Gating-and-inpainting perspective on GW150914 ringdown overtone: Understanding the data analysis systematics

Yi-Fan Wang (王一帆),[1,2,3,*] Collin D. Capano,[4,5] Jahed Abedi,[6,7,2,3] Shilpa Kastha,[8,9,10,2,3] Badri Krishnan,[11,2,3] Alex B. Nielsen,[7] Alexander H. Nitz,[4] and Julian Westerweck[2,3,12]

[1]*Max-Planck-Institut für Gravitationsphysik (Albert-Einstein-Institut), Am Mühlenberg 1, D-14476 Potsdam, Germany*
[2]*Max-Planck-Institut für Gravitationsphysik (Albert-Einstein-Institut), Callinstraße 38, D-30167 Hannover, Germany*
[3]*Leibniz Universität Hannover, D-30167 Hannover, Germany*
[4]*Department of Physics, Syracuse University, Syracuse, New York 13244, USA*
[5]*Department of Physics, University of Massachusetts, Dartmouth, Massachusetts 02747, USA*
[6]*Beijing Institute of Mathematical Sciences and Applications (BIMSA), Huairou District, Beijing 101408, People's Republic of China*
[7]*Department of Mathematics and Physics, University of Stavanger, NO-4036 Stavanger, Norway*
[8]*Saha Institute of Nuclear Physics, 1/AF Bidhannagar, Kolkata 700064, India*
[9]*Homi Bhabha National Institute, Training School Complex, Anushaktinagar, Mumbai 400094, India*
[10]*Niels Bohr International Academy, Niels Bohr Institute, Blegdamsvej 17, 2100 Copenhagen, Denmark*
[11]*Institute for Mathematics, Astrophysics and Particle Physics, Radboud University, Heyendaalseweg 135, 6525 AJ Nijmegen, The Netherlands*
[12]*Institute for Gravitational Wave Astronomy and School of Physics and Astronomy, University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom*

(Received 7 November 2023; accepted 22 September 2025; published 10 October 2025)

We revisit the recent debate on the evidence for an overtone in the black hole ringdown of GW150914 using an independent data-analysis pipeline. By gating and inpainting the data, we discard the contamination from earlier parts of the gravitational wave signal before ringdown. This enables parameter estimation to be conducted in the frequency domain, which is mathematically equivalent to the time domain method. We keep the settings as similar as possible to the previous studies by Cotesta *et al.* [Phys. Rev. Lett. **129**, 111102 (2022)] and Isi *et al.* [Phys. Rev. Lett. **123**, 111102 (2019); arXiv:2202.02941] which yielded conflicting results on the Bayes factor of the overtone. Our aim is to understand how different data analysis systematics, including sampling rates, erroneous timestamps, and the frequency resolution of the noise power spectrum, would influence the statistical significance of an overtone. Our main results indicate the following: (i) a low-resolution estimation of the noise power spectrum tends to diminish the significance of overtones, (ii) adjusting the start time to a later digitized point reduces the significance of overtones, and (iii) overtone evidence varies with different sampling rates if the start time is too early, indicating that the overtone is a poor model, hence we propose a convergence test to verify the validity of an overtone model. With these issues addressed, we find the Bayes factors for the overtone to range from 10 to 26 in a range of times centered at the best-fit merger time of GW150914, which supports the existence of an overtone in agreement with the conclusions of Isi *et al.* in the above papers. These results are obtained by keeping the start time and sky location fixed, enabling a direct comparison with other work. Marginalizing over these parameters would lower the Bayes factor to 1 for the evidence of an overtone.

DOI: 10.1103/3gqn-297f

I. INTRODUCTION

The gravitational waves (GWs) emitted during the black hole (BH) ringdown consists of a superposition of damped sinusoids known as quasinormal modes (QNMs) [1]. According to the no-hair theorem [2,3], the characteristic frequencies and damping times are exclusively determined by the astrophysical BH's mass and spin. When multiple modes are identified from the ringdown's GW, the BH mass

*Contact author: yifan.wang@aei.mpg.de

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Open access publication funded by the Max Planck Society.

and spin can be inferred independently and cross-checked. This is often known as BH spectroscopy [4], and offers an unequivocal way to test the validity of general relativity (GR).

The first evidence of a QNM is reported for GW150914 [5–7], in which the single $(\ell, m, n) = (2, 2, 0)$ mode is found with a frequency and decay time consistent with the GR expectation from the full signal analysis. Theoretical studies [8] suggest the detection of a secondary QNM would likely only occur once Advanced LIGO [9], Advanced Virgo [10] and KAGRA [11] have reached their design sensitivities. Nevertheless, the event GW190521 is discovered to have an unexpectedly high redshifted remnant mass $\sim 260 M_\odot$ [12,13]. The evidence of a subdominant mode (3,3,0) from GW190521 is reported with a Bayes factor of 56 [14–17]; also see [18] for an alternative interpretation involving the (2,1,0) mode. By fitting numerical relativity data, Refs. [19,20] shows that the QNM description can be valid as early as the merger stage, provided that overtones are considered, however, Ref. [21] reports that the overtones identified in this region are not physical, leaving it as an open question. Starting the analysis from the merger time, Isi *et al.* [22] reports the first detection of a (2,2,1) mode from GW150914 with a significance of 3.6σ and shows the parameters to be consistent with the prediction of GR.

However, Cotesta *et al.* [23] claims that the detection of the overtone in Isi *et al.* [22] is noise-dominated. Their reanalysis shows that the Bayes factor in favor of the overtone compared to only the fundamental mode is less than 1 around the merger time, hence no evidence for the overtone is found. Nevertheless, Isi and Farr [24] revisits the analysis and claims to be unable to reproduce the results in Cotesta *et al.* [23] and shows that the Bayes factor of the model including the (2,2,1) mode indicates the presence of the overtone.

The data analysis frameworks employed in [23] and [22,24] are fundamentally similar. To remove the influence of the preringdown GW signal, both compute the likelihood in the time domain [25,26], as diagonalizing the covariance matrix with a Fourier transform used by conventional GW parameter estimation is no longer applicable. In practice, data analysis systematics, such as different data length and noise power spectral density (PSD) estimation [27,28], can result in the inconsistent findings between [23] and [22,24], highlighting the need for a comprehensive understanding of the data analysis techniques employed in ringdown overtone analysis. In this study, we utilize an independent gating-and-inpainting framework to revisit GW150914, aiming to offer a new perspective to understand the analysis for overtones. Further research into the GW150914 overtone, such as [29–33], utilizes various methodologies, however, we do not directly compare our findings with these studies since they essentially explore a different likelihood. We limit our comparative analysis to [22–24,34], given that the underlying likelihoods are mathematically equivalent (see proof in Appendix A). As a result, any observed discrepancies are attributable to data analysis systematics.

In [14] a new approach is proposed to excise the contamination from the GW signal prior to the ringdown by gating and inpainting the data following the original idea of [35]. This is mathematically equivalent to the time domain method of [22–24] and enables us to keep the analysis in the frequency domain. Thereby, we can take advantage of the parameter estimation package PyCBCinference [36] by using several existing modules, such as data conditioning, PSD estimation with Welch's method [37,38], frequency-domain likelihood calculation, which are well tested and were used in numerous previous studies, e.g., [39,40].

We explore the impacts of different sampling rates on the GW150914 overtone significance as the sampling rate appears to differ between [23] and [22,24] (16384 Hz vs 2048 Hz). In addition, the extremely short decay time of the overtone, typically $\lesssim 1$ ms, motivates us to examine the matched-filtering of high frequency contents of the signal, which is straightforward by examining the waveform in the frequency domain (see Sec. IV). As emphasized by [23], the merger time of GW150914 is subject to uncertainty, and we follow [22–24] to select a set of discrete times centered around the best-fit merger time as the ringdown starting time.[1] We also carefully ensure the starting time is precisely implemented in PyCBCinference, rather than rounding up to that of a LIGO data sample, in light of our findings that the evidence of an overtone is highly sensitive to such approximations (see Sec. V) because of the overtone's rapid decay. We also find that the frequency resolution in the noise PSD estimate affects the statistical significance of the (2,2,1) overtone (refer to Sec. VI). We verify our method with numerical relativity simulations and report the results in Sec. VII.

II. GATED GAUSSIAN LIKELIHOOD

We briefly review the gated Gaussian likelihood that employs data gating and inpainting [14,26,35]. The conventional likelihood used in GW parameter estimation for Gaussian and stationary noise is

$$\mathcal{L}(\mathbf{n}) = \frac{1}{\sqrt{(2\pi)^N |C|}} \exp\left[-\frac{1}{2}\mathbf{n}^T C^{-1} \mathbf{n}\right] \tag{1}$$

where $\mathbf{n}$ is the noise vector with N elements, C is the covariance matrix of $\mathbf{n}$. By the stationary assumption, C is a Toeplitz matrix, and can be further diagonalized by a discrete Fourier transform basis matrix if the noise data is circulant. Therefore, the likelihood can be greatly simplified in the frequency domain as

[1] See [41] which reports the analysis for GW150914 overtone by sampling and marginalizing over the sky location and starting time. This marginalization would lower the Bayes factor to 1.

$$\ln \mathcal{L}(\mathbf{n}) = -\frac{1}{2}\langle \mathbf{n}|\mathbf{n}\rangle - \ln\sqrt{(2\pi)^N |C|} \tag{2}$$

where the inner product is defined as

$$\langle \mathbf{a}|\mathbf{b}\rangle = 4\Re\left\{\frac{1}{T}\sum_{p=1}^{\lfloor (N-1)/2\rfloor}\frac{\tilde{\mathbf{a}}_p^\dagger \tilde{\mathbf{b}}_p}{S_{n,p}}\right\} \tag{3}$$

in which $\tilde{\mathbf{a}}$ is the discrete Fourier transform of $\mathbf{a}$ with N samples across time T, and $\dagger$ means complex conjugation, $S_{n,p}(f)$ is the one-sided PSD of the noise and the subscript p indicates the pth element. When a GW is present, the noise can be obtained by subtracting the GW waveform $\mathbf{h}$ from the detector measurement $\mathbf{d}$, so that the likelihood of a GW waveform is $\mathcal{L}(\mathbf{n}) = \mathcal{L}(\mathbf{d}-\mathbf{h})$.

Ringdown inference aims to exclusively analyze signals after the remnant BH enters a linear perturbation regime, hence the preringdown contamination should be excised. However, this would break the circularity condition due to the abrupt onset of the ringdown signal, thus the covariance matrix cannot be diagonalized simply by a Fourier transform [26]. One needs to numerically invert the noncirculant covariance matrix in Eq. (1), as implemented by the time domain analysis [22–24].

Alternatively, Ref. [35] proposes, and Ref. [14] applies in ringdown analysis, a relation between the inversion of the covariance matrix from truncated data, $\mathbf{n}_{\rm tr}$, and that from the complete data, $\mathbf{n}$, by replacing (inpainting) the excised data with $\mathbf{x}$. Without loss of generality, we express the complete data $\mathbf{n}$ as the concatenation of three vectors, which is $\mathbf{n} = \mathbf{n}_1 \oplus \mathbf{n}_2 \oplus \mathbf{n}_3$, where $\oplus$ denotes the concatenation operation; the border of $\mathbf{n}_2$ and $\mathbf{n}_3$ delineates the preringdown and ringdown stage, and $\mathbf{n}_2$ is long enough to cover the entire preringdown GW signals. The truncated data can be expressed by $\mathbf{n}_{\rm tr} = \mathbf{n}_1 \oplus \mathbf{n}_3$. The gated Gaussian likelihood aims to replace $\mathbf{n}_2$ with a vector $\mathbf{x}$, as $\mathbf{n}_{\rm inpaint} = \mathbf{n}_1 \oplus \mathbf{x} \oplus \mathbf{n}_3$, such that

$$\mathbf{n}_{\rm tr} C_{\rm tr}^{-1} \mathbf{n}_{\rm tr} = \mathbf{n}_{\rm inpaint}{}^T C^{-1} \mathbf{n}_{\rm inpaint} \tag{4}$$

where $C_{\rm tr}^{-1}$ denotes the covariance matrix of $\mathbf{n}_{\rm tr}$. The solution is obtained by solving the Toeplitz linear equation (for a proof see Appendix A)

$$[C^{-1}(\mathbf{n}_1 \oplus \mathbf{x} \oplus \mathbf{n}_3)]_{\rm inpaint} = \mathbf{0}_{\rm inpaint} \tag{5}$$

where the subscript asserts this equation is valid only in the rows corresponding to the data being inpainted. Given an M-dimensional inpainting vector $\mathbf{x}$, Eq. (5) is an M-dimensional Toeplitz linear equation with the time complexity scaling as M^2. Since the right-hand side of Eq. (4) resumes the use of C^{-1}, one can diagonalize it with a discrete Fourier transform and thus perform the analysis in the frequency domain once $\mathbf{x}$ is obtained via Eq. (5).

III. RESULTS OF THE OVERTONE EVIDENCE

We reanalyze the overtone of GW150914 data [42] using similar settings to [22–24]. Our waveform model is

$$h_+ + i h_\times = \sum_{\ell mn} {}_{-2}S_{\ell mn}(\iota,\varphi;\chi_f) A_{\ell mn} e^{i(\Omega_{\ell mn} t + \phi_{\ell mn})}, \tag{6}$$

where ${}_{-2}S_{\ell mn}$ are the spin-weighted spheroidal harmonics [43,44], ι and φ are the inclination angle and azimuthal angle; $\Omega_{\ell mn} = 2\pi f_{\ell mn} + i/\tau_{\ell mn}$ is the complex frequency, $f_{\ell mn}$ and $\tau_{\ell mn}$ are the characteristic frequency and decay time exclusively determined by M_f and χ_f, the mass and spin of the remnant BH; $A_{\ell mn}$ and $\phi_{\ell mn}$ are the amplitude and initial phase, which in principle can be determined by the initial conditions of the BH perturbation, however, due to lack of concrete knowledge, we treat them as free parameters to be inferred from the data.

We follow Ref. [23] and use the reference GPS time $t_{\rm ref} = 1126259462.42323$ s as a median of GW150914's merger time recorded by LIGO Hanford, and expand the analysis by scanning different starting times within $t_{\rm ref} \pm 1.5$ ms, corresponding to 2σ uncertainty around $t_{\rm ref}$. We ensure that the ringdown analysis starts precisely at the designated time by reconstructing the subsampling data point by time-shifting in the frequency domain, instead of rounding it up to the nearest available discrete data point. A comparison with not doing so can be found in Sec. V. The amplitude priors on A_{220} and A_{221} are uniform in $[0, 5\times10^{-20}]$; the phase priors on ϕ_{220} and ϕ_{221} are uniform in $[0, 2\pi]$; all of which are identical to [23]. The prior of final mass and final dimensionless spin is chosen to be uniform in [35,140] $M_\odot$ and [0,0.99]; the inclination angle, azimuthal angle, polarization angle and sky localization are fixed to the values given by [22–24], which in turn are obtained from the maximum likelihood value from the analysis of the complete signal of GW150914. We use the Dynesty sampler [45] to sample the likelihood. The inspiral-merger-ringdown time length of the dominant mode of a 30-30$M_\odot$ binary is 0.94 s with a lower frequency 20 Hz estimated using the waveform model IMRPhenomD [46,47], hence we choose the gating-and-inpainting window to be 1 s to remove the preringdown signal. We examine 8 s of data centered on the starting time, padded with an additional 4 s of data both at the beginning and end to address the boundary wraparound from the whitening filter, and taper the whitening filter to zero within the 4 s duration [38]. The autocorrelation function of the LIGO data generally decays to zero within a few seconds, as quantitatively demonstrated in Fig. 9 of [26]. The Appendix B demonstrates the robustness of these choices by changing the gating length to 2 s or changing the data analysis duration from 8 s to 4 s (in addition to the padding time); we obtain consistent results. We also use four different sampling rates f_s from 1024 Hz to 8192 Hz to study the impact of different upper frequency limits on overtone

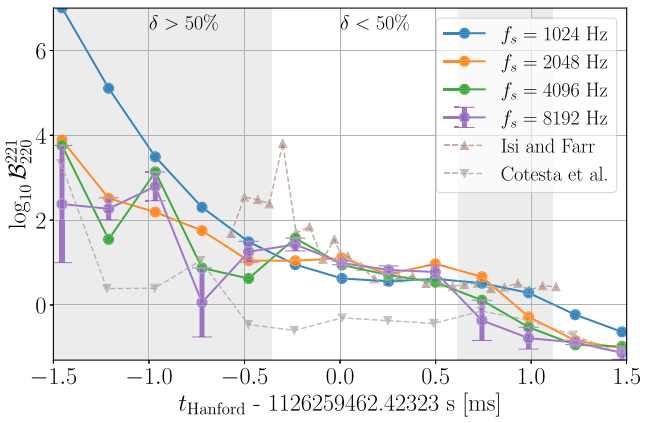

FIG. 1. The logarithm of Bayes factors comparing the $(2,2,0)+(2,2,1)$ model and the (2,2,0)-only model with respect to a variety of starting times for sampling rates $f_s = 1024/2048/4096/8192$ Hz. As a comparison, we plot the Bayes factors obtained from [23,24]. We also plot symmetric error bars with length $\mathcal{B}_{f_s=8192} \times \delta$. The shaded regions depict where the results from different sampling rates have not converged, quantified by $\delta > 50\%$, while in the nonshaded region all have $\delta < 50\%$.

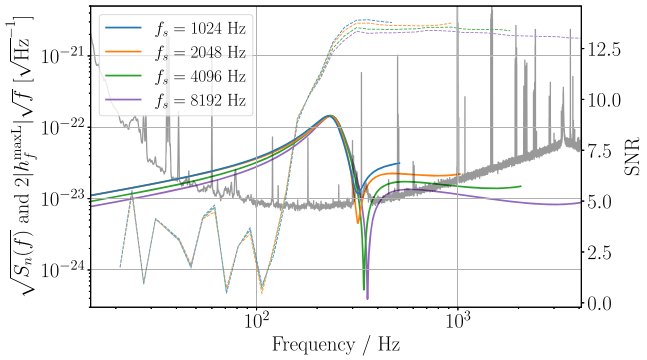

FIG. 2. The solid lines present the ringdown overtone waveform generated with the maximum likelihood parameters for four different sampling rates, with the scale shown on the left y-axis. The dashed lines present the matched-filtering SNR as a function of the frequency upper limit of the integration. The lower frequency limit is 20 Hz. The ringdown starting time is $t_{\rm ref} - 0.75$ ms. The gray curve shows the ASD $\sqrt{S_n(f)}$.

inference. No higher sampling rate is considered because the LIGO data calibration is only valid from 10 Hz to 5 kHz [42,48].

The findings of our study are presented in Fig. 1, which shows the logarithm of the Bayes factor $\log_{10} \mathcal{B}^{221}_{220}$, comparing the waveform model with modes $(2,2,0)+(2,2,1)$ to that with only the (2,2,0) mode at various starting times. We choose a stopping criterion for the Dynesty sampler where the change in the natural logarithm of the Bayes factor is less than 0.1. Consequently, the uncertainty in the Bayes evidence ratio is approximately $2\log_{10} e^{0.1} \sim 0.08$ in Fig. 1, and therefore negligible. For comparison, we plot the Bayes factors reported by [23,24], respectively. Note that Ref. [28] later corrects the Bayesian evidence given in Ref. [23], showing an improved agreement with Refs. [22,24] but the Bayesian log evidence for an overtone is still negative throughout almost the entire time range. We consider the $f_s = 8192$ Hz runs as our fiducial results and quantify the convergence of Bayes factors by measuring the fractional difference between the highest and second highest sampling rate results, i.e., $\delta = |\mathcal{B}_{f_s=8192} - \mathcal{B}_{f_s=4096}|/\mathcal{B}_{f_s=8192}$, as shown in Fig. 1.

We notice the intriguing trend that the Bayes factors from different sampling rates only start to converge after $t_{\rm ref} - 0.25$ ms, where we quantify convergence by the criterion $\delta < 50\%$. Prior to that, there are noticeable disagreements from different sampling rates. Results with $f_s = 1024$ Hz yield the strongest evidence for the (2,2,1) mode (we will discuss in more details in the next section). We regard this divergence as an indicator that the overtone model does not consistently match the data across different frequency bands, implying that it is not an adequate model to describe the preringdown phase. This approach is conceptually similar to the chi-square signal consistency test commonly used in GW searches [49].

In Fig. 2, we further examine the behavior of the matched-filtering signal-to-noise ratio (SNR), defined as $\langle d|h\rangle/\sqrt{\langle h|h\rangle}$. A well-fitting template should exhibit a monotonic increase in the accumulated SNR as the high-frequency cutoff increases until the noise-dominated regime. Our results show that the majority of the SNR is accumulated below ~300 Hz. Extending the integration beyond this frequency only introduces inconsistencies between the overtone template and the data, further supporting the model's inadequacy. However, we note that identifying a more suitable model for this region falls outside the scope of this work.

In light of this observation, we propose a discriminator that utilizes the (non)convergence of results from various sampling rates to determine the region where the ringdown overtone model is applicable, as opposed to the region where preringdown contamination is present. At the late time around $t_{\rm ref} + 1$ ms, we again observed discrepancies of different sampling rates, suggesting the overtone model is again not applicable.

Around the best-fit merger time of GW150914, specifically in $[-0.25, 0.5]$ ms, we obtain converged Bayes factors from four different sampling rates consistently greater than 1 in favor of the existence of the (2,2,1) mode. In particular, at $t_{\rm ref} - 0.25$ ms, which was considered as the merger time by Isi *et al.* [22], we find $\mathcal{B}^{221}_{220} = 26$, the median of A_{221} deviates from zero with 2.8σ; at $t_{\rm ref}$, we find $\mathcal{B}^{221}_{220} = 10$, and a 2.5σ nonzero A_{221}, which indicates positive but moderate evidence for the presence of the (2,2,1) mode.

Notably, our Bayes factors agree with those from Isi and Farr [24] in and only in the convergence region $[-0.25, 0.5]$ ms. Nevertheless, there is a discrepancy at -0.3 ms with a notable outlier identified to have

$\log_{10}\mathcal{B} \sim 4$ by [24]. Given their finer time stride, we further perform additional analyses with $f_s = 2048$ Hz with more finely spaced starting times, but cannot reproduce the significant Bayes factor. After $t_{\rm ref} + 0.75$ ms, our results are consistent with Cotesta *et al.* [23], indicating no overtone is found at a late time. To understand how various methods affect the statistical significance of the overtone, we note that one of the differences is that [22–24] use a 0.2 s and 0.1 s duration for data analysis, respectively, while we use 16 s to account for the nonzero whitening filter over a few seconds; thus we conclude a sufficiently long analysis duration can be one of the factors to enhance the statistical significance of finding an overtone.

IV. UNDERSTANDING THE DISCREPANCY FROM DIFFERENT SAMPLING RATES

Prior to $t_{\rm ref} - 0.25$ ms, we notice a divergence of results from four different sampling rates. To better understand its origin, we choose a particular time $t_{\rm ref} - 0.75$ ms, which shows discrepancies, and analyze the results of parameter estimation in depth. In Fig. 2, we present the Fourier transform of the ringdown overtone waveforms, $h_f^{\rm maxL}$, from parameters corresponding to the maximum likelihood sample (note they are slightly different for different sampling rates). We plot $2|h_f^{\rm maxL}|\sqrt{f}$ (with scale shown on the left Y-axis) such that the area under the square ratio between the waveform and the amplitude spectral density (ASD) indicates the optimal SNR, $\sqrt{\langle h|h\rangle}$. We also explicitly plot the matched-filtering SNR accumulated from a lower frequency of 20 Hz up to a frequency upper limit shown in the x-axis, with the scale shown on the right y-axis. Also note that the waveforms shown in Fig. 2 are obtained by numerically Fourier transforming the ringdown overtone signals. The Fourier transforms of the inspiral-merger-ringdown signals would fall off faster than any power law at high frequency, see, e.g., Ref. [46].

We notice that the $f_s = 1024$ Hz result tends to favor a waveform with higher amplitude at the Nyquist frequency 512 Hz. Examining the overtone amplitude A_{221} reveals that a stronger (2,2,1) mode is favored, which manifests as the tilt at high frequency due to the short decay time (∼1.5 ms) of the overtone. The rapid decay leads to a broader frequency-domain representation of the waveform. However, when the data analysis is extended to higher frequency bands, this strong (2,2,1) mode is no longer preferred by the standard of Bayes factors. The SNR also gradually decays from 13.76 to 13.71, 13.32, and 13.01 for sampling rates increasing from 1024 Hz to 8192 Hz. The discrepancies suggest the starting time is too early and overtone templates do not match the data well. The low sampling rate result tends to be more affected by the contamination from the preringdown to produce a high SNR.

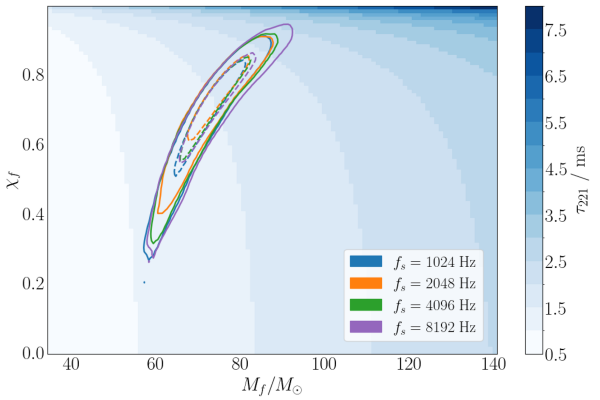

FIG. 3. The marginal posterior of M_f and χ_f for four different sampling rates with analysis starting time $t_{\rm ref} - 0.75$ ms. In the background, we plot the value of τ_{221} as expected in GR as a function of the mass and spin of a Kerr BH. The solid and dashed lines show the 90% and 50% credible regions, respectively.

To further illustrate this with the entire posterior instead of a single point from maximum likelihood, we plot the mass and spin posterior distribution in Fig. 3. For all (M_f, χ_f), we also compute the characteristic decay time of the overtone, τ_{221}, predicted from GR as the background of the figure. Figure 3 shows that the 8192 Hz result extends to a region with a longer decay time, while the low sampling rate results are more restricted to a shorter decay time. The posterior of τ_{221} and A_{221} is plotted in Fig. 4, showing the parameters' negative correlation. The 8192 Hz result favors a longer decay time τ_{221}, hence a higher posterior density at

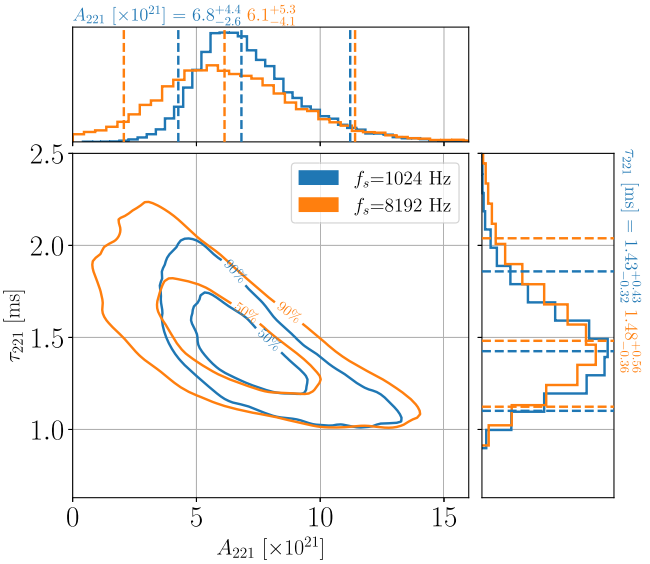

FIG. 4. A comparison of the posteriors for A_{221} and τ_{221} using two different sampling rates, 1024 Hz and 8192 Hz, with analysis starting time $t_{\rm ref} - 0.75$ ms. The contours denote the 90% and 50% credible regions. The shaded region shows the posterior probability density of the $f_s = 1024$ Hz result.

$A_{221} = 0$. Using the Savage-Dickey density ratio, the density of $A_{221} = 0$ directly determines $\mathcal{B}^{221}_{220}$. We thus conclude that the (dis) favoring of shorter τ_{221}, which in turn is directly related to the signal intensity in high-frequency bands, by an inappropriate sampling rate can bias the posterior of A_{221}, and thus enhance or weaken the evidence for a (2,2,1) mode. The findings underscore the importance of examining the frequency spectrum in order to determine an appropriate time range within which a ringdown overtone model is applicable.

V. IMPACT OF AN INACCURATE STARTING TIME

The original gating and inpainting formalism is potentially subject to a subtle caveat that the starting and ending time of the inpainting can only land on a specific data point due to the discrete nature of the sampled data. As illustrated by Fig. 1 and Sec. III, the evidence of an overtone is sensitive to the ringdown starting time at a submillisecond level because of the rapid decay of an overtone. This issue is particularly severe for a lower sampling rate with a coarser time resolution. Consequently, it is necessary to ensure a precise starting time for the ringdown analysis.

We have addressed this issue by reconstructing subdata points from the sampled data at the starting time of ringdown, $t_{\rm ringdown}$, which is achieved by time shifting in the frequency domain by an offset between $t_{\rm ringdown}$ and the time stamp of the floor-nearest data sample, $t_{\rm nearest}$. To visualize, the discrete data samples from the LIGO Hanford with $f_s = 2048$ Hz and those being reconstructed are plotted in Fig. 5. To verify this procedure, we plot the absolute value of the data before and after being time shifted in the frequency domain, they are expected to be identical and this is indeed the case as shown in the inset plot of Fig. 5.

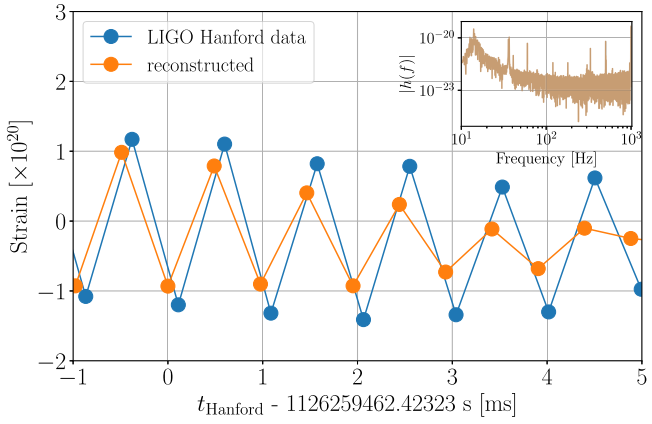

FIG. 5. Reconstructing the subsample corresponding to the starting time of ringdown, $t_{\rm ringdown}$. This is done by time shifting the data by an offset of the difference of $t_{\rm ringdown}$ and $t_{\rm nearest}$ which corresponds to the nearest data sample from LIGO Hanford. In the inset we also plot the absolute value of the data before and after being time shifted in the frequency domain. The figure shows they are identical as expected.

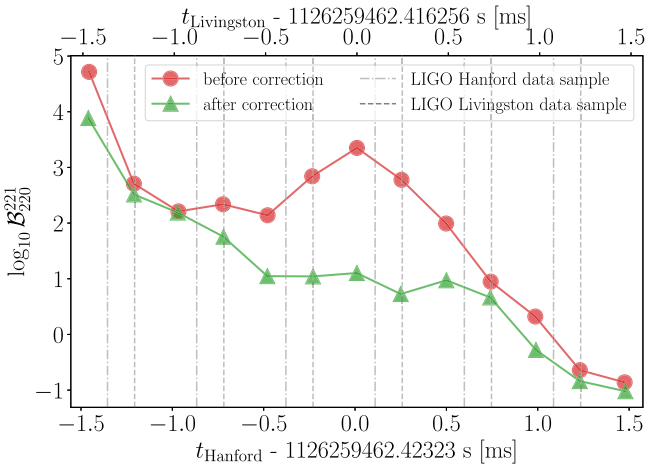

FIG. 6. The results of logarithm Bayes factor from sampling rate 2048 Hz before and after accounting for the issue when the ringdown starting time lands on a subsample. The vertical (dot) dashed lines indicate the time stamp of the data samples from LIGO Hanford or Livingston.

In Fig. 6 we plot the Bayes factor $\log_{10} \mathcal{B}^{221}_{220}$ from $f_s = 2048$ Hz before and after accounting for this issue. To guide the eyes, we also plot vertical dashed lines for the time stamps for the sampled data of LIGO Hanford and Livingston. Before addressing this problem, the starting time of ringdown, or equivalently, the ending time of inpainting, is rounded up to the floor-nearest time of a data sample. This effectively results in an earlier and incoherent starting time between LIGO Hanford and Livingston, biasing the Bayes factors toward higher values due to the contamination from preringdown signals. Likewise, rounding up to a subsequent discretized data sample would reduce the overtone evidence. At a sampling rate of 16384 Hz, this effect should be insignificant (refer to Refs. [27,28] for further discussions on time discretization in time-domain data analysis pipelines). Nonetheless, our Fig. 6 illustrates that Bayes factors can be significantly biased by several orders of magnitude if an accurate starting time is not considered when using a 2048 Hz sampling rate.

VI. IMPACT OF NOISE POWER SPECTRUM ESTIMATION

This section investigates the impact of different PSD estimation techniques within our pipeline. According to Ref. [28], the choice of PSD estimation method affects the overtone evidence utilizing either the frequency-domain Welch method [37] or a time-domain auto correlation function (ACF) based approach. However, we identified an additional systematic error: a lower frequency resolution in the noise PSD estimation significantly diminishes the evidence for the (2,2,1) mode.

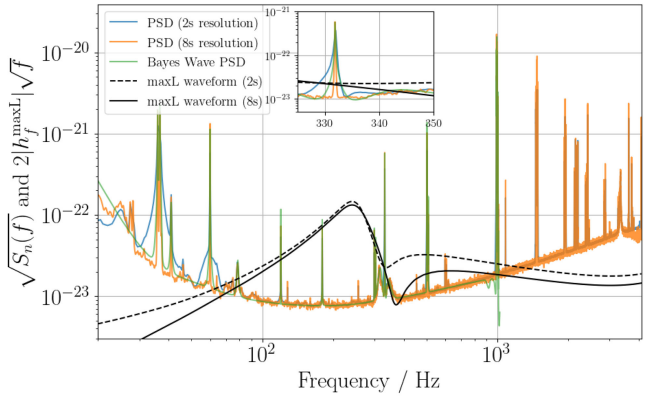

FIG. 7. Comparison of the different PSD estimations obtained by this work using a resolution of 8 s, that from [23] with a 2 s resolution, and the PSD obtained by BayesWave used for GW150914 parameter estimation. The maximum likelihood waveforms of an overtone model corresponding to the two PSDs are also plotted.

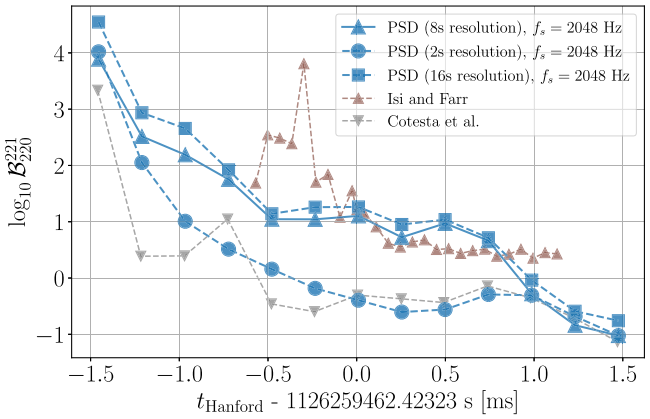

FIG. 8. The Bayes factors of a overtone model by using the PSD of 16 s, 8 s, and 2 s resolution, respectively. Here only shows the results from a sampling rate of 2048 Hz, and the results from other sampling rates are consistent. The Bayes factors from [23,24] are plotted for comparison.

Figure 7 shows the PSD used in this study, estimated using the Welch method, which involves Fourier transforming several 8-second data segments, totaling 512 seconds, as described in Ref. [38]. For comparison, the PSD from Ref. [23] is plotted, which was derived using the authors' released configuration [50], with 2-second data segments in the Fourier transform. Additionally, we present the PSD estimated by the BayesWave algorithm [51–53], employed by the LIGO and Virgo collaboration for the GW150914 parameter estimation and publicly available [54]. A notable feature is the enhanced resolution of a noise line structure in the PSD used in this study, which we primarily attribute to the use of 8-second data segments among other technical factors such as window functions. On the other hand, we observed no visual differences between the PSD obtained via the Welch method and that from numerically Fourier transforming a time-domain ACF, so we plot only the result from the Welch method.

We conduct Bayesian parameter estimation for models both with and without an overtone, replacing the PSD with an estimate from a 2 s resolution as depicted in Fig. 7. The Bayes factors for the overtone are shown in Fig. 8. Using a PSD with a less resolved line structure significantly reduces evidence for the overtone. Only results using a 2048 Hz sampling rate are presented, but similar conclusions hold across other sampling rates, where the logarithm of the Bayes factors remains below 0 around the merger time. A less resolved PSD effectively implies a higher noise level, decreasing the likelihood of an overtone model. This is evident in Fig. 7, which shows the maximum likelihood waveforms for modes (2,2,0) + (2,2,1). A 2 s resolution PSD tends to favor a lower amplitude for (2,2,1).

We also perform an inference using a PSD derived from the Welch method with a 16 s resolution, the results are consistent with that of the 8 s resolution as shown in Fig. 8. We also plot the Bayes factors reported in Refs. [23,24] for comparison. The Bayes factors derived from the 2 s resolution PSD exhibit a higher level of consistency with those from Ref. [23]. Consequently, we determined that utilizing a shorter data segment, resulting in a less resolved line structure for noise PSD estimation, can be one of the contributing factors in diminishing the statistical significance of a ringdown overtone analysis.

VII. VERIFICATION BY NUMERICAL RELATIVITY INJECTIONS

Finally, we use injections of numerical relativity simulations to validate our QNM overtone inference based on gating-and-inpainting. It is worth mentioning that statistical validation for similar purposes has been conducted in Ref. [15], using injections of the numerical relativity surrogate model NRSur7dq4 [55], and in Ref. [17] using IMRPhenomTPHM [56], a time-domain phenomenological inspiral-merger-ringdown waveform model, to verify the capability of the gating-and-inpainting to effectively recover the QNM (3,3,0) mode. Reference [41] has also used injections of a frequency-domain phenomenological waveform model IMRPhenomXPHM [57] to demonstrate the effectiveness when marginalizing over the QNM start time.

In this study, we utilize the numerical relativity simulation SXS:BBH:0305 from the Simulating eXtreme Spacetimes catalog [58,59] converted to the convention of LIGO Algorithms Library's coordinate system [60]. This simulation is designed to resemble the properties of the first detection GW150914. The detector frame total mass is set to $72M_\odot$, with the source located at a luminosity distance of 410 Mpc; the sky location, polarization, inclination, and coalescence phase are the same as the numerical relativity injections in Ref. [23]. The component spins, aligned with the orbital angular momentum, are 0.33 and −0.44,

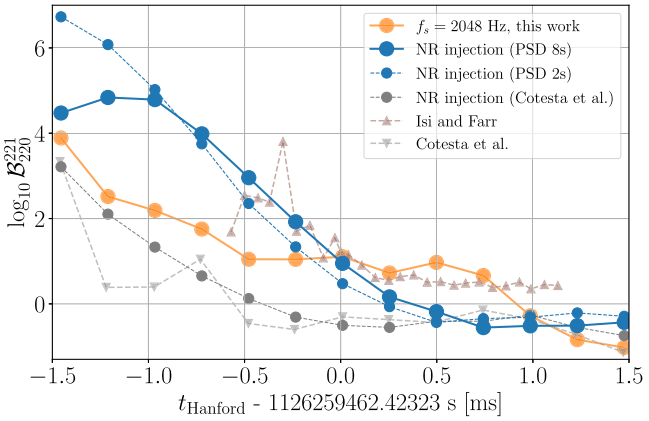

FIG. 9. The Bayes factors of an overtone drawn from the zero noise injection with a numerical relativity simulation SXS:BBH:0305, using a PSD estimation with 8 s and 2 s resolution, respectively. For comparison, we also plot the results extracted from Ref. [23] (their Fig. 1) using the same numerical relativity injection into zero noise. We also replot the Bayes factor reported by this work and that from Refs. [22–24] for a reference.

respectively, and the mass ratio is 1.22. We inject the simulation in zero noise, hence the inference results should be interpreted as an average over an infinite set of noise realizations, each drawn from the same underlying Gaussian distribution. We compute the likelihood using a sampling rate of 2048 Hz and the same noise PSD as in Sec. III. To further verify our observation about the impact of PSD estimation, we also use the PSD from a 2s resolution as is discussed in Sec. VI.

We report our results in Fig. 9. As a consequence of zero noise injection, we find that the Bayes factor for an overtone detected in the numerical relativity signal, utilizing a 8s PSD resolution, is much smoother than in the real data. It monotonically decreases as the starting time advances and stabilizes ∼0.5 ms after the merger with no evidence of an overtone. At the merger epoch, the Bayes factor is ∼9, consistent with the inferences obtained from the real data. Just before and after the merger, the Bayes factor decays approximately exponentially, reducing by an order of magnitude when the start time advances by ∼0.25 ms, again demonstrating the rapid decay of the evidence of an overtone. For comparison, using a PSD with 2s resolution, the Bayes factor is found to be slightly lower around the merger epoch, aligned with our observations in Sec. VI. We also replot the Bayes factor from real data drawn by this work and from Refs. [22–24] in Fig. 9 for a reference.

For additional comparison, we extract the results from numerical relativity injection from Ref. [23] (their Fig. 1), which were derived using the identical SXS simulation injected into zero noise. These results should be directly comparable to ours obtained with the 2 s PSD. Indeed, we notice a remarkable similarity between the two sets of results, except for an apparent time delay. If we manually time shift the Bayes factor from [23] by about 0.8 ms, the Bayes factor results align strikingly well. While we cannot explain the exact numeric value of the apparent time shift, this provides positive evidence that our analysis using PyCBCinference would converge with the results from [23], aside from an apparent shift of the time stamps.

VIII. CONCLUSION AND DISCUSSION

We revisit the GW150914 overtone using an independent gating-and-inpainting analysis, with settings as similar as possible to [22–24]; notably we use a 8 s data analysis duration with another 4 s both at the head and the rear accounting for the length of the whitening filter. We examine the frequency spectral content of the recovered waveforms and interpret the divergence of results from different sampling rates as evidence for where the overtone model is not valid to be matched with the data. When the results converge around the merger time of GW150914, we find Bayes factor values fall within the range of 10 to 26, which supports the existence of an overtone, in agreement with the conclusion of [22,24]. We show that starting the analysis too early or too late will lead to discrepancies for different sampling rates. At too early times, an inappropriately low sampling rate tends to favor a signal with a shorter decay time and thus biases the estimation toward stronger evidence for the (2,2,1) mode. In light of these discoveries, we propose a new strategy, from the data analysis perspective, by analyzing the convergence of different sampling rates to determine the validity of the overtone model, complementary to the efforts of, e.g., Refs. [20,21,61], which address whether the overtone is valid physically.

We also investigate the systematics of data analysis. The statistical significance of an overtone would be influenced by an imperfect sampling rate, incorrect starting time, and poor resolution of the PSD estimation. Furthermore, we test the robustness of the Bayes factor results using a numerical relativity simulation. Additional checks, including changing to another sampling parameter parametrization, changing the gating length or the data analysis duration, are reported in Appendix B, the results show the Bayes factors remain consistent.

We summarize our findings below.

(i) Using a low sampling rate, such as 1024 Hz, leads to inconsistent overtone significance when starting the analysis before the merger, as demonstrated in Fig. 1. Therefore, we propose a consistency test from different sampling rates to identify the valid region for an overtone. A sampling rate of 2048 Hz or higher is sufficient for overtone analysis when starting at or after the merger.

(ii) Given the extremely short time of an overtone (e.g., with a final mass of $60M_\odot$ and a final spin of 0.8 leading to a decay time of 1.3 ms for the (2,2,1) overtone), rounding to the nearest discretized data sample yields an inconsistent and inaccurate starting time across two or more GW detectors. This, in turn, biases the statistical evidence, as

illustrated in Fig. 5 with a 2048 Hz sampling rate example. Hence, we accurately determine the starting time by reconstructing the subsample.

(iii) Using a short data segment to estimate the PSD of noise does not resolve the line structure well in the frequency domain, effectively increasing the noise level and reducing the evidence for the overtone model, as demonstrated in Fig. 8. Therefore, we adopt an 8 s long data segment to resolve the PSD line structure for overtone analysis.

Recently, the authors of Refs. [22,24] published a comment showing that increasing the analysis duration and correcting the approximation of starting time discretization can alleviate the discrepancies [27]. However, the authors of [23] replied that the logarithmic Bayes factors are still negative after addressing the comments [28]. Our method in the current work does not have any of the aforementioned limitations as we have used a long analysis duration (16 s) and reconstructed the subsampling data point to ensure the starting time is precise (Sec. V). Our work shows that using these more robust choices will affect the results in the direction of enhancing the statistical significance of an overtone.

We release the scripts to reproduce this work and the posterior files at [62].

ACKNOWLEDGMENTS

Y.-F. W. acknowledges the Max Planck Gesellschaft, the Atlas computing team, and AEI Hannover where the majority of this work was done; he also thanks Gregorio Carullo, Max Isi and Will Farr for insightful comments. J. A. was supported by ROMFORSK Grant Project No. 302640. C. D. C. acknowledges support from NSF Award No. PHY-2309356. A. H. N. acknowledges support from NSF Grant No. PHY-2309240. S. K. acknowledges support from the Villum Investigator program supported by the VILLUM Foundation (Grant No. VIL37766) and the DNRF Chair program (Grant No. DNRF162) by the Danish National Research Foundation. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant Agreement No. 101131233. This research has made use of data or software obtained from the Gravitational Wave Open Science Center (gwosc.org), a service of the LIGO Scientific Collaboration, the Virgo Collaboration, and KAGRA. This material is based upon work supported by NSF's LIGO Laboratory which is a major facility fully funded by the National Science Foundation, as well as the Science and Technology Facilities Council (STFC) of the United Kingdom, the Max-Planck-Society (MPS), and the State of Niedersachsen/Germany for support of the construction of Advanced LIGO and construction and operation of the GEO600 detector. Additional support for Advanced LIGO was provided by the Australian Research Council. Virgo is funded, through the European Gravitational Observatory (EGO), by the French Centre National de Recherche Scientifique (CNRS), the Italian Istituto Nazionale di Fisica Nucleare (INFN) and the Dutch Nikhef, with contributions by institutions from Belgium, Germany, Greece, Hungary, Ireland, Japan, Monaco, Poland, Portugal, Spain. K. A. G. R. A. is supported by Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan Society for the Promotion of Science (JSPS) in Japan; National Research Foundation (NRF) and Ministry of Science and ICT (MSIT) in Korea; Academia Sinica (AS) and National Science and Technology Council (NSTC) in Taiwan.

DATA AVAILABILITY

The data that support the findings of this article are openly available [62].

APPENDIX A: PROOF FOR THE SOLUTION OF INPAINTING

Recall that the Gaussian likelihood for the noise $\mathbf{n}$ is

$$\ln \mathcal{L} = -\frac{1}{2}\mathbf{n}^T C^{-1}\mathbf{n} - \ln\sqrt{(2\pi)^N |C|}. \tag{A1}$$

As first introduced in [35], they construct an inpainting operator

$$F = \mathbf{1} - AM^{-1}A^T C^{-1} \tag{A2}$$

where $M = A^T C^{-1} A$. The matrix A is an "extraction matrix" with the size $N \times M$, where N and M are the numbers of elements of $\mathbf{n}$, and $\mathbf{x}$, the bad data to be inpainted, respectively. Explicitly, A is an identity matrix in the rows corresponding to $\mathbf{x}$ and zeros elsewhere

$$A = \begin{pmatrix} & 0 & & \\ 1 & \cdots & & \\ & 1 & \cdots & \\ & & 1 & \cdots \\ & \cdots & & \\ & & \cdots & 1 \\ & 0 & & \end{pmatrix} \tag{A3}$$

Such construction will have the desirable property that, after acting F on $\mathbf{n}$, any elements in the gating region will not impact the computation of $\mathbf{n}^T C^{-1} F \mathbf{n}$.

We offer another perspective, which is mathematically equivalent to [35], by considering the inverse of the covariance from the truncated data. Without loss of generality, we express $\mathbf{n}_{\text{inpaint}}$ as the concatenation of the truncated data and the (yet unknown) inpainting data,

$\mathbf{n}_{\rm inpaint} = \mathbf{n}_{\rm tr} \oplus \mathbf{x}$ (the more general case that $\mathbf{x}$ is in the middle of $\mathbf{n}_{\rm inpaint}$ can be obtained by acting permutation matrix on it, and the following derivation remains the same). Hence the covariance matrix can be formally expressed by a block matrix, its inversion is

$$C^{-1} = \begin{pmatrix} C_{\rm tr} & B \\ B^T & D \end{pmatrix}^{-1} = \begin{pmatrix} a & b \\ b^T & d \end{pmatrix} \tag{A4}$$

where B, D, a, b, d are all block matrices yet unknown. Because of the inversion relation, we have

$$\begin{aligned} C_{\rm tr} a + B b^T &= \mathbf{1} \\ C_{\rm tr} b + B d &= \mathbf{0} \end{aligned} \tag{A5}$$

where 1 and 0 are the unity and zero matrix, respectively. Express B by the second line of Eq. (A5) and insert to the first line, one gets

$$C_{\rm tr}^{-1} = a - b d^{-1} b^T \tag{A6}$$

Hence the likelihood of truncated data can be written as

$$\mathbf{n}_{\rm tr}^T C_{\rm tr}^{-1} \mathbf{n}_{\rm tr} = \mathbf{n}_{\rm tr}^T a \mathbf{n}_{\rm tr} - \mathbf{n}_{\rm tr}^T b d^{-1} b^T \mathbf{n}_{\rm tr} \tag{A7}$$

In the main text we have introduced the solution to be

$$C^{-1}(\mathbf{n}_{\rm tr} \oplus \mathbf{x})_{\rm inpaint} = \mathbf{0}_{\rm inpaint} \tag{A8}$$

This can be expressed as

$$b^T \mathbf{n}_{\rm tr} + d\mathbf{x} = \mathbf{0} \tag{A9}$$

Hence

$$\mathbf{x} = -d^{-1} b^T \mathbf{n}_{\rm tr} \tag{A10}$$

Therefore

$$\begin{aligned} &(\mathbf{n}_{\rm tr} \oplus \mathbf{x})^T C^{-1} (\mathbf{n}_{\rm tr} \oplus \mathbf{x}) \\ &= (\mathbf{n}_{\rm tr} \oplus \mathbf{x})^T \begin{pmatrix} a & b \\ b^T & d \end{pmatrix} (\mathbf{n}_{\rm tr} \oplus \mathbf{x}) \\ &= \mathbf{n}_{\rm tr}^T a \mathbf{n}_{\rm tr} - \mathbf{n}_{\rm tr}^T b d^{-1} b^T \mathbf{n}_{\rm tr} \end{aligned} \tag{A11}$$

Together with Eq. (A7) we have proved that inpainting with $\mathbf{x}$ will resume the use of C^{-1} in the likelihood. As discussed in [26], constructing the inpainting filter directly as in Eq. (A2) invokes inverting the $A^T C^{-1} A$ which requires M^3 time complexity. However, as we use Levinson-Durbin Recursion [63,64] in SCIPY [65] to solve the Toeplitz linear equation for inpainting data in Eq. (A8), it only requires the M^2 time complexity.

Moreover, we verified the robustness of our gating-and-inpainting solver by confirming that the over-whitened waveform, $h(f)/S_h(f)$, is indeed zero within the inpainted area, as expected by solving Eq. (5).

APPENDIX B: CHECKS ON THE IMPACTS OF RINGDOWN AMPLITUDE PARAMETRIZATION, DATA ANALYSIS DURATION, AND GATING LENGTH

To verify the robustness of our inference configurations, we performed various checks by sampling over another overtone amplitude parametrization, adjusting the data analysis duration, and the length of gating-and-inpainting. The results for the evidence of an overtone are consistent with the baseline models used to present our main results in Sec. III.

We first consider another overtone amplitude parametrization. When sampling the likelihood distribution, we notice that the label switching issue between the (2,2,0) and (2,2,1) mode would sometimes occur, in which the sampler would explore where the A_{220} is almost zero, and A_{221} is favored associated with much heavier remnant mass. This is because the (2,2,1) mode of the template is locked on the (2,2,0) signal in the data.

In light of this issue, we choose to sample the relative amplitude between the (2,2,1) and (2,2,0) mode, $A_{221}^{\rm rel} = A_{221}/A_{220}$, instead of the absolute amplitude. We choose $A_{221}^{\rm rel}$ to be uniform in [0, 5], and $\log_{10} A_{220}$ uniformly distributed in $[-24, -19]$.

The results of Bayes factors are shown in Fig. 10. We only show the results from $f_s = 2048$ Hz; other sampling rates present consistent conclusions. We note that the two parametrizations agree with each other well. At zero epoch, the relative amplitude parametrization slightly prefers a lower $\mathcal{B}_{220}^{221}$, which can be attributed to a slightly higher

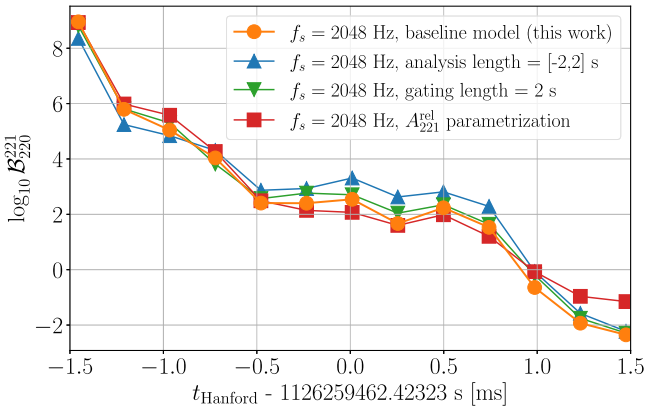

FIG. 10. Comparison of Bayes factors for a different parametrization that samples $A_{221}^{\rm rel}$ and $\log_{10} A_{220}$, a different data analysis duration and a different gating-and-inpainting length. The results with adjusted settings are consistent the baseline settings used to report the main results of this work.

weight in the prior space for $A_{221} = 0$. At a sufficiently late time, the relative parametrization favors a Bayes factor being 1 between the $(2, 2, 1) + (2, 2, 0)$ and (2,2,0) modes, i.e., no preference for any one of the models.

We have also changed the data analysis duration from $[-4, 4]$ s around the merger epoch of GW150914 used in Sec. III to $[-2, 2]$ s. Separately, we changed the gating-and-inpainting length from 1 s to 2 s. The Bayes factors for an overtone are presented in Fig. 10 and broadly consistent with the baseline settings. Overall, the results demonstrate the robustness of the Bayes factor results in the main text.

[1] C. V. Vishveshwara, Scattering of gravitational radiation by a schwarzschild black-hole, Nature (London) **227**, 936 (1970).

[2] Werner Israel, Event horizons in static vacuum space-times, Phys. Rev. **164**, 1776 (1967).

[3] B. Carter, Axisymmetric black hole has only two degrees of freedom, Phys. Rev. Lett. **26**, 331 (1971).

[4] Olaf Dreyer, Bernard J. Kelly, Badri Krishnan, Lee Samuel Finn, David Garrison, and Ramon Lopez-Aleman, Black hole spectroscopy: Testing general relativity through gravitational wave observations, Classical Quantum Gravity **21**, 787 (2004).

[5] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Observation of gravitational waves from a binary black hole merger, Phys. Rev. Lett. **116**, 061102 (2016).

[6] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Tests of general relativity with GW150914, Phys. Rev. Lett. **116**, 221101 (2016); **121**, 129902(E) (2018).

[7] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Tests of general relativity with binary black holes from the second LIGO-Virgo gravitational-wave transient catalog, Phys. Rev. D **103**, 122002 (2021).

[8] Miriam Cabero, Julian Westerweck, Collin D. Capano, Sumit Kumar, Alex B. Nielsen, and Badri Krishnan, Black hole spectroscopy in the next decade, Phys. Rev. D **101**, 064044 (2020).

[9] J. Aasi *et al.* (LIGO Scientific Collaboration), Advanced LIGO, Classical Quantum Gravity **32**, 074001 (2015).

[10] F. Acernese *et al.* (VIRGO Collaboration), Advanced Virgo: A second-generation interferometric gravitational wave detector, Classical Quantum Gravity **32**, 024001 (2015).

[11] Homare Abe *et al.* (KAGRA Collaboration), The current status and future prospects of kagra, the large-scale cryogenic gravitational wave telescope built in the kamioka underground, Galaxies **10**, 3 (2022).

[12] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), GW190521: A binary black hole merger with a total mass of $150M_{\odot}$, Phys. Rev. Lett. **125**, 101102 (2020).

[13] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Properties and astrophysical implications of the $150M_{\odot}$ binary black hole merger GW190521, Astrophys. J. Lett. **900**, L13 (2020).

[14] Collin D. Capano, Miriam Cabero, Julian Westerweck, Jahed Abedi, Shilpa Kastha, Alexander H. Nitz, Yi-Fan Wang, Alex B. Nielsen, and Badri Krishnan, Multimode quasinormal spectrum from a perturbed black hole, Phys. Rev. Lett. **131**, 221402 (2023).

[15] Collin D. Capano, Jahed Abedi, Shilpa Kastha, Alexander H. Nitz, Julian Westerweck, Yi-Fan Wang, Miriam Cabero, Alex B. Nielsen, and Badri Krishnan, Estimating false alarm rates of sub-dominant quasi-normal modes in GW190521, Classical Quantum Gravity **41**, 245009 (2024).

[16] Xisco Jiménez Forteza, Swetha Bhagwat, Sumit Kumar, and Paolo Pani, Novel ringdown amplitude-phase consistency test, Phys. Rev. Lett. **130**, 021001 (2023).

[17] Jahed Abedi, Collin D. Capano, Shilpa Kastha, Alexander H. Nitz, Yi-Fan Wang, Julian Westerweck, Alex B. Nielsen, and Badri Krishnan, Spectroscopy for asymmetric binary black hole mergers, Phys. Rev. D **108**, 104009 (2023).

[18] Harrison Siegel, Maximiliano Isi, and Will M. Farr, Ringdown of GW190521: Hints of multiple quasinormal modes with a precessional interpretation, Phys. Rev. D **108**, 064008 (2023).

[19] Matthew Giesler, Maximiliano Isi, Mark A. Scheel, and Saul Teukolsky, Black hole ringdown: The importance of overtones, Phys. Rev. X **9**, 041060 (2019).

[20] Matthew Giesler *et al.*, Overtones and nonlinearities in binary black hole ringdowns, Phys. Rev. D **111**, 084041 (2025).

[21] Vishal Baibhav, Mark Ho-Yeuk Cheung, Emanuele Berti, Vitor Cardoso, Gregorio Carullo, Roberto Cotesta, Walter Del Pozzo, and Francisco Duque, Agnostic black hole spectroscopy: Quasinormal mode content of numerical relativity waveforms and limits of validity of linear perturbation theory, Phys. Rev. D **108**, 104020 (2023).

[22] Maximiliano Isi, Matthew Giesler, Will M. Farr, Mark A. Scheel, and Saul A. Teukolsky, Testing the no-hair theorem with GW150914, Phys. Rev. Lett. **123**, 111102 (2019).

[23] Roberto Cotesta, Gregorio Carullo, Emanuele Berti, and Vitor Cardoso, Analysis of ringdown overtones in GW150914, Phys. Rev. Lett. **129**, 111102 (2022).

[24] Maximiliano Isi and Will M. Farr, Revisiting the ringdown of GW150914, arXiv:2202.02941.

[25] Gregorio Carullo, Walter Del Pozzo, and John Veitch, Observational black hole spectroscopy: A time-domain multimode analysis of GW150914, Phys. Rev. D **99**, 123029 (2019); **100**, 089903(E) (2019).

[26] Maximiliano Isi and Will M. Farr, Analyzing black-hole ringdowns, arXiv:2107.05609.

[27] M. Isi and W. M. Farr, Comment on analysis of ringdown overtones in GW150914, Phys. Rev. Lett. **131**, 169001 (2023).

[28] Gregorio Carullo, Roberto Cotesta, Emanuele Berti, and Vitor Cardoso, Carullo *et al.* reply:, Phys. Rev. Lett. **131**, 169002 (2023).

[29] Juan Calderón Bustillo, Paul D. Lasky, and Eric Thrane, Black-hole spectroscopy, the no-hair theorem, and GW150914: Kerr versus Occam, Phys. Rev. D **103**, 024041 (2021).

[30] Eliot Finch and Christopher J. Moore, Searching for a ringdown overtone in GW150914, Phys. Rev. D **106**, 043005 (2022).

[31] Sizheng Ma, Ling Sun, and Yanbei Chen, Using rational filters to uncover the first ringdown overtone in GW150914, Phys. Rev. D **107**, 084010 (2023).

[32] Sizheng Ma, Ling Sun, and Yanbei Chen, Black hole spectroscopy by mode cleaning, Phys. Rev. Lett. **130**, 141401 (2023).

[33] Marco Crisostomi, Kallol Dey, Enrico Barausse, and Roberto Trotta, Neural posterior estimation with guaranteed exact coverage: The ringdown of GW150914, Phys. Rev. D **108**, 044029 (2023).

[34] Hai-Tian Wang and Lijing Shao, Effect of noise estimation in time-domain ringdown analysis: A case study with GW150914, Phys. Rev. D **108**, 123018 (2023).

[35] Barak Zackay, Tejaswi Venumadhav, Javier Roulet, Liang Dai, and Matias Zaldarriaga, Detecting gravitational waves in data with non-stationary and non-Gaussian noise, Phys. Rev. D **104**, 063034 (2021).

[36] C. M. Biwer, Collin D. Capano, Soumi De, Miriam Cabero, Duncan A. Brown, Alexander H. Nitz, and V. Raymond, PyCBCinference: A Python-based parameter estimation toolkit for compact binary coalescence signals, Publ. Astron. Soc. Pac. **131**, 024503 (2019).

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

[38] Bruce Allen, Warren G. Anderson, Patrick R. Brady, Duncan A. Brown, and Jolien D. E. Creighton, FINDCHIRP: An algorithm for detection of gravitational waves from inspiraling compact binaries, Phys. Rev. D **85**, 122006 (2012).

[39] Alexander H. Nitz, Collin D. Capano, Sumit Kumar, Yi-Fan Wang, Shilpa Kastha, Marlin Schäfer, Rahul Dhurkunde, and Miriam Cabero, 3-OGC: Catalog of gravitational waves from compact-binary mergers, Astrophys. J. **922**, 76 (2021).

[40] Alexander H. Nitz, Sumit Kumar, Yi-Fan Wang, Shilpa Kastha, Shichao Wu, Marlin Schäfer, Rahul Dhurkunde, and Collin D. Capano, 4-OGC: Catalog of gravitational waves from compact binary mergers, Astrophys. J. **946**, 59 (2023).

[41] Alex Correia, Yi-Fan Wang, Julian Westerweck, and Collin D. Capano, Low evidence for ringdown overtone in GW150914 when marginalizing over time and sky location uncertainty, Phys. Rev. D **110**, L041501 (2024).

[42] Rich Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Open data from the first and second observing runs of Advanced LIGO and Advanced Virgo, SoftwareX **13**, 100658 (2021).

[43] Subrahmanyan Chandrasekhar, *The Mathematical Theory of Black Holes* (Clarendon Press, 1988).

[44] Emanuele Berti, Vitor Cardoso, and Marc Casals, Eigenvalues and eigenfunctions of spin-weighted spheroidal harmonics in four and higher dimensions, Phys. Rev. D **73**, 024013 (2006); **73**, 109902(E) (2006).

[45] Joshua S Speagle, dynesty: A dynamic nested sampling package for estimating Bayesian posteriors and evidences, Mon. Not. R. Astron. Soc. **493**, 3132 (2020).

[46] Sascha Husa, Sebastian Khan, Mark Hannam, Michael Pürrer, Frank Ohme, Xisco Jiménez Forteza, and Alejandro Bohé, Frequency-domain gravitational waves from nonprecessing black-hole binaries. I. New numerical waveforms and anatomy of the signal, Phys. Rev. D **93**, 044006 (2016).

[47] Sebastian Khan, Sascha Husa, Mark Hannam, Frank Ohme, Michael Pürrer, Xisco Jiménez Forteza, and Alejandro Bohé, Frequency-domain gravitational waves from nonprecessing black-hole binaries. II. A phenomenological model for the advanced detector era, Phys. Rev. D **93**, 044007 (2016).

[48] Benjamin P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), A guide to LIGO–Virgo detector noise and extraction of transient gravitational-wave signals, Classical Quantum Gravity **37**, 055002 (2020).

[49] Bruce Allen, χ^2 time-frequency discriminator for gravitational wave detection, Phys. Rev. D **71**, 062001 (2005).

[50] https://github.com/rcotesta/GW150914_ringdown.

[51] Neil J. Cornish and Tyson B. Littenberg, Bayeswave: Bayesian inference for gravitational wave bursts and instrument glitches, Classical Quantum Gravity **32**, 135012 (2015).

[52] Tyson B. Littenberg and Neil J. Cornish, Bayesian inference for spectral estimation of gravitational wave detector noise, Phys. Rev. D **91**, 084034 (2015).

[53] Neil J. Cornish, Tyson B. Littenberg, Bence Bécsy, Katerina Chatziioannou, James A. Clark, Sudarshan Ghonge, and Margaret Millhouse, BayesWave analysis pipeline in the era of gravitational wave observations, Phys. Rev. D **103**, 044006 (2021).

[54] Power Spectral Densities (PSD) release for GWTC-1, https://dcc.ligo.org/LIGO-P1900011/public.

[55] Vijay Varma, Scott E. Field, Mark A. Scheel, Jonathan Blackman, Davide Gerosa, Leo C. Stein, Lawrence E. Kidder, and Harald P. Pfeiffer, Surrogate models for precessing binary black hole simulations with unequal masses, Phys. Rev. Res. **1**, 033015 (2019).

[56] Héctor Estellés, Marta Colleoni, Cecilio García-Quirós, Sascha Husa, David Keitel, Maite Mateu-Lucena, Maria de Lluc Planas, and Antoni Ramos-Buades, New twists in compact binary waveform modeling: A fast time-domain model for precession, Phys. Rev. D **105**, 084040 (2022).

[57] Geraint Pratten *et al.*, Computationally efficient models for the dominant and subdominant harmonic modes of precessing binary black holes, Phys. Rev. D **103**, 104056 (2021).

[58] Michael Boyle *et al.*, The SXS Collaboration catalog of binary black hole simulations, Classical Quantum Gravity **36**, 195006 (2019).

[59] Mark A. Scheel *et al.*, The SXS Collaboration's third catalog of binary black hole simulations, arXiv:2505.13378.

[60] Patricia Schmidt, Ian W. Harry, and Harald P. Pfeiffer, Numerical relativity injection infrastructure, arXiv:1703.01076.

[61] Peter James Nee, Sebastian H. Völkel, and Harald P. Pfeiffer, Role of black hole quasinormal mode overtones for ringdown analysis, Phys. Rev. D **108**, 044032 (2023).

[62] Data release for "Gating-and-inpainting perspective on GW150914 ringdown overtone: Understanding the data analysis systematics" systematics, https://github.com/gwastro/gw150914-overtone.

[63] Norman Levinson, The wiener (root mean square) error criterion in filter design and prediction, J. Math. Phys. (N.Y.) **25**, 261 (1946).

[64] James Durbin, The fitting of time-series models, Revue de l'Institut International de Statistique (1960), p. 233.

[65] Pauli Virtanen *et al.*, SCIPY1.0–Fundamental algorithms for scientific computing in Python, Nat. Methods **17**, 261 (2020).