credible intervals for amplitude and phase at time  $t$  overlap with the credible intervals at all subsequent times. We test a set of tones including the first four overtones of the fundamental mode and the 320 tone and find that the 220 and 221 tones can be measured consistently with the inclusion of additional overtones. The 222 tone measurements can be stabilized when we include the 223 tone, but only in a narrow time window, after which it is too weak to measure. The 223 tone recovery appears to be unstable, and does not become stable with the introduction of the 224 tone. We find that  $N = 3$  tones can be stably recovered simultaneously. However, when analyzing  $N \geq 4$  tones, the amplitude of one tone is consistent with zero. Thus, within our framework, one can identify only  $N = 3$  tones with nonzero amplitude that are simultaneously stable.

DOI: [10.1103/PhysRevD.109.124030](https://doi.org/10.1103/PhysRevD.109.124030)

### I. INTRODUCTION

The final stage of a binary black hole coalescence, called the ringdown, consists of a perturbed remnant black hole emitting gravitational waves. In general relativity, the gravitational waves from the ringdown can be decomposed using spin-weighted spheroidal harmonics into quasinormal modes [QNMs, see [1–5]]. Each spheroidal harmonic mode is labeled with indices  $\ell \geq 2$  and  $|m| \leq \ell$ . There are an infinite number of tones associated with each angular mode, each denoted with  $n \geq 0$ . The frequency and damping time of each tone depends only on the mass and spin of the remnant black hole (assuming zero charge) according to the no-hair theorem [e.g., [6]].

Understanding how to measure black hole tones can allow us to undertake tests of general relativity and the

no-hair theorem for black holes [7–12]. However, the start of the ringdown—defined as the time when the signal can be described with black hole perturbation theory—is ambiguous [e.g., [13–16]]. Understanding how early the perturbative prescription can be applied to the signal is key to correctly performing tests of general relativity and the no-hair theorem. Beginning the analysis too early could result in overfitting to nonlinear features in the signal [e.g., [17–19]]. If one waits too long to begin the analysis, the strain amplitude will have decayed exponentially, making spectroscopic tests impossible given the finite sensitivity of our instruments [14,20]. While black hole spectroscopy is part of a broader effort to test general relativity with gravitational waves, along with e.g., inspiral-merger-ringdown tests [e.g., [7,21–23]], the ringdown provides one of the most direct ways to test the no-hair theorem. Binary parameter-informed analyses of the ringdown provide an alternative

\*teagan.clarke@monash.edu

approach to damped sinusoid models for QNM extraction and tests of the no-hair theorem [e.g., [24,25]].

Numerical relativity simulations provide numerical solutions to the field equations [26–28], and are the state of the art for investigating the QNM decomposition of the ringdown. However, the components contained in these solutions remain unclear. Black hole perturbation theory provides predictions for the frequency and damping times of each tone, but the optimal time to fit them and how many tones can be meaningfully extracted remains unclear due to, e.g., nonlinearities, nonorthogonal QNM decomposition, and error in the simulations themselves. Much effort has been devoted to determining how many tones measured at what time provide the optimal fit of physically motivated tones to the numerical simulations.

Many studies [e.g., [29–34]] find that several overtones of the  $\ell m = 22$  mode can be extracted at early times after the merger and that they are needed to infer the correct mass and spin of the black hole. Reference [35] suggests that the linear QNM model may be valid as early as the strain peak time. They show that higher overtones of the 22 mode up to  $n = 7$  improve fits to simulated gravitational-waveforms from the strain peak time. Reference [36] fit away these seven tones from the NR data using frequency domain filters, uncovering evidence for second-order effects and spherical-spheroidal mode mixing.

References [37,38] find that seven tones of the  $\ell = m = 2$  as well as many tones from additional QNMs can be included to improve the fits. Tones up to  $n = 9$  have also been explored for their potential to further improve fits at the strain peak time [39]. The apparent linearity of the signal at such early times could be explained if the nonlinear effects are hidden behind the apparent horizon of the black hole [e.g., [40,41]], unable to reach the observer at infinity. On the other hand, nonlinear effects like second order QNMs, may be required to accurately model the ringdown of higher harmonics [e.g., [42,43]], although the magnitude of such second-order contributions are by far subdominant for existing detections.

While it seems clear that the early ringdown can be modeled with a large number of overtones, there is debate as to whether the associated fits are physical and to which extent [e.g., [32,44]], i.e., are we overfitting tones to produce what amounts to a phenomenological model?

References [45,46] construct analytic fits for stable tones as a function of binary parameters. They point out that in the purely perturbative QNM regime, tone amplitudes and phases should be measured consistently at different start times. Reference [46] finds that the amplitudes for tones with  $n > 0$  are not consistent in time and omit these from their analytic fits. However, [47] finds that analytic fits for tones are biased when fitting at early times, possibly

indicating the presence of higher tones and nonlinearities near the strain peak.

Reference [19] further explores whether or not the ringdown tones are consistent when measured at different start times. They argue: if the overtone parameters do not yield consistent fits at different start times, they are not physical, but rather show evidence of overfitting. Reference [48] finds that overtones beyond 223 are always inconsistent by this metric when fitting in the time domain, while [49] uses analytic modeling to caution that increasing the number of tones may end up overfitting to a misspecified model. Reference [19] similarly finds that adding tones improves the fits, even to unphysical hybrid waveforms, which they point out as a sign of overfitting. Reference [50] attempts to resolve this issue by measuring the stability of the tones as they are being fitted, iteratively removing inconsistent tones until only stably recovered tones are left. For the 22 extraction of a numerical-relativity waveform, they find that five tones are stably recovered, with 221 being the highest stably measured overtone. The other robust tones consist of retrograde modes or modes with the same  $m$  but different  $\ell$ , due to other angular QNMs mixing into the 22 spherical harmonic. Reference [51] uses the Bayesian evidence to algorithmically select tones to be included in their model, and to search for nonlinear contributions. They find that nonlinear contributions may be important for ringdown analysis, although the importance decreases as the black hole spin increases.

In parallel to the debate about the physicality of the linear perturbation fits, a number of studies have attempted to measure black hole overtones in gravitational-wave data with mixed results. Assuming a QNM decomposition, Refs. [17,52] search for overtones in the late-time ringdown of GW150914, only finding evidence of the fundamental 220 tone. By studying at earlier times in the ringdown, Refs. [9,16,53,54] find evidence for the 221 tone in GW150914. References [55–57] find only weak support for the 221 tone in GW150914. However, [11,58] do not find evidence for any higher tones beyond the fundamental in GW150914 (although see Refs. [59,60] for further discussion on these results).

In this paper, we seek to arrive at a *self-consistent*, perturbative model of the post-merger ringdown signal. We present a Bayesian-inference, forward-modeling procedure as a method for extracting the ringdown QNMs from numerical-relativity simulations. Using a numerical-relativity waveform, we start at the end of the ringdown when the strain should be dominated by the 220 (and 320) tones and use this to fit the  $n = 0$  amplitude and phase.

We also fit a noise amplitude—a phenomenological tool, which we introduce in order to account for various unmodeled physics—to cover anything that causes the late-time ringdown to deviate from a pure 220 tone, apart from some small mode-mixing contribution from the

320 tone. Having established the asymptotic behavior of the ringdown signal, we work backward, carrying out Bayesian inference to see if there is support for additional tones at earlier times.

We assess whether each tone in a set is stably recovered by checking whether the fits obtained from different times are consistent. In doing so, we aim to determine if a numerical-relativity waveform can be self-consistently modeled using a superposition of tones, and how many tones can be said to be present.

The remainder of this paper is organized as follows. In Sec. II we describe our model and fitting procedure. In Sec. III we present our results before discussing their implications in Sec. IV.

## II. FRAMEWORK

We model the ringdown strain as a sum of damped sinusoids, writing each spin-weighted spherical harmonic mode of the strain as a complex-valued time series:  $h = h_+ - ih_\times$  as

$$h_{\ell m}^{N_{\text{tones}}}(t) = \sum_{n=0}^{N_{\text{tones}}} A_{\ell mn} e^{-i[\omega_{\ell mn}(t-t_0) + \phi_{\ell mn}]}, \quad (1)$$

where  $\omega_{\ell mn}(M_f, \chi_f)$  are the complex frequencies determined by the remnant mass and spin through the no-hair theorem [61,62]. The negative imaginary component of  $\omega_{\ell mn}$  is the inverse tone damping time  $\tau_{\ell mn}$ . The tone amplitudes  $A_{\ell mn}$  and phases,  $\phi_{\ell mn}$  depend on the initial conditions of the perturbations on the black hole. The variable  $t_0$  is the reference time of the fit, which we take as the peak time of the 22 mode strain,  $t = 0M$ . We use geometric units in this study and measure  $t$  in units of the initial binary mass  $M$ , which is set to unity.

We fit the  $\ell = m = 2$  mode alone and vary  $N_{\text{tones}}$  to change the number of tones in the fit. We order the tones in descending order of damping time: 220, 221, ...,  $22N_{\text{tones}}$ . We fit numerical relativity data that has been decomposed into spin-weighted spherical-harmonics  $_{-2}Y_{\ell m}$ . However, the basis for a perturbed black hole is actually described by spin-weighted *spheroidal* harmonics  $_{-2}S_{\ell m}$  [e.g., [2,63]]. Because of this, mode-mixing occurs between QNMs with the same  $m$  but different  $\ell$ . For the fundamental 22 mode, the dominant source of mode-mixing comes from the 320 [e.g., [29]], which has a higher frequency than the 220 and a comparable damping time. To account for this, we also include the 320 in our fits.

The data for our fit is from the simulating extreme space-times (SXS) catalog [64,65]. We use the Cauchy characteristic evolved (CCE) waveform simulation [66–68], which was generated using the `SpECTRE` code’s CCE module [69]. The waveform is mapped to the superrest frame of the remnant black hole with Bondi-van der Burg-Metzner-Sachs

(BMS) frame fixing [38,70,71]. This is the correct frame mapping for QNM extraction and eliminates unphysical frame effects. The frame fixing is performed using the `SCRI` Python module [72–75]. We use the `SXS:0305` waveform, which corresponds to the most likely parameters for the GW150914 LIGO-Virgo observation [76]. The simulation produces a remnant mass  $M_f = 0.952M$  and dimensionless spin  $\chi_f = 0.692$ . We use the highest available resolution for this study, which has a time resolution of  $\approx 0.1M$  around the strain peak time. We do not interpolate to a uniform time grid for this study.

We use the MCMC sampler `EMCEE` [77] to fit a damped sinusoid model according to Eq. (1), fitting the tone amplitude and phase as a function of time. We fix the frequency and damping time of each tone as a function of the known asymptotic remnant mass and spin from the numerical relativity simulation waveform metadata. It has been pointed out [e.g., [29,44,78–80]] that the mass and spin of the black hole may still evolve at early times in the ringdown, potentially impacting QNM fits. We do not consider this effect in this study. We calculate the frequency and damping times using the `QNM` Python package for Kerr black holes [81]. We parametrize the amplitude and phase as  $x = A \cos(\phi)$  and  $y = A \sin(\phi)$  and sample in normal distributions of  $x$  and  $y$  with mean 0 and standard deviation 1.

In order to carry out Bayesian inference, we must introduce some notion of uncertainty. One path forward could be to set the noise at the level of estimated numerical-relativity precision. However, this numerical noise does not capture the systematic uncertainty in our calculation. Our premise is that the very end of the numerical-relativity waveform should be consistent with a pure 220 tone with small contributions from the 320 tone and late-time polynomial tails [e.g., [82,83]], although the latter are yet to be seen in numerical simulations from the SXS catalog. We use this late-time fit as a point of reference for understanding the earlier part of the waveform. If the end of the waveform is *inconsistent* with a set of tones comprising of only the 220 and 320 tones, this additional structure represents a systematic error within our framework.

In order to quantify this systematic uncertainty, we fit the late waveform (starting from  $t = 100M$ ), allowing for a mixture of the 220, 320 and 221 tones. We choose  $100M$  as the earliest time when the 221 is beginning to become unresolved for an analytic waveform test consisting of three tones with the same frequency and damping times of the 220, 320 and 221 in the numerical-relativity simulation and amplitudes and phases consistent with our fits to the numerical-relativity data.

Ideally one would wait longer than  $100M$ . However, the numerical-relativity error starts to increase from  $100M$ , suggesting the numerical relativity error starts to become

more relevant from this point.<sup>1</sup> We model the systematic uncertainty so that the *artificial* noise on the strain is drawn from a Gaussian distribution with width  $\sigma_{\text{sys}}$ . The artificial noise is implemented by dividing the data by the exponential term of the 220 tone,  $e^{-t/\tau_{220}}$ , before performing our fits.  $\tau_{220}$  is the damping time of the 220 tone  $\approx 11.74M$ , calculated using the simulation remnant mass and spin in the QNM package [81].

We vary  $\sigma_{\text{sys}}$  until the posterior for the amplitude of the 221 tone  $A_{221}$  is consistent with zero at one-sigma credibility. This enforces our requirement that the late-time ringdown should not contain any contribution from overtones higher than the 220. We obtain a value of  $\sigma_{\text{sys}} = 0.06$ , which is  $\approx 70$  times higher than the expected numerical-relativity error at  $t = 100M$ . This suggests that our ability to understand the late-time behavior of the ringdown is likely limited not by numerical-relativity noise, but by theoretical uncertainty about the relative contribution of various tones and/or nonlinearities. Additional contributions from mode-mixing and power-law tails may also be included in our measurement of  $\sigma_{\text{sys}}$ .<sup>2</sup> The quantity  $\sigma_{\text{sys}}$  has no physical meaning; it is a purely phenomenological tool that we introduce in order to quantify our present theoretical uncertainty about the behavior of the late-time ringdown. We find that the value of  $\sigma_{\text{sys}}$  measured is not sensitive to the chosen time from approximately  $t = 70M$  with  $\sigma_{\text{sys}}$  changing by only  $\approx 0.01$  between  $70M$  and  $105M$ . The exact value of  $\sigma_{\text{sys}}$  and the time at which it is measured is not important for our investigation into the stability of sets of tones. Our decision to model  $\sigma_{\text{sys}}$  with an artificial uncertainty (rather than an absolute uncertainty) is motivated by experimentation. We find that our fits are relatively consistent in size and magnitude and that our sampling is better behaved when we assume an artificial uncertainty as a function of time, while they become difficult to compare assuming a fixed absolute uncertainty. Our model for  $\sigma_{\text{sys}}$  is shown in Fig. 1. The top panel shows the real part of the 22 mode numerical-relativity waveform  $h_+(t)$  and fit. We include a gray band representing our choice for the artificial noise, while the middle panel shows  $h_+(t)$  multiplied by  $e^{t/\tau_{220}}$ ; the artificial noise  $\sigma_{\text{sys}}$  can be visualized as a horizontal line at 0.06 on this panel. The bottom panel is a time series showing the difference between the numerical-relativity waveform and our fit. We provide more details on this approach in Appendix A.

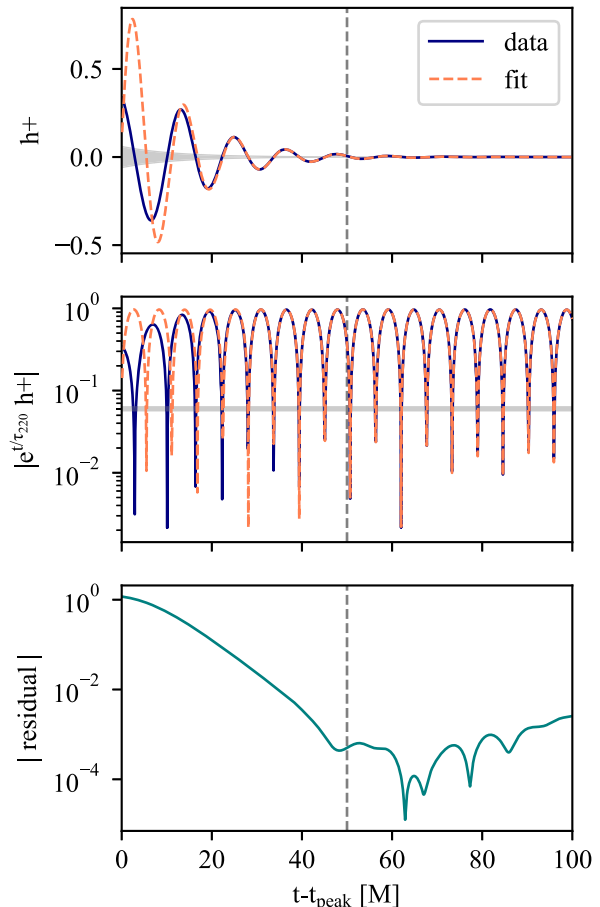


FIG. 1. Ringdown time series. The top panel shows the 22 mode numerical-relativity data  $h_+(t)$ , which is the real part of the 22 mode strain, in solid blue alongside a fit obtained using just the 220 and 320 tones predicted by perturbation theory. The dashed vertical line represents the start time for this fit ( $50M$ ). The gray band represents the artificial noise  $\sigma_{\text{sys}}$ . The second panel shows the same time series, but multiplied by  $e^{t/\tau_{220}}$  in order to counteract the exponential decay of the 220 tone. The horizontal line represents our model for the uncertainty  $\sigma_{\text{sys}}$  in this regime, which in this representation is constant. The bottom panel shows the difference between the numerical-relativity waveform and our fit as a function of time.

### III. RESULTS

#### A. Consistency of tones with time

We seek to find a set of tones that can produce a self-consistent measurement of the ringdown signal across time. To this end, it is useful to introduce a criterion to determine if each tone in a set is stably recovered. For a tone to be recovered stably at time  $t_0$ , we require that the 95% credible intervals for the amplitude and phase of that tone overlap with the 95% credible intervals of *all* subsequent fits (corresponding to larger values of  $t_0$ ). Thus, e.g., the fit at  $10M$  must be consistent with all the fits between  $10M$  and  $70M$  to be considered stable at  $10M$ . We are interested in tones that are both consistent and resolved. In order for a tone to be considered stable and resolved at time  $t_0$ , we

<sup>1</sup>The numerical relativity error is the difference between the two highest resolution waveforms after they are mapped to the same BMS frame.

<sup>2</sup>Presumably, if one could obtain a sufficiently long and accurate numerical-relativity waveform,  $\sigma_{\text{sys}}$  would approach the level of the numerical-relativity noise.

further require that the 95% credible interval for the amplitude excludes zero. This creates three possible classifications:

- (1) *Stable recovery*. The amplitude of the tone is non-zero and the amplitude and phase of the tone are consistent with subsequent fits.
- (2) *Unstable recovery*. The amplitude of the tone is nonzero, but the amplitude and/or phase are *inconsistent* with subsequent fits.
- (3) *Unresolved*. The amplitude of the tone is consistent with zero.

This classification method is illustrated in Fig. 2. Each panel shows amplitude of a different tone as a function of start time  $t_0$ . There are five tones in our fit: 220, 320, 221, 222 and 223. This amounts to four tones of the  $\ell = m = 2$  QNM plus a contribution from the 320 due to mode-mixing. At regularly spaced values of  $t_0$ , we plot the posterior for each amplitude in teal. The amplitudes are all measured at a reference time of  $t = t_{\text{peak}}$  while the fit is performed at different start times  $t_0$ . The background is shaded according to the stability of the tone. The light-gray regions indicate that the tone is unstably recovered, the white regions indicate it is stably recovered, and the dark-gray regions indicate that the tone is unresolved. To avoid clutter, we do not show the accompanying plots of phase, although we also require phase consistency in order for a tone to be classified as stable. We provide the corresponding plot of the phase posteriors in Appendix B.

From the first two panels, we see that the 220 and 320 tones are stably recovered from  $t_0 = 0M$ . In Appendix C we compare the 220 posteriors for the lowest and highest  $n$  fits, demonstrating how the 220 stabilizes with the inclusion of more tones. The 221 tone is initially unstably recovered, stabilizes at  $10M$ , and then becomes unresolved at  $30M$ . The 222 tone is stably recovered within a narrow window around  $5 - 15M$  while the 223 tone is unresolved. With our framework, the  $\ell = m = 2$  ringdown signal is well described by four tones which are stable from approximately  $10M$ , when considering a model allowing 220, 320, 221, 222 and 223 contributions.

We investigate how the stability changes with the number of tones in a set. First we investigate the 220 alone. We find that the 220 is not stably recovered until approximately  $30M$  unless other tones are included in the fit. This is consistent with the expectation that the 220 should become dominant at late times. Adding the 320 does not improve the stability of the recovery significantly. When we add the 221, the 220 becomes stably recovered at  $15M$ . The earliest stable time for the 220 changes to  $5M$  when we add the 222 and  $0M$  when we add the 223.<sup>3</sup>

<sup>3</sup>The fits in Sec. III are all carried out assuming the frequency and damping times for each tone predicted by perturbation theory using the remnant mass and spin of the simulation. In Appendix D, we provide the results of investigations where we treat the tone frequency as a free parameter.

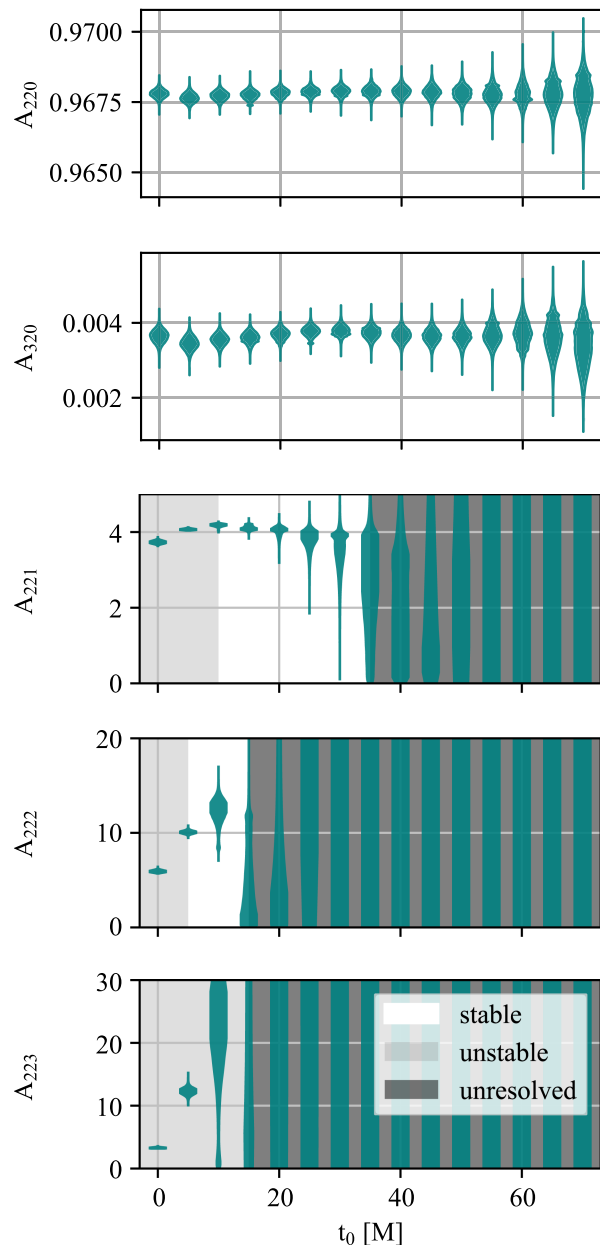


FIG. 2. Posterior distributions for amplitude as a function of  $t_0$ . Each panel represents a different tone. We show the regimes where each fit is stably recovered (white), unstably recovered (light gray) and unresolved (dark gray). We also fit the phase of each tone and require the phase to also be stable for that tone to be considered stable. For simplicity we only show the amplitude posteriors here. We show the corresponding phase posteriors in Fig. 6. The 221, 222 and 223 tones are unstably recovered at early times and become unresolved at late times, which implies that they do not significantly improve the fit at late times. The 223 tone is not stably recovered at any time but may stabilize with additional tones.

This analysis suggests that the 22 mode ringdown signal is well described by five tones. Adding the 224 tone may slightly improve the recovery stability of the 222, but does not improve that of the 223 and decays too quickly to be

TABLE I. Summary of when each tone is stably recovered for different values of  $N_{\text{tones}}$  in a given set. The variable  $t_{\text{stable}}$  indicates the time at which the amplitude and phase of each tone become consistent with subsequent fits using later parts of the waveform (and which the amplitude of the tone is inconsistent with zero). This time (and all other times in this table are quoted in units of  $M$ ). If the tone is not stably recovered for any time, this cell is marked “...”. The variable  $t_{\text{unresolved}}$  indicates the time at which the amplitude becomes consistent with zero. If the tone is never consistent with zero, this cell is marked “...”  $\tau$  denotes the damping time for each tone.

$N_{\text{tones}}$	Tone	$\tau$	$t_{\text{stable}}$	$t_{\text{unresolved}}$	Stable interval
1	220	11.74	30	...	30–70
	320	11.27	0	...	
3	220	11.74	15	...	30–45
	320	11.27	5	...	
	221	3.88	30	45	
4	220	11.74	5	...	...
	320	11.27	0	...	
	221	3.88	20	40	
	222	2.30	...	20	
5	220	11.74	0	...	...
	320	11.27	0	...	
	221	3.88	10	35	
	222	2.30	5	15	
	223	1.62	...	15	
6	220	11.74	0	...	...
	320	11.27	0	...	
	221	3.88	10	35	
	222	2.30	0	15	
	224	1.26	...	5	

stably recovered itself. The 221 behavior remains unchanged with the inclusion of the 224. Additional tones produce fits consistent with zero amplitude. Table I summarizes the recovery stability of different tones as we vary the number of tones  $N$ .

We test the impact of reducing the spacing in start times  $t_0$  on our results. We test the set of five-tones fit using a spacing of  $2M$  between 0 and  $20M$  and find the tones behave consistently with the more coarsely spaced test. However, the 221 recovery is slightly improved, stabilizing at  $6M$ , rather than  $10M$ . The 222 recovery stabilizes at  $8M$ —slightly later than the previous test—and becomes unresolved at  $14M$ . The 223 recovery remains unstable at all times and becomes unresolved at  $12M$ . We emphasize that the qualitative results are not significantly changed by the spacing of time steps.

We investigate the contribution of other potential sources of mode-mixing: the counterrotating 220 tone and the 321 and 420 tones which were found to be visible in Refs. [50,51]. We test the counterrotating 220 (r220) in

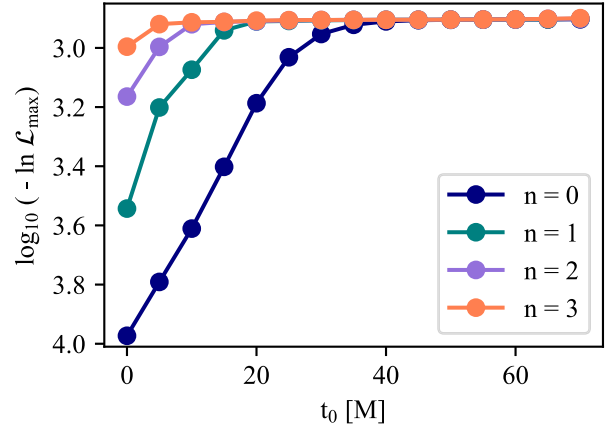


FIG. 3. Maximum natural log likelihood as a function of  $t_0$  for different values of  $N_{\text{tones}}$ . (To handle the large dynamic range of likelihood values, we actually plot the  $\log_{10}$  of  $-\ln \mathcal{L}_{\text{max}}$ .) All four models include the 320 tone. At early times the fits improve as we add tones to the model before saturating at around  $30M$ , lending further support to the conclusion that the fit is dominated by the 220 past  $30M$ .

a set of tones with the 220 and 320 and find that the r220 tone is resolved only until  $15M$  and its amplitude is smaller than the 320 ( $<0.003$ ). We test a set of tones including the five tones measured in Fig. 2 and the 321 tone. The 321 can be resolved until  $20M$  with an amplitude of  $\approx 0.1$  in the stable region of  $5 - 15M$ . The qualitative results of the other tones in this set remain unchanged, although the posteriors are in general broadened due to increased statistical uncertainty from introducing another parameter. We find that the 420 amplitude is consistent with 0 from  $t = 0M$  in a set of four tones (220, 320, 420, 221).

As a consistency check, we assess the goodness-of-fit for each of our models by comparing the maximum log likelihood of each model as a function of time. Figure 3 shows the log likelihood of each model as a function of start time  $t_0$ . The likelihood increases as the number of tones is increased, suggesting that the addition of tones serves to improve the fit, although the improvements are minimal from  $30 - 40M$  after the strain peak.

## B. How many tones can be resolved at the waveform peak?

Reference [35] suggests that the perturbative regime can be applied as early as the waveform peak ( $t_0 = 0M$ ), and that seven tones produce the best fit at this time. We investigate the number of tones that can be resolved at  $0M$ , when the strain amplitude is maximal. We fit our numerical-relativity data with a set of eight tones, consisting of the first seven tones of the 22 mode and the 320 tone. We find that overtones higher than 224 are not resolved because their amplitude posteriors are consistent with zero. This appears to be consistent with Fig. 9 in [35].

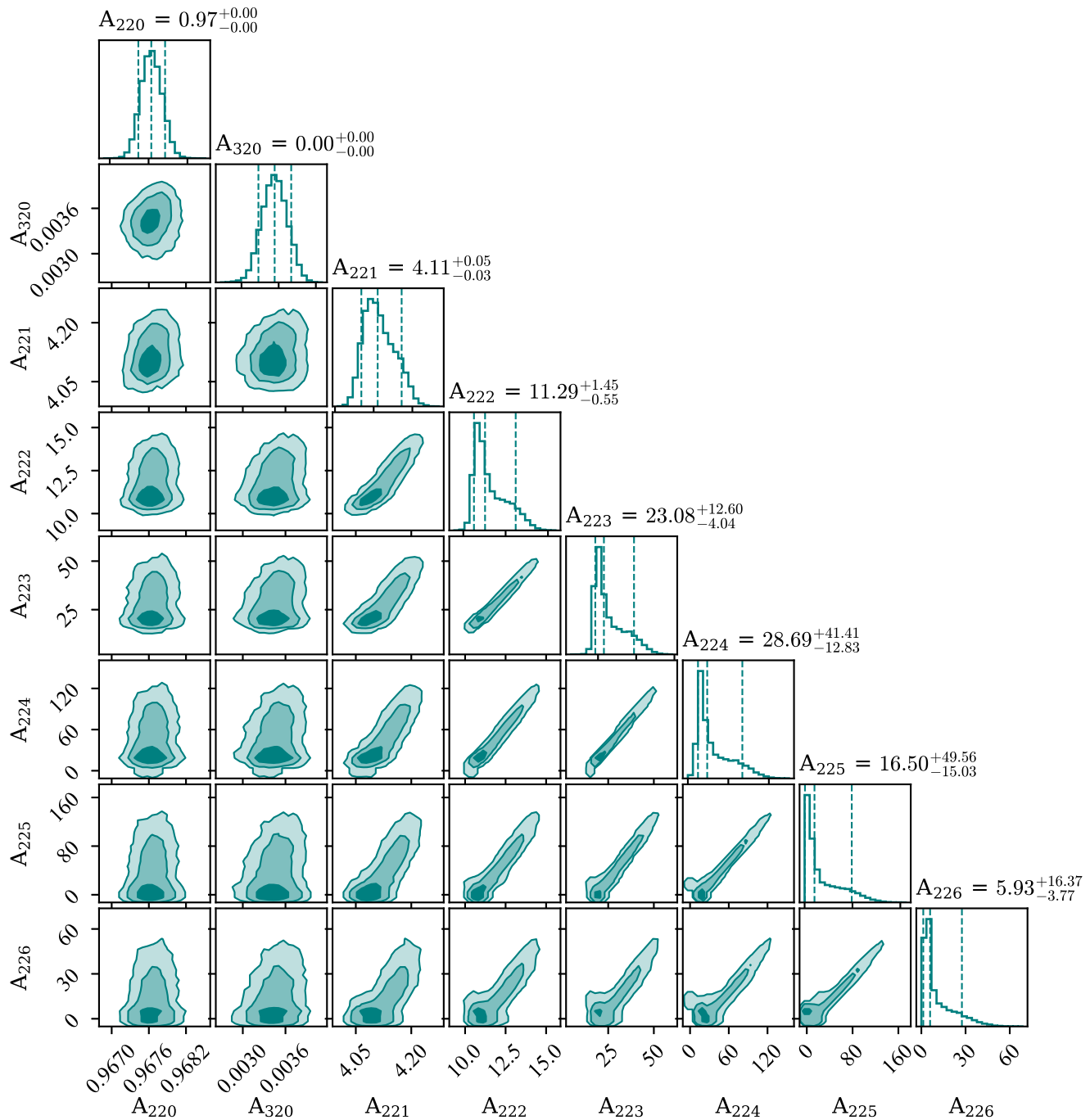


FIG. 4. The amplitude posteriors for a seven-tone fit at the strain peak ( $t_0 = 0M$ ). Tones up to  $n = 4$  are measured confidently with amplitudes consistent with [35], higher overtones return posteriors that support zero. This suggests that including tones higher than the 224 may be over-fitting. We notice that the correlation between tones increases as  $n$  increases. This is consistent with the findings of [18], and reflects the increasing difficulty to distinguish between higher tones that decay quickly and may not be stably recovered.

We also test a model with seven tones (six tones of the 22 mode and the 320) and find that when the 225 is the highest tone, the 225 can be resolved away from zero, but is not resolved with the addition of the 226. It becomes consistent with zero at the next time step. This suggests that at least

some over-fitting is likely occurring for  $n = 7$  within the framework of our phenomenological noise model, since those higher overtones are not resolved for this fit. Figure 4 shows the posteriors for the amplitudes of each tone included in this fit.

#### IV. DISCUSSION AND CONCLUSIONS

In this work we introduce a framework to determine under what circumstances a perturbative description can be self-consistently applied to a binary black hole ringdown obtained from numerical relativity simulations. We start near the end of the waveform where the 220 tone is expected to be dominant and work backward, using the late-time waveform to set the scale of systematic uncertainty in our model. Similarly to [e.g., [45,50]] we employ a criterion for stability to ensure that the fits obtained at different start times are consistent. We find that it is possible to achieve a self-consistent perturbative description within our framework from  $\approx 10M$  after peak strain. The perturbative description does not appear to be stable back to the peak strain given the specific set of tones considered here.

The 220 is not stably recovered without the addition of the 221 until late times. However, the 221 is only stably recovered and resolvable with this method for a short time interval between  $10 - 35M$ . Including the 222 appears to improve the fits at early times and allows the 220 and 221 to be recovered stably at earlier times. The 222 recovery can be briefly stabilized with the addition of the 223. However the 223 is not stably recovered at any of the times we investigate and does not seem to stabilize with the addition of the 224.

The 320 tone, which we add as the dominant contribution from mode mixing, is always stable from the waveform peak. Reference [35] suggests that the perturbative model can be applied at or before  $t_0 = t_{\text{peak}}$ . We show that, in our framework, at  $t_0 = t_{\text{peak}} = 0M$  the numerical-relativity data can be explained in a self-consistent way with a set of five tones (four tones of the 22 mode along with the 320 tone). Any higher tones included in the fit are consistent with zero amplitude.

None of the sets of tones beyond  $N_{\text{tones}} = 3$  include a region where *all* tones included in the set are recovered stably. When  $N_{\text{tones}} > 3$ , the highest tone transitions from unstable to unresolved with no stable region. If one requires that all the tones in a set are stably recovered over some finite interval, then the largest available set of tones is  $N_{\text{tones}} = 3$ .

We have shown that a superposition of quasinormal tones can be used to produce a self-consistent description of the ringdown. However, this does not prove that these fits are “physical” (as opposed to phenomenological) or that the signal is consistent with a perturbed black hole. This study brings into focus a great challenge at the heart of the black-hole spectroscopy program: it is unclear how one can even answer the question of whether the perturbative description is physical or phenomenological. The stability of the quasinormal tone fits is a requirement for their

physical interpretability, but further work is required to understand the relation of these observables to the underlying nature of the spacetime and our ability to probe it.

*Note added.* As we were preparing this manuscript, we became aware of work by [84] and [85]. Reference [84] also attempts to extract tones in a self-consistent framework—by iteratively subtracting away the tone with the longest damping time. They find that this technique improves the stability of the extracted amplitude and phases for up to five tones of SXS:0305. This is broadly consistent with our result that only a limited number of tones can be fitted in a self-consistent framework, although we find this is limited to four tones rather than five. They also include the 320 contribution from mode-mixing and find that this improves the stability of the early tones but introduces more instability of tones higher than  $n = 2$ . Reference [85] uses Bayesian inference to compare the performance of nonlinear inspiral-merger-ringdown (IMR) models with linear QNM ringdown models. They find that IMR models produce better fits with higher Bayes factors at early times. They find that overtones may be measurable in high SNR events consistent with third-generation observatories. They caution that the instability of tones may make linear QNM models less reliable than nonlinear models for performing tests of general relativity.

#### ACKNOWLEDGMENTS

We thank the referee for their helpful suggestions to improve the manuscript. We thank Dana Jones and Gregorio Carullo for their helpful comments on this work. We also thank Saul Teukolsky, Vishal Baibhav, Will Farr, Harrison Siegel and Ben Farr for helpful advice and discussions. This work is supported through Australian Research Council (ARC) Centre of Excellence CE170100004, Discovery Projects No. DP220101610 and No. DP230103088, and LIEF Project No. LE210100002. This work was supported in part by the Sherman Fairchild Foundation and NSF Grants No. PHY-2011968, No. PHY-2011961, No. PHY-2309211, No. PHY-2309231, No. OAC-2209656 at Caltech, as well as NSF Grants No. PHY-2207342 and No. OAC-2209655 at Cornell. T. A. C. receives support from the Australian Government Research Training Program. The authors are grateful for computational resources provided by the LIGO Laboratory computing cluster at California Institute of Technology supported by National Science Foundation Grants No. PHY-0757058 and No. PHY-0823459, and the Ngarrgu Tindebeek / OzSTAR Australian national facility at Swinburne University of Technology.



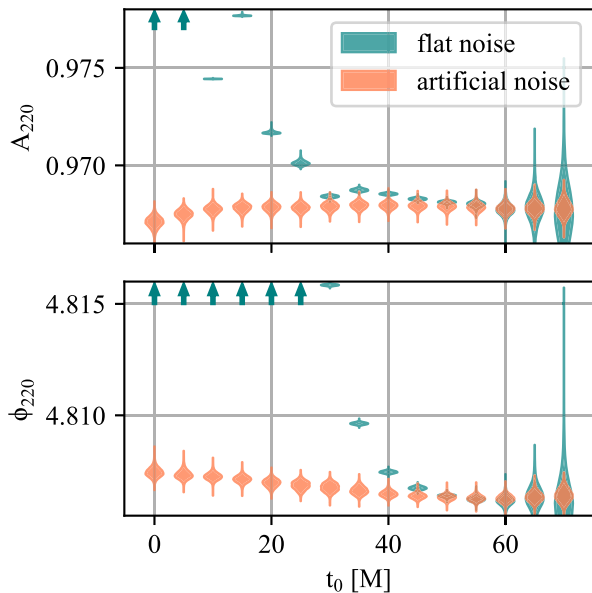


FIG. 5. Posterior for amplitude and phase as a function of start time  $t_0$  for the 220 tone. (The fits also include the 320 tone; not shown here.) The “flat-noise” model (teal) assumes the systematic error does not change with respect to  $t_0$ . The “artificial-noise” model (orange) assumes that systematic error is a constant fraction of the 220 amplitude. The artificial-noise model produces more consistent fits than the flat-noise model.

### APPENDIX A: NOISE MODEL

In the body of this manuscript, we employ an “artificial noise” model in which the systematic uncertainty is a constant error relative to the 220 tone; see Sec. II. We also test a “flat noise” model in which the systematic uncertainty is constant at all times. In Fig. 5 we show how the flat-noise fits compare to the artificial-noise fits. (For this analysis, we include the 220 and 320 tones.) We find that the posteriors are more comparable in magnitude and size over time using the artificial noise approach. We also find better sampler convergence, which is likely due to the SNR being comparable at early and late times rather than orders of magnitude different. Thus, in order to facilitate a self-consistent overtone model, we employ the artificial noise model in the main body of the manuscript. However we emphasize that this noise model is entirely phenomenological and not physically motivated.

### APPENDIX B: PHASE POSTERIORIORS

In Fig. 6 we plot the phase posteriors of the five-tone fit we present in the main body of the manuscript. The corresponding posteriors for the tone amplitudes are presented in Fig. 2.

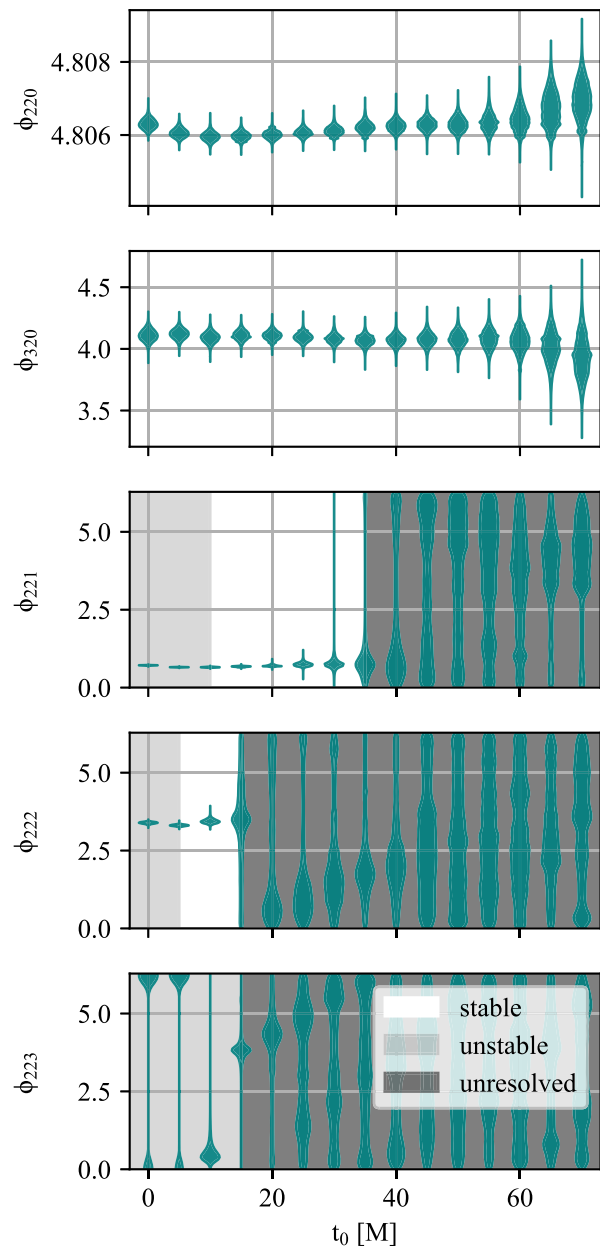


FIG. 6. Posterior distributions for phase as a function of  $t_0$ . Each panel represents a different tone. We show the regimes where each fit is stably recovered (white), unstably recovered (light gray) and unresolved (dark gray). The 221, 222 and 223 tones are unstably recovered at early times and become unresolved at late times, which implies that they do not significantly improve the fit at late times. The 223 tone is not stably recovered at any time but may stabilize with additional tones. We show the corresponding amplitude posteriors in Fig. 2.

### APPENDIX C: 220 STABILITY

Figure 7 shows the posterior distributions recovered for the 220 tone with the lowest and highest dimensional models included in the study. We also show the median values that are common to all posteriors in the 5-tone fit.

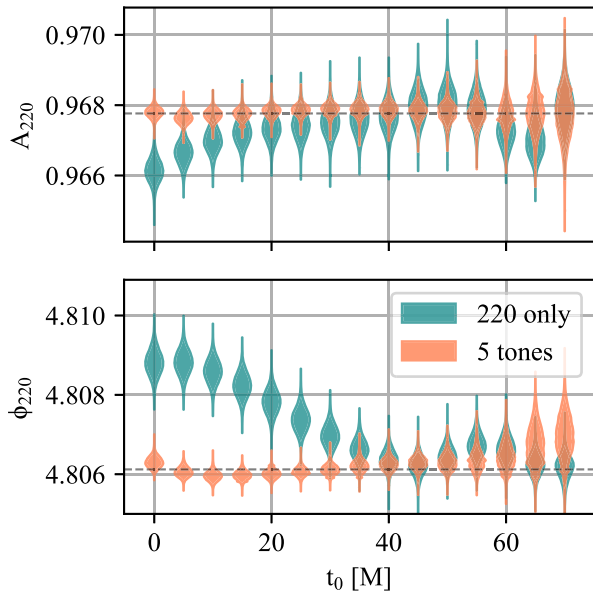


FIG. 7. The amplitude and phase posteriors of the 220 tone when measured with a 1 and 5-tone model. The 5-tone model is stable at all times, while the 1-tone model becomes stable at  $30M$ . The dashed horizontal lines show the median of the overlap at 95% confidence for the 5-tone model.

#### APPENDIX D: VARYING THE FREQUENCY OF OVERTONES

In the main body of this manuscript, we assume that the frequency and damping time of each tone are fixed to the values expected from black-hole perturbation theory. Here, we relax this assumption so that the frequency can be treated as a free parameter. We test a set of three tones including the 220, 320 and 221 tones and allowing the frequency of the 221 to vary. We use a uniform prior between 0.5 and 0.6 for the 221 frequency. Figure 8 shows the result of this fit. The predicted frequency value is shown as a red line in the bottom panel. The predicted frequency is consistent with the posterior only for start times  $t_0 > 15M$ . The posteriors for other parameters do not change significantly when  $f_{221}$  is treated as a free parameter. The fact that the numerical-relativity data prefer the wrong value of  $f_{221}$  before  $15M$  could mean that nonlinearities are present in the data at early times that are influencing the fit. Alternatively, it may be that more tones are required at early times to avoid this behavior, which is what [e.g., [35]] would imply.

We also measure the frequency of the 220 tone in a two-mode fit with the 220 and 320 tones. We find that the posterior moves away from the predicted frequency value while the phase becomes unstable. The frequency predicted by general relativity is only recovered for three time steps between  $35$  and  $45M$ . In the three- and four-tone models, varying the 221 frequency slightly improves the stability of the first three tones. The general relativity frequency for the

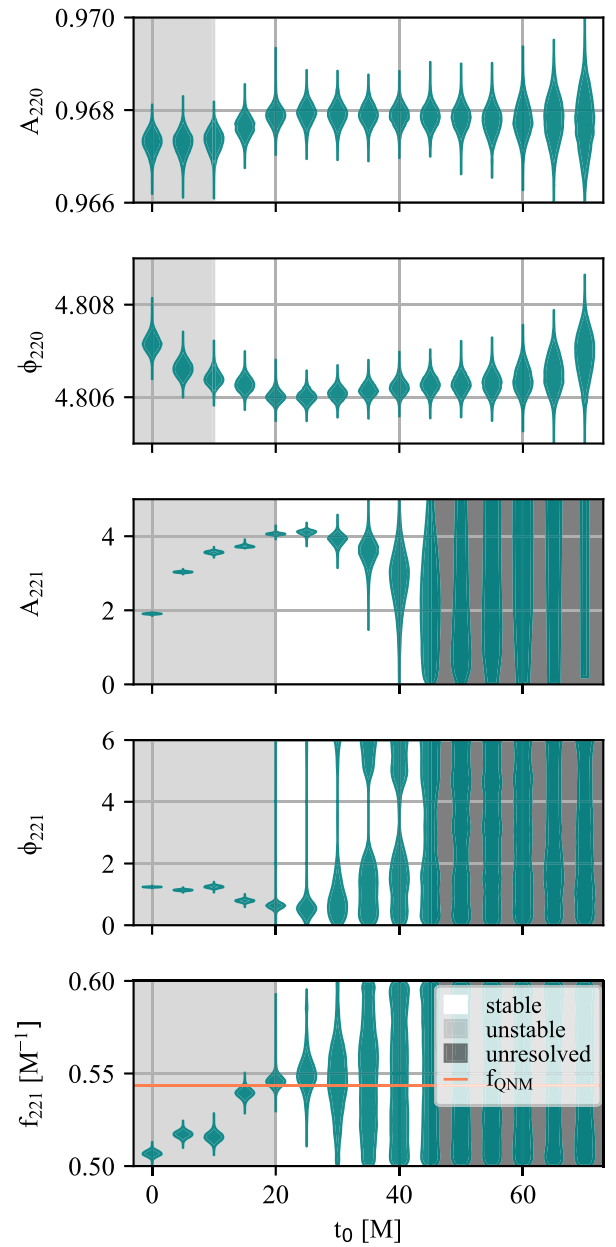


FIG. 8. Violin plots showing the posterior distributions for a three-tone fit where  $f_{221}$  is treated like a free parameter. Including the 221 makes the 220 fit stable at earlier times. The fit becomes stable at  $20M$ , which is  $10M$  earlier than when we keep  $f_{221}$  fixed. Similarly, the 220 becomes stably recovered  $5M$  earlier than when  $f_{221}$  was fixed.

221 is recovered at relatively late times:  $15M$  for a three-mode fit and  $10M$  for a four-mode fit. This further highlights that caution is required when fitting the tone frequencies due to the uncertain number of tones required, potential nonlinearities and nonoptimal fits, especially when doing so to test the no-hair theorem as shown in [e.g., [16,35]].

Table II summarizes the stability of the recovery of each of the tones when we sample the frequency as well as the

TABLE II. The stability of our tones when we vary either the  $f_{220}$  or  $f_{221}$  as free parameters. In the two-tone fit, we allow  $f_{220}$  to vary. The variable  $t_{\text{stable}}$  indicates the time at which the amplitude and phase of each tone become consistent with subsequent fits using later parts of the waveform (and which the amplitude of the tone is inconsistent with zero). This time (and all other times in this table are quoted in units of  $M$ . If the tone is not stably recovered for any time, this cell is marked “...” The variable  $t_{\text{unresolved}}$  indicates the time at which the amplitude becomes consistent with zero. If the tone is never consistent with zero, this cell is marked “...”  $\tau$  denotes the damping time for each tone.

$N_{\text{tones}}$	Tone	$\tau$	$t_{\text{stable}}$	$t_{\text{unresolved}}$	Stable interval
2	220	11.74	60	...	60–70
	320	11.27	5	...	
3	220	11.74	10	...	20–45
	320	11.27	0	...	
	221	3.88	20	45	
4	220	11.74	5	...	...
	320	11.27	0	...	
	221	3.88	15	35	
	222	2.30	...	20	

amplitude and phase. Treating  $f_{221}$  as a free parameter helps stabilize the fits at earlier times than when we sample in the amplitude and phase alone. The 221 also remains distinct from zero for an extra time step. However, we see

the opposite effect when we treat  $f_{220}$  as a free parameter for the two-tone fit. The 220 parameters do not become stable until  $60M$  and the 320 tone is slightly affected, stabilizing at  $5M$  instead of  $0M$ .

- 
- [1] C. V. Vishveshwara, Stability of the Schwarzschild metric, *Phys. Rev. D* **1**, 2870 (1970).
- [2] S. A. Teukolsky, Perturbations of a rotating black hole. I. Fundamental equations for gravitational, electromagnetic, and neutrino-field perturbations, *Astrophys. J.* **185**, 635 (1973).
- [3] S. Chandrasekhar and S. Detweiler, The quasi-normal modes of the Schwarzschild black hole, *Proc. R. Soc. A* **344**, 441 (1975).
- [4] E. W. Leaver, An analytic representation for the quasi normal modes of Kerr black holes, *Proc. R. Soc. A* **402**, 285 (1985).
- [5] K. D. Kokkotas and B. G. Schmidt, Quasi-normal modes of stars and black holes, *Living Rev. Relativity* **2**, 2 (1999).
- [6] B. Carter, Axisymmetric black hole has only two degrees of freedom, *Phys. Rev. Lett.* **26**, 331 (1971).
- [7] O. Dreyer, B. Kelly, B. Krishnan, L. S. Finn, D. Garrison, and R. Lopez-Aleman, Black-hole spectroscopy: Testing general relativity through gravitational-wave observations, *Classical Quantum Gravity* **21**, 787 (2004).
- [8] E. Berti, V. Cardoso, and C. M. Will, Gravitational-wave spectroscopy of massive black holes with the space interferometer LISA, *Phys. Rev. D* **73**, 064030 (2006).
- [9] M. Isi, M. Giesler, W. M. Farr, M. A. Scheel, and S. A. Teukolsky, Testing the no-hair theorem with GW150914, *Phys. Rev. Lett.* **123**, 111102 (2019).
- [10] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Tests of general relativity with binary black holes from the second LIGO-Virgo gravitational-wave transient catalog, *Phys. Rev. D* **103**, 122002 (2021).
- [11] R. Cotesta, G. Carullo, E. Berti, and V. Cardoso, Analysis of ringdown overtones in GW150914, *Phys. Rev. Lett.* **129**, 111102 (2022).
- [12] R. Abbott *et al.* (LIGO Scientific, Virgo, KAGRA Collaborations), Tests of general relativity with GWTC-3, [arXiv: 2112.06861](https://arxiv.org/abs/2112.06861).
- [13] H.-P. Nollert, About the significance of quasinormal modes of black holes, *Phys. Rev. D* **53**, 4397 (1996).
- [14] E. Thrane, P. D. Lasky, and Y. Levin, Challenges for testing the no-hair theorem with current and planned gravitational-wave detectors, *Phys. Rev. D* **96**, 102004 (2017).
- [15] S. Bhagwat, M. Okounkova, S. W. Ballmer, D. A. Brown, M. Giesler, M. A. Scheel, and S. A. Teukolsky, On choosing the start time of binary black hole ringdowns, *Phys. Rev. D* **97**, 104065 (2018).
- [16] M. Isi and W. M. Farr, Analyzing black-hole ringdowns, [arXiv:2107.05609](https://arxiv.org/abs/2107.05609).
- [17] B. P. Abbott *et al.*, Tests of general relativity with GW150914, *Phys. Rev. Lett.* **116**, 221101 (2016).
- [18] S. Bhagwat, X. J. Forteza, P. Pani, and V. Ferrari, Ringdown overtones, black hole spectroscopy, and no-hair theorem tests, *Phys. Rev. D* **101**, 044033 (2020).
- [19] V. Baibhav, M. H.-Y. Cheung, E. Berti, V. Cardoso, G. Carullo, R. Cotesta, W. Del Pozzo, and F. Duque, Agnostic black hole spectroscopy: Quasinormal mode content of numerical relativity waveforms and limits of validity of

- linear perturbation theory, *Phys. Rev. D* **108**, 104020 (2023).
- [20] J. C. Bustillo, P. D. Lasky, and E. Thrane, Black-hole spectroscopy, the no-hair theorem, and GW150914: Kerr versus Occam, *Phys. Rev. D* **103**, 024041 (2021).
- [21] S. A. Hughes and K. Menou, Golden binary gravitational-wave sources: Robust probes of strong-field gravity, *Astrophys. J.* **623**, 689 (2005).
- [22] A. Ghosh, N. K. Johnson-McDaniel, A. Ghosh, C. K. Mishra, P. Ajith, W. D. Pozzo, C. P. L. Berry, A. B. Nielsen, and L. London, Testing general relativity using gravitational wave signals from the inspiral, merger and ringdown of binary black holes, *Classical Quantum Gravity* **35**, 014002 (2017).
- [23] N. K. Johnson-McDaniel, A. Ghosh, S. Ghonge, M. Saleem, N. V. Krishnendu, and J. A. Clark, Investigating the relation between gravitational wave tests of general relativity, *Phys. Rev. D* **105**, 044020 (2022).
- [24] R. Brito, A. Buonanno, and V. Raymond, Black-hole spectroscopy by making full use of gravitational-wave modeling, *Phys. Rev. D* **98**, 084038 (2018).
- [25] V. Gennari, G. Carullo, and W. Del Pozzo, Searching for ringdown higher modes with a numerical relativity-informed post-merger model, *Eur. Phys. J. C* **84**, 233 (2024).
- [26] F. Pretorius, Numerical relativity using a generalized harmonic decomposition, *Classical Quantum Gravity* **22**, 425 (2005).
- [27] M. Campanelli, C. O. Lousto, P. Marronetti, and Y. Zlochower, Accurate evolutions of orbiting black-hole binaries without excision, *Phys. Rev. Lett.* **96**, 111101 (2006).
- [28] J. G. Baker, J. Centrella, D.-I. Choi, M. Koppitz, and J. van Meter, Gravitational-wave extraction from an inspiraling configuration of merging black holes, *Phys. Rev. Lett.* **96**, 111102 (2006).
- [29] A. Buonanno, G. B. Cook, and F. Pretorius, Inspiral, merger, and ring-down of equal-mass black-hole binaries, *Phys. Rev. D* **75**, 124018 (2007).
- [30] V. Baibhav, E. Berti, V. Cardoso, and G. Khanna, Black hole spectroscopy: Systematic errors and ringdown energy estimates, *Phys. Rev. D* **97**, 044048 (2018).
- [31] I. Ota and C. Chirenti, Overtones or higher harmonics? Prospects for testing the no-hair theorem with gravitational wave detections, *Phys. Rev. D* **101**, 104005 (2020).
- [32] P. Mourier, X. Jiménez Forteza, D. Pook-Kolb, B. Krishnan, and E. Schnetter, Quasinormal modes and their overtones at the common horizon in a binary black hole merger, *Phys. Rev. D* **103**, 044054 (2021).
- [33] N. Sago, S. Isoyama, and H. Nakano, Fundamental tone and overtones of quasinormal modes in ringdown gravitational waves: A detailed study in black hole perturbation, *Universe* **7**, 357 (2021).
- [34] X. Li, L. Sun, R. K. L. Lo, E. Payne, and Y. Chen, Angular emission patterns of remnant black holes, *Phys. Rev. D* **105**, 024016 (2022).
- [35] M. Giesler, M. Isi, M. A. Scheel, and S. A. Teukolsky, Black hole ringdown: The importance of overtones, *Phys. Rev. X* **9**, 041060 (2019).
- [36] S. Ma, K. Mitman, L. Sun, N. Deppe, F. Hébert, L. E. Kidder, J. Moxon, W. Throwe, N. L. Vu, and Y. Chen, Quasinormal-mode filters: A new approach to analyze the gravitational-wave ringdown of binary black-hole mergers, *Phys. Rev. D* **106**, 084036 (2022).
- [37] G. B. Cook, Aspects of multimode Kerr ringdown fitting, *Phys. Rev. D* **102**, 024027 (2020).
- [38] L. M. Zertuche, K. Mitman, N. Khera, L. C. Stein, M. Boyle, N. Deppe, F. Hébert, D. A. B. Izzo, L. E. Kidder, J. Moxon, H. P. Pfeiffer, M. A. Scheel, S. A. Teukolsky, W. Throwe, and N. Vu, High precision ringdown modeling: Multimode fits and BMS frames, *Phys. Rev. D* **105**, 104015 (2022).
- [39] X. J. Forteza and P. Mourier, High-overtone fits to numerical relativity ringdowns: Beyond the dismissed  $n = 8$  special tone, *Phys. Rev. D* **104**, 124072 (2021).
- [40] M. Okounkova, Revisiting non-linearity in binary black hole mergers, [arXiv:2004.00671](https://arxiv.org/abs/2004.00671).
- [41] Y. Chen, P. Kumar, N. Khera, N. Deppe, A. Dhani, M. Boyle, M. Giesler, L. E. Kidder, H. P. Pfeiffer, M. A. Scheel, and S. A. Teukolsky, Multipole moments on the common horizon in a binary-black-hole simulation, *Phys. Rev. D* **106**, 124045 (2022).
- [42] K. Mitman, M. Lagos, L. C. Stein, S. Ma, L. Hui, Y. Chen, N. Deppe, F. Hébert, L. E. Kidder, J. Moxon, M. A. Scheel, S. A. Teukolsky, W. Throwe, and N. L. Vu, Nonlinearities in black hole ringdowns, *Phys. Rev. Lett.* **130**, 081402 (2023).
- [43] M. H.-Y. Cheung, V. Baibhav, E. Berti, V. Cardoso, G. Carullo, R. Cotesta, W. Del Pozzo, F. Duque, T. Helfer, E. Shukla, and K. W. Wong, Nonlinear effects in black hole ringdown, *Phys. Rev. Lett.* **130**, 081401 (2023).
- [44] D. Pook-Kolb, O. Birnholtz, J. L. Jaramillo, B. Krishnan, and E. Schnetter, Horizons in a binary black hole merger II: Fluxes, multipole moments and stability, [arXiv:2006.03940](https://arxiv.org/abs/2006.03940).
- [45] L. London, D. Shoemaker, and J. Healy, Modeling ringdown: Beyond the fundamental quasinormal modes, *Phys. Rev. D* **90**, 124032 (2014).
- [46] L. London, Modeling ringdown. II. Aligned-spin binary black holes, implications for data analysis and fundamental theory, *Phys. Rev. D* **102**, 084052 (2020).
- [47] X. Jiménez Forteza, S. Bhagwat, P. Pani, and V. Ferrari, Spectroscopy of binary black hole ringdown using overtones and angular modes, *Phys. Rev. D* **102**, 044053 (2020).
- [48] H. Zhu, J. L. Ripley, A. Cárdenas-Avedaño, and F. Pretorius, Challenges in quasinormal mode extraction: Perspectives from numerical solutions to the Teukolsky equation, *Phys. Rev. D* **109**, 044010.
- [49] P. J. Nee, S. H. Völkel, and H. P. Pfeiffer, Role of black hole quasinormal mode overtones for ringdown analysis, *Phys. Rev. D* **108**, 044032 (2023).
- [50] M. Ho-Yeuk Cheung, E. Berti, V. Baibhav, and R. Cotesta, Extracting linear and nonlinear quasinormal modes from black hole merger simulations, *Phys. Rev. D* **109**, 044069 (2024).
- [51] J. Redondo-Yuste, G. Carullo, J. L. Ripley, E. Berti, and V. Cardoso, Spin dependence of black hole ringdown nonlinearities, [arXiv:2308.14796](https://arxiv.org/abs/2308.14796).
- [52] G. Carullo, W. D. Pozzo, and J. Veitch, Observational black hole spectroscopy: A time-domain multimode analysis of GW150914, *Phys. Rev. D* **99**, 123029 (2019).
- [53] Y.-F. Wang, C. D. Capano, J. Abedi, S. Kastha, B. Krishnan, A. B. Nielsen, A. H. Nitz, and J. Westerweck,

- A frequency-domain perspective on GW150914 ringdown overtone, [arXiv:2310.19645](https://arxiv.org/abs/2310.19645).
- [54] S. Ma, L. Sun, and Y. Chen, Using rational filters to uncover the first ringdown overtone in GW150914, *Phys. Rev. D* **107**, 084010 (2023).
- [55] E. Finch and C. J. Moore, Searching for a ringdown overtone in GW150914, *Phys. Rev. D* **106**, 043005 (2022).
- [56] H.-T. Wang and L. Shao, Effect of noise estimation in time-domain ringdown analysis: A case study with GW150914, *Phys. Rev. D* **108**, 123018 (2023).
- [57] M. Crisostomi, K. Dey, E. Barausse, and R. Trotta, Neural posterior estimation with guaranteed exact coverage: The ringdown of GW150914, *Phys. Rev. D* **108**, 044029 (2023).
- [58] A. Correia, Y.-F. Wang, and C. D. Capano, Low evidence for ringdown overtone in GW150914 when marginalizing over time and sky location uncertainty, [arXiv:2312.14118](https://arxiv.org/abs/2312.14118).
- [59] M. Isi and W. M. Farr, Comment on “Analysis of ringdown overtones in GW150914”, *Phys. Rev. Lett.* **131**, 169001 (2023).
- [60] G. Carullo, R. Cotesta, E. Berti, and V. Cardoso, Reply to comment on “Analysis of ringdown overtones in GW150914”, *Phys. Rev. Lett.* **131**, 169002 (2023).
- [61] E. Berti, V. Cardoso, and A. O. Starinets, Topical review: Quasinormal modes of black holes and black branes, *Classical Quantum Gravity* **26**, 163001 (2009).
- [62] E. Berti, Ringdown (2023), <https://pages.jh.edu/eberti2/ringdown/>.
- [63] S. A. Teukolsky, Rotating black holes: Separable wave equations for gravitational and electromagnetic perturbations, *Phys. Rev. Lett.* **29**, 1114 (1972).
- [64] A. H. Mroué, M. A. Scheel, B. Szilágyi, H. P. Pfeiffer, M. Boyle, D. A. Hemberger, L. E. Kidder, G. Lovelace, S. Ossokine, N. W. Taylor, A. Zenginoğlu, L. T. Buchman, T. Chu, E. Foley, M. Giesler, R. Owen, and S. A. Teukolsky, Catalog of 174 binary black hole simulations for gravitational wave astronomy, *Phys. Rev. Lett.* **111**, 241104 (2013).
- [65] M. Boyle *et al.*, The SXS Collaboration catalog of binary black hole simulations, *Classical Quantum Gravity* **36**, 195006 (2019).
- [66] K. Mitman, J. Moxon, M. A. Scheel, S. A. Teukolsky, M. Boyle, N. Deppe, L. E. Kidder, and W. Throwe, Computation of displacement and spin gravitational memory in numerical relativity, *Phys. Rev. D* **102**, 104007 (2020).
- [67] J. Moxon, M. A. Scheel, and S. A. Teukolsky, Improved Cauchy-characteristic evolution system for high-precision numerical relativity waveforms, *Phys. Rev. D* **102**, 044052 (2020).
- [68] J. Moxon, M. A. Scheel, S. A. Teukolsky, N. Deppe, N. Fischer, F. Hébert, L. E. Kidder, and W. Throwe, The `SPECTRE` Cauchy-characteristic evolution system for rapid, precise waveform extraction, *Phys. Rev. D* **107**, 064013 (2023).
- [69] N. Deppe *et al.*, `SPECTRE` (Zenodo, 2023), [10.5281/zenodo.8431874](https://zenodo.org/record/8431874).
- [70] K. Mitman, N. Khera, D. A. B. Izzo, L. C. Stein, M. Boyle, N. Deppe, L. E. Kidder, J. Moxon, H. P. Pfeiffer, M. A. Scheel, S. A. Teukolsky, and W. Throwe, Fixing the BMS frame of numerical relativity waveforms, *Phys. Rev. D* **104**, 024051 (2021).
- [71] K. Mitman, L. C. Stein, M. Boyle, N. Deppe, F. Hébert, L. E. Kidder, J. Moxon, M. A. Scheel, S. A. Teukolsky, W. Throwe, and N. L. Vu, Fixing the BMS frame of numerical relativity waveforms with BMS charges, *Phys. Rev. D* **106**, 084029 (2022).
- [72] M. Boyle, Angular velocity of gravitational radiation from precessing binaries and the corotating frame, *Phys. Rev. D* **87**, 104006 (2013).
- [73] M. Boyle, L. E. Kidder, S. Ossokine, and H. P. Pfeiffer, Gravitational-wave modes from precessing black-hole binaries, [arXiv:1409.4431](https://arxiv.org/abs/1409.4431).
- [74] M. Boyle, Transformations of asymptotic gravitational-wave data, *Phys. Rev. D* **93**, 084031 (2016).
- [75] M. Boyle, D. Izzo, L. Stein, A. Khairnar, H. Rüter, M. Scheel, V. Varma, and K. Mitman, `SCRI` (Zenodo, 2023), [10.5281/zenodo.10081312](https://zenodo.org/record/10081312).
- [76] B. P. Abbott *et al.*, Observation of gravitational waves from a binary black hole merger, *Phys. Rev. Lett.* **116**, 061102 (2016).
- [77] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, `EMCEE: The MCMC Hammer`, *Publ. Astron. Soc. Pac.* **125**, 306 (2013).
- [78] D. Pook-Kolb, O. Birmholtz, J. L. Jaramillo, B. Krishnan, and E. Schnetter, Horizons in a binary black hole merger II: Fluxes, multipole moments and stability, [arXiv:2006.03940](https://arxiv.org/abs/2006.03940).
- [79] L. Sberna, P. Bosch, W. E. East, S. R. Green, and L. Lehner, Nonlinear effects in the black hole ringdown: Absorption-induced mode excitation, *Phys. Rev. D* **105**, 064046 (2022).
- [80] J. Redondo-Yuste, D. Pereñíguez, and V. Cardoso, Ringdown of a dynamical spacetime, *Phys. Rev. D* **109**, 044048 (2024).
- [81] L. C. Stein, `QNM: A Python package for calculating Kerr quasinormal modes, separation constants, and spherical-spheroidal mixing coefficients`, *J. Open Source Software* **4**, 1683 (2019).
- [82] E. W. Leaver, Spectral decomposition of the perturbation response of the Schwarzschild geometry, *Phys. Rev. D* **34**, 384 (1986).
- [83] G. Carullo and M. De Amicis, Late-time tails in nonlinear evolutions of merging black hole binaries, [arXiv:2310.12968](https://arxiv.org/abs/2310.12968).
- [84] K. Takahashi and H. Motohashi, Iterative extraction of overtones from black hole ringdown, [arXiv:2311.12762](https://arxiv.org/abs/2311.12762).
- [85] Y. Qiu, X. Jiménez Forteza, and P. Mourier, Linear vs. nonlinear modelling of black hole ringdowns, *Phys. Rev. D* **109**, 064075 (2024).