Black Hole Ringdown: The Importance of Overtones

Matthew Giesler[1,*] Maximiliano Isi,[2,3] Mark A. Scheel,[1] and Saul A. Teukolsky[1,4]

[1]*TAPIR, Walter Burke Institute for Theoretical Physics, California Institute of Technology, Pasadena, California 91125, USA*
[2]*LIGO Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[3]*LIGO Laboratory, California Institute of Technology, Pasadena, California 91125, USA*
[4]*Cornell Center for Astrophysics and Planetary Science, Cornell University, Ithaca, New York 14853, USA*

It is possible to infer the mass and spin of the remnant black hole from binary black hole mergers by comparing the ringdown gravitational wave signal to results from studies of perturbed Kerr spacetimes. Typically, these studies are based on the fundamental quasinormal mode of the dominant $\ell = m = 2$ harmonic. By modeling the ringdown of accurate numerical relativity simulations, we find, in agreement with previous findings, that the fundamental mode alone is insufficient to recover the true underlying mass and spin, unless the analysis is started very late in the ringdown. Including higher overtones associated with this $\ell = m = 2$ harmonic resolves this issue and provides an unbiased estimate of the true remnant parameters. Further, including overtones allows for the modeling of the ringdown signal for all times beyond the peak strain amplitude, indicating that the linear quasinormal regime starts much sooner than previously expected. This result implies that the spacetime is well described as a linearly perturbed black hole with a fixed mass and spin as early as the peak. A model for the ringdown beginning at the peak strain amplitude can exploit the higher signal-to-noise ratio in detectors, reducing uncertainties in the extracted remnant quantities. These results should be taken into consideration when testing the no-hair theorem.

I. INTRODUCTION

The end state of astrophysical binary black hole (BBH) mergers is a perturbed single BH characterized by two parameters: the final remnant mass M_f and spin angular momentum S_f [1–3]. The perturbed BH radiates gravitational waves at a specific set of frequencies over characteristic timescales completely determined by the mass and spin. The segment of the gravitational wave signal associated with the single BH's oscillations is known as the "ringdown" phase, as the perturbed BH rings down analogous to a struck bell. The set of frequencies and damping times associated with a given BH are known as quasinormal modes (QNMs), the damped oscillations connected to the underlying BH geometry. The modes can be decomposed into spin-weighted spheroidal harmonics with angular indices (ℓ, m) [4–6]. For each (ℓ, m), there exists a discrete set of complex frequencies denoted $\omega_{\ell mn}$, where n is the "overtone" index. The oscillatory behavior is described by $\mathrm{Re}(\omega_{\ell mn})$, while $\mathrm{Im}(\omega_{\ell mn})$ is related to the damping timescale by $\tau_{\ell mn} = -\mathrm{Im}(\omega_{\ell mn})^{-1}$. For a given (ℓ, m), the overtone index sorts the QNMs in order of decreasing damping timescales, so $n = 0$ corresponds to the least-damped mode (i.e., the longest-lived mode), which is often referred to as the fundamental mode.

The recent detections of merging BBHs [7–13] by Advanced LIGO [14] and Virgo [15], including the ringdown phase, have stimulated significant interest in measuring the QNMs from the observations [16–22]. Accurately determining the QNMs allows for precise tests of general relativity (GR) [23–29]. In Ref. [16], the frequency and damping time of the fundamental mode were inferred from the ringdown data of the first event (GW150914). The analysis was performed at several time offsets with respect to the time of peak strain amplitude. For sufficiently late values of this start time, the frequency and damping time were found to be in agreement with the prediction from GR for a remnant consistent with the full waveform. The multiple start times used in the analysis reflect an uncertainty about when the fundamental mode becomes a valid description for the ringdown, as there is noticeable disagreement between the measured mode and the GR prediction at early times. This result raises the following question: At what point in the ringdown does perturbation theory become relevant?

In this paper, we consider the contribution of QNM overtones to the ringdown. Including overtones allows for

*mgiesler@tapir.caltech.edu

an excellent description of the waveform well before the fundamental mode becomes dominant and extends the regime over which perturbation theory is applicable to times even before the peak strain amplitude of the waveform. Moreover, an improved model for the ringdown through the inclusion of overtones can provide more accurate estimates of the remnant mass and spin [30,31]. Furthermore, the inclusion of higher overtones provides a means to test GR at a more stringent level because the QNM frequencies of all included overtones are independently constrained by GR for any given M_f and S_f.

We begin by demonstrating the benefits of including overtones, in agreement with Refs. [30,31], by analyzing a numerical relativity (NR) waveform. We then show how overtones can improve the extraction of information from noisy LIGO or Virgo data. We show that the overtones are not subdominant, as is often assumed, but are instead critically necessary to properly model the linear ringdown regime. The inclusion of QNM overtones provides a high-accuracy description of the ringdown as early as the time of the peak strain amplitude, where the high signal-to-noise ratio (SNR) can be exploited to significantly reduce the uncertainty in the extracted remnant properties.

II. PREVIOUS STUDIES

There have been numerous attempts to identify the start time of ringdown, that is, the point in time where a transition has occurred from the nonlinear regime into one where the signal can be described by a linear superposition of damped sinusoids [19,21,31–34]. To highlight the existing disagreement in the literature, the following studies, each using NR waveforms as a test bed, come to different conclusions regarding this transition time. In Ref. [32], the start of the ringdown phase is inferred to be $10M$ (where M is the total binary mass, and $G = c = 1$) after the peak luminosity of the $\ell = m = 2$ component of the strain h; this is the time at which the frequency of the $\ell = m = 2$ mode roughly agrees with that of the fundamental QNM. In Ref. [33], the ringdown portion of the waveform is considered to be $10M$ after the peak luminosity of the Newman-Penrose scalar Ψ_4 (related to two time derivatives of h). A ringdown model with the fundamental and the first two overtones was built under this assumed start time and employed in Ref. [21], which concluded that a start time of $16M$ after the peak strain amplitude is optimal. The peak of Ψ_4 is implicitly used as the start time for the ringdown in Ref. [31], where a superposition of the fundamental mode plus the first two overtones provides an accurate representation of the remnant properties and the fundamental frequency expected from perturbation theory. Interestingly, in one of the earliest analyses of BBH waveforms using NR simulations, despite the limited numerical accuracy available for simulations at that time, Buonanno, Cook, and Pretorius [30] were able to fit three overtones to the NR ringdown waveform by

extending their analysis to times *before* the peak amplitude of Ψ_4. A superposition of QNMs, including overtones and pseudo-QNMs, became an integral part of modeling the merger-ringdown regime in earlier effective one body models [35–38].

A likely cause of confusion is that start times are defined with respect to the peak of some waveform quantity, and different authors choose different waveform quantities for this purpose. To illuminate the implicit time offsets incurred by differences in this choice, consider as a specific example the GW150914-like NR waveform SXS:BBH:0305 in the Simulating eXtreme Spacetimes (SXS) catalog [39,40]. For this waveform, the peak of h occurs first, followed by the peak luminosity of h, then the peak of Ψ_4, and finally the peak luminosity of Ψ_4. These last 3 times are about $7M$, $10M$, and $11M$ after the peak of h. As we will show, overtones beyond $n \sim 2$ are expected to have significantly decayed by the peak of Ψ_4, so relying on the peak of Ψ_4 to begin a ringdown analysis may be problematic.

The miscellany of start times above can be reconciled, to some extent, by considering the contribution of overtones to the ringdown. Relying solely on the fundamental mode as a description for the ringdown should result in only late-time agreement. Additional consideration of overtones at late times should result in finding significantly reduced amplitudes in any overtones that remain. As we demonstrate below, the reduced amplitudes at late times is because the overtones decay more quickly for larger n; each additional overtone included leads to a superposition of QNMs that provides a description of the ringdown at earlier times. Ignoring the contribution of overtones, by considering them to be negligible as in Ref. [19], indirectly leads to the conclusion that remnant properties remain unconstrainable even in the infinite SNR limit—which we find to be untrue.

III. MODEL

We use the fundamental QNM and a varying number of overtones to determine when the linear QNM solution best describes the (ℓ, m) mode extracted from NR simulations. Throughout, we focus on the aforementioned astrophysically relevant NR waveform SXS:BBH:0305 in the SXS catalog, which is modeled after the GW150914 event. The waveform represents a simulated system with a mass ratio of 1.22, where the larger BH has dimensionless spin $\vec{\chi} = 0.33\hat{z}$ and the smaller companion BH has dimensionless spin $\vec{\chi} = -0.44\hat{z}$. The resulting remnant in this simulation has a final mass $M_f = 0.9520M$ and dimensionless spin $\chi_f = S_f/M_f^2 = 0.6921$. We explore at what time the linear QNM description provides not only an optimal fit for the resulting ringdown waveform but also an optimal estimate of the remnant mass and spin.

We model the ringdown radiation as a sum of damped sinusoids [41–44] by writing each angular mode of the complex strain, $h = h_+ - ih_\times$, as

$$h_{\ell m}^N(t) = \sum_{n=0}^{N} C_{\ell mn} e^{-i\omega_{\ell mn}(t-t_0)} \quad t \geq t_0, \qquad (1)$$

with complex frequencies $\omega_{\ell mn} = \omega_{\ell mn}(M_f, \chi_f)$ as determined by perturbation theory [45,46]. Here, t_0 corresponds to a specifiable "start time" for the model, and times before t_0 are not included in the model. The complex coefficients $C_{\ell mn}$, which are not known *a priori* as they depend on the binary configuration and dynamics near merger, are determined using unweighted linear least squares in the time domain. The complex-valued amplitudes can be factored into a real-valued amplitude and phase, $C_{\ell mn} = |A_{\ell mn}|e^{-i\phi_{\ell mn}}$, of which we make direct use in Sec. IV C.

Throughout this work, we focus on describing the dominant *spherical* harmonic mode in the NR simulation, the $\ell = m = 2$ mode [47]. The natural angular basis in perturbation theory is spin-weighted *spheroidal* harmonics [4–6], which can be written as an expansion in spin-weighted spherical harmonics [6,48–50]. Decomposing the ringdown into spherical harmonics results in mixing of the spheroidal and spherical bases between the angular functions with the same m but different ℓ's, and this mixing increases with χ_f [6,51]. For the SXS:BBH:0305 waveform, the $\ell = m = 2$ spherical harmonic remains a good approximation for the $\ell = m = 2$ spheroidal harmonic. The amplitudes of the spheroidal and spherical $\ell = m = 2$ modes differ by a maximum of only 0.4%, which occurs roughly $15M$ after the peak of h. This difference is significantly smaller at the peak. The mixing is small because higher (ℓ, m) harmonics are subdominant for this waveform, but in a more general case, these higher harmonics may play a more important role.

IV. RESULTS

A. QNM overtone fits

The linear superposition of the fundamental QNM and N overtones is an excellent description of the waveform around and before the peak strain. To demonstrate this case, we begin by fixing the remnant properties to the final values provided by the NR simulation. With the mass M_f and dimensionless spin χ_f fixed, the set of frequencies $\omega_{22n}(M_f, \chi_f)$ is fully specified by perturbation theory. The only remaining free parameters in Eq. (1) are the complex coefficients C_{22n} and the model start time t_0. For N included overtones, and a given choice of t_0, we determine the $(N+1)$ complex C_{22n}'s using unweighted linear least squares, thus obtaining a model waveform given by Eq. (1). We construct such a model waveform for $t \geq t_0$ at many start times beginning at $t_0 = t_{\text{peak}} - 25M$ and extending to times $t_0 = t_{\text{peak}} + 60M$, where t_{peak} is the peak amplitude of the complex strain. For each start time t_0, we compute the mismatch $\mathcal{M}$ between our model waveform h_{22}^N and the NR waveform h_{22}^{NR} through

$$\mathcal{M} = 1 - \frac{\langle h_{22}^{\text{NR}}, h_{22}^N \rangle}{\sqrt{\langle h_{22}^{\text{NR}}, h_{22}^{\text{NR}} \rangle \langle h_{22}^N, h_{22}^N \rangle}}. \qquad (2)$$

In the above, the inner product between two complex waveforms, say, $x(t)$ and $y(t)$, is defined by

$$\langle x(t), y(t) \rangle = \int_{t_0}^{T} x(t)\overline{y(t)}dt, \qquad (3)$$

where the bar denotes the complex conjugate, the lower limit of the integral is the start time parameter t_0 in Eq. (1), and the upper limit of the integral T is chosen to be a time before the NR waveform has decayed to numerical noise. For the aforementioned NR simulation, we set $T = t_{\text{peak}} + 90M$.

This procedure results in mismatches as a function of t_0 for each set of overtones; these mismatches are presented in Fig. 1. The figure shows that $N = 7$ overtones provide the minimum mismatch at the earliest of times, as compared to the other overtone models. The waveform corresponding to the $N = 7$ overtone model and $t_0 = t_{\text{peak}}$ is visualized in Fig. 2, where the model waveform is compared to the NR waveform along with the fit residual.

At face value, Fig. 1 provides us with a guide for determining the times where a linear ringdown model with N QNM overtones is applicable. However, relying on the mismatch alone can be deceiving. The $n = 7$ overtone decays away very quickly, yet Fig. 1 shows that retaining this overtone still produces small mismatches at times beyond when this mode should no longer be numerically resolvable. This result is due to overfitting to numerical
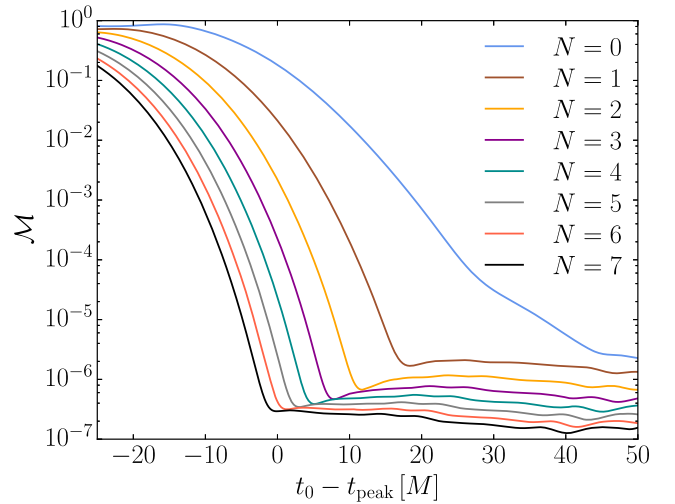

FIG. 1. Mismatches as a function of time for the eight models, each including up to N QNM overtones. The mismatch associated with each model at a given t_0 corresponds to the mismatch computed using Eq. (2), between the model and the NR waveform for $t \geq t_0$, where t_0 specifies the lower limit used in Eq. (3). Each additional overtone decreases the minimum achievable mismatch, with the minimum consistently shifting to earlier times.
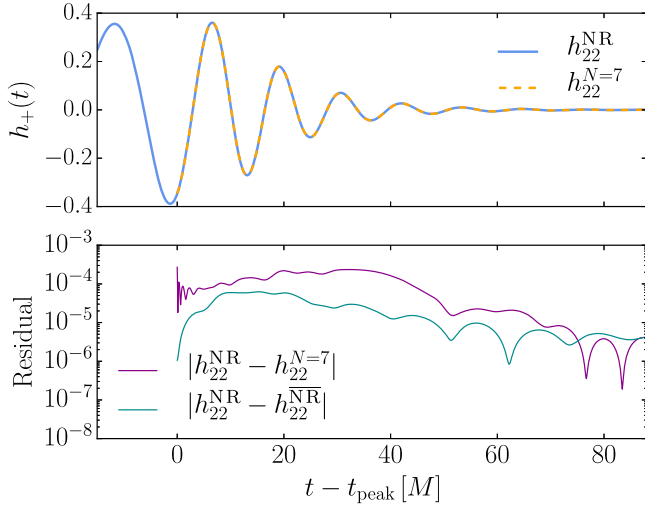
FIG. 2.   Comparison between the plus polarization of the $\ell = m = 2$ mode of the NR waveform and the $N = 7$ linear QNM model. The QNM model begins at $t_0 = t_{\text{peak}}$. The upper panel shows both waveforms, and the lower panel shows the residual for $t \geq t_{\text{peak}}$. For reference, the lower panel also shows an estimate of the error in the NR waveform, $|h_{22}^{\text{NR}} - h_{22}^{\overline{\text{NR}}}|$, where h_{22}^{NR} refers to the highest resolution waveform of SXS:BBH:0305 and $h_{22}^{\overline{\text{NR}}}$ refers to the next highest resolution waveform for this same system. The two NR waveforms are aligned at $t_0 = t_{\text{peak}}$, in both time and phase.

noise after the higher overtones in each model have sufficiently decayed. We find that the turnover subsequent to the first mismatch minimum in Fig. 1 is a good approximation for when each overtone has a negligible amplitude.

It is important that the model not only minimizes the residual in the waveform quantity but also provides faithful estimates of the underlying system parameters. In particular, we may demand that the inferred mass and spin agree with the true values known from the NR simulation. To check that the model does indeed faithfully represent the NR waveform with the correct final mass and spin, we repeat the fits but we allow M_f and χ_f to vary, and we set the frequencies of each overtone to their GR-consistent values through the perturbation-theory formula for $\omega_{22n}(M_f, \chi_f)$. As a measure of error, we use

$$\epsilon = \sqrt{(\delta M_f/M)^2 + (\delta \chi_f)^2}, \qquad (4)$$

where δM_f and $\delta \chi_f$ are the differences between the best-fit estimates for M_f and χ_f as compared to the remnant values from the NR simulation. Using a model with $N = 7$ overtones and $t_0 = t_{\text{peak}}$, the best-fit estimates for M_f and χ_f yield a value of $\epsilon \sim 2 \times 10^{-4}$. For reference, by comparing the two highest resolutions of this simulation, we estimate the error in the NR measured remnant mass and spin to be $\delta M_f \sim 1.3 \times 10^{-5} M$ and $\delta \chi_f \sim 2.1 \times 10^{-5}$, which corresponds to a value of $\epsilon \sim 2 \times 10^{-5}$.
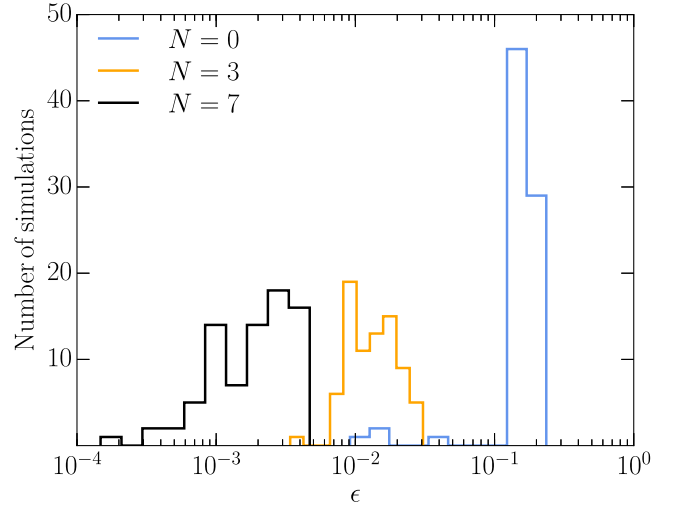

FIG. 3.   The distribution of ϵ, Eq. (4), for a range of simulations in the SXS catalog. The distribution includes systems with mass ratios up to 8 and orbit-aligned component spins with $|\vec{\chi}| \leq 0.8$. The distributions shown are for $N = \{0, 3, 7\}$ overtones at the peak of the strain amplitude. For the best performing model, $N = 7$, the median value is 2×10^{-3}, and the maximum error in estimating the mass and spin is about 5×10^{-3}.

Furthermore, the difference in the recovered M_f and χ_f as compared to the NR values increases as we drop overtones from the model. This behavior appears to be robust. Repeating the above analysis on roughly 80 additional waveforms in the SXS catalog with aligned spins and mass ratios up to 8 [39,52] yields similar results, with a median value of $\epsilon \sim 10^{-3}$. The full distribution of ϵ for this part of parameter space, with $N = 7$ overtones at $t_0 = t_{\text{peak}}$, is shown in Fig. 3.

Returning to our analysis of SXS:BBH:0305, to highlight the worst-fit and best-fit cases and to visualize the mismatch as a function of mass and spin, we compute the mismatch between NR and the model in Eq. (1) with $t_0 = t_{\text{peak}}$ and the C_{22n}'s determined by a least-squares fit for a grid of M_f and χ_f values. In Fig. 4, we see that with $N = 7$ overtones, the mismatch has a deep minimum associated with the true remnant quantities. However, using solely the fundamental mode, $N = 0$, with $t_0 = t_{\text{peak}}$, provides largely biased estimates for the remnant M_f and χ_f, as is visible in Fig. 5. This result is not surprising in light of Fig. 1, where at this time the $N = 0$ model provides the poorest mismatch; this result is a consequence of the higher overtones dominating the waveform at this time. The bias can be overcome by waiting a sufficiently long time, which allows the overtones to decay away and the fundamental mode to become dominant. This case can be seen in Fig. 6, where we repeat the same procedure with $N = 0$ and $t_0 = t_{\text{peak}} + 47M$. Here, the resulting distribution of mismatches in the $M_f - \chi_f$ plane is on par with the distribution associated with including $N = 7$ overtones and
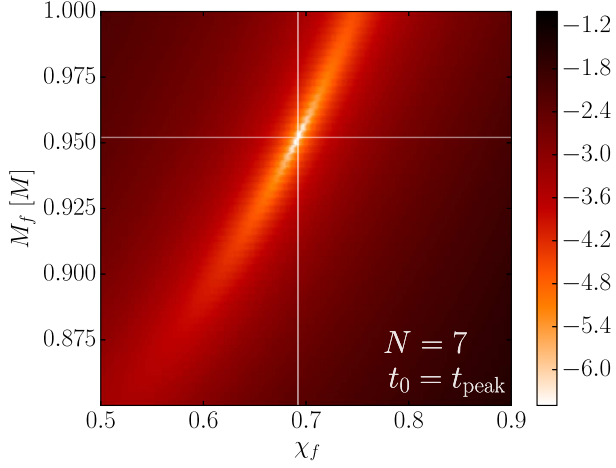
FIG. 4. Mismatches for the $\ell = m = 2$ mode between NR and Eq. (1) for a grid of M_f and χ_f with $N = 7$ and $t_0 = t_{peak}$. The white horizontal and vertical lines correspond to the NR values and are in good agreement with the M_f and χ_f mismatch distribution using the maximum number of overtones considered.

$t_0 = t_{peak}$, with the $N = 7$ case producing a smaller absolute mismatch than the $N = 0$ case. The key point is that we can recover similar information about the underlying remnant at the peak, through the inclusion of overtones, as we can by analyzing the waveform at late times. As discussed in more detail in Sec. IV C, extending the ringdown model to earlier times allows us to access higher signal-to-noise ratios and can significantly reduce uncertainties in parameter estimation.
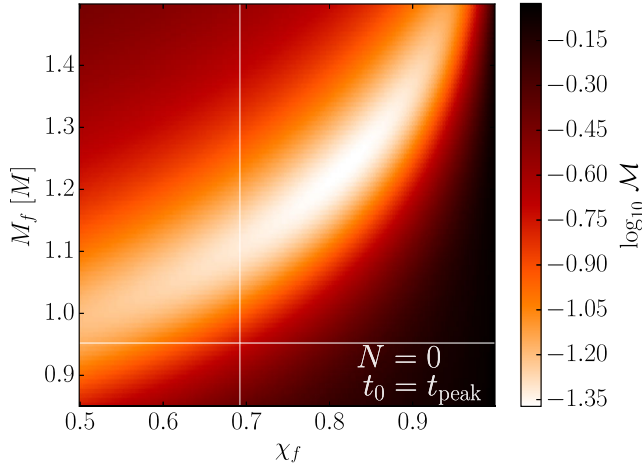

FIG. 5. The mismatches for the $\ell = m = 2$ mode between NR and Eq. (1) over a grid of χ_f and M_f with $N = 0$, the fundamental mode only, and $t_0 = t_{peak}$. The white horizontal and vertical lines correspond to the remnant values from NR. As the fundamental mode is subdominant at this time, this single-mode model is a poor probe of the underlying remnant mass and spin. Note that the mass and mismatch scales used in this figure are significantly different from Fig. 4 because of the discrepant single-mode fit at early times.
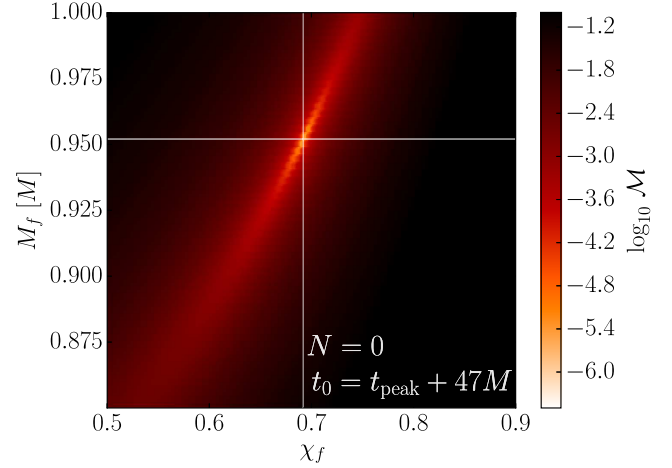

FIG. 6. Mismatches for the $\ell = m = 2$ mode between NR and Eq. (1) for a grid of M_f and χ_f with $N = 0$, the fundamental mode, at a late time $t_0 = t_{peak} + 47M$. Here, the fundamental mode is dominant since the overtones have decayed away by this time. Again, the white horizontal and vertical lines correspond to the remnant values from NR and now; at this late time, we find good agreement between the true values and those recovered by using only the fundamental mode as a probe for M_f and χ_f.

One might be concerned that the additional free parameters in the fit, introduced by including the overtones, simply allow for fitting away any nonlinearities that may be present, making the fundamental mode more easily resolvable and therefore better determining the underlying remnant mass and spin. A simple test of this idea is to repeat the fit while still setting the fundamental frequencies $\omega_{220}(M_f, \chi_f)$ according to perturbation theory but to intentionally set the frequencies of the overtones to incorrect values. The fit will then have the same number of degrees of freedom (d.o.f.) as previously but without the correct physics. Let $\omega_{22n}(M_f, \chi_f)$ be the set of frequencies determined by perturbation theory, and take $\tilde{\omega}_{22n}(M_f, \chi_f)$ to be the set of frequencies with the fundamental unmodified, but with $\tilde{\omega}_{22n}(M_f, \chi_f) = \omega_{22n}(M_f, \chi_f)(1 + \delta)$, for $n > 0$. As a measure of error, we rely on ϵ, Eq. (4), the root-mean-squared error in the estimated mass and spin as compared to the known NR values.

For demonstration purposes, we let δ take on values from the set $\pm \{0.01, 0.05, 0.2\}$ and fit to the spherical $\ell = m = 2$ mode with $t_0 = t_{peak}$ for different numbers of included overtones N. A comparison between the unmodified and modified models with the same number of d.o.f. is presented in Fig. 7. From Fig. 7, it is evident that the unmodified set of QNMs, $\omega_{22n}(M_f, \chi_f)$, remains true to the underlying mass and spin and converges to smaller errors as the number of included overtones is increased. In the case where the overtones are given slightly incorrect frequencies by the δ parameter introduced above, including higher overtones yields fits that remain biased away from the true values, leading to larger values of ϵ.
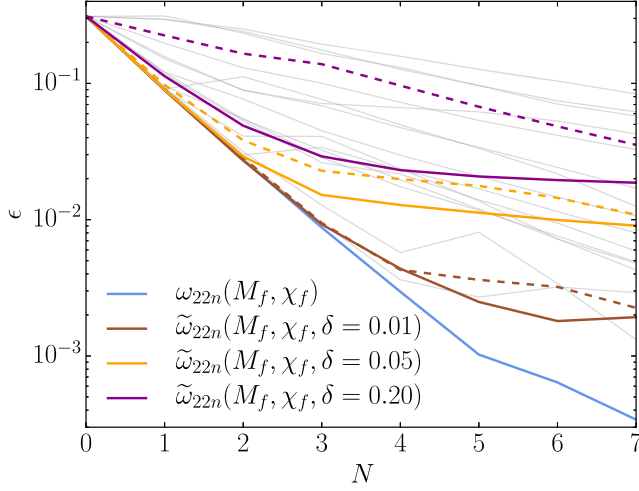
FIG. 7. The root-mean-squared error ϵ defined in Eq. (4), for different sets of frequencies, as a function of the number of included overtones N. The fits are performed on the spherical $\ell = m = 2$ mode at $t = t_0 = t_{\rm peak}$. The label $\omega_{22n}(M_f, \chi_f)$ represents the set of frequencies consistent with perturbation theory, while $\tilde{\omega}_{22n}(M_f, \chi_f)$ represents the set of frequencies with the fundamental mode, $n = 0$, unmodified but with a slight modification to the overtone frequencies by a factor of $(1 + \delta)$. For each δ, there is an associated dashed line of the same color that corresponds to $\delta \to -\delta$. The faint grey lines correspond to frequencies with a random δ for each n, as explained in the last paragraph of Sec. IV A. The results suggest that there is information present in the overtones that contributes to extracting the remnant properties at the peak, as these outperform a similar set of functions, with the same d.o.f. for each N but with frequencies inconsistent with the asymptotic remnant.

Furthermore, in an additional test, we have allowed for different values of δ for each n, each independently sampled from a normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 0.2$. In this test, each overtone frequency is randomly modified to a different extent about $\omega_{22n}(M_f, \chi_f)$. In all 100 cases randomly generated from the above distribution, the ϵ's associated with the modified frequencies always remain bounded from below by the ϵ associated with the GR frequencies of the asymptotic remnant. A random, representative, subset of these 100 cases is shown as faint grey traces in Fig. 7. The results suggest that the overtones associated with the asymptotic remnant provide a sufficiently good linear description of the perturbations for all times beyond the peak of this mode, while a similar set of overtones that are inconsistent with the asymptotic remnant do not.

B. Characterizing the overtones

The behavior in the previous section can be explained by carefully understanding how the overtones contribute to the ringdown. As briefly touched on in Sec. I, the overtones are those modes with $n > 0$, where n orders the modes based on decreasing damping time. While these modes are the least important in a time-weighted sense, describing them as "overtones" is somewhat of a misnomer. In a classical description of harmonics, overtones are at higher frequencies than the fundamental, typically multiples of the first harmonic, and are usually subdominant. However, for QNMs, the overtones decrease in frequency and are not necessarily subdominant. As briefly mentioned in Sec. III, the amplitude of each QNM overtone in the ringdown depends on the binary configuration and the dynamics leading up to merger. The dynamics specify the "initial data" for the ringdown, determining which QNMs are excited and to what extent. As such, the overtone amplitudes for waveform SXS:BBH:0305 will differ from those with different initial data, i.e., binary configurations with different mass ratios or different spin vectors.

To provide a qualitative understanding of the relative amplitudes of different overtones, we decompose the ringdown waveform of SXS:BBH:0305 into its constituent overtones. Using $t_0 = t_{\rm peak}$ and $N = 7$ overtones, we determine the C_{22n}'s as in Sec. IV A, with M_f and χ_f fixed to the NR simulation values. The corresponding values $A_n = |C_{22n}(t = t_0 = t_{\rm peak})|$ form the entries in the bottom row of Table I. For $N = 6$, we keep $t_0 = t_{\rm peak}$, so the amplitudes are measured with respect to the peak; however, we include in our fit only data for $t \geq t_{\rm fit}$, where $t_{\rm fit}$ corresponds to the earliest minimum in Fig. 1 for this N. These amplitudes correspond to the penultimate row of Table I, and the fit time $t_{\rm fit}$ is stated, with respect to $t_{\rm peak}$, in the last column. The result of this procedure for the remaining N is given in Table I, where we provide our best estimate of the amplitudes at $t = t_0 = t_{\rm peak}$ associated with each overtone. The values in Table I are computed for the highest numerical resolution of the NR waveform

TABLE I. Best-fit estimates of the amplitudes A_n of the fundamental mode and overtones in the ringdown of NR simulation SXS:BBH:0305, with $t_0 = t_{\rm peak}$. Amplitudes are computed for various values of N, the total number of overtones included in the fit. Also shown is the time $t_{\rm fit}$ where the fit is performed for each N, stated with respect to $t_{\rm peak}$. Note that A_n are always the amplitudes at $t = t_0 = t_{\rm peak}$, even if the fit is performed at a later time. The amplitude values are truncated such that the last significant figure agrees with the two highest resolutions for the NR simulation.

N	A_0	A_1	A_2	A_3	A_4	A_5	A_6	A_7	$t_{\rm fit} - t_{\rm peak}$
0	0.971	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	47.00
1	0.974	3.89	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	18.48
2	0.973	4.14	8.1	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	11.85
3	0.972	4.19	9.9	11.4	$\cdots$	$\cdots$	$\cdots$	$\cdots$	8.05
4	0.972	4.20	10.6	16.6	11.6	$\cdots$	$\cdots$	$\cdots$	5.04
5	0.972	4.21	11.0	19.8	21.4	10.1	$\cdots$	$\cdots$	3.01
6	0.971	4.22	11.2	21.8	28	21	6.6	$\cdots$	1.50
7	0.971	4.22	11.3	23.0	33	29	14	2.9	0.00

SXS:BBH:0305, but they are truncated at a level such that the estimates agree with the next highest resolution.

The initial amplitude of the fundamental mode A_0 is consistently recovered for all models, each model having a different N and a different fit time that is optimal for that N. The first few overtones show a similar behavior, while the higher overtones display larger uncertainties in the recovered amplitudes and are increasingly sensitive to the fit time and the number of included overtones. This sensitivity is a consequence of the strong exponential time dependence in the overtones and is recognized as the time-shift problem [53]. But perhaps the most important thing to notice is that the overtones can have significantly higher amplitudes than the fundamental mode. As discussed above, the initial amplitudes of the overtones depend on the details of the nonlinear binary coalescence, which ultimately depend on the binary parameters. Consequently, the amplitudes of the overtones relative to the fundamental mode will vary across parameter space. The complex amplitudes $C_{\ell mn}$, also known as the QNM excitation coefficients, can be written as $C_{\ell mn} = B_{\ell mn} I_{\ell mn}$, where $B_{\ell mn}$ is a purely geometric piece determined by the remnant BH, referred to as the QNM excitation factor, and $I_{\ell mn}$ is the source term that depends on the binary dynamics [54–56]. Excitation factors have been computed for the first three overtones for Kerr BHs in Refs. [56,57]; these QNM excitation factors can provide some insight into how the relative amplitudes might behave for different remnant spins.

The NR waveform SXS:BBH:0305 has a dimensionless remnant spin $\chi_f \sim 0.7$, for which the relative excitation factors $|B_{22n}|/|B_{220}|$ of the fundamental and the first three $\ell = m = 2$ QNM overtones are roughly 1.0, 3.53, 5.23, 5.32. However, for a remnant of $\chi_f = 0$, the excitation factors $|B_{22n}/B_{220}|$ of these same QNMs are 1.0, 1.28, 1.06, 0.62, which indicates that the overtones may be relatively less important for lower remnant spins. Using Ref. [57], we have computed the excitation factors for the next two highest overtones of the remnant of SXS:BBH:0305, and we find that $|B_{224}|/|B_{220}| \sim 15.21$ and $|B_{225}|/|B_{220}| \sim 29.31$. Additional excitation factors are difficult to compute, but the trend is not expected to continue as it is conjectured that, for Kerr BHs, $B_{\ell mn} \sim 1/n$ for large n [56].

The overtone amplitudes in Table I increase with overtone number, peak around $n = 4$, and then decrease. Therefore, we expect that the rapidly decaying overtones beyond about $n = 7$ are subdominant; this expectation justifies truncating the expansion in the vicinity of $n = 7$. Preliminary studies indicate that $n = 8$ does not improve the fit at $t_0 = t_{\text{peak}}$. An additional caveat is that the amplitudes in Table I are those recovered from the $\ell = m = 2$ spherical harmonic as opposed to the $\ell = m = 2$ spheroidal harmonic. However, the spherical-spheroidal mixing is small (cf. Sec. III) and should not significantly change the qualitative behavior of the relative amplitudes in Table I.

Using our results from the last row of Table I, and using the analytic decay rates corresponding to the true M_f and χ_f, we can reconstruct the expected individual contributions of each overtone to the total $\ell = m = 2$ ringdown signal at any given t; in other words, we can compute the time-dependent amplitudes $A_{22n}(t)$ of each overtone. These are related to the A_n in Table I by $A_{22n}(t) = A_n e^{-(t-t_0)/\tau_{22n}}$. These amplitudes are shown in Fig. 8. This result establishes why one has to wait until $10M$–$20M$ after the peak before the fundamental becomes the dominant contribution.

Note that Fig. 8 uses a single fit over the range $t \geq t_{\text{peak}}$ and assumes the expected analytic time dependence of each overtone amplitude for $t \geq t_{\text{peak}}$. Alternatively, we can attempt to reconstruct each $A_{22n}(t)$ numerically by performing a different fit for the amplitudes at each time t. For each time t, we choose $t_0 = t_{\text{fit}} = t$, and we fit data only for times greater than or equal to t_0. The numerically extracted time dependence of the overtone amplitudes, $A_{22n}(t)$, is shown in Fig. 9. Obtaining an accurate fit in this way is difficult because of various numerical complications, such as the small differences in frequencies and amplitudes between neighboring overtones, the poor resolution of overtones with small amplitudes, and the risk of overfitting at late times after some overtones have decayed away. At later times, there is significantly less power in the highest overtones—making them more difficult to resolve. To mitigate some of these difficulties, when performing the fit at each time t, we exclude overtones whose fitted amplitude has increased relative to that at the previous time. This method is motivated by the fact that the model is one of exponentially damped sinusoids. Therefore, if at any
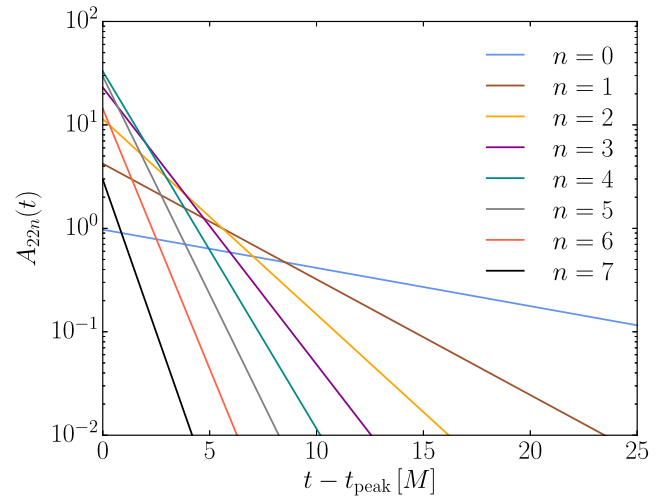

FIG. 8. A decomposition of the ringdown in terms of the overtones for $N = 7$. The evolution of the overtones is computed from the analytic decay rates with initial amplitudes at $t = t_0 = t_{\text{peak}}$ specified by the bottom row of Table I. Notice that the fundamental mode does not dominate the ringdown of SXS:BBH:0305 until roughly $10M$ after t_{peak}.
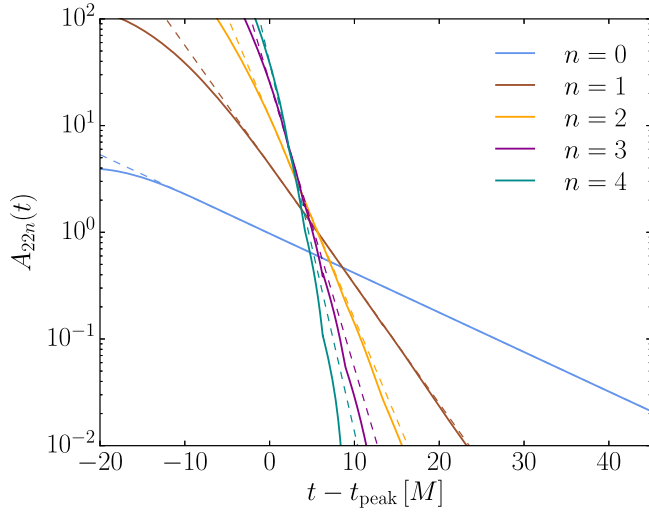
FIG. 9. The numerically recovered amplitudes for the fundamental QNM and the first few overtones at each fit time, t (solid curves). Dashed lines are the same as the lines in Fig. 8. The numerically extracted amplitudes across t agree very well with the expected decay for the longest-lived modes, while modes that decay more quickly are more susceptible to fitting issues. Interestingly, the fundamental mode is in excellent agreement with the expected decay rate at times preceding the peak amplitude of the strain.

time an overtone has a larger amplitude than the amplitude recovered at a previous time, we consider that overtone to no longer be of physical relevance, and we permanently remove it from the allowed set of modes for future fit times. It is always the highest overtone available in the remaining set of modes that gets dropped, as this mode decays more quickly than the other ones. Although we only show up to $N = 4$ in Fig. 9 because numerically extracting amplitudes is difficult at late times, the benefit of using overtones up to $N = 7$ in estimating the remnant mass and spin is apparent in Fig. 7. Consequently, more advanced fitting methods should allow for an improvement in numerically recovering higher-order overtones as a function of time, which will be explored further in future work.

Finally, it is worth pointing out that there is good agreement between the model and NR even at times before t_{peak}, as indicated by the mismatches in Fig. 1, as well as by the early agreement between the numerically extracted amplitude of the fundamental mode and the expected analytic behavior visible in Fig. 9. Since the QNMs are solutions to perturbed single BH spacetimes, the agreement could be interpreted as an indication that the region of the prepeak waveform already begins to behave as a perturbed single BH to observers at infinity. This observed behavior will be explored further in future work.

C. Observing overtones with GW detectors

Overtones can enhance the power of gravitational wave detectors to probe the ringdown regime. We illustrate this

by studying the simulated output of a LIGO-like detector in response to the same GW considered above, the NR simulation SXS:BBH:0305. For simplicity, we assume the orbital plane of the source faces the instrument head-on (no inclination). We choose a sky location for which the detector has optimal response to the plus polarization but none to cross, with polarizations defined in the same frame implicitly assumed in Eq. (1). To mimic GW150914, we rescale the NR template to correspond to a total initial binary mass of 72 $M_\odot$, in the detector frame, and a source distance of 400 Mpc. We inject the $\ell = m = 2$ mode of the signal into simulated Gaussian noise corresponding to the sensitivity of Advanced LIGO in its design configuration [58]. This method yields a postpeak optimal SNR of about 42 [59].

To extract information from the noisy data, we carry out a Bayesian analysis similar to that in Refs. [16,60] but based on the overtone ringdown model of Eq. (1), with $\ell = m = 2$ and varying N. For any given start time t_0, we obtain a posterior probability density over the space of remnant mass and spin, as well as the amplitudes and phases of the set of QNMs included in the template. We parametrize start times via $\Delta t_0 = t_0 - t_{h\text{-peak}}$, where $t_{h\text{-peak}}$ refers to the signal peak at the detector ($t_{h\text{-peak}} \approx t_{\text{peak}} - 0.48$ ms $\approx t_{\text{peak}} - 1.3M$). Unlike Ref. [16], we sample over the amplitudes and phases directly, instead of marginalizing over them analytically, and we place uniform priors on all parameters. In particular, we consider masses and orbit-aligned spins within $[10, 100]M_\odot$ and $[0, 1]$, respectively. We allow the QNM phases to cover their full range, $[0, 2\pi]$, but restrict the amplitudes (measured at $t = t_{h\text{-peak}}$) to $[0.01, 250]h_{\text{peak}}$, where $h_{\text{peak}} = 2 \times 10^{-21}$ is the total signal peak. This arbitrary amplitude interval fully supports the posterior in all cases we consider. We assume all extrinsic parameters, like sky location and inclination, are perfectly known. We sample posteriors using the Markov chain Monte Carlo (MCMC) implementations in KOMBINE [61] and, for verification, EMCEE [62].

The highest N we consider in our inference model is $N = 3$, as that is the most we can hope to resolve given the SNR of our simulation. A guiding principle for two waveforms to be indistinguishable is $\mathcal{M} < \text{SNR}^{-2}/2$, in terms of the mismatch $\mathcal{M}$ defined in Eq. (2) but with a noise-weighted inner product [63–65]. For the system at hand, this principle implies that postmerger templates with mismatches $\mathcal{M} \lesssim 3 \times 10^{-4}$ are effectively identical. If fitting from the peak on, Fig. 1 then implies that differences between $N \geq 3$ templates are unmeasurable. We confirmed this empirically by checking that $N = 4$ does not lead to inference improvements with respect to $N = 3$ and only seems to introduce degenerate parameters. By the same token, we have also verified that, at this SNR, our results are largely unaffected by the presence or absence of the next dominant angular mode (3,2) in the injected NR waveform, as its amplitude is an order of magnitude weaker than that of the dominant (2,2) mode for the chosen system.
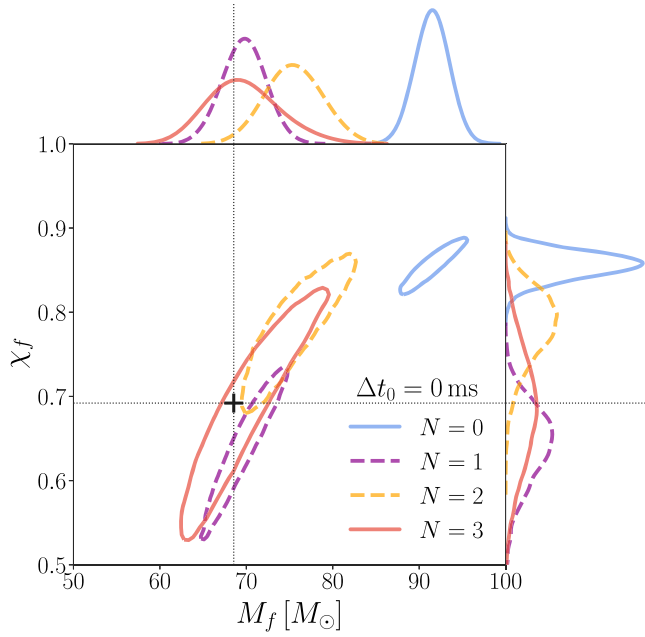
FIG. 10.   Remnant parameters inferred with different numbers of overtones, using data starting at the peak amplitude of the injected strain. Contours represent 90%-credible regions on the remnant mass (M_f) and dimensionless spin (χ_f), obtained from the Bayesian analysis of a GW150914-like NR signal injected into simulated noise for a single Advanced LIGO detector at design sensitivity. The inference model was as in Eq. (1), with ($\ell = m = 2$) and different numbers of overtones N: 0 (solid blue), 1 (dashed purple), 2 (dashed yellow), and 3 (solid red). In all cases, the analysis uses data starting at peak strain ($\Delta t_0 = t_0 - t_{h\text{-peak}} = 0$). The top and right panels show 1D posteriors for M_f and χ_f, respectively. Amplitudes and phases are marginalized over. The intersection of the dotted lines marks the true value ($M_f = 68.5\ M_\odot$, $\chi_f = 0.69$).
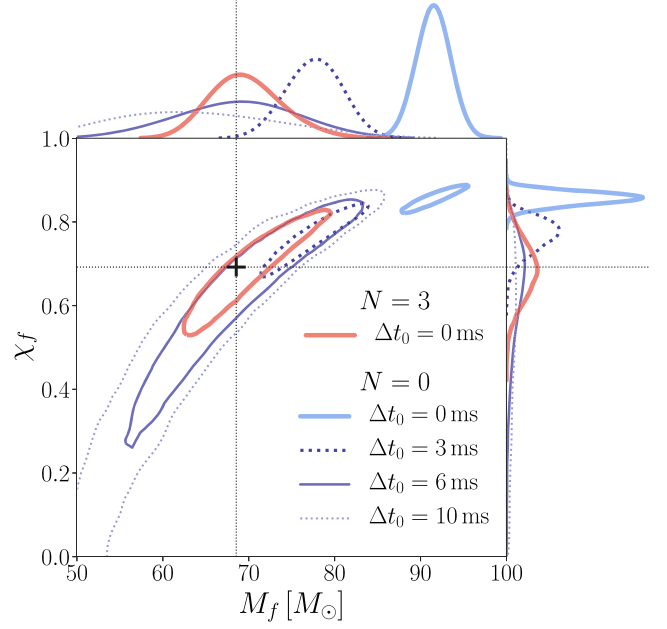
FIG. 11.   Remnant parameters inferred using only the fundamental mode, using data starting at different times relative to the peak amplitude of the injected strain. Contours represent 90%-credible regions on the remnant mass (M_f) and dimensionless spin (χ_f), obtained from the Bayesian analysis of a GW150914-like NR signal injected into simulated noise for a single Advanced LIGO detector at design sensitivity. For the blue contours, the inference model included no overtones ($N = 0$) and used data starting at different times after the peak: $\Delta t_0 = t_0 - t_{h\text{-peak}} \in [0, 3, 6, 10]$ ms (blue contours). For the red contour, the analysis was conducted with three overtones ($N = 3$) starting at the peak ($\Delta t_0 = 0$), as in Fig. 10. The top and right panels show 1D posteriors for M_f and χ_f, respectively. Amplitudes and phases are marginalized over. The intersection of the dotted lines marks the true value ($M_f = 68.5\ M_\odot$, $\chi_f = 0.69$).

At higher SNRs, additional (2,2) overtones and/or angular modes (potentially, with their respective overtones) are necessary to keep the modeling error below the statistical error.

Our findings are summarized in Figs. 10 and 11. In Fig. 10, we show the posteriors recovered for the remnant mass and spin under the assumption that the ringdown begins at the peak of the signal strain and for models with different numbers of overtones. For each case, the main panel displays contours enclosing 90% of the posterior probability, while the curves on the top and right represent the corresponding marginalized distributions for the mass and spin. As expected, the fundamental mode ($N = 0$) is insufficient to describe the signal near the peak, yielding an estimate of the remnant properties that is far from the true values determined from the NR simulation (dotted lines). As the number of overtones is increased, the inferred mass and spin become increasingly more accurate, with $N = 3$ producing the best results (true value within top 40%-credible region). This result illustrates how the overtones

can provide an independent measurement of the remnant properties by studying the signal near the peak.

We find that the estimate of the mass and spin obtained with overtones at the peak is more accurate than the one obtained with only the fundamental mode at later times. We illustrate this in Fig. 11, which shows the 90%-credible regions on M_f and χ_f inferred using only the fundamental mode ($N = 0$) at different times after the peak strain (blue contours), as well as the $N = 3$ result from Fig. 10 for comparison (red contour). As anticipated in Ref. [16], the fundamental mode is a faithful representation of the signal only at later times, which in our case means that the true values are enclosed in the 90%-credible region only for $\Delta t_0 \geq 5$ ms. The penalty for analyzing the signal at later times is a reduction in SNR that results in increased uncertainty, as evidenced by the large area of the blue contours in Fig. 11. We obtain a more precise estimate by taking advantage of the overtones at the peak. We suspect that the observed agreement at 3 ms in Ref. [16] is a

consequence of the lower SNR of GW150914. At lower SNRs, the statistical errors outweigh the systematic errors associated with including only the fundamental mode.

V. DISCUSSION AND CONCLUSIONS

For a given mass M_f and spin χ_f, perturbation theory precisely predicts the spectrum of QNMs associated with a ringing single BH, including the characteristic frequencies for these QNMs. The QNM frequencies are denoted $\omega_{\ell mn}(M_f, \chi_f)$, where ℓ and m describe the angular dependence of a mode and n, the often-ignored integer overtone index, sorts QNMs with the same angular dependence by how quickly they decay. The slowest decaying fundamental mode, $n = 0$, is often considered to be of primary importance, while the more quickly decaying overtones are often disregarded. However, we find that the overtones are not necessarily subdominant as is often assumed, but instead, they can dominate the early part of the ringdown.

Using a superposition of QNMs, we model the ringdown portion of the $\ell = m = 2$ mode of the numerical relativity waveform SXS:BBH:0305, which is consistent with GW150914. We find that with enough included overtones, the QNMs provide an excellent description for the GW strain for all times beyond the peak amplitude of the complex strain h. For the GW150914-like NR waveform we analyzed, the overtones dominate the early part of the perturbations but decay away much more quickly than the fundamental mode, which eventually becomes dominant roughly $10M$ after the peak amplitude (Fig. 8). This later time where the fundamental dominates is sometimes referred to in the literature as the start of the ringdown, the time of a transition to the linear regime, or the beginning of the domain of applicability of perturbation theory. However, this time is merely the time at which one may ignore the contribution of overtones, which play a key role in the early ringdown. Including the QNM overtones extends the reach of perturbation theory back to the time of the peak strain amplitude, indicating that the linear ringdown regime begins much earlier than one would conclude by ignoring these additional modes. As mentioned in Sec. IV A, we have verified, on a sizable set of aligned-spin waveforms in the SXS catalog, that the inclusion of overtones provides an accurate model for the postpeak strain. Not only do the overtones provide excellent mismatches, but the best-fit mass and spin are accurately recovered with median absolute errors in M_f/M and χ_f of about 10^{-3}. We therefore expect the early dominance of overtones to be a generic feature of the ringdown.

The QNM overtones can enhance the power of GW detectors to probe the ringdown regime. They can be used to extract information about QNMs at the peak of the signal, where the SNR is high. In contrast, the usual approach relies solely on the later portion of the signal that is dominated by the (initially weaker) fundamental mode, paying the price of larger statistical errors and uncertainty in the appropriate time where this mode dominates [16–22,25–29]. This effect is visible in Fig. 11, where a model with $N = 3$ overtones remains faithful to the true remnant mass and spin with less uncertainty than one with $N = 0$ at later times. The resolvability of these overtones provides a set of independent modes, each with unique frequencies, that can potentially be used to constrain deviations from GR.

Studies of the ringdown GW spectrum can provide a direct way to experimentally determine whether compact binary coalescences result in the Kerr BHs predicted by GR [23,24]. This includes tests of the no-hair theorem and the area law, as well as searches for BH mimickers. The program, sometimes known as "black hole spectroscopy," generally requires independent measurement of at least two modes, which are conventionally taken to be the fundamentals of two different angular harmonics (e.g., Refs. [25,29]). However, such a choice is only available for systems that present a sufficiently strong secondary angular mode, which only tends to occur under some specific conditions (e.g., for high mass ratios) [66–70]. Furthermore, as we have observed, these fundamental modes should dominate only at late times, being subject to significantly more noise than modes that can be extracted near the peak of the waveform. The extraction of an overtone, in addition to the fundamental mode, could potentially serve as an alternative two-mode test of the no-hair theorem.

The impact of overtones on ringdown tests of GR can already be seen from Fig. 11: By studying the QNMs at early and late times, we may obtain two independent measurements of the remnant parameters, enabling powerful consistency checks. Unlike tests that rely on a multiplicity of angular modes, studies of overtones should be feasible at SNRs achievable with existing detectors, as we demonstrate by our study of a GW150914-like signal seen at design sensitivity by Advanced LIGO (Sec. IV C). For signals in which they are measurable, higher angular modes and their overtones could make these tests even more powerful. Overtones can therefore enable a whole new set of precision studies of the ringdown and make black hole spectroscopy realizable with current detectors.

[1] W. Israel, *Event Horizons in Static Vacuum Space-Times*, Phys. Rev. **164**, 1776 (1967).

[2] B. Carter, *Axisymmetric Black Hole Has Only Two Degrees of Freedom*, Phys. Rev. Lett. **26**, 331 (1971).

[3] S. W. Hawking, *Black Holes inGeneral Relativity*, Commun. Math. Phys. **25**, 152 (1972).

[4] S. A. Teukolsky, *Rotating Black Holes—Separable Wave Equations for Gravitational and Electromagnetic Perturbations*, Phys. Rev. Lett. **29**, 1114 (1972).

[5] S. A. Teukolsky, *Perturbations of a Rotating Black Hole. I. Fundamental Equations for Gravitational, Electromagnetic, and Neutrino-Field Perturbations*, Astrophys. J. **185**, 635 (1973).

[6] W. H. Press and S. A. Teukolsky, *Perturbations of a Rotating Black Hole. II. Dynamical Stability of the Kerr Metric*, Astrophys. J. **185**, 649 (1973).

[7] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), *Observation of Gravitational Waves from a Binary Black Hole Merger*, Phys. Rev. Lett. **116**, 061102 (2016).

[8] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), *GW151226: Observation of Gravitational Waves from a 22-Solar-Mass Binary Black Hole Coalescence*, Phys. Rev. Lett. **116**, 241103 (2016).

[9] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), *Binary Black Hole Mergers in the First Advanced LIGO Observing Run*, Phys. Rev. X **6**, 041015 (2016).

[10] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), *GW170104: Observation of a 50-Solar-Mass Binary Black Hole Coalescence at Redshift 0.2*, Phys. Rev. Lett. **118**, 221101 (2017).

[11] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), *GW170608: Observation of a 19 Solar-Mass Binary Black Hole Coalescence*, Astrophys. J. **851**, L35 (2017).

[12] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), *GW170814: A Three-Detector Observation of Gravitational Waves from a Binary Black Hole Coalescence*, Phys. Rev. Lett. **119**, 141101 (2017).

[13] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaboration), *GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo During the First and Second Observing Runs*, Phys. Rev. X **9**, 031040 (2019).

[14] J. Aasi *et al.* (LIGO Scientific), *Advanced LIGO*, Classical Quantum Gravity **32**, 115012 (2015).

[15] F. Acernese *et al.*, *Advanced Virgo: A Second-Generation Interferometric Gravitational Wave Detector*, Classical Quantum Gravity **32**, 024001 (2015).

[16] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), *Tests of General Relativity with GW150914*, Phys. Rev. Lett. **116**, 221101 (2016).

[17] W. Del Pozzo and A. Nagar, *Analytic Family of Post-Merger Template Waveforms*, Phys. Rev. D **95**, 124034 (2017).

[18] M. Cabero, C. D. Capano, O. Fischer-Birnholtz, B. Krishnan, A. B. Nielsen, A. H. Nitz, and C. M. Biwer, *Observational Tests of the Black Hole Area Increase Law*, Phys. Rev. D **97**, 124069 (2018).

[19] E. Thrane, P. D. Lasky, and Y. Levin, *Challenges Testing the No-Hair Theorem with Gravitational Waves*, Phys. Rev. D **96**, 102004 (2017).

[20] R. Brito, A. Buonanno, and V. Raymond, *Black-Hole Spectroscopy by Making Full Use of Gravitational-Wave Modeling*, Phys. Rev. D **98**, 084038 (2018).

[21] G. Carullo *et al.*, *Empirical Tests of the Black Hole No-Hair Conjecture Using Gravitational-Wave Observations*, Phys. Rev. D **98**, 104020 (2018).

[22] G. Carullo, W. Del Pozzo, and J. Veitch, *Observational Black Hole Spectroscopy: A Time-Domain Multimode Analysis of GW150914*, Phys. Rev. D **99**, 123029 (2019).

[23] O. Dreyer, B. J. 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).

[24] E. Berti, V. Cardoso, and C. M. Will, *On Gravitational-Wave Spectroscopy of Massive Black Holes with the Space Interferometer LISA*, Phys. Rev. D **73**, 064030 (2006).

[25] S. Gossan, J. Veitch, and B. S. Sathyaprakash, *Bayesian Model Selection for Testing the No-Hair Theorem with Black Hole Ringdowns*, Phys. Rev. D **85**, 124056 (2012).

[26] J. Meidam, M. Agathos, C. Van Den Broeck, J. Veitch, and B. S. Sathyaprakash, *Testing the No-Hair Theorem with Black Hole Ringdowns Using TIGER*, Phys. Rev. D **90**, 064009 (2014).

[27] E. Berti *et al.*, *Testing General Relativity with Present and Future Astrophysical Observations*, Classical Quantum Gravity **32**, 243001 (2015).

[28] E. Berti, A. Sesana, E. Barausse, V. Cardoso, and K. Belczynski, *Spectroscopy of Kerr black holes with Earth- and Space-Based Interferometers*, Phys. Rev. Lett. **117**, 101102 (2016).

[29] V. Baibhav and E. Berti, *Multimode Black Hole Spectroscopy*, Phys. Rev. D **99**, 024005 (2019).

[30] 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).

[31] V. Baibhav, E. Berti, V. Cardoso, and G. Khanna, *Black Hole Spectroscopy: Systematic Errors and Ringdown Energy Estimates*, Phys. Rev. D **97**, 044048 (2018).

[32] I. Kamaretsos, M. Hannam, S. Husa, and B. S. Sathyaprakash, *Black-Hole Hair Loss: Learning about Binary Progenitors from Ringdown Signals*, Phys. Rev. D **85**, 024018 (2012).

[33] L. London, D. Shoemaker, and J. Healy, *Modeling Ringdown: Beyond the Fundamental Quasinormal Modes*, Phys. Rev. D **90**, 124032 (2014); *Erratum*, Phys. Rev. D **94**, 069902(E) (2016).

[34] 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).

[35] Y. Pan, A. Buonanno, A. Taracchini, L. E. Kidder, A. H. Mroue, H. P. Pfeiffer, M. A. Scheel, and B. Szilagyi, *Inspiral-Merger-Ringdown Waveforms of Spinning, Precessing Black-Hole Binaries in the Effective-One-Body Formalism*, Phys. Rev. D **89**, 084006 (2014).

[36] A. Taracchini, A. Buonanno, Y. Pan, T. Hinderer, M. Boyle, D. A. Hemberger, L. E. Kidder, G. Lovelace, A. H. Mroue, H. P. Pfeiffer, M. A. Scheel, B. Szilagyi, N. W. Taylor, and A. Zenginoglu, *Effective-One-Body Model for Black-Hole Binaries with Generic Mass Ratios and Spins*, Phys. Rev. D **89**, 061502(R) (2014).

[37] S. Babak, A. Taracchini, and A. Buonanno, *Validating the Effective-One-Body Model of Spinning, Precessing Binary Black Holes Against Numerical Relativity*, Phys. Rev. D **95**, 024010 (2017).

[38] A recent extension of EOB, referred to as parameterized EOBNR [20], was designed for future tests of the no-hair theorem by measuring the frequencies of the $\ell = m = 2$ and $\ell = m = 3$ fundamental modes. Restricted to nonspinning binaries, pEOBNR models the full inspiral and merger with an attached ringdown model (including overtones), in order to avoid deciding at what time the QNMs alone provide an accurate description of the waveform.

[39] See http://www.black-holes.org/waveforms.

[40] A. H. Mroue, M. A. Scheel, B. Szilagyi, H. P. Pfeiffer, M. Boyle, D. A. Hemberger, L. E. Kidder, G. Lovelace, S. Ossokine, N. W. Taylor, A. Zenginoglu, L. T. Buchman, T. Chu, E. Foley, M. Giesler, R. Owen, and S. A. Teukolsky, *A Catalog of 174 Binary Black Hole Simulations for Gravitational Wave Astronomy*, Phys. Rev. Lett. **111**, 241104 (2013).

[41] C. V. Vishveshwara, *Stability of the Schwarzschild Metric*, Phys. Rev. D **1**, 2870 (1970).

[42] W. Press, *Long Wave Trains of Gravitational Waves from a Vibrating Black Hole*, Astrophys. J. Lett. **170**, L105 (1971).

[43] S. A. Teukolsky, *Perturbations of a Rotating Black Hole. 1. Fundamental Equations for Gravitational Electromagnetic and Neutrino Field Perturbations*, Astrophys. J. **185**, 635 (1973).

[44] S. Chandrasekhar and S. Detweiler, *The Quasi-normal Modes of the Schwarzschild Black Hole*, Proc. R. Soc. A **344**, 441 (1975).

[45] E. Berti, V. Cardoso, and A. O. Starinets, *Topical Review: Quasinormal Modes of Black Holes and Black Branes*, Classical Quantum Gravity **26**, 163001 (2009).

[46] See http://pages.jh.edu/~eberti2/ringdown.

[47] We have verified the presence and early dominance of overtones in other resolvable (ℓ, m)'s in the NR waveform.

[48] E. T. Newman and R. Penrose, *Note on the Bondi–Metzner–Sachs Group*, J. Math. Phys. (N.Y.) **7**, 863 (1966).

[49] J. N. Goldberg, A. J. Macfarlane, E. T. Newman, F. Rohrlich, and E. C. G. Sudarshan, *Spin-s Spherical Harmonics and ð*, J. Math. Phys. (N.Y.) **8**, 2155 (1967).

[50] K. S. Thorne, *Multipole Expansions of Gravitational Radiation*, Rev. Mod. Phys. **52**, 299 (1980).

[51] E. Berti and A. Klein, *Mixing of Spherical and Spheroidal Modes in Perturbed Kerr Black Holes*, Phys. Rev. D **90**, 064012 (2014).

[52] V. Varma, S. E. Field, M. A. Scheel, J. Blackman, L. E. Kidder, and H. P. Pfeiffer, *Surrogate Model of Hybridized Numerical Relativity Binary Black Hole Waveforms*, Phys. Rev. D **99**, 064045 (2019).

[53] E. N. Dorband, E. Berti, P. Diener, E. Schnetter, and M. Tiglio, *A Numerical Study of the Quasinormal Mode Excitation of Kerr Black Holes*, Phys. Rev. D **74**, 084028 (2006).

[54] E. W. Leaver, *An Analytic Representation for the Quasi Normal Modes of Kerr Black Holes*, Proc. R. Soc. A **402**, 285 (1985).

[55] E. W. Leaver, *Spectral Decomposition of the Perturbation Response of the Schwarzschild Geometry*, Phys. Rev. D **34**, 384 (1986).

[56] E. Berti and V. Cardoso, *Quasinormal Ringing of Kerr Black Holes. I. The Excitation Factors*, Phys. Rev. D **74**, 104020 (2006).

[57] Z. Zhang, E. Berti, and V. Cardoso, *Quasinormal Ringing of Kerr Black Holes. II. Excitation by Particles Falling Radially with Arbitrary Energy*, Phys. Rev. D **88**, 044018 (2013).

[58] D. H. Shoemaker *et al.*, Advanced LIGO Anticipated Sensitivity Curves, Tech. Rep. LIGO-T0900288, LIGO Laboratory, 2009.

[59] Defined as the SNR in frequencies above 154.68 Hz, the instantaneous frequency at the peak of the time-domain signal.

[60] R. Prix, Bayesian QNM Search on GW150914, Tech. Rep. LIGO-T1500618, LIGO Scientific Collaboration, 2016.

[61] B. Farr and W. M. Farr, *Kombine: A Kernel-Density-Based, Embarrassingly Parallel Ensemble Sampler* (to be published).

[62] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, *emcee: The MCMC Hammer*, Publ. Astron. Soc. Pac. **125**, 306 (2013).

[63] E. E. Flanagan and S. A. Hughes, *Measuring Gravitational Waves from Binary Black Hole Coalescences: 2. The Waves' Information and Its Extraction, with and without Templates*, Phys. Rev. D **57**, 4566 (1998).

[64] L. Lindblom, B. J. Owen, and D. A. Brown, *Model Waveform Accuracy Standards for Gravitational Wave Data Analysis*, Phys. Rev. D **78**, 124020 (2008).

[65] S. T. McWilliams, B. J. Kelly, and J. G. Baker, *Observing Mergers of Non-spinning Black-Hole Binaries*, Phys. Rev. D **82**, 024014 (2010).

[66] V. Varma and P. Ajith, *Effects of Nonquadrupole Modes in the Detection and Parameter Estimation of Black Hole Binaries with Nonprecessing Spins*, Phys. Rev. D **96**, 124024 (2017).

[67] C. Capano, Y. Pan, and A. Buonanno, *Impact of Higher Harmonics in Searching for Gravitational Waves from Nonspinning Binary Black Holes*, Phys. Rev. D **89,** 102003 (2014).

[68] T. B. Littenberg, J. G. Baker, A. Buonanno, and B. J. Kelly, *Systematic Biases in Parameter Estimation of Binary Black-Hole Mergers*, Phys. Rev. D **87**, 104003 (2013).

[69] J. C. Bustillo, P. Laguna, and D. Shoemaker, *Detectability of Gravitational Waves from Binary Black Holes: Impact of Precession and Higher Modes*, Phys. Rev. D **95**, 104038 (2017).

[70] V. Varma, P. Ajith, S. Husa, J. C. Bustillo, M. Hannam, and M. Pürrer, *Gravitational-Wave Observations of Binary Black Holes: Effect of Nonquadrupole Modes*, Phys. Rev. D **90**, 124004 (2014).