Statistical identification of ringdown modes with rational filters

Neil Lu [ID],[1, *] Sizheng Ma [ID],[2, †] Ornella J. Piccinni [ID],[1, ‡] Ling Sun [ID],[1, §] and Eliot Finch [ID][3]

[1]*OzGrav-ANU, Centre for Gravitational Astrophysics, College of
Science, The Australian National University, ACT 2601, Australia*
[2]*Perimeter Institute for Theoretical Physics, Waterloo, ON N2L2Y5, Canada*
[3]*TAPIR, California Institute of Technology, Pasadena, CA 91125, USA*
(Dated: May 24, 2025)

Measuring quasinormal modes (QNMs) during the ringdown phase of binary black hole coalescences provides key insights into merger dynamics and enables tests of the no-hair theorem. The QNM rational filter has recently been introduced as a technique to identify specific QNMs in ringdown signals without sampling over mode amplitudes and phases. In this work, we extend the QNM rational filter framework to quantify the statistical confidence of subdominant mode detections in real gravitational wave (GW) observations. We employ a frequentist approach to estimate false-alarm probabilities and propose a workflow for robust identification of specific QNMs. We first validate our methodology using synthetic signals generated from numerical relativity waveforms. We then reanalyze the first GW event, GW150914, finding a marginal detection of an overtone, but at time when the applicability of constant amplitude QNM fits is not fully understood. This extended methodology provides a systematic approach to improving the reliability of QNM detections, paving the way for more precise tests of strong-field gravity with current and future GW observations.

I. INTRODUCTION

The ringdown phase of a binary black hole (BH) coalescence corresponds to the relaxation of the deformed remnant BH as it settles into an equilibrium state. At sufficiently late times, the dynamics are expected to be linear, with the emitted gravitational waves (GWs) well described by a superposition of quasinormal modes (QNMs), each contributing as a damped sinusoid to the signal [1–6]. Due to the no-hair theorem in general relativity (GR) [7–11], the complex frequencies of these QNMs are uniquely determined by the mass and spin of the remnant BH. Independently measuring multiple QNMs and inferring the corresponding mass and spin — a process known as BH spectroscopy — provides a means to test the no-hair theorem [12–14] and the BH area law [15]. This is made possible by recent detections of binary BH coalescences with ground-based detectors, like LIGO [16], Virgo [17], and KAGRA [18], which allow for the possibility to measure multiple QNMs from individual events.

At present, various methods are used to extract QNMs from observational data, e.g., time-domain [14, 19–23], a wavelet-based frequency-domain [24, 25], using "gating and in-painting" [26, 27], parameterized effective one-body [28–31] and machine-learning based techniques [32]. Different methods are used to avoid spectral leakage while processing short ringdown signals under acyclic boundary conditions. They have been applied to GW events such as GW150914 [19, 25, 27, 33–41], GW190521 [26, 42–44], and a selected catalog of GW events [36, 45]. While individual pipelines claim the detection of an overtone

for GW150914 [14] or a higher-order mode for GW190521 [26, 43], there is a lack of universal consensus about these detections, see particularly [43, 46–48].

The QNM rational filter has lately been proposed as a new method to identify QNMs through mode cleaning [21, 37, 38]. The filter removes particular QNM(s) from a ringdown signal and depends only on the mass and spin of the remnant BH (by assuming that GR is correct). Crucially, the filter does not sample over the QNM amplitude(s) and phase(s). Under certain prior assumptions [38], the filter is related to the time-domain analysis in [14, 19, 20] with maximum likelihood estimation for the mode amplitudes. Previous studies with QNM rational filters have developed a semi-Bayesian framework by formulating a novel likelihood function and have tentatively applied the framework to GW150914 in a search for the first overtone [37, 38]. However, the statistical properties and robustness of the filter in the low signal-to-noise ratio (SNR) regime remain unexplored.

In this paper, we extend our previous studies [21, 37, 38] to quantify the statistical significance of QNM observations made with the rational filter. In particular, we focus on the detection of subdominant ringdown modes with low SNRs. We take a frequentist approach to estimate false-alarm probabilities (FAPs) caused by background noise. Building on this, we formulate a workflow to robustly identify QNMs in ringdown signals and apply the workflow to simulations using numerical-relativity (NR) waveforms as well as the real event, GW150914.

The structure of the paper is organized as follows. In Sec. II, we review the background of QNMs, the rational filter, and define a detection statistic. In Sec. III, we quantify the significance of the observation of a QNM via a FAP and describe ways to incorporate information obtained from inspiral-merger-ringdown (IMR) analysis that can increase the sensitivity to subdominant modes. In Sec. IV, we demonstrate a procedure of estimating

* neil.lu@anu.edu.au
† sma2@perimeterinstitute.ca
‡ ornellajuliana.piccinni@anu.edu.au
§ ling.sun@anu.edu.au

ringdown SNR using a relationship between the detection statistic and SNR. In Sec. V, we formulate a full workflow for analyzing GW events to identify the QNM contents. In Sec. VI, we apply the workflow to NR simulations and demonstrate the expected modes can be successfully identified following such a procedure. We re-analyze GW150914 in Sec. VII to quantify the statistical significance of the first overtone. Finally, we conclude and discuss future work in Sec. VIII.

II. QNM RATIONAL FILTER

The real-valued time-series of the (ℓ, m, n) QNM in a GW detector is given by:

$$h(t) = A_{\ell mn} e^{-(t-t_0)/\tau_{\ell mn}} \times \cos\left[2\pi f_{\ell mn}(t - t_0) + \phi_{\ell mn}\right], \tag{1}$$

where t_0 is a given reference time, $f_{\ell mn}$ is the oscillation frequency and $\tau_{\ell mn}$ is the damping time. In GR, $(f_{\ell mn}, \tau_{\ell mn})$ are uniquely determined by the mass and spin of the remnant Kerr BH, whereas the amplitude $A_{\ell mn}$ and phase $\phi_{\ell mn}$ encode information about the merger dynamics.

Throughout this work, we use three indices (ℓ, m, n) to label QNMs[1], adopting the convention $f_{\ell mn} > 0$. In Eq. (1) and throughout this paper, we consider only the prograde modes[2] and neglect retrograde modes, which implies $m > 0$ (although see [50] for notations). To avoid word clutter, we refer to modes by their mode numbers, i.e. the 220 mode refers to the $(\ell = 2, m = 2, n = 0)$ mode. We denote time t in units of the final remnant BH mass $M_{\rm rem}$.

The QNM rational filter removes QNM in Eq. (1) by multiplying the signal with the frequency-domain filter [21, 37, 38]:

$$\mathcal{F}_{\ell mn}(\omega; M_f, \chi_f) = \frac{\omega - \omega_{\ell mn}}{\omega - \omega_{\ell mn}^*} \frac{\omega + \omega_{\ell mn}^*}{\omega + \omega_{\ell mn}}, \tag{2}$$

where ω is the real-valued frequency, $\omega_{\ell mn} = 2\pi f_{\ell mn} - i/\tau_{\ell mn}$ is the complex QNM frequency, M_f and χ_f, are a given remnant BH mass and spin, and $*$ represents the complex conjugate. A linear superposition of multiple QNMs can be simultaneously removed using a total filter:

$$\mathcal{F}_{\rm tot}(\omega; M_f, \chi_f) = \prod_{\ell mn} \mathcal{F}_{\ell mn}(\omega; M_f, \chi_f). \tag{3}$$

Because the filter is applied to the data in the frequency domain, we transform the GW time-series data, d, into

the frequency domain (denoted by $\tilde{d}$) using a discrete Fourier transform:

$$\tilde{d}_k = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} d_n e^{-i2\pi \frac{k}{N} n}, \tag{4}$$

where n is the index of the time-series, N is the total number of data points, and we use the orthogonal normalization convention. After applying the filter in the frequency domain, we convert the filtered data back to the time domain using an inverse Fourier transform:

$$d_n^F = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \mathcal{F}_{\ell mn}\left(2\pi \frac{k}{N}\right) \tilde{d}_k e^{i2\pi \frac{k}{N} n}, \tag{5}$$

where the superscript "F" denotes the filtered data. The rational filter completely removes the mode(s) from the time-series if the correct BH parameters are used to construct the filter.

In Refs. [37, 38], analysis using the rational filter was conducted in a semi-Bayesian framework using a likelihood function:

$$\ln \mathcal{L}(d_t | M_f, \chi_f, \mathcal{F}_{\rm tot}, \Delta t_0) = -\frac{1}{2} \sum_{i,j>t_0} d_i^F C_{ij}^{-1} d_j^F, \tag{6}$$

where $\Delta t_0 = t_{\rm analysis} - t_0$ with $t_{\rm analysis}$ as the starting time of the analysis and we take t_0 to be the merger time inferred by the IMR analysis of an event. C_{ij} is the autocovariance matrix of the noise estimated using the Welch method [51], as is calculated in other ringdown pipelines [19, 20, 22]. Note, especially, that the filter does not depend on the QNM amplitude(s) and phase(s), so it does sample over them but also cannot infer their values.

We define a hypothesis $\mathcal{H}$ as the set of ringdown modes present in a signal at an analysis starting time. The evidence for a hypothesis is given by marginalizing the likelihood over the parameters M_f and χ_f:

$$\mathcal{Z}(d|\mathcal{H}) = \sum_{M_f, \chi_f} \mathcal{L}(d|M_f, \chi_f, \mathcal{F}_{\rm tot}, \Delta t_0)\Pi(M_f, \chi_f)\Delta M_f \Delta \chi_f, \tag{7}$$

where $\Pi(M_f, \chi_f)$ is the prior for M_f and χ_f, and we have evaluated the integral using a sum with a resolution of $(\Delta M_f, \Delta \chi_f)$. Throughout this work, we use a uniform prior over mass and spin. Given an observation, we define a detection statistic ($\mathcal{D}$) used for hypothesis testing to quantify how well an observation prefers a hypothesis $\mathcal{H}$ over an alternative $\mathcal{H}'$:

$$\mathcal{D}(\mathcal{H} : \mathcal{H}') = \log_{10} \frac{\mathcal{Z}(d|\mathcal{H})}{\mathcal{Z}(d|\mathcal{H}')}. \tag{8}$$

Here $\mathcal{D}$ is analogous to a log Bayes factor but formally differs from the Bayes factors computed by other time-domain ringdown analyses (for a further details see Appendix A). We label a hypothesis by its QNM mode

---

[1] QNMs are typically labeled by four indices; see the Appendix of [49] for a brief summary.
[2] The strain in Eq. (1) contains two prograde modes, emitting towards both North and South directions with respect to the BH spin axis.

content, e.g. $\{220, 221\}$. For a hypothesis, the joint posterior quantile, i.e. the credible contour of which a given set of mass and spin (M_0, χ_0) falls on, can be computed by integrating the likelihoods returned by the rational filter:

$$p(M_0, \chi_0) = \frac{\sum_{\mathcal{L}(d|M_f, \chi_f) > \mathcal{L}(d|M_0, \chi_0)} \mathcal{L}(d|M_f, \chi_f)}{\sum_{M_f, \chi_f} \mathcal{L}(d|M_f, \chi_f)} . \quad (9)$$

Where we have omitted the $\mathcal{F}_{\text{tot}}$ and Δt_0 parameters in the likelihood for compactness and a lower $p(M_0, \chi_0)$ value indicates a better match between (M_0, χ_0) and the estimated (M_f, χ_f).

Finally, the matched-filter SNR of an observation is:

$$\text{SNR} = \frac{\langle h_t | d_t \rangle}{\sqrt{\langle h_t | h_t \rangle}} , \quad (10)$$

with

$$\langle h_t | d_t \rangle = \sum_{i, j > t_0} h_i C_{ij}^{-1} d_j . \quad (11)$$

Note that the matched-filter SNR differs from the optimal SNR, which assumes no additive noise in the data and is given by $\sqrt{\langle h_t | h_t \rangle}$. Throughout the paper, we use SNR to denote the matched filter SNR of a signal for conciseness.

III. FALSE-ALARM PROBABILITY AND $\mathcal{D}$ THRESHOLD

For real GW events, detector noise can mimic ringdown signals and lead to a positive $\mathcal{D}$ when fitting with additional ringdown modes. In the low SNR regime, where the $\mathcal{D}$ for subdominant modes are not large, it can be difficult to determine whether a positive $\mathcal{D}$ arises due to the presence of signal or because of noise fluctuations. We, therefore, quantify the probability that a given $\mathcal{D}$ occurs due to noise using a frequentist approach, empirically determining the distribution of $\mathcal{D}$ that arise from noise. In Sec. III A, we use a toy model to describe the procedure for detecting a single ringdown mode. In Sec. III B, we discuss a more practical scenario about the detection of a secondary ringdown mode after the dominant mode has already been identified. In Sec. III C, we discuss modified procedures that can improve the detectability of modes by using knowledge of the remnant BH mass and spin obtained from the IMR analysis.

A. Single mode detection

We first consider a simplified scenario and show how we would claim the detection of a single ringdown mode. While it is not a real challenge to detect a dominant ringdown mode in GW observations, we use this relatively simple case to illustrate the procedure of computing the FAP for a single-mode detection. We use data containing

only pure noise and compute the $\mathcal{D}$ between $\mathcal{H}$, a $\{220\}$ mode hypothesis, and $\mathcal{H}'$, the null hypothesis (no filter applied). I.e., we compute $\mathcal{D}$ between an overfiltered hypothesis and the correct hypothesis. By doing so, we empirically find the distribution of $\mathcal{D}$, which, if positive, falsely prefers the existence of a 220 mode. We compute $\mathcal{D}$ over the prior space described in Table I. The likelihood function depends only on the remnant mass and spin, and we compute it over the entire parameter space without relying on Markov Chain Monte Carlo (MCMC) or nested sampling algorithms. When computing the likelihood over the prior space, ΔM_f and $\Delta \chi_f$ are the grid spacing between points in M_f and χ_f, respectively.

We perform the analysis using both a single detector, and two detectors with independent noise. We generate simulated colored Gaussian noise at the Advanced LIGO (aLIGO) design sensitivity [16][3]. We generate a 16-s pure-noise data stretch sampled at 4096 Hz and estimate the power spectral density (PSD) using the Welch method [51]. In this pure-noise study, we analyze a 0.2-s segment of data from a start time halfway through the 16-s data stretch (i.e., 8 s after the beginning of the data stretch), searching for a 220 ringdown mode. We perform the analysis for 200 realizations of noise, after which the analysis yields a stabilized distribution of $\mathcal{D}$ (see Appendix B). Throughout the rest of the paper, we refer to the procedure described here as the "$\mathcal{H}$:$\mathcal{H}'$ background study."

M_f prior	ΔM_f	χ_f prior	$\Delta \chi_f$
[10, 150] $M_\odot$	0.1$M_\odot$	[0, 0.99]	0.005

Table I: Prior ranges of M_f and χ_f, and corresponding resolutions, ΔM_f and $\Delta \chi_f$, used in the analysis.

Figure 1 shows the relationship between FAP and $\mathcal{D}$ when comparing a $\{220\}$ mode hypothesis with the null hypothesis. It shows that many of the simulations recover a positive $\mathcal{D}$, which indicates false positives for the presence of a ringdown mode. We define a threshold on $\mathcal{D}$ for a statistical significance corresponding to a 1% FAP, denoted by $\mathcal{D}^{\text{thr}}$. The thresholds for a single detector (two detectors) is $\mathcal{D}^{\text{thr}, 1-\text{ifo}}_{\{220\}:\{\text{null}\}} = 1.68$ ($\mathcal{D}^{\text{thr}, 2-\text{ifo}}_{\{220\}:\{\text{null}\}} = 2.20$). The two-detector case has a higher threshold because analyzing additional noise from an extra detector increases the probability that statistical fluctuations mimic a QNM signal. The filter is only partially coherent between detectors; signals in different detectors are filtered with coherent real frequencies and damping times, but the filtering does not enforce coherent amplitudes and phases of the ringdown modes. Thus, noise with incoherent amplitudes and phases in two detectors is not intrinsically penalized in the analysis, leading to an increased threshold. For the rest of the paper, we compute the two-detector $\mathcal{D}^{\text{thr}, 2-\text{ifo}}$

---

[3] The amplitude spectral density is given by `aligo_O4high.txt` available at [52].

by default unless noted otherwise. For real events, since GW detector noise is non-stationary over long timescales, we recompute the threshold for each observed GW event using the detector noise close to the event ($\mathcal{O}$(hour) before and after the event). When computing the threshold, we ensure the analysis settings (sampling rate, segment length, etc.) are the same as those used when analyzing the event itself.
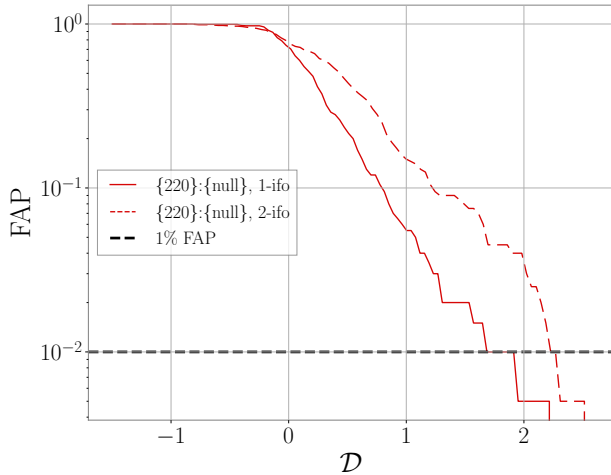

Figure 1: False-alarm probability (FAP) as a function of detection statistic $\mathcal{D}$ when comparing the {220} and null hypotheses in pure-noise simulations using a single detector (solid) and two detectors (dotted). Results are obtained from 200 injections in pure colored Gaussian noise at the aLIGO design sensitivity [52].

To verify that astrophysical ringdown signals are detectable with a 1% FAP threshold, we estimate the minimum required SNR for a ringdown signal to yield $\mathcal{D} > \mathcal{D}^{\mathrm{thr}}_{\{220\}:\{\mathrm{null}\}}$. We find that the SNR of a single mode required for an observation above threshold is $\mathrm{SNR}_{1-\mathrm{ifo}} = 2.8$ and $\mathrm{SNR}_{2-\mathrm{ifo}} = 3.2$ (see Sec. IV for details). Many existing GW observations have ringdown SNRs above such levels (e.g. ~ 14 for GW150914 [14]). Thus, the dominant mode can be confidently detected in GW events.

We further evaluate the performance of the method using a receiver operating characteristic (ROC) curve shown in Fig. 2, plotting the true-positive probability (TPP) against the FAP. To compute the TPP, we inject a signal with the 220 mode and ringdown parameters consistent with GW150914, given in Table II. We set the mode amplitude to two different values, resulting in $\mathrm{SNR}^{2-\mathrm{ifo}} = 4$ and 5, to evaluate the recovery rate. For each choice of the mode amplitude, the waveform is injected into 200 different noise realizations. Then, we compute $\mathcal{D}_{\{220\}:\{\mathrm{null}\}}$ for each injection and plot the TPP for each given FAP. Fig. 2 shows that, as expected, signals with a larger SNR are easier to recover. At both of these relatively low SNRs, the signal can achieve TPP $\gtrsim 90\%$

for FAP $\gtrsim 20\%$, demonstrating that the rational filter can efficiently distinguish the signal from noise. Note, however, that the ROC curve depends on the injection parameters (remnant BH mass, spin, and QNM phase) even when the SNR is fixed. Therefore, these results should only be interpreted as indicative.

ϕ_{220}	M_{inj}	χ_{inj}
4.81	$68.5 M_\odot$	0.692

Table II: Parameters of the {220} waveform injected to calculate the ROC curve shown in Fig. 2. These ringdown parameters are consistent with GW150914 [53, 54].
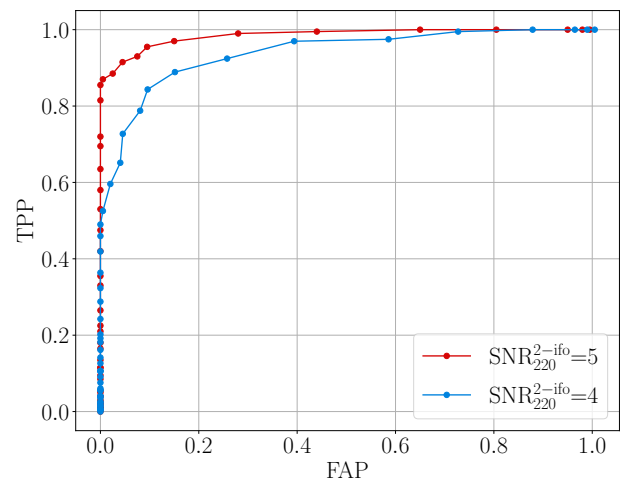

Figure 2: ROC curves for {220} mode identification at two different SNR values.

B. Identifying a second mode

We now extend this method to detecting a subdominant mode, assuming the dominant mode has been confidently identified. Conceptually, the idea is now to inject a single QNM into noise and compute $\mathcal{D}$ between a hypothesis with two QNMs and the hypothesis with the single correct QNM. Specifically, we follow the procedure described in Sec. III A to create 16 s of Gaussian noise, colored with the aLIGO design PSD and sampled at 4096 Hz. We inject a "ringup-ringdown" QNM signal into the generated noise:

$$x(t) = h(|t - 8 \text{ s}|), \tag{12}$$

where $|\cdot|$ denotes the absolute value function, $h(t)$ is given in Eq. (1), and we include the ringup component to avoid the impact of spectral leakage. The QNM is injected with a start time halfway through the timeseries (i.e., 8 s after the start). Here we always inject the 220 mode, which

is expected to be dominant for quasi-circular binary BH coalescences, but the technique is readily applicable if a different QNM is dominant. The parameters of the injected QNM are randomly sampled from the ranges given in Table III. Note that the complex frequencies of the injected QNMs are determined by the properties of the remnant BHs, where M_{inj} and χ_{inj} are the mass and spin of the remnant, respectively. We inject a range of SNRs, but the computed background threshold does not depend on the SNR of the injection because the injected mode is always completely filtered out at the maximum likelihood values (which is the dominant term of the computed evidence) for both hypotheses we consider. We start our analysis 3 M_{inj} after the QNM start time to ensure no impact is introduced to the analysis by the ringup part of the injection, but the computed background threshold is insensitive to the start time of the analysis (for $\Delta t_0 \geq 0$). We analyze 0.2 s of data as a standard setup. For real GW events, the injected signal parameters for the background study could be adjusted to better represent the ranges that are relevant to the true event, and the segment length and downsampling effects would be verified to not affect the estimated background threshold [43].

M_{inj}	χ_{inj}	Mode phase	SNR
[30, 120] $M_\odot$	[0, 0.95]	[0, 2π]	[4, 200]

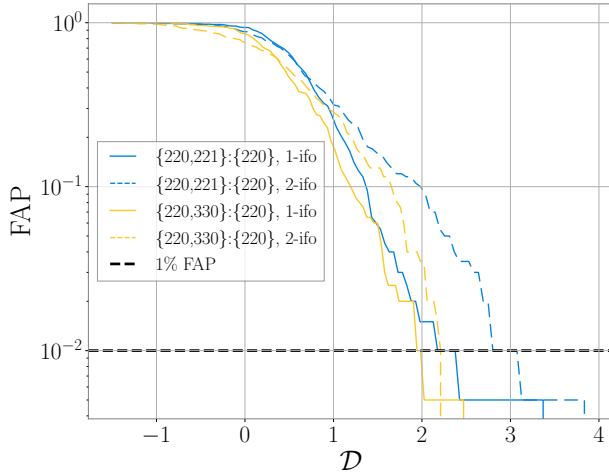Table III: Ranges of injected signal parameters when computing the threshold in Fig. 3

Figure 3: FAP as a function of $\mathcal{D}$ when comparing the {220,221}:{220} (blue) and {220,330}:{220} (yellow) for injections of {220}-only signals. Solid and dotted curves are for a single detector and two detectors, respectively. The results are based on 200 simulations in colored Gaussian noise at aLIGO design sensitivity.

We compute $\mathcal{D}$ for a particular two-mode hypothesis (e.g., {220,221} or {220,330}) compared to the correct {220} hypothesis. We use the same prior region as those in Table I and analyze 200 injections to determine the empirical $\mathcal{D}$ threshold. Fig. 3 shows the results. They are qualitatively similar to the pure-noise case (Sec. III A) where the majority of injections recover positive $\mathcal{D}$ values, and the two-detector threshold ($\mathcal{D}^{\text{thr, 2-ifo}}_{\{220,221\}:\{220\}} = 2.79$) is larger than the single-detector one ($\mathcal{D}^{\text{thr,1-ifo}}_{\{220,221\}:\{220\}} = 2.17$). The $\mathcal{D}$ thresholds also depend on the identity of the secondary mode. In fact, when overfiltering the same set of {220}-only injections with the {220,330} and {220,221} models, the thresholds are $\mathcal{D}^{\text{thr, 2-ifo}}_{\{220,330\}:\{220\}} = 2.18$ and $\mathcal{D}^{\text{thr, 2-ifo}}_{\{220,221\}:\{220\}} = 1.94$, respectively. This is because the distinguishability of different modes depend on how close their complex frequencies are. The $\mathcal{D}^{\text{thr}}$, therefore, needs to be computed for each specific secondary mode being tested.

Following the approach in Sec. III A, we construct the ROC curve for the two-mode scenario to quantify the sensitivity and robustness of our method. We inject a {220,221} signal with parameters given in Table IV, consistent with GW150914, with the waveform multiplied by different scale factors so that the SNR of the subdominant mode is $\text{SNR}_{221} = 4, 8, 16$. Each waveform is injected into 200 independent noise realizations to compute the TPP. As shown in Fig. 4, the ROC curves for $\text{SNR}_{221} = 4$ and 8 do not show ideal recovery. However, signals with $\text{SNR}_{221} = 16$ are confidently detectable.

Conceptually, the idea is to inject a set of QNMs in pure noise and compare two hypotheses: one that matches the true mode content of the injection, and another that overfilters the data by assuming an additional QNM.

A_{220}	A_{221}	ϕ_{220}	ϕ_{221}	M_{inj}	χ_{inj}
1.0	4.24	4.81	0.676	68.5	0.692

Table IV: Synthetic signal parameters of the two-mode {220,221} ringdown waveform (GW150914-like) used in simulations to calculate the ROC curve in Fig. 4.

C. IMR-conditional injections

The background study described in Secs. III A and III B does not rely on the IMR analysis of the event. Any ringdown mode surpassing the threshold as described above can be tested for consistency with the remnant BH parameters obtained independently from the IMR parameter estimation (PE). In this section, however, we introduce techniques that can improve the sensitivity of the rational filter to subdominant modes by including the information from the IMR PE. These techniques could be useful for current GW observations where the ringdown SNRs are generally not high.
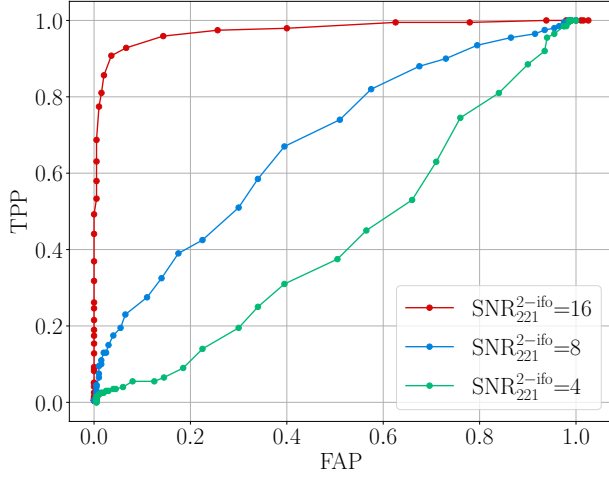
Figure 4: ROC curves when comparing the the true {220,221} hypothesis to the single-mode {220} hypothesis with three different SNRs of the 221 mode.

1. IMR-conditional injections with full prior

When evaluating the background, we have two options of selecting injection parameters: 1) randomize the $M_{\rm inj}$ and $\chi_{\rm inj}$ in the full prior range, as described above, and 2) sample $M_{\rm inj}$ and $\chi_{\rm inj}$ from the posterior obtained from the IMR analysis. In both cases the amplitude and phase of the QNMs injected are randomized. These two options correspond to two different motivations: a) *IMR-agnostic* analyses for IMR consistency tests or independent tests of GR via BH spectroscopy, and b) *IMR-conditional* studies, assuming that the IMR PE results are reliable and aiming for enhanced sensitivity to identify subdominant modes, which are useful for investigating initial conditions of the binary [55].

To study how conditional injections impact the background distribution and the threshold, we repeat the analysis in Sec. III B—injecting a 220 QNM into a two-detector network and overfiltering with a {220,221} mode model. This time, however, the remnant BH mass and spin for the injections are sampled from the posterior distribution of GW150914 inferred from the IMR [56, 57]. The analysis still uses the full prior space given in Table III. The background distribution from the conditional injections is shown in Fig. 5, where the thresholds for the IMR-agnostic injections and conditional injections are $\mathcal{D}^{\rm thr,\,agnostic}_{\{220,221\}:\{220\}} = 2.79$ and $\mathcal{D}^{\rm thr,\,conditional}_{\{220,221\}:\{220\}} = 2.11$, respectively. Therefore, using conditional injections for a background study lowers the threshold required for detecting subdominant modes and increases the sensitivity to these modes.
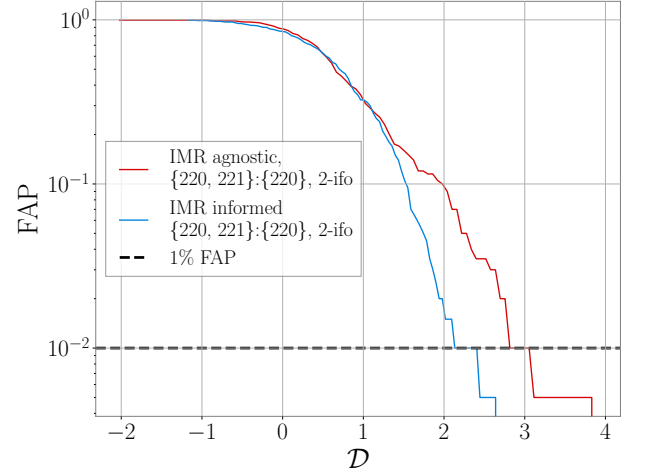

Figure 5: FAP as a function of $\mathcal{D}$ when comparing the {220,221}:{220} hypotheses using IMR-agnostic and IMR-conditional injections of {220}-only signals. The parameters for the conditional injections are sampled from the posterior distribution of GW150914 inferred from IMR [56]. For each injection configuration, the results are obtained from 200 synthetic signals injected into colored Gaussian noise at aLIGO design sensitivity in two detectors.

2. IMR-conditional injections with restricted priors

For certain mode configurations, the likelihoods computed over the whole parameter space contain bimodal features. Specifically, for signals with a 220 fundamental mode and a subdominant 330 or 440 mode, the likelihood contours in the mass and spin plane show bimodality. This is because, for a given χ_f, the damping times of the 220 and 330 (440) are similar, but their real frequencies differ by a ratio of approximately 2:3 (1:2). Since the real frequency of a QNM is scaled by the remnant BH mass, a 220 mode can almost entirely be filtered out by a 330 filter constructed with a similar χ_f value and a M_f value ~ 1.5 times higher than the true M_f. An example of this bimodality is shown in Fig. 6a, where we inject a {220,330} waveform with parameters given in Table V (consistent with those predicted for a high-mass-ratio system which excites a non-negligible 330 mode). The signal is injected into colored Gaussian noise, and its amplitude is scaled so that the total ringdown SNR and the 330-mode SNR are $\text{SNR}^{2-\rm ifo}_{\rm tot} = 20.5$ and $\text{SNR}^{2-\rm ifo}_{330} = 4$, respectively. Fig. 6a shows the likelihood contours obtained from the rational filter using the correct {220,330} QNM model. The secondary peak is found at a mass approximately 1.5 times higher than the mass corresponding to the main peak.

If the prior is restricted to $M_f = [30, 85] M_\odot$, $\chi_f = [0.35, 0.95]$, only a single peak appears in the likelihood contours, as shown in Fig. 6b. The restricted prior is chosen to only include the peak that agrees with the

A_{220}	A_{330}	ϕ_{220}	ϕ_{330}	$M_{\rm inj}$	$\chi_{\rm inj}$	$\text{SNR}_{\rm tot}$	SNR_{330}
1.0	0.081	0.529	1.457	72	0.64	20.5	4

Table V: Synthetic signal parameters of the two-mode {220,330} ringdown waveform. These parameters are consistent with a binary system with where $m_1/m_2 = 10$, $\chi_1 = \chi_2 = 0.5$, where m_1, m_2 and χ_1, χ_2 are the initial masses and spins of the two progenitor BHs, and the system has an inclination angle, $\iota = \pi/2$, and polarization angle $\psi = 0$, as given by the `postmerger` package [58].

remnant BH parameters used for injections (or IMR posterior if analyzing real GW events). This restricted prior also corresponds to the expectation that the 220 QNM should be dominant over the 330 mode, as is the case for quasi-circular binary BH mergers (although see [55] for precessing binaries).

Restricting the prior range may reduce the $\mathcal{D}$ of the signal due to the fact that the likelihoods are integrated over the whole prior range. However, the threshold obtained by analyzing the background over the restricted prior range also decreases. To determine whether using the restricted prior improves the detectability of a subdominant 330 mode, we construct a ROC curve for the {220,330}:{220} analysis, similar to Sec. III B. In this case, we compute the TPP from 200 injections of the {220,330} waveform in colored Gaussian noise (all simulated with the parameters given in Table V). The threshold is determined by performing a {220,330}:{220} background study where the 220-only injection parameters are fixed (Table V). The study is repeated for the two prior choices, full and restricted, and the results are shown in Fig. 7. The restricted prior improves the sensitivity to the subdominant 330 mode.[4] In this analysis, the restricted prior range is chosen such that only the main peak of the likelihood in the mass and spin plane, which agrees with IMR, is included. There are, however, alternative choices for the restricted prior range, and an optimal prior range may exist for a given system that maximizes the detectability of a subdominant mode. We defer the investigation of an optimized choice of priors to future work. In this example, we only demonstrate the application of this technique for waveforms with the {220,330} modes. It is, in principle, applicable to any mode combination which shows multimodality or features in the likelihoods which are a-priori known to be unphysical based on numerical relativity or the IMR analysis of the event.

---

[4] Both ROC curves are better than the $\text{SNR}_{221}^{2-\rm ifo} = 4$ case in Fig. 4 because the 330 QNM is more separated in frequency and damping time from the 220 mode than the 221 mode.
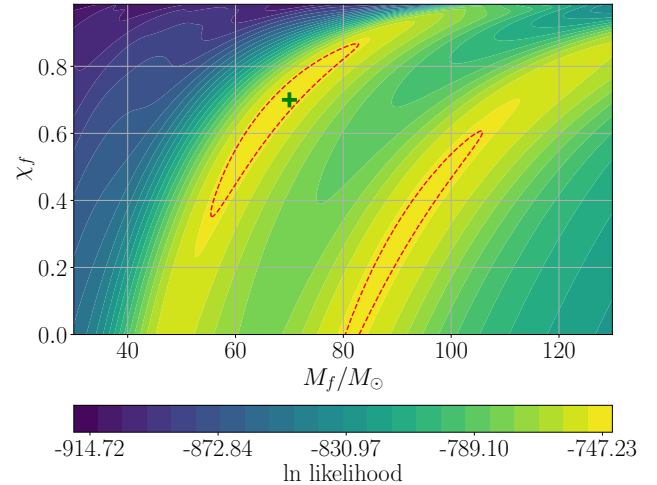
(a) Full prior
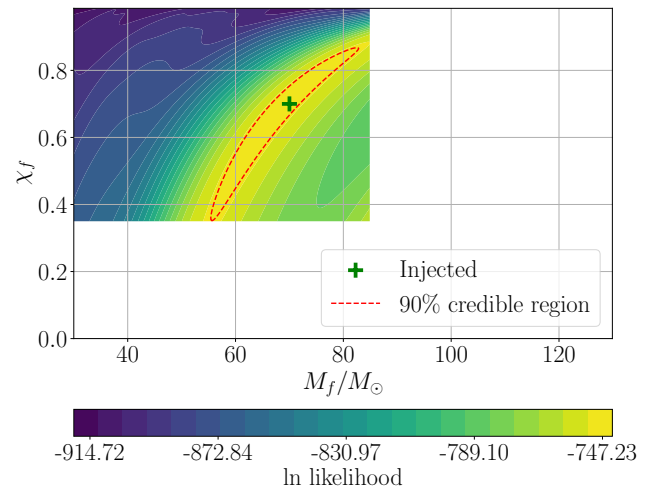
(b) Restricted prior

Figure 6: Likelihood contours of M_f and χ_f for a synthetic waveform with {220,330} modes obtained using the correct {220,330} rational filter. Panels (a) and (b) are for the two choices of the prior range. The green plus mark indicates the true injection parameters given in Table V.

IV. SNR ESTIMATION

In this section, we describe how to use the rational filter to estimate the ringdown SNR after all detectable modes are identified. We also explain the limitations of SNR estimates for individual modes.

A generic relationship for the Bayes factors (BFs) of two hypotheses, $\mathcal{H}$ and $\mathcal{H}$', corresponding to signal templates,
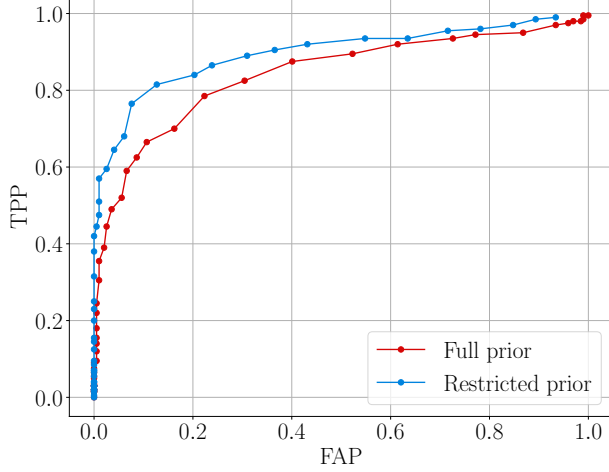
Figure 7: ROC curves for detecting the subdominant 330 mode using the full prior and the restricted prior. The results are obtained from 200 injections in colored Gaussian noise with $\mathrm{SNR}_{330}^{2-\mathrm{ifo}} = 4$.
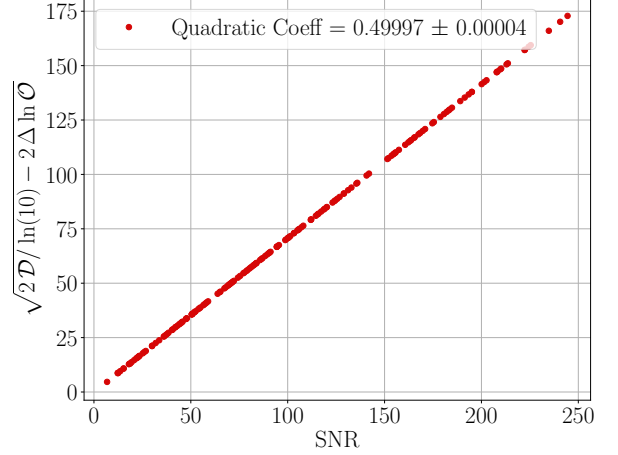

Figure 8: The scaling between $\mathcal{D}$ (comparing the injected mode hypothesis and the null hypothesis) and the SNR of the injected signal. Results are calculated from 200 injections (see Table III for injection parameter ranges).

h and h', versus the SNR of a signal is [59]:

$$\ln(\mathrm{BF}\,[h':h]) = (1 - \mathrm{FF}^2)\frac{\mathrm{SNR}^2}{2} + \Delta\ln\mathcal{O}, \qquad (13)$$

where FF is the fitting factor, i.e., the best match that can be achieved between the two hypotheses, defined by:

$$\mathrm{FF} = \max\frac{(h|h')}{\sqrt{(h|h)(h'|h')}}. \qquad (14)$$

The maximum is taken over all the parameters of h and h', and $\mathcal{O}$ is the Occam's factor approximated by:

$$\mathcal{O} = \frac{V_{\mathrm{post},95\%}}{V_{\mathrm{prior}}}, \qquad (15)$$

where $V_{\mathrm{post},\,95\%}$ is the volume of the 95% credible region of the posterior, and V_{prior} is the volume of the prior. Since $\mathcal{D}$ is analogous to $\ln(\mathrm{BF})$, we should see the same relation in Eq. (13) between $\mathcal{D}$ and SNR.

To verify this we inject 220 QNM signals into colored Gaussian noise with randomized parameters (remnant BH mass and spin, and mode amplitude and phase) drawn over the same ranges as listed in Table III. We compute $\mathcal{D}$ between the correct 220-mode hypothesis and the null hypothesis using the same prior range as that described in Sec. III. Since the null hypothesis implies no signals are present (i.e. $h(t) = 0$), we can see from Eq. (14) that FF = 0. For each injection, we also directly compute the Occam's factor $\mathcal{O}$ from the posterior using Eq. (15). The results are shown in Fig. 8, we fit the data points with a second-order polynomial and find that the quadratic coefficient is 0.49997 ± 0.00004, which agrees precisely with Eq. (13) when FF = 0.

We now use Eq. (13) when FF = 0 to find an expression for SNR:

$$\mathrm{SNR} = \sqrt{2\,\mathcal{D}/\ln(10) - 2\,\Delta\ln\mathcal{O}}. \qquad (16)$$

To quantify the accuracy of this estimation, we consider the error in the estimated SNR for an injection:

$$\Delta\,\mathrm{SNR} = \mathrm{SNR}_{\mathrm{inj}} - \mathrm{SNR}, \qquad (17)$$

where $\mathrm{SNR}_{\mathrm{inj}}$ is the matched filter SNR of the injected signal. We inject ringdown signals with $\mathrm{SNR}_{\mathrm{inj}} \in [5, 35]$, which is consistent with the expected SNR range for current-generation GW detectors, and compute the recovered $\Delta\,\mathrm{SNR}$. Fig. 9 shows the result and demonstrates that the computed SNRs are broadly reliable with $|\Delta\mathrm{SNR}| \lesssim 0.6$ and the error further decreases for signals with larger SNRs. The error arises because Eq. (13) is derived using Laplace's approximation for the likelihood distribution [59, 60]. Particularly, the definition of Occam's factor in Eq. (15) assumes that the likelihoods are Gaussian distributed and peaked at the maximum likelihood value. However, for the rational filter, this assumption is generally not true, which leads to residual errors in the estimated SNR. This effect is more significant in the low-SNR regime where the Occam's factor is larger, but we still find that the SNR estimation is sufficiently accurate. We repeat this verification, that $\Delta\mathrm{SNR} \lesssim 1$, for injections of up to three modes in noise generated with different PSDs and find the result remains robust.

This technique can compute the total SNR of an arbitrary number of identified modes, but cannot estimate the SNR of individual modes when the signal contains more than one mode. This is because the FF in Eq. (13) between non-trivial hypotheses cannot be calculated. See Appendix D for a more detailed discussion.
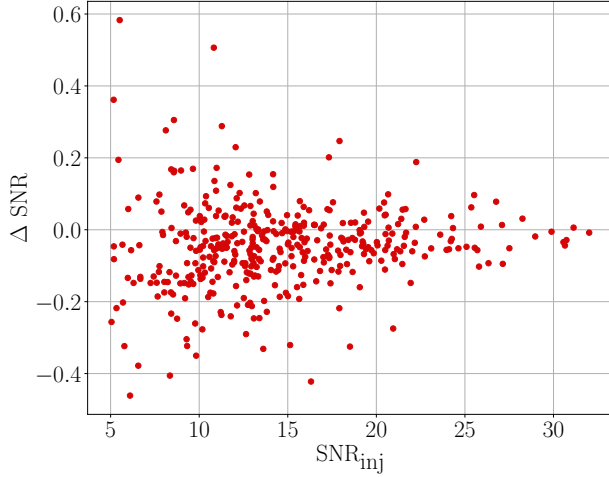
Figure 9: Error in the estimated SNR versus the true SNR for 200 injections of a {220}-only ringdown signal in colored Gaussian noise at the aLIGO design sensitivity.

V. ANALYSIS PROCEDURE

In this section, we describe the analysis procedure for analyzing GW events and identifying their mode composition. We discuss how to identify the correct QNM if multiple subdominant modes are potential candidates (Sec. V A) and describe the complete workflow (Sec. V B).

A. Identify the correct mode

The background study described in Sec. II offers an approach to differentiate a true subdominant mode from fluctuations due to detector noise and, therefore, quantify the statistical significance of a potential subdominant mode. However, a subdominant mode present in the signal can also be falsely identified as a different QNM. To demonstrate this, we construct a {220,221} signal with parameters in Table IV and tune the amplitude so that the subdominant mode has $\mathrm{SNR}_{221} = 16$. We inject the signal into 200 noise realizations and compute $\mathcal{D}$ for both the correct mode model $\{220, 221\} : \{220\}$ and two incorrect mode models, $\{220, 210\} : \{220\}$ and $\{220, 200\} : \{220\}$. Fig 10a show that all three hypotheses yield similar distributions and all have $\mathcal{D} > \mathcal{D}^{\mathrm{thr}}$ in over 95% realizations. The similarity between distributions occurs because all mode models are able to fit a majority of the SNR of the subdominant mode. We, therefore, cannot determine the correct mode on the basis of $\mathcal{D}$ alone.

We use the posterior quantile of the injected parameters $p(M_{\mathrm{inj}}, \chi_{\mathrm{inj}})$ [Eq. (9)] as an additional criterion to identify the true subdominant mode when multiple hypotheses are above threshold. We expect the true mode hypothesis to result in lower posterior quantiles compared to other mode
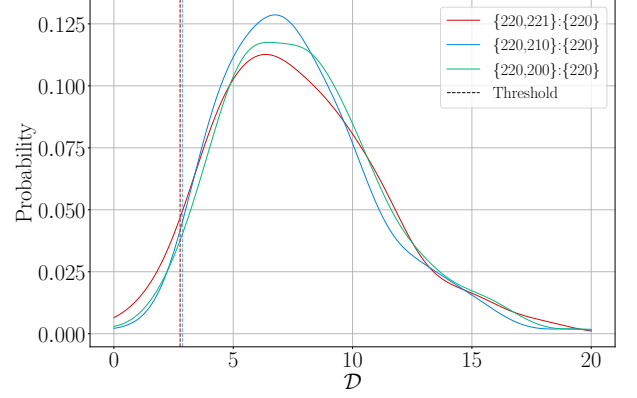
(a) Distribution of $\mathcal{D}$
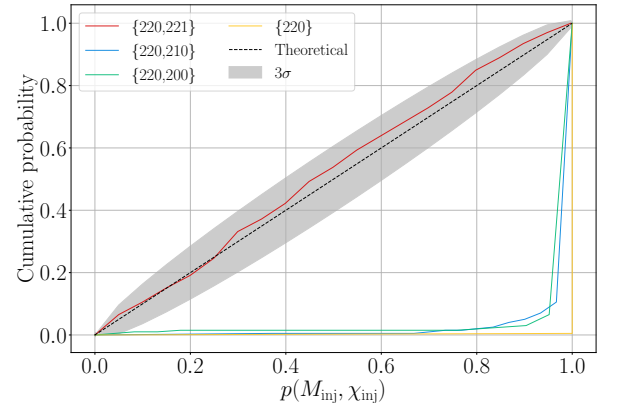
(b) Distribution of $p(M_{\mathrm{inj}}, \chi_{\mathrm{inj}})$

Figure 10: Recovered $\mathcal{D}$ and $p(M_{\mathrm{inj}}, \chi_{\mathrm{inj}})$ from 200 injections of a {220,221} ringdown signal (see parameters in Table IV; with amplitudes scaled to yield $\mathrm{SNR}_{221} = 16$). Different colors correspond to different mode models (see legend). Panel (a) shows the probability distribution of $\mathcal{D}$ with different mode models. Panel (b) plots the cumulative probability distributions of $p(M_{\mathrm{inj}}, \chi_{\mathrm{inj}})$; the theoretical prediction of an accurate model with 3σ bounds is plotted for comparison. Each synthetic signal is injected into colored Gaussian noise of one aLIGO detector at the design sensitivity.

hypotheses. We verify this in Fig 10b which shows the cumulative probability of $p(M_{\mathrm{inj}}, \chi_{\mathrm{inj}})$, i.e. the probability that $p(M_{\mathrm{inj}}, \chi_{\mathrm{inj}})$ is less than a given value on the x-axis: $P\left[p(M_{\mathrm{inj}}, \chi_{\mathrm{inj}}) < x\right]$ where P denotes probability. The plot demonstrates that the true {220,221} hypothesis is more likely to recover a lower $p(M_{\mathrm{inj}}, \chi_{\mathrm{inj}})$ than the other mode hypotheses which are biased to higher $p(M_{\mathrm{inj}}, \chi_{\mathrm{inj}})$. The plot also shows that the $\{220, 221\}$ model hypothesis agrees with the theoretical expectation [61, 62]:

$$P\left[p(M_{\mathrm{inj}}, \chi_{\mathrm{inj}}) < x\right] = x, \tag{18}$$

within the 3σ level. Eq. (18) holds for any signal contain-

ing a specific set of QNMs when it is recovered with the correct hypothesis. Incorrect models, however, are always biased to higher $p(M_{\rm inj}, \chi_{\rm inj})$. Using $p(M_{\rm inj}, \chi_{\rm inj})$ is, therefore, a robust way to identify the correct subdominant mode when there are multiple candidates.

For real GW observations, we use the IMR-inferred remnant BH mass and spin $p(M_{\rm rem}, \chi_{\rm rem})$ to compute the posterior quantile. While this makes the analysis conditional on the IMR analysis, it is necessary to identify the correct subdominant mode at current ringdown SNRs $[\sim \mathcal{O}(10)]$. In a higher SNR regime, the distributions of $\mathcal{D}$ between the true model and an incorrect model are more distinct, which allows us to conduct IMR-agnostic analysis. We defer the investigation of the SNRs required for IMR-agnostic analysis to future work.

B. Workflow

We now formulate a complete workflow to analyze a GW event and identify all detectable ringdown modes using the rational filter. The workflow is outlined in Fig. 11, and we describe it in detail below.

Step 1. Pre-processing: We first *obtain the IMR PE* results, which are used as inputs by the rational filter. In particular, the sky position of the event (right ascension and declination) and geocentric start time are used to align the merger time at different detectors. The remnant BH mass and spin inferred by the IMR analysis are also needed in later steps when we compare the results from the rational filter analysis to those from the IMR. We then *list detected modes:* $\mathcal{M}_{\rm detected}$. This list contains the ringdown modes identified in the data by the rational filter and initially consists only of the 220 mode (the dominant QNM for quasi-circular binaries and so assumed to be always detected, although see [55] for precessing binaries), but is updated throughout the procedure. In principle, the analysis can simply start with an empty list and test the 220 mode as the first dominant mode. However, unless there are reasons to do this (e.g., analyzing a high-eccentricity merging system, obtaining unexpected results for initial analysis with the rational filter, etc.), we start by assuming that the 220 mode is the existing dominant mode. We now *list candidate modes* $\mathcal{M}_{\rm candidate}$ which could be the next most dominant mode. While the number of QNMs excited after merger can be arbitrarily large, in practice, only a small number of subdominant modes are potentially detectable with high confidence for current-generation detectors. We start with an exhaustive list of these modes, e.g., {221, 210, 200, 330, 440} and any others as needed for each individual event suggested by other ringdown pipelines or surrogate models [58]. The rational filter method is computationally efficient and thus allows us to test an exhaustive list of mode candidates.

Step 2. Analysis: We now iterate through the modes in $\mathcal{M}_{\rm candidate}$. For each candidate mode, we test if it is the next most dominant QNM by comparing two hypotheses; $\mathcal{H}$ where the modes in $\mathcal{M}_{\rm detected}$ exist and additionally

the chosen candidate mode also exists, and $\mathcal{H}'$ where only $\mathcal{M}_{\rm detected}$ exist. As we iterate through $\mathcal{M}_{\rm candidate}$, the identity of the extra added mode changes, but we only ever test one additional mode at a time. Given the chosen candidate mode, we perform the $\mathcal{H} : \mathcal{H}'$ background study and compute the *threshold* $\mathcal{D}^{\rm thr}$ as described in Sec III. We perform the *event analysis* as described in Sec. II and compute $\mathcal{D}_{\mathcal{H}:\mathcal{H}'}$. For both hypotheses, we also compute the joint posterior quantile on which the maximum a-posteriori value of the remnant BH mass and spin obtained from the IMR analysis $p(M_{\rm IMR}, \chi_{\rm IMR})$ lies. After these are computed, for each candidate mode being tested, we check *is it a possible detection?* A candidate is classified as a potential detection if $\mathcal{D} > \mathcal{D}^{\rm thr}$ *and* the posterior quantile decreases when including this candidate mode, i.e. adding the candidate mode makes the analysis more consistent with IMR [5]. If the candidate mode satisfies these criteria, it is added to a potential detections mode list $\mathcal{M}_{\rm potential}$. Otherwise, it is discarded. Regardless of whether the candidate is a potential detection or not, we repeat the procedure in Step 2 for each of the candidate modes in $\mathcal{M}_{\rm candidate}$ before continuing.

Step 3. Post-processing: If $\mathcal{M}_{\rm potential}$ has at least one mode, an additional ringdown mode has been identified. In this case, we go back to Step 1 and add the top mode—the one yields the largest $\mathcal{D}$ in Step 2 and leads to the largest decrease in $p(M_{\rm IMR}, \chi_{\rm IMR})$—to $\mathcal{M}_{\rm detected}$. We then look for the next subdominant mode and use the remaining (if any) modes in $\mathcal{M}_{\rm potential}$ as the new candidate modes in $\mathcal{M}_{\rm candidate}$. We additionally add the next overtone corresponding to the newly confirmed mode to $\mathcal{M}_{\rm candidate}$ (i.e., if the 330 mode is confirmed, add the 331 mode to the candidate mode list). If, however, $\mathcal{M}_{\rm potential}$ does not contain any modes, then no additional subdominant modes can be detected. The final confirmed ringdown modes list is then $\mathcal{M}_{\rm detected}$, and we compute the total ringdown SNR for these modes (as described in Sec. IV).

The procedure described above can alternatively be performed agnostic of the remnant mass and spin inferred by IMR, which means skipping the posterior quantile calculation in Step 2. In this case, the background study and threshold calculation can only be done using fully agnostic injections. However, to achieve better sensitivity, especially in the low-SNR regime, the full flow chart should be followed, and the thresholds should be computed using either IMR-conditional injections, restricted priors, or both, as described in Sec. III C.

---

[5] While other statistics aside from the posterior quantile may be a better indicator of the agreement between results from ringdown and IMR (e.g. how much of the 90% credible regions obtained from the rational filter and the IMR analyses overlap), we defer a detailed study of different statistics to future work.
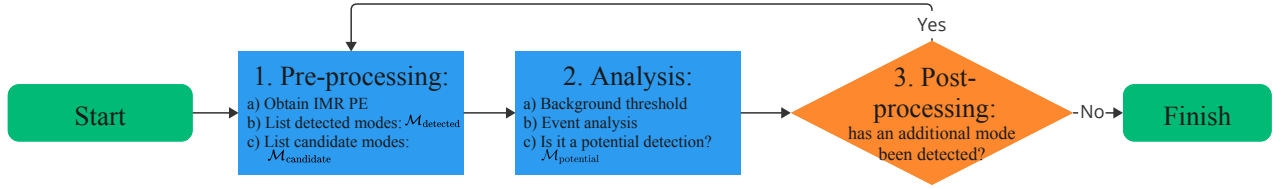
Figure 11: Flow chart describing the ringdown analysis steps for a GW event using the QNM rational filter.

VI. NR SIMULATIONS

We apply the procedure described in Sec. V to analyze two NR waveforms from the Simulating eXtreme Spacetimes (SXS) catalog [63, 64]. We validate the method assuming the QNMs in the signal are not fully known a-priori.

A. SXS:BBH:0305

We first analyze the GW150914-like NR waveform, SXS:BBH:0305 [53]. We inject the $h_{2,\pm2}$ harmonics into colored Gaussian noise at the aLIGO design sensitivity [52] for two LIGO detectors (with inclination angle, ι, and polarization angle, ψ, given by the maximum likelihood values inferred for GW150914 by the GWTC-2.1 PE [56, 57]). We set the remnant mass in the detector frame to $M_{\rm inj} = 68.5 M_\odot$ and tune the amplitude such that the signal has a network ringdown $\rm SNR^{2-ifo} = 59.96$. The signal of 4 s is injected into noise with a sampling rate of 16384 Hz. We select $\Delta t_0 \in [0, 15] M_{\rm inj}$[6], take 0.2 s of data starting at each Δt_0, and downsample it to 4096 Hz.

Following Fig. 11, in Step 1, we assume that the dominant QNM has already been detected and start with $\mathcal{M}_{\rm detected} = \{220\}$. The candidate mode list for the second ringdown mode is: $\mathcal{M}_{\rm candidate} = \{221, 210, 200, 330, 440\}$. In Step 2, we iterate through all candidates for the secondary mode. The resulting $\mathcal{D}$ and $p(M_{\rm inj}, \chi_{\rm inj})$ are shown in Fig. 12. The dotted lines are the thresholds calculated as described in Sec. III. The colored bands around $\mathcal{D}$ indicate the error bars due to potential noise fluctuations (i.e., corresponding to the 1-σ range of the background $\mathcal{D}$ distribution); see Appendix C for more details. All of the candidate modes (independently analyzed) yield $\mathcal{D} > \mathcal{D}^{\rm thr}$ and also show a decrease of $p(M_{\rm inj}, \chi_{\rm inj})$ for $\Delta t_0 \leq 13 M_{\rm inj}$. Thus, they are accepted to the potential detection list, i.e., $\mathcal{M}_{\rm potential} = \{221, 210, 200, 330, 440\}$. In Step 3, we compare the different modes in $\mathcal{M}_{\rm potential}$ and find that the $\mathcal{D}$ for the secondary mode to be the 330 or 440 QNM

---

[6] We note that the precise onset of ringdown is ambiguous and remains a topic of ongoing study (e.g., [49, 65–69]). A detailed investigation of this issue is beyond the scope of this work.

are markedly lower than those for the 221, 210, and 200 modes. In the lower panel of Fig 12, we also find that the posterior quantile of the {220,221} hypothesis is lower than the other modes for all times and leads to $p(M_{\rm inj}, \chi_{\rm inj}) \approx 0$ at $\Delta t_0 = 10M$. Hypotheses with other candidate modes only have $p(M_{\rm inj}, \chi_{\rm inj}) \sim 0$ at much later times ($\Delta t_0 \sim 20M$), when $\mathcal{D}$ values of the candidate modes have already fallen below the threshold and the posterior quantiles are no longer reliable. Therefore, the 221 QNM is the identified secondary mode. We return to Step 1 and update the detected and candidate mode lists to be $\mathcal{M}_{\rm detected} = \{220, 221\}$ and $\mathcal{M}_{\rm candidate} = \{222, 210, 200, 330, 440\}$.

We now perform Step 2 for a third ringdown mode and show the results in Fig. 13. In the upper panel, we see that the 440 candidate mode is not a potential detection as its $\mathcal{D}$ falls below the threshold at any start time. All other modes are added to $\mathcal{M}_{\rm potential}$. In the lower panel, we compare all potential modes following Step 3. We identify the third ringdown mode to be the 222 QNM because its $p(M_{\rm inj}, \chi_{\rm inj})$ are significantly lower than the others at all times, although the $\mathcal{D}$ values are comparable among all hypotheses. We now return to Step 1 and update the mode lists $\mathcal{M}_{\rm detected} = \{220, 221, 222\}$ and $\mathcal{M}_{\rm candidate} = \{223, 210, 200, 330\}$ to search for a fourth mode. However, no candidate yields $\mathcal{D} > \mathcal{D}^{\rm thr}$. Thus, we complete the analysis and estimate the network ringdown SNR, yielding 59.81. We compare the estimated SNR to the true injected SNR and find $\Delta \rm SNR = 0.05$, demonstrating that we have recovered the majority of the signal power.

B. SXS:BBH:1107

We now study a high-mass-ratio system (using the SXS:BBH:1107 waveform) with initial parameters $m_1/m_2 = 10$, $\chi_1 = \chi_2 = 10^{-7}$, where m_1, m_2 and χ_1, χ_2 are the initial masses and spins of the two progenitor BHs. We inject a synthetic signal using the remnant BH properties $M_{\rm inj} = 68.5 M_\odot$, $\chi_{\rm inj} = 0.261$ with a ringdown $\rm SNR^{2-ifo} = 38.41$. In such a high-mass-ratio system, we expect that the higher-order angular modes are more strongly excited. We inject the $h_{2,\pm2}$ and $h_{3,\pm3}$ modes with ($\iota = \pi/4, \psi = 0$) into colored Gaussian noise (two aLIGO detectors) as described in Sec. VI A.

Again, we carry out the analysis following the steps
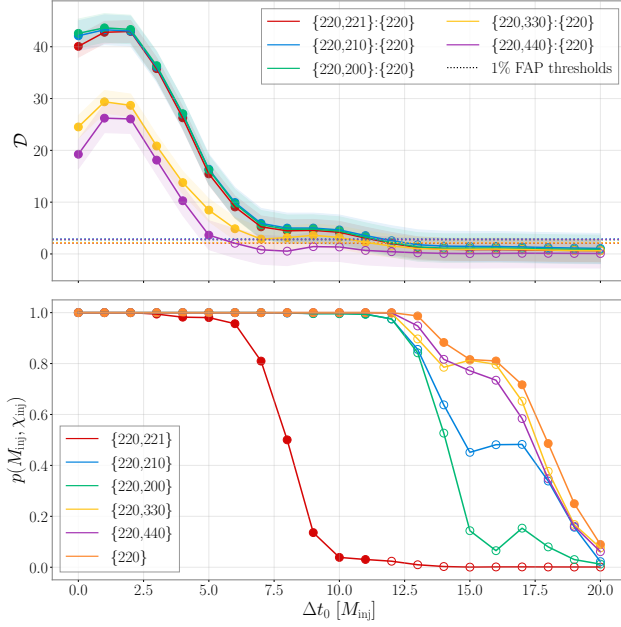
Figure 12: Resulting $\mathcal{D}$ and $p(M_{\text{inj}}, \chi_{\text{inj}})$ when testing different candidates of the second ringdown mode against the {220}-only hypothesis. The filled and unfilled markers correspond to $\mathcal{D} > \mathcal{D}^{\text{thr}}$ and $\mathcal{D} < \mathcal{D}^{\text{thr}}$, respectively. The shaded bands around $\mathcal{D}$ indicate the 99% error bars due to potential noise fluctuations. The synthetic signal is constructed using the $h_{2,\pm2}$ harmonic of the SXS:BBH:0305 waveform and injected into colored Gaussian noise with a network ringdown SNR of 59.86.

Figure 13: (Similar to Fig. 12) Resulting $\mathcal{D}$ and $p(M_{\text{inj}}, \chi_{\text{inj}})$ when testing different candidates of the third ringdown mode compared to a {220,221} hypothesis.

modes is 38.15 with an error $\Delta\,\text{SNR} = -0.14$.

VII. GW150914 ANALYSIS

In this section, we analyze the real event GW150914 which was previously studied using the rational filter in Refs. [37, 38], where evidence for the 221 mode was found. We re-analyze the event by taking into account the noise background and quantifying the statistical significance of the overtone mode. Here, we follow Fig. 11 and repeat the procedure as demonstrated in the NR simulations. We use the PE results from the IMR analysis (Step 1), set the geocentric start time of the signal to be $t_0 = 1126259462.4083$, and use the inferred sky location (right ascension $\alpha = 1.95$ rad and declination $\delta = -1.27$ rad) to account for the signal travel time between the two aLIGO detectors. We start the analysis with $\mathcal{M}_{\text{detected}} = \{220\}$ and $\mathcal{M}_{\text{candidate}} = \{221\}$. We only consider the 221 mode as a candidate because GW150914 has been extensively studied, and no analysis finds evidence of any other mode (including using the rational filter).

In Step 2, we calculate the $\mathcal{D}$ threshold when comparing the {220,221} and {220} hypotheses. Since the detector noise is non-stationary and non-Gaussian, we take detector data in the 1.5 hours around the event time to evaluate the noise background. Specifically, we exclude 1 s of data around the event time and select 150 data chunks before and another 150 after the event, each spanning 16 s. For example, the first chunk after the event

in Fig. 11. In Step 1, we start with $\mathcal{M}_{\text{detected}} = \{220\}$ and $\mathcal{M}_{\text{candidate}} = \{221, 210, 200, 330, 440\}$. The results of Step 2 are shown in Fig. 14. The upper panel shows all candidate modes yield $\mathcal{D} > \mathcal{D}^{\text{thr}}$. The lower panel shows all candidate modes also reduce $p(M_{\text{inj}}, \chi_{\text{inj}})$ compared to the $\mathcal{M}_{\text{detected}}$, so they are all added to $\mathcal{M}_{\text{potential}}$. In Step 3, we identify 330 mode as the top candidate for two reasons. First, $\mathcal{D}$ for 330 (yellow) is only slightly lower than that for 200 (green), and they are comparable within the uncertainty range. Second, adding the 330 QNM yields $p(M_{\text{inj}}, \chi_{\text{inj}}) \sim 0$ for $\Delta t_0 > 11 M_{\text{inj}}$, but adding the 200 QNM generally does not produce low $p(M_{\text{inj}}, \chi_{\text{inj}})$ values. We conclude that 330 mode is the secondary QNM and return to Step 1, updating the mode lists to be $\mathcal{M}_{\text{detected}} = \{220, 330\}$ and $\mathcal{M}_{\text{candidate}} = \{331, 221, 210, 200, 440\}$.

Fig. 15 shows the results from the next iteration to identify the third mode. Although all $\mathcal{D}$ values are comparable and above threshold for the majority of the time, the 221 mode yields the lowest $p(M_{\text{inj}}, \chi_{\text{inj}})$ at times $\Delta t \leq 7 M_{\text{inj}}$. We update $\mathcal{M}_{\text{detected}} = \{220, 330, 221\}$ and $\mathcal{M}_{\text{candidate}} = \{331, 222, 210, 200, 440\}$. In the next iteration, none of the candidates satisfy $\mathcal{D} > \mathcal{D}^{\text{thr}}$, so we conclude that all the detectable ringdown modes are identified. The total SNR estimated for $\{220, 330, 221\}$
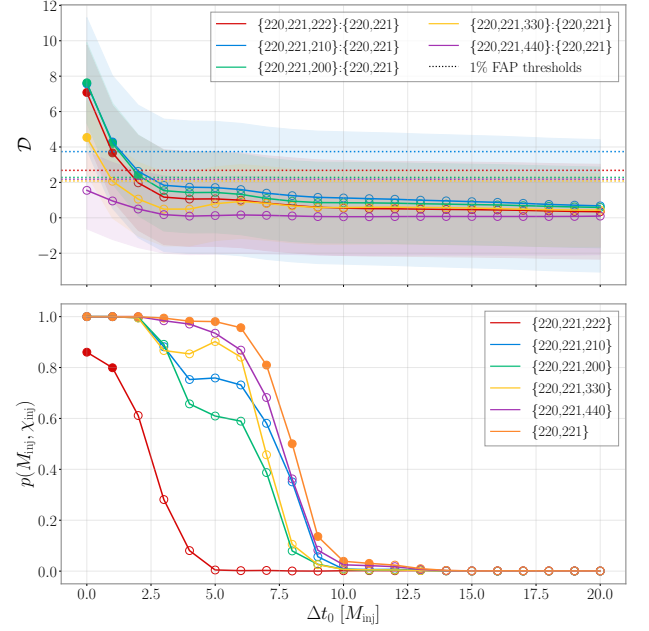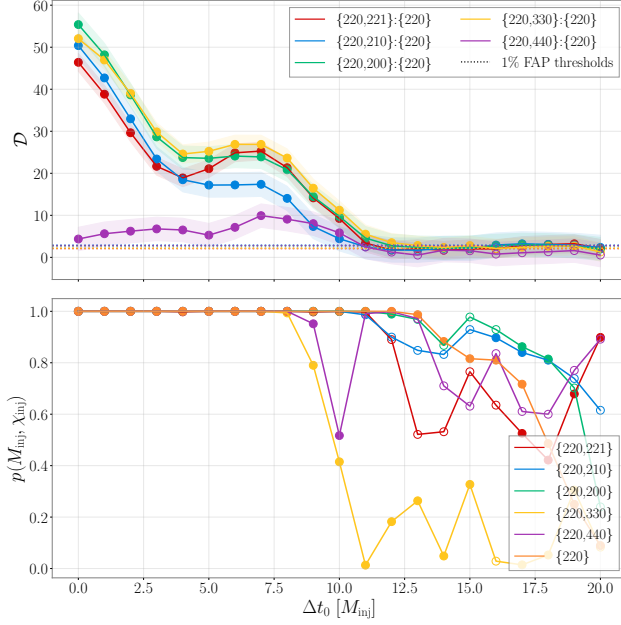
Figure 14: Resulting $\mathcal{D}$ and $p(M_{\mathrm{inj}}, \chi_{\mathrm{inj}})$ when testing different candidates of the second ringdown mode against the {220}-only hypothesis. The synthetic signal is constructed using the $h_{2,\pm 2}$ and $h_{3,\pm 3}$ harmonics of the SXS:BBH:1107 waveform and injected into colored Gaussian noise with a network ringdown SNR of 38.15.

Figure 15: (Similar to Fig. 14) Resulting $\mathcal{D}$ and $p(M_{\mathrm{inj}}, \chi_{\mathrm{inj}})$ when testing different candidates for the third ringdown mode compared to a {220,330} hypothesis.

spans $[t_{\mathrm{start}}+0.5\,\mathrm{s},\ t_{\mathrm{start}}+16.5\,\mathrm{s}]$, and the next one spans $[t_{\mathrm{start}}+16.5\,\mathrm{s},\ t_{\mathrm{start}}+32.5\,\mathrm{s}]$. Similarly, we get all 300 noise data chunks. For each data chunk, we use the full 16 s to estimate the PSD and inject a {220} signal so that the peak strain of the injected signal occurs at the center of the data chunk, i.e., 8 s after the start. The $\mathcal{D}$ distribution of a {220,221} hypothesis compared to a {220} hypothesis is shown in Fig. 16. The background is estimated both in an agnostic and a conditional way, as described in Sec. III. The parameters for the conditional injections are sampled from the posterior distribution obtained in the IMR analysis [56, 57]. As expected, the threshold obtained is lower with the conditional injections than with the agnostic injections.

We then analyze the GW150914 event itself, with a similar analysis configuration as that presented in Refs. [37, 38]. The analysis is performed using a 0.2 s long data stretch (i.e., the "window width" in [37, 38]), and the data are conditioned in the same way (see Sec. II B of [38]). However, this time we estimate the PSD by taking 16 s of data starting from 0.5 s after the peak of the event (instead of 32 s around the event, including the event itself as described in [37, 38]). We find this approach more robust in estimating PSDs where the ringdown signal has decayed sufficiently, but the noise behaviour remains similar to that during the event. In addition, we use a sampling rate of 4096 Hz instead of 2048 Hz used in Refs. [37, 38] to further reduce the impact
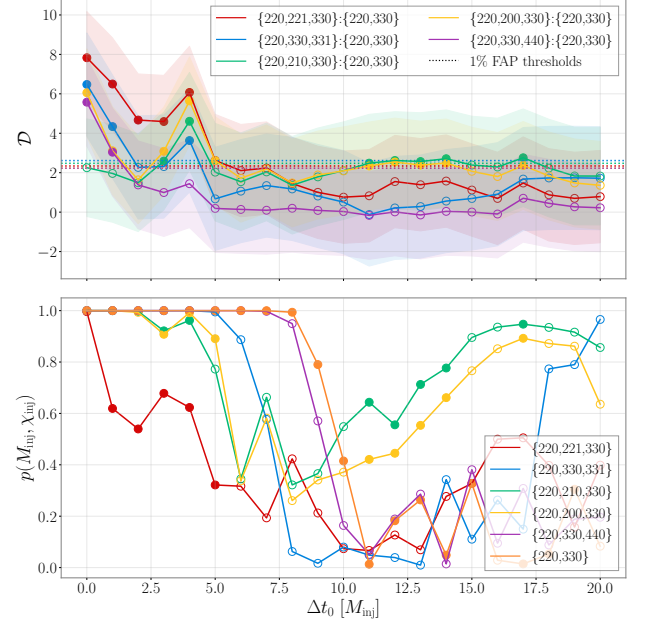
of downsampling.

Fig. 17 shows the results. The resulting $\mathcal{D}$ never exceeds the $\mathcal{D}^{\mathrm{thr,\,agnostic}}_{\{220,221\}:\{220\}}$ and is marginally above $\mathcal{D}^{\mathrm{thr,\,conditional}}_{\{220,221\}:\{220\}}$ only at $\Delta t_0 = 0.5 M_{\mathrm{IMR}}$. For such a marginal detection, we do not test for the presence of a third ringdown mode. The estimated SNR for the {220,221} analysis at $\Delta t_0 = 0.5 M_f$ is 12.96. Thus, with our method, we conclude that there is only marginal evidence of a 221 overtone mode in GW150914 at an early time of the ringdown. These results are consistent with previous analysis of GW150914 with the rational filter which identified a $\mathcal{D}_{\{220,221\}:\{220\}} \approx 2.8$ (presented in [38] as a linear Bayes factor of ≈ 600).

VIII. CONCLUSION

In this work, we extend the QNM rational filter framework to improve the statistical identification of subdominant ringdown modes in GW signals. By employing a frequentist approach, we quantified the threshold based on a given FAP and proposed a systematic workflow for robustly identifying specific QNMs. The methodology is validated using synthetic signals built from NR waveforms, and we apply it to the real event GW150914 to assess the statistical significance of its first overtone.

The QNM rational filter provides an efficient tool for identifying ringdown modes without sampling over amplitudes and phases. The frequentist approach to back-
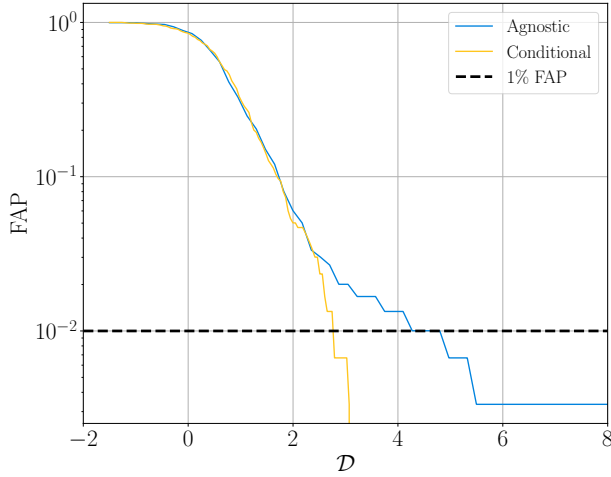
Figure 16: Background study of {220,221}:{220} for GW150914. Results are obtained by injecting a {220}-only waveform into 300 chunks of noise-only data around GW150914. As expected, the threshold corresponding to the 1% FAP is lower with conditional injections than with agnostic injections.

ground noise analysis enhances the robustness of mode identification by reducing false positives, thereby increasing confidence in subdominant mode detections. The improved workflow presented in this study paves the way for more robust tests of strong-field gravity and BH spectroscopy, offering new opportunities to verify the no-hair theorem and probe deviations from GR.
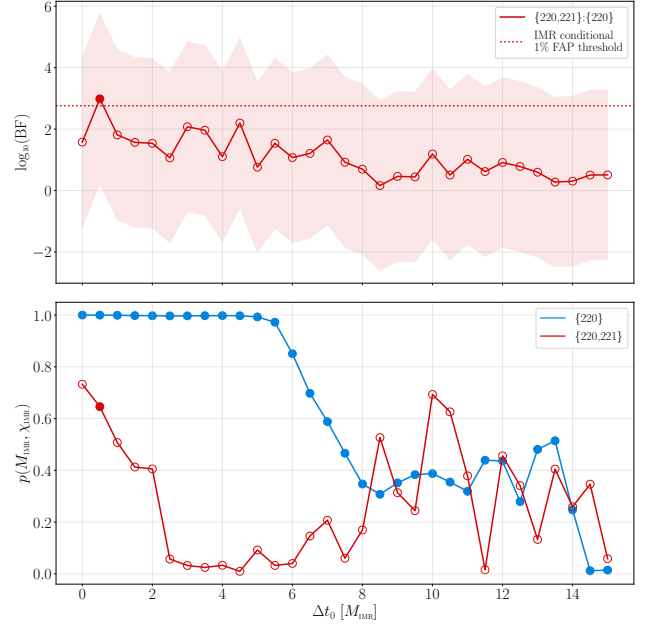
ACKNOWLEDGMENTS

Figure 17: Resulting $\mathcal{D}$ and $p(M_{\mathrm{IMR}}, \chi_{\mathrm{IMR}})$ when comparing the {220, 221}:{220} hypotheses for the ringdown signal of GW150914. The dotted line is the threshold computed from IMR-conditional injections. The filled and unfilled markers correspond to $\mathcal{D} > \mathcal{D}^{\mathrm{thr}}$ and $\mathcal{D} < \mathcal{D}^{\mathrm{thr}}$, respectively. The shaded bands around $\mathcal{D}$ indicate the 99% error bars due to potential noise fluctuations.

Appendix A: The rational filter as a semi-Bayesian analysis

Previous work on the rational filter described it as a Bayesian analysis [21, 37, 38]. Here we provide clarification about the extent to which the rational filter is Bayesian.

The likelihood in Eq. (6) depends on the parameters of the rational filter: M_f and χ_f. The rational filter analysis and the contours computed in the M_f and χ_f plane are Bayesian in this respect. However, when comparing the rational filter analysis to the usual time-domain analysis, it is shown in the supplementary material of Ref. [37] that, for a single detector, the rational filter is related to the usual time-domain likelihood when using the maximum likelihood estimation for mode amplitudes under the assumption of white noise. The maximum-likelihood estimation for amplitudes is a frequentist method[7] which

---

[7] It is further shown in the supplementary material of Ref. [37]

is then combined with the Bayesian inference on M_f and χ_f. For multiple detectors, the rational filter is additionally incoherent in amplitudes between detectors (as discussed in Sec. III A). We thus refer to the rational filter as "semi-Bayesian" to distinguish it from the usual time-domain analysis. The detection statistic $\mathcal{D}$ defined in this work [Eq. (8)] is thus directly analogous to a log Bayes factor but is distinct from the Bayes factor of time-domain analyses. It is still, however, capable of being used to distinguish between different mode hypotheses (as verified throughout the paper and specifically in Sec. VI).

Within this paper, we provide a method to quantify the statistical significance of a given $\mathcal{D}$ in order to distinguish true signal observations from noise more robustly. This improves on more generalized criteria for the necessary Bayes factors required for a statistical significant detection, such as those suggested in [70, 71]. Specifically, the procedure outlined in Sec. III better accounts for the detector noise behavior around a GW event and the impact from any non-Gaussian noise features. However, this work does not comment on the known issues of using Bayes factors when well-motivated priors are not known [72, 73]. This work also does not comment on the time at which constant amplitude QNM fittings are valid, as discussed in [49, 68].
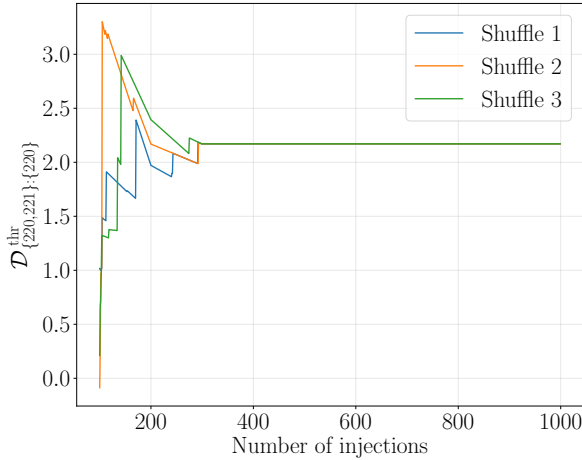
Appendix B: Convergence of $\mathcal{D}^{\mathrm{thr}}$

Figure 18: $\mathcal{D}^{\mathrm{thr}}$ for the {220,221}:{220} background study estimated using different numbers of injections (from 50 to 1000) for three different shuffles .

that the maximum likelihood estimation for mode amplitudes is equivalent to marginalizing over the amplitudes when using an infinitely large flat prior—although this is formally an improper prior.

To determine how many simulations are required in the background study to obtain a convergent $\mathcal{D}^{\mathrm{thr}}$, we carry out the {220,221}:{220} background analysis using 1000 injections. We consider how the threshold $\mathcal{D}^{\mathrm{thr,\,2-ifo}}_{\{220+221\}:\{220\}}$ varies as we consider subsets of the injections of increasing size, starting from 50 simulations up to the full set of 1000. We also shuffle the order of the injections three times to check the robustness of the calculated threshold. Fig. 18 shows that 200 injections are able to obtain a fairly robust threshold, which remains constant when using > 300 injections.

The noise around real GW events is not guaranteed to be stationary or Gaussian which may affect the number of simulations required for a convergent estimate of $\mathcal{D}^{\mathrm{thr}}$. For real GW events, therefore, we would check convergence using the Gelman-Rubin statistic [74] or other statistical tests (e.g., those described in [75]).

Appendix C: Error bars of $\mathcal{D}$

Here, we seek to quantify the error bar of a calculated $\mathcal{D}$. Recall that $\mathcal{D}$ is the ratio between the evidence for two different mode hypotheses [Eq. (8)]. It is common to approximate the evidence of a given hypothesis by its maximum likelihood value [59] (here we neglect the Occam's factor). Let us now quantify the variance of the maximum likelihood for a given hypothesis using Eq. (6) and Eq. (11):

$$\sigma^2 = \mathrm{Var}\left(-\frac{1}{2}\left\langle (s+n)^F \middle| (s+n)^F \right\rangle\right). \tag{C1}$$

For simplicity, we factor out the $-\frac{1}{2}$ term, use the linearity of the filter and the linearity of the inner product [defined in Eq. (11)], and get:

$$4\sigma^2 = \mathrm{Var}\left(\left\langle s^F \middle| s^F \right\rangle + 2\left\langle s^F \middle| n^F \right\rangle + \left\langle n^F \middle| n^F \right\rangle\right). \tag{C2}$$

Using the identity $\mathrm{Var}(A + B) = \mathrm{Var}(A) + \mathrm{Var}(B) + 2\,\mathrm{Cov}(A, B)$ (where Cov denotes the covariance operator) we get:

$$4\sigma^2 = \mathrm{Var}\left(\left\langle s^F \middle| s^F \right\rangle\right) + \mathrm{Var}\left(2\left\langle s^F \middle| n^F \right\rangle + \left\langle n^F \middle| n^F \right\rangle\right) + 2\,\mathrm{Cov}\left(\left\langle s^F \middle| s^F \right\rangle, 2\left\langle s^F \middle| n^F \right\rangle + \left\langle n^F \middle| n^F \right\rangle\right). \tag{C3}$$

The first term $\mathrm{Var}\left(\left\langle s^F \middle| s^F \right\rangle\right) = 0$ because s^F is a constant (not a distribution). Similarly, the covariance term is zero because its first argument $\left\langle s^F \middle| s^F \right\rangle$ is a constant. We are, therefore, left with:

$$4\sigma^2 = \mathrm{Var}\left(2\left\langle s^F \middle| n^F \right\rangle + \left\langle n^F \middle| n^F \right\rangle\right) \tag{C4}$$

$$= 4\,\mathrm{Var}\left(\left\langle s^F \middle| n^F \right\rangle\right) + \mathrm{Var}\left(\left\langle n^F \middle| n^F \right\rangle\right) + 2\,\mathrm{Cov}\left(\left\langle s^F \middle| n^F \right\rangle, \left\langle n^F \middle| n^F \right\rangle\right). \tag{C5}$$

Here we can use the Cauchy-Schwarz inequality for random variables to bound the covariance term:

$$4\sigma^2 \leq 4 \, \mathrm{Var}\left(\langle s^F | n^F \rangle\right) + \mathrm{Var}\left(\langle n^F | n^F \rangle\right)$$
$$+ 2 \sqrt{\mathrm{Var}\left(\langle s^F | n^F \rangle\right) \mathrm{Var}\left(\langle n^F | n^F \rangle\right)}. \quad \text{(C6)}$$

Let us consider the first term and rewrite it in terms of the expectation E:

$$\mathrm{Var}\left(\langle s^F | n^F \rangle\right) = E\left[\langle s^F | n^F \rangle^2\right] - E\left[\langle s^F | n^F \rangle\right]^2. \quad \text{(C7)}$$

The second term is zero because $\langle s^F | n^F \rangle$ is the inner product between a constant (s^F) and a random distribution with zero mean n^F so it's expectation value is zero. We re-express the first term using the commutativity of the inner product for real-valued inputs:

$$\mathrm{Var}\left(\langle s^F | n^F \rangle\right) = E\left[\langle s^F | n^F \rangle \langle n^F | s^F \rangle\right]$$
$$= \langle s^F | s^F \rangle, \quad \text{(C8)}$$

where we use an identity from Eq. (3) of [76]. Eq. (C8) is exactly the residual optimal SNR of the misfiltered signal, denoted by $\mathrm{SNR}_{\mathrm{res}}$. This is related to the fitting factor using an alternative definition [59]:

$$\mathrm{FF}^2 = \max \left(1 - \frac{\mathrm{SNR}_{\mathrm{res}}^2}{\mathrm{SNR}^2}\right), \quad \text{(C9)}$$

where the maximum is taken over the prior parameter space and corresponds to the maximum likelihood values used to approximate the evidence. We are interested in finding the FF between the two hypotheses for calculating $\mathcal{D}$. While we cannot compute the FF analytically, as explained in more detail in Appendix D, we can numerically verify that the FFs between the hypotheses we have tested are all within the range of [0.98, 1) for a variety of systems. Since the ringdown SNR of current GW observations is $\mathcal{O}(10)$, the residual SNR of the misfiltered waveforms are $\ll 1$. When substitute this into Eq. (C6) it eliminates the first and third terms and that remains is:

$$\sigma^2 \leq \mathrm{Var}\left(-\frac{1}{2}\langle n^F | n^F \rangle\right). \quad \text{(C10)}$$

This is exactly the quantity computed in Sec. III to determine the FAP. Therefore, the 99% error bar around the computed $\mathcal{D}$ is exactly $\mathcal{D}^{\mathrm{thr}}$.

Appendix D: FF between non-trivial hypotheses

Here, we provide a detailed explanation about why the rational filter cannot compute the SNR of individual ringdown modes. We first review the effect of applying a rational filter built to remove the (ℓ', m', n') mode to the (ℓ, m, n) QNM [$h(t)$ in Eq. (1)]. The filtered time-series becomes [38]:

$$h^F(t) = A_{\ell m n} B_{\ell' m' n'}^{\ell m n} e^{-(t-t_0)/\tau_{\ell m n}}$$
$$\times \cos\left[2\pi f_{\ell m n}(t - t_0) + \phi_{\ell m n} + \varphi_{\ell' m' n'}^{\ell m n}\right] \text{(D1)}$$

where the amplitude $B_{\ell' m' n'}^{\ell m n}$ and phase $\varphi_{\ell' m' n'}^{\ell m n}$ terms can be computed from the filter:

$$B_{\ell' m' n'}^{\ell m n} e^{i\varphi_{\ell' m' n'}^{\ell m n}} = \mathcal{F}_{\ell' m' n'}(\omega_{\ell m n}). \quad \text{(D2)}$$

So the application of an incorrect filter reduces the amplitude of the signal by a factor of $B_{\ell' m' n'}^{\ell m n}$ and introduces a phase shift of $\varphi_{\ell' m' n'}^{\ell m n}$.

Let us now consider an injection of {220,440} QNMs and analyze it under two hypotheses: a {220,440} mode hypothesis and a {220} hypothesis. For simplicity, we inject a weak 440 mode so that there is little systematic bias in the inferred remnant mass and spin when using the incorrect {220} model, but relax this condition later. The {220} filter with the maximum likelihood completely removes the 220 mode but also partially filters out the 440 mode. From Eq. (13), we could compute the SNR of the 440 mode if we could determine the FF between the {220,440}:{220} models. We can try to do so by considering the alternative definition of the FF in Eq. (C9) [59]. The maximum value corresponds to the filter that produces the maximum likelihood in the prior parameter space. In the specific scenario considered here, $\mathrm{SNR}_{\mathrm{res}}$ is the residual 440 mode SNR after it is partially filtered out by the maximum likelihood {220} filter. Since the rational filter does not affect the noise power or spectrum [37], we can consider the energy of the waveforms instead of the SNR:

$$\frac{\mathrm{SNR}_{\mathrm{res}}^2}{\mathrm{SNR}^2} = \frac{\int_0^\infty |h_t^F|^2 dt}{\int_0^\infty |h(t)|^2 dt}, \quad \text{(D3)}$$

where recall that h_t^F is the misfiltered waveform in the time domain and is computed analogously using Eq. (5). In the case of a (ℓ, m, n) mode being incorrectly filtered by the rational filter of the mode (ℓ', m', n'), we can analytically compute Eq. (D3) using Eq. (D1) to get:

$$\frac{\mathrm{SNR}_{\mathrm{res}}^2}{\mathrm{SNR}^2} = \frac{\left(B_{440}^{220}\right)^2 \left(1 + \omega_{440}^2 \tau_{440}^2 + \cos[2(\phi_{440} + \varphi_{440}^{220})] - \omega_{440}\tau_{440}\sin[2(\phi_{440} + \varphi_{440}^{220})]\right)}{\left(\sum_{\ell m n = \{220,440\}} 1 + \omega_{\ell m n}^2 \tau_{\ell m n}^2 + \cos(2\phi_{\ell m n}) - \omega_{\ell m n}\tau_{\ell m n}\sin(2\varphi_{\ell m n}^{\ell' m' n'})\right)}. \quad \text{(D4)}$$
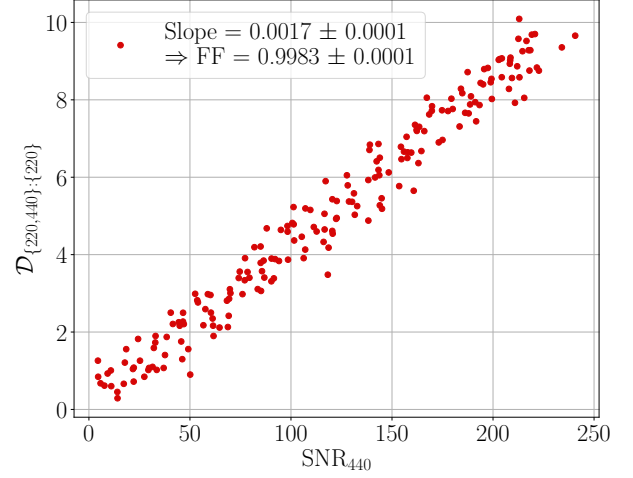
Eq. (D4) depends on ϕ_{440}, and M_{inj} and χ_{inj} of the system that control $B_{\ell m n}^{\ell' m' n'}$ and $\varphi_{\ell m n}^{\ell' m' n'}$ through Eq. (D2).

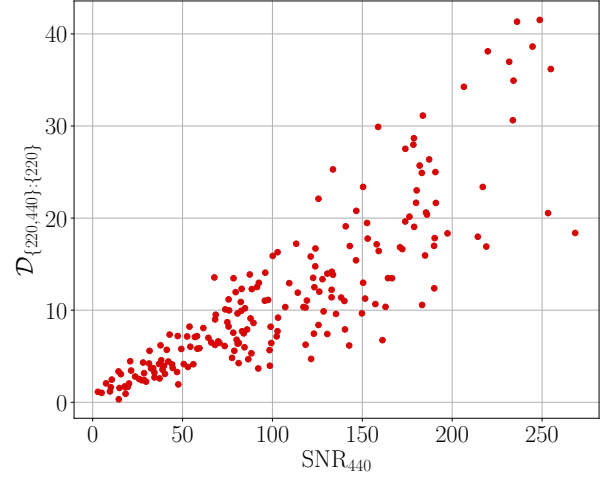No information about the phase ϕ_{440} is inferred by the rational filter, and the constraints on M_{inj} and χ_{inj}, for most realistic GW observations, are imprecise. If we relax our condition that the 440 mode is weak and hence does not bias the maximum likelihood parameter estimate from the {220} filter, then Eq. (D4) becomes even more complex with a summation over the {220,440} modes in the denominator. Therefore, we cannot determine the values of Eq. (D4) and thus the FF between any non-trivial hypotheses. This means that even when comparing the same hypotheses, the FF remain constant only when ϕ_{440}, M_{inj}, and χ_{inj} are fixed, as illustrated in Fig. 19. Fig. 19a shows that if the injection parameters are fixed, there is a linear relationship between SNR and $\sqrt{\mathcal{D}/\ln(10)}$—the FF between the hypotheses is a well-defined value. In Fig. 19b, however, when the injection parameters are randomized, there is a large uncertainty in the slope (i.e. the FF between hypotheses varies) so that there is no clear mapping relationship between $\sqrt{\mathcal{D}/\ln(10)}$ and the single mode SNR.

(a) Fixed injection parameters.

(b) Randomized injection parameters.

Figure 19: $\mathcal{D}_{\{220,440\}:\{220\}}$ as a function of the 440-mode SNR.

REFERENCES

[1] C. V. Vishveshwara, Stability of the schwarzschild metric, Physical Review D: Particles and Fields **1**, 2870 (1970).

[2] S. A. Teukolsky, Perturbations of a rotating black hole. I. Fundamental equations for gravitational, electromagnetic, and neutrino-field perturbations, The Astrophysical Journal **185**, 635 (1973).

[3] S. Chandrasekhar and S. Detweiler, The quasi-normal modes of the Schwarzschild black hole, Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences **344**, 441 (1975).

[4] E. W. Leaver and S. Chandrasekhar, An analytic representation for the quasi-normal modes of Kerr black holes, Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences **402**, 285 (1985).

[5] K. D. Kokkotas and B. G. Schmidt, Quasi-Normal Modes of Stars and Black Holes, Living Reviews in Relativity **2**, 2 (1999).

[6] E. Berti, V. Cardoso, and A. O. Starinets, Quasinormal modes of black holes and black branes, Classical and Quantum Gravity **26**, 163001 (2009).

[7] W. Israel, Event Horizons in Static Vacuum Space-Times, Physical Review **164**, 1776 (1967).

[8] B. Carter, Axisymmetric black hole has only two degrees of freedom, Physical Review Letters **26**, 331 (1971).

[9] S. W. Hawking, Black holes in general relativity, Communications in Mathematical Physics **25**, 152 (1972).

[10] D. C. Robinson, Uniqueness of the Kerr Black Hole, Physical Review Letters **34**, 905 (1975).

[11] P. O. Mazur, Proof of uniqueness of the Kerr-Newman black hole solution, Journal of Physics A: Mathematical and General **15**, 3173 (1982).

[12] O. Dreyer, B. Kelly, B. Krishnan, *et al.*, Black-hole spectroscopy: Testing general relativity through gravitational-wave observations, Classical and Quantum Gravity **21**, 787 (2004).

[13] E. Berti, V. Cardoso, and C. M. Will, Gravitational-wave spectroscopy of massive black holes with the space interferometer LISA, Physical Review D: Particles and Fields **73**, 064030 (2006).

[14] M. Isi, M. Giesler, W. M. Farr, *et al.*, Testing the No-Hair Theorem with GW150914, Physical Review Letters **123**, 111102 (2019).

[15] M. Isi, W. M. Farr, M. Giesler, *et al.*, Testing the black-hole area law with GW150914, Physical Review Letters **127**, 011103 (2021), arXiv:2012.04486 [astro-ph, physics:gr-qc].

[16] T. L. S. Collaboration, J. Aasi, B. P. Abbott, *et al.*, Advanced LIGO, Classical and Quantum Gravity **32**, 074001 (2015).

[17] F. Acernese, M. Agathos, K. Agatsuma, *et al.*, Advanced Virgo: A second-generation interferometric gravitational wave detector, Classical and Quantum Gravity **32**, 024001 (2014).

[18] T. Akutsu, M. Ando, K. Arai, *et al.*, Overview of KAGRA: Detector design and construction history, Progress of Theoretical and Experimental Physics **2021**, 05A101 (2020), https://academic.oup.com/ptep/article-pdf/2021/5/05A101/37974994/ptaa125.pdf.

[19] G. Carullo, W. Del Pozzo, and J. Veitch, Observational black hole spectroscopy: A time-domain multimode analysis of GW150914, Physical Review D **99**, 123029 (2019).

[20] M. Isi and W. M. Farr, Analyzing black-hole ringdowns (2021), arXiv:2107.05609 [astro-ph, physics:gr-qc].

[21] S. Ma, K. Mitman, L. Sun, *et al.*, Quasinormal-mode filters: A new approach to analyze the gravitational-wave ringdown of binary black-hole mergers, Physical Review D **106**, 084036 (2022), arXiv:2207.10870 [gr-qc].

[22] H. Siegel, M. Isi, and W. M. Farr, Analyzing black-hole ringdowns. II. Data conditioning, Physical Review D **111**, 044070 (2025).

[23] Y. Dong, Z. Wang, H.-T. Wang, *et al.*, A practical Bayesian method for gravitational-wave ringdown analysis with multiple modes (2025), arXiv:2502.01093 [gr-qc].

[24] E. Finch and C. J. Moore, Frequency-domain analysis of black-hole ringdowns, Physical Review D: Particles and Fields **104**, 123034 (2021).

[25] E. Finch and C. J. Moore, Searching for a Ringdown Overtone in GW150914, Physical Review D **106**, 043005 (2022), arXiv:2205.07809 [astro-ph, physics:gr-qc].

[26] C. D. Capano, M. Cabero, J. Westerweck, *et al.*, Multimode Quasinormal Spectrum from a Perturbed Black Hole, Physical Review Letters **131**, 221402 (2023).

[27] Y.-F. Wang, C. D. Capano, J. Abedi, *et al.*, A frequency-domain perspective on GW150914 ringdown overtone (2023), arXiv:2310.19645 [astro-ph, physics:gr-qc].

[28] R. Brito, A. Buonanno, and V. Raymond, Black-hole spectroscopy by making full use of gravitational-wave modeling, Physical Review D **98**, 084038 (2018).

[29] E. Maggio, H. O. Silva, A. Buonanno, *et al.*, Tests of general relativity in the nonlinear regime: A parametrized plunge-merger-ringdown gravitational waveform model, Physical Review D **108**, 024043 (2023).

[30] D. P. Mihaylov, S. Ossokine, A. Buonanno, *et al.*, py-SEOBNR: A software package for the next generation of effective-one-body multipolar waveform models (2023), arXiv:2303.18203 [gr-qc].

[31] M. van de Meent, A. Buonanno, D. P. Mihaylov, *et al.*, Enhancing the SEOBNRv5 effective-one-body waveform model with second-order gravitational self-force fluxes, Physical Review D **108**, 124038 (2023).

[32] C. Pacilio, S. Bhagwat, and R. Cotesta, Simulation-based inference of black hole ringdowns in the time domain, Physical Review D **110**, 083010 (2024).

[33] LIGO Scientific Collaboration and Virgo Collaboration, B. P. Abbott, *et al.*, Observation of Gravitational Waves from a Binary Black Hole Merger, Physical Review Letters **116**, 061102 (2016).

[34] LIGO Scientific and Virgo Collaborations, B. P. Abbott, *et al.*, Tests of General Relativity with GW150914, Physical Review Letters **116**, 221101 (2016).

[35] J. C. Bustillo, P. D. Lasky, and E. Thrane, Black-hole spectroscopy, the no-hair theorem, and GW150914: Kerr versus Occam, Physical Review D **103**, 024041 (2021).

[36] A. Ghosh, R. Brito, and A. Buonanno, Constraints on quasinormal-mode frequencies with LIGO-Virgo binary–black-hole observations, Physical Review D **103**, 124041 (2021).

[37] S. Ma, L. Sun, and Y. Chen, Black hole spectroscopy by mode cleaning, Physical Review Letters **130**, 141401 (2023), arXiv:2301.06705 [astro-ph, physics:gr-qc].

[38] S. Ma, L. Sun, and Y. Chen, Using rational filters to uncover the first ringdown overtone in GW150914, Physical Review D **107**, 084010 (2023).

[39] M. Crisostomi, K. Dey, E. Barausse, *et al.*, Neural posterior estimation with guaranteed exact coverage: The ringdown of GW150914, Physical Review D **108**, 044029 (2023).

[40] H.-T. Wang, Z. Wang, Y. Dong, *et al.*, Reanalyzing the ringdown signal of GW150914 using the F-statistic method (2024), arXiv:2411.13333 [gr-qc].

[41] A. Correia, Y.-F. Wang, J. Westerweck, *et al.*, Low evidence for ringdown overtone in GW150914 when marginalizing over time and sky location uncertainty (2024), arXiv:2312.14118 [gr-qc].

[42] R. Abbott, T. D. Abbott, S. Abraham, *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), GW190521: A Binary Black Hole Merger with a Total Mass of 150 $M_\odot$, Physical Review Letters **125**, 101102 (2020).

[43] H. Siegel, M. Isi, and W. M. Farr, Ringdown of GW190521: Hints of multiple quasinormal modes with a precessional interpretation, Physical Review D **108**, 064008 (2023).

[44] C. D. Capano, J. Abedi, S. Kastha, *et al.*, Estimating false alarm rates of sub-dominant quasi-normal modes in GW190521, Classical and Quantum Gravity **41**, 245009 (2024).

[45] T. L. S. Collaboration, the Virgo Collaboration, the KAGRA Collaboration, *et al.*, Tests of General Relativity with GWTC-3 (2021), arXiv:2112.06861 [astro-ph, physics:gr-qc, physics:hep-th].

[46] R. Cotesta, G. Carullo, E. Berti, *et al.*, Analysis of Ringdown Overtones in GW150914, Physical Review Letters **129**, 111102 (2022).

[47] G. Carullo, R. Cotesta, E. Berti, *et al.*, Carullo *et al.* Reply:, Physical Review Letters **131**, 169002 (2023).

[48] M. Isi and W. M. Farr, Comment on "Analysis of Ringdown Overtones in GW150914", Physical Review Letters **131**, 169001 (2023).

[49] M. Giesler, S. Ma, K. Mitman, *et al.*, Overtones and Nonlinearities in Binary Black Hole Ringdowns (2024), arXiv:2411.11269 [gr-qc].

[50] X. Li, L. Sun, R. K. L. Lo, *et al.*, Angular emission patterns of remnant black holes, Physical Review D **105**, 024016 (2022).

[51] P. Welch, The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms, IEEE Transactions on Audio and Electroacoustics **15**, 70 (1967).

[52] LVK noise curves, https://dcc.ligo.org/LIGO-T2000012/public.

[53] G. Lovelace, C. O. Lousto, J. Healy, *et al.*, Modeling the source of GW150914 with targeted numerical-relativity simulations, Classical and Quantum Gravity **33**, 244002 (2016).

[54] T. A. Clarke, M. Isi, P. D. Lasky, *et al.*, Striking the right tone: Towards a self-consistent framework for measuring black hole ringdowns (2024), arXiv:2402.02819 [astro-ph, physics:gr-qc].

[55] H. Zhu, H. Siegel, K. Mitman, *et al.*, Black Hole Spectroscopy for Precessing Binary Black Hole Coalescences (2023), arXiv:2312.08588 [astro-ph, physics:gr-qc].

[56] The LIGO Scientific Collaboration and the Virgo Collab-

oration, R. Abbott, *et al.*, GWTC-2.1: Deep extended catalog of compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run, Physical Review D **109**, 022001 (2024).

[57] L. S. Collaboration and V. Collaboration, GWTC-2.1: Deep extended catalog of compact binary coalescences observed by LIGO and virgo during the first half of the third observing run - parameter estimation data release (2022).

[58] C. Pacilio, S. Bhagwat, F. Nobili, *et al.*, Postmerger: A flexible mapping of ringdown amplitudes for non-precessing binary black holes (2024), arXiv:2408.05276.

[59] N. Cornish, L. Sampson, N. Yunes, *et al.*, Gravitational Wave Tests of General Relativity with the Parameterized Post-Einsteinian Framework, Physical Review D **84**, 062003 (2011), arXiv:1105.2088 [astro-ph, physics:gr-qc].

[60] J. D. Romano and Neil. J. Cornish, Detection methods for stochastic gravitational-wave backgrounds: A unified treatment, Living Reviews in Relativity **20**, 2 (2017).

[61] S. R. Cook, A. Gelman, and D. B. Rubin, Validation of Software for Bayesian Models Using Posterior Quantiles, Journal of Computational and Graphical Statistics **15**, 675 (2006), 27594203.

[62] S. Talts, M. Betancourt, D. Simpson, *et al.*, Validating Bayesian Inference Algorithms with Simulation-Based Calibration (2020), arXiv:1804.06788 [stat].

[63] A. H. Mroué, M. A. Scheel, B. Szilágyi, *et al.*, Catalog of 174 Binary Black Hole Simulations for Gravitational Wave Astronomy, Physical Review Letters **111**, 241104 (2013).

[64] M. Boyle, D. Hemberger, D. A. B. Iozzo, *et al.*, The SXS collaboration catalog of binary black hole simulations, Classical and Quantum Gravity **36**, 195006 (2019).

[65] E. Thrane, P. D. Lasky, and Y. Levin, Challenges for testing the no-hair theorem with current and planned gravitational-wave detectors, Physical Review D **96**, 102004 (2017).

[66] S. Bhagwat, M. Okounkova, S. W. Ballmer, *et al.*, On choosing the start time of binary black hole ringdowns, Physical Review D **97**, 104065 (2018).

[67] M. Giesler, M. Isi, M. Scheel, *et al.*, Black hole ringdown: The importance of overtones, Physical Review X **9**, 041060 (2019), arXiv:1903.08284 [gr-qc].

[68] V. Baibhav, M. H.-Y. Cheung, E. Berti, *et al.*, Agnostic black hole spectroscopy: Quasinormal mode content of numerical relativity waveforms and limits of validity of linear perturbation theory, Physical Review D **108**, 104020 (2023).

[69] P. J. Nee, S. H. Völkel, and H. P. Pfeiffer, Role of black hole quasinormal mode overtones for ringdown analysis, Physical Review D **108**, 044032 (2023).

[70] R. E. Kass and A. E. and Raftery, Bayes Factors, Journal of the American Statistical Association **90**, 773 (1995).

[71] H. Jeffreys, *The Theory of Probability*, Oxford Classic Texts in the Physical Sciences (OUP Oxford, 1998).

[72] A. Gelman, D. Simpson, and M. Betancourt, The Prior Can Often Only Be Understood in the Context of the Likelihood, Entropy **19**, 555 (2017).

[73] C. P. Robert, *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*, 2nd ed., Springer Texts in Statistics (Springer, New York, 2007).

[74] A. Gelman and D. B. Rubin, Inference from Iterative Simulation Using Multiple Sequences, Statistical Science **7**, 457 (1992).

[75] Model checking and diagnostics — PyMC 2.3.6 documentation, https://pymcmc.readthedocs.io/en/latest/modelchecking.html.

[76] M. Vallisneri, Use and abuse of the Fisher information matrix in the assessment of gravitational-wave parameter-estimation prospects, Physical Review D **77**, 042001 (2008).