This is the accepted manuscript made available via CHORUS. The article has been published as:

Low evidence for ringdown overtone in GW150914 when marginalizing over time and sky location uncertainty

Alex Correia, Yi-Fan Wang, Julian Westerweck, and Collin D. Capano

Phys. Rev. D **110**, L041501 — Published 7 August 2024

DOI: 10.1103/PhysRevD.110.L041501

Low evidence for ringdown overtone in GW150914 when marginalizing over time and sky location uncertainty

Alex Correia,[1] Yi-Fan Wang,[2] Julian Westerweck,[3,4,5] and Collin D. Capano[6,1,a]

[1]*Physics Department, University of Massachusetts Dartmouth, North Dartmouth, MA 02747*
[2]*Max-Planck-Institut für Gravitationsphysik (Albert-Einstein-Institut), Am Mühlenberg 1, D-14476 Potsdam, Germany*
[3]*Max-Planck-Institut für Gravitationsphysik (Albert-Einstein-Institut), Callinstraße 38, D-30167 Hannover, Germany*
[4]*Leibniz Universität Hannover, D-30167 Hannover, Germany*
[5]*Institute for Gravitational Wave Astronomy and School of Physics and Astronomy, University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom*
[6]*Department of Physics, Syracuse University, Syracuse, NY 13244*

(Dated: May 8, 2024)

Tests of the no-hair theorem using astrophysical black holes involve the detection of at least two quasi-normal modes (QNMs) in the gravitational wave emitted by a perturbed black hole. A detection of two modes — the dominant, $(\ell, m, n) = (2, 2, 0)$, mode and its first overtone, the $(2, 2, 1)$ mode — in the post-merger signal of the binary black hole merger GW150914 was claimed in Isi et al. [1], with further evidence provided in Isi & Farr [2]. However, Cotesta et al. [3] disputed this claim, finding that evidence for the overtone only appeared if the signal was analyzed before merger, when a QNM description of the signal is not valid. Due to technical challenges, both of these analyses fixed the merger time and sky location of GW150914 when estimating the evidence for the overtone. At least some of the contention can be attributed to fixing these parameters. Here, we surmount this difficulty and fully marginalize over merger time and sky-location uncertainty while doing a post-merger QNM analysis of GW150914. We find that marginalizing over all parameters yields low evidence for the presence of the overtone, with a Bayes factor of 1.10 ± 0.03 in favor of a QNM model with the overtone versus one without. The arrival time uncertainty of GW150914 is too large to definitively claim detection of the $(2, 2, 1)$ mode.

INTRODUCTION

The direct detection of gravitational waves has made it possible to perform "Black hole spectroscopy" on real astrophysical black holes. Black holes that are formed from a binary black hole (BBH) merger are initially perturbed. The perturbation is radiated away in a gravitational-wave "ringdown" as the black hole settles down to a Kerr space time. The ringdown is a superposition of damped sinusoids, or quasi-normal modes (QNMs), that are uniquely determined by three indices: two angular indices $\ell \geq 2$ and $m \leq |\ell|$, and an overtone number $n \geq 0$ [4, 5]. As a consequence of the no-hair theorem, we expect that the frequencies and damping times of all the QNMs are fully characterized by the final black hole's mass and spin. The Kerr nature of the final object can be tested if more than one QNM can be observed in a post-merger gravitational wave (GW) [6]. Should the QNMs be found to need more than two parameters to characterize their frequency and damping times it would indicate that the final object is not a Kerr black hole as expected from General Relativity (GR).

Black hole spectroscopy is challenging to accomplish in practice. The QNMs are exponentially damped, and so the signal-to-noise ratio (SNR) of the signal falls off quickly after merger. This would favor searching for QNMs immediately after merger, but it was thought that one had to wait $\sim 10\,M$ (where M is the mass of the black hole, in seconds) or longer (depending on the sensitivity of the detectors [7, 8]) after the time at which the GW amplitude peaks (t_c) for non-linearities to radiate away and a QNM description of the signal to be valid [9–13]. The $10\,M$ rule was arrived at by studying the stability of the "fundamental" $n = 0$ modes in numerical signals. However, in Giesler et al. [14] it was shown that including overtones of the dominant, $(\ell, m, n) = (2, 2, 0)$ mode — i.e., modes with $(2, 2, n \geq 1)$ — yielded surprisingly good fits to a numerical relativity (NR) simulation at t_c. This led Giesler et al. to claim that the spacetime could be described by a linearly perturbed black hole *at the merger*. Isi et al. [1] then performed an overtone analysis of the first detected binary black hole merger, GW150914 [15], and claimed detection of the $(2, 2, 1)$ mode at $3.6\,\sigma$. They performed a no-hair test using the overtone by allowing its frequency and damping time to vary in a Bayesian analysis, finding that it was broadly consistent with GR.

The validity of using a QNM description of the signal at t_c is controversial. Although it was known that overtones yielded a better fit to the GW at merger prior to Giesler et al. [9], it was not clear whether this indicated that the spacetime was truly described by a linearly perturbed BH (as claimed in Giesler et al.) or if it was simply due to the overtones overfitting the signal (meaning that any sufficiently compatible signal could be fit by them). While Giesler et al. provided some evidence for the former, subsequent analyses have challenged their interpretation [16–22], leaving the question unresolved.

Even the claim of a detection of an overtone in GW150914 by Isi et al. [1] has been challenged. Cotesta et al. [3] performed an overtone analysis at multiple times around the purported t_c of GW150914. At each time they

calculated the Bayes factor for a signal model consisting of the $(2,2,0)+(2,2,1)$ QNMs versus one consisting of just the $(2,2,0)$ mode. Assuming equal prior weight for the two models, the Bayes factor $\mathcal{B} = Z_{220+221}/Z_{220}$ (where Z_x is the Bayesian evidence for model x) gives the odds ratio that the overtone model is a better description of the signal. Cotesta et al. found that the Bayes factor was ~ 1 at t_c, indicating that the data were uninformative as to the presence of the $(2,2,1)$ mode. They claimed that positive evidence for the overtone only appeared if the QNM model was used *prior* to the merger. Since a QNM model is not valid there, this would indicate that the overtone model was simply overfitting the data. The low Bayes factor was also consistent with what had been reported in Bustillo et al. [23] and in publications from the LIGO and Virgo Collaborations (LVC) [24].[25] However, in a rebuttal, Isi & Farr [2] found that the Bayes factor for the overtone was nearly 100 at t_c, reinforcing the initial finding in Isi et al. [1].

The disagreement between Isi et al. and Cotesta et al. may be due to several factors, including data conditioning choices and the manner in which the likelihood function is evaluated. The effects of these choices are investigated in Wang et al. [26]. It is evident, however, that the choice of start time for the ringdown is one of the most important factors — if not the defining factor [27] — in determining if the overtone is present. While they may disagree on the magnitude of the Bayes factor at a given time, both Cotesta et al. and Isi & Farr found that its value changes rapidly as the ringdown start time is varied around t_c, increasing by several orders of magnitude in less than $5M$. For GW150914, $5M$ corresponds to ~ 1.7 ms [15]. This is strikingly small; for comparison, the statistical uncertainty on t_c when using an inspiral-merger-ringdown (IMR) template to model the full signal is ~ 1.2 ms [28, 29]

The statistical uncertainty in t_c arises from an overall uncertainty in the arrival time of the signal in a single detector, and the uncertainty in the sky location of GW150914. The latter affects the relative offset in arrival times between the Hanford and Livingston detectors, which observed the event [15]. A full accounting of the evidence for or against the overtone should marginalize over the time and sky location uncertainty of GW150914. However, the analysis pipelines used in Isi et al. and Cotesta et al. fix the arrival time and sky location of the signal to their maximum likelihood values from IMR studies.

Even when changing the start time to recalculate the Bayes factor, the sky location was still fixed in the previous analyses. This was done to keep the ringdown start time unchanged during these pipelines' Bayesian inference analysis. Fixing the start time is necessary due to technical hurdles, discussed below, when calculating the likelihood. For this reason, almost every ringdown analysis that has been performed to date — both on GW150914 and other events — has fixed the sky location and merger time [1–3, 24, 26, 30–34].

Finch & Moore were able to marginalize over the sky location and arrival time in their post-merger analysis of GW150914 [35]. They accomplished this by attaching an arbitrary signal model to the start of the ringdown. This had the effect of smoothly tapering the signal model to zero, thereby allowing them to use the traditional frequency-domain likelihood function that is used in standard parameter estimation analyses. Doing so, they recovered Bayes factors B^{221}_{220} between 0.1 and 10, depending on the prior chosen on t_c. Their prior choices consisted of Gaussian distributions of width $1M$ centered at various t_c values.

By using the traditional frequency-domain likelihood function Finch & Moore coupled information from the pre-merger part of the signal with the post-merger QNM model. The frequency-domain likelihood is equivalent to the convolution of the "whitened" data with the "whitened" signal model. The whitening filter couples adjacent points in time over several milliseconds. This means, for example, that the beginning of the whitened ringdown will be a weighted sum over a time span starting before the merger that is on order with the damping time of the overtone (~ 1.4 ms). Consequently, the recovered QNM parameters will be influenced by both the choice of signal model and the data just prior to the start of the ringdown.

Here, we overcome the technical hurdles that arise when marginalizing over time and sky location uncertainty while doing a purely QNM analysis of the post-merger portion of GW150914. We simultaneously analyze the pre-merger data using the inspiral-merger portion of an IMR template waveform. The pre- and post-merger portions of the analysis only share a common sky location and geocentric t_c. This allows us to use the full signal to isolate the binary, giving the most accurate assessment of the timing uncertainty, while still maintaining independence between the pre- and post-merger analyses. In a companion paper [36] we show that it is necessary to model the entire observable signal in order to marginalize over timing uncertainties. Consequently, our analysis here is the most complete accounting of all uncertainties that can be preformed when using a QNM-only model of the post-merger signal.

METHODS

An in-depth description of our analysis is presented in a companion paper [36]. Here, we give a brief overview.

The Bayesian evidence for a signal model is given by marginalizing the likelihood function times the prior over all the model parameters $\boldsymbol{\vartheta}$. In GW astronomy it is common to assume that each interferometer outputs independent, wide-sense stationary Gaussian noise with zero mean in the absence of a GW signal. Under this assumption, the likelihood function is given by

$$p(\mathbf{s}|\boldsymbol{\vartheta}, h) = \prod_{D=H,L,\ldots} \frac{\exp[-\frac{1}{2}\mathbf{r}_D^{\mathsf{T}}\boldsymbol{\Sigma}_D^{-1}\mathbf{r}_D]}{\sqrt{(2\pi)^N \det \boldsymbol{\Sigma}_D}}. \quad (1)$$

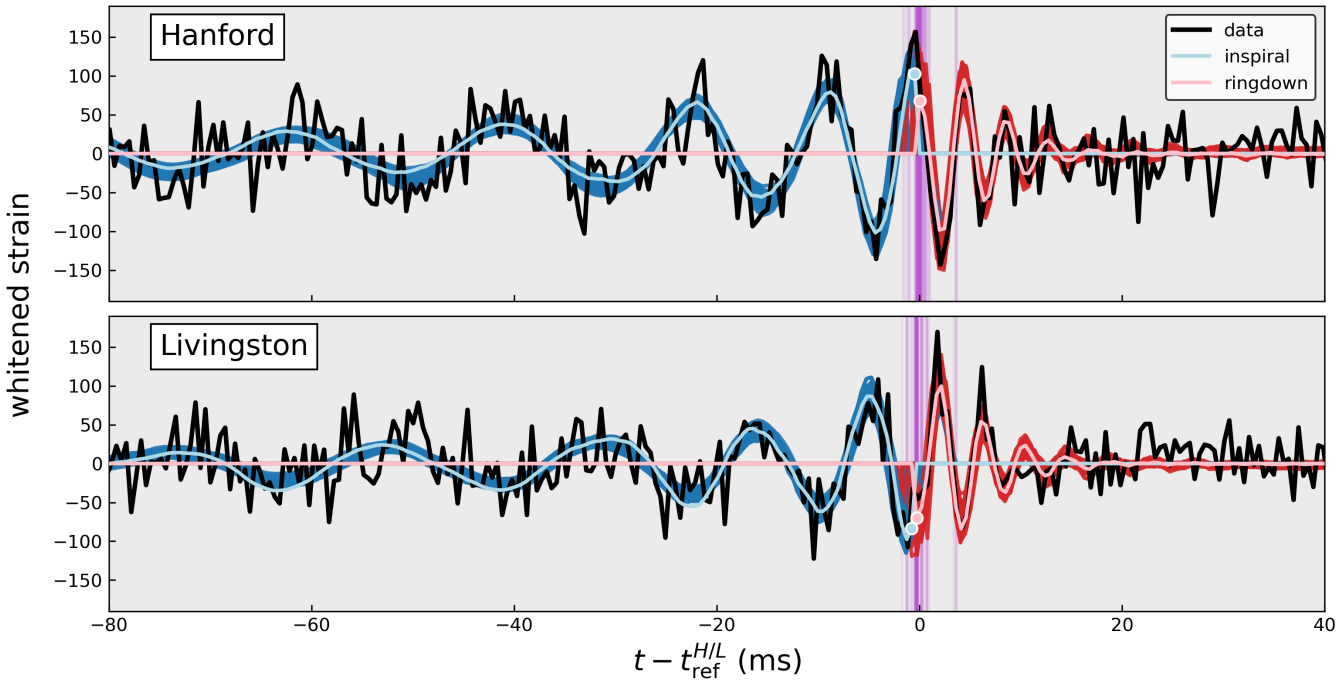


FIG. 1. Whitened strain in the Hanford (top) and Livingston (bottom) detectors around GW150914. Plotted is the original data (black line) along with the maximum likelihood waveform, which consists of an inspiral portion (blue lines) and ringdown portion (red lines). The inspiral and ringdown models are independent of each other, and only share a common sky location and geocentric t_c. Ringdown (inspiral) waveforms are gated and in-painted before (after) the t_c in each detector; the first (last) non-zero point in the waveform is indicated by the solid dot. The data are also gated and in-painted before (after) the detector t_c for the ringdown (inspiral) model, so that neither model simultaneously includes the same data point from the signal in the likelihood. Individual waveforms drawn from the upper 90th percentile of the posterior are also plotted to indicate the uncertainty region. Vertical purple lines indicate the gate time for each of the waveforms plotted. The time axis is zeroed with respect to the maximum likelihood t_c in each detector $t_{\rm ref}^{H/L}$ from the 4-OGC IMR analysis [29].

Here, the product is taken over the number of detectors, $\mathbf{r}_D = \mathbf{s}_D - \mathbf{h}_D(\boldsymbol{\vartheta})$ is the residual after subtracting the signal model $\mathbf{h}(\boldsymbol{\vartheta})$ from the time-domain data $\mathbf{s}_D$, which contains N samples. The elements of the covariance matrix $\boldsymbol{\Sigma}_D$ are given by the autocorrelation function of the data.

In standard GW parameter estimation (in which $\mathbf{h}$ models the entire observable signal), $\boldsymbol{\Sigma}_D$ can be approximated as a circulant matrix. That yields an analytic solution for $\boldsymbol{\Sigma}_D^{-1}$, which in turn yields a weighted inner product for the log likelihood that is most efficiently calculated in the frequency domain. However, since a QNM template only models the ringdown portion of the signal, data prior to the start of the ringdown must be excised from the analysis, else biases will arise. Excising data breaks the circulant approximation of $\boldsymbol{\Sigma}_D$, and so the standard frequency-domain likelihood function cannot be used.

There are two approaches around this. One is to numerically invert $\boldsymbol{\Sigma}_D$ and explicitly calculate the matrix multiplication in Eq. (1). This is what is done by the "ringdown" [37] and "pyring" [38, 39] analyses used by Isi et al. and Cotesta et al., respectively. Numerically inverting $\boldsymbol{\Sigma}_D$ is computationally expensive, but need only to be done once per analysis if the start time of the ringdown does not change. If the start time of the ringdown does change (meaning different times need to be excised from the data), then $\boldsymbol{\Sigma}_D^{-1}$ and $\det \boldsymbol{\Sigma}_D$ need to be recalculated, making the analysis substantially more expensive.

The other approach is to use "gating and in-painting" to solve the likelihood [40]. This is mathematically equivalent to the brute force numerical method, but is more efficient for longer stretches of data [26, 41]. In this approach, a "gate" is applied to a stretch of data, within which the data is zeroed out. The gated portion is then "in-painted" so that it contributes nothing to the likelihood. For a ringdown-only analysis the gate ends at the start of the ringdown; we use a 1 s gate to ensure that the rest of the observable signal is removed. The gate shifts around in time as the start time of the ringdown is varied, thereby removing more or less of the signal from the analysis.

The cost of computing the numerator in Eq. (1) with the gating and in-painting method is the same regardless of the start time of the ringdown. However, $\det \boldsymbol{\Sigma}_D$ still needs to be calculated, as its value will change for different ringdown start times. For this reason, studies that have used gating and in-painting have still fixed the start time and sky

location when doing QNM analyses [26, 30, 32–34, 42].

In Ref. [36], we show that to high precision $\det \mathbf{\Sigma}_D$ can be approximated by a simple linear interpolation. This allows us to quickly evaluate all terms in Eq. (1) using gating and in-painting, regardless of the ringdown start time, thereby allowing for time and sky-location marginalization.

We also show in Ref. [36] that it is not possible to only model part of a signal when marginalizing over t_c and sky location. Marginalizing over these parameters while only modelling part of the signal means that the portion of data that is excised from the analysis is allowed to shift around in time. If the shift is such that the entire observable signal is excised from the analysis, then a relatively large likelihood can be obtained for any parameter value by reducing the overall amplitude of the template to noise level. Consequently, the posterior probability will favor excising as much of the signal as possible, even the part that the template is meant to model. This means that for a ringdown-only analysis the start time of the template will be pushed to as *late* as the prior will allow. Likewise, if an inspiral-only analysis is performed (in which the post-merger is excised and the pre-merger is modelled with a template) then the end time of the template will be pushed to as *early* as the prior will allow. It is necessary to have a signal model for all times that a signal is observable in the data when doing any sort of marginalization over arrival time.

We overcome this problem by simultaneously yet independently modelling the pre- and post-merger data. We treat the pre- and post-merger signal as separate analyses that share a common set of hyperparameters — the sky location and geocentric t_c. We use gating and in-painting for both. For the pre-merger analysis, we excise the data *after* t_c while for the post-merger analysis we excise the data *before* t_c. The pre-merger portion of the signal is modelled using the IMRPhenomXPHM approximant [43], which we zero out for all times after t_c. The post-merger signal is modelled using a QNM model consisting of the $(2,2,0)+(2,2,1)$ modes, which is zeroed out for all times before t_c. The final mass and final spin in the QNM model are varied over uniform priors such that $M_f \in [10, 200] M_\odot$ and $\chi_f \in [-0.99, 0.99]$. For the coalescence time we use a uniform prior spanning $t_{\rm ref} \pm 0.1$ s, where $t_{\rm ref}$ is the maximum likelihood geocentric coalescence time from Ref. [29]. This is the same prior width on t_c as used in standard IMR analyses.

For the $(2,2,0)$ amplitude A_{220} we sample over the range $[10^{-25}, 8\times10^{-17}]$ uniformly in log base 10. For the $(2,2,1)$ amplitude A_{221} we use a uniform prior on $A_{221}/A_{220} \in [0,5)$. We additionally impose the constraint that the $(2,2,0)$ mode's contribution to the post-merger SNR be > 2. This is to ensure that at least the fundamental mode is present in the post-merger signal. Otherwise, we obtain samples where the $(2,2,1)$ mode appears to have large amplitude, but is in fact matching the $(2,2,0)$ mode in the signal (the "label switching" problem).

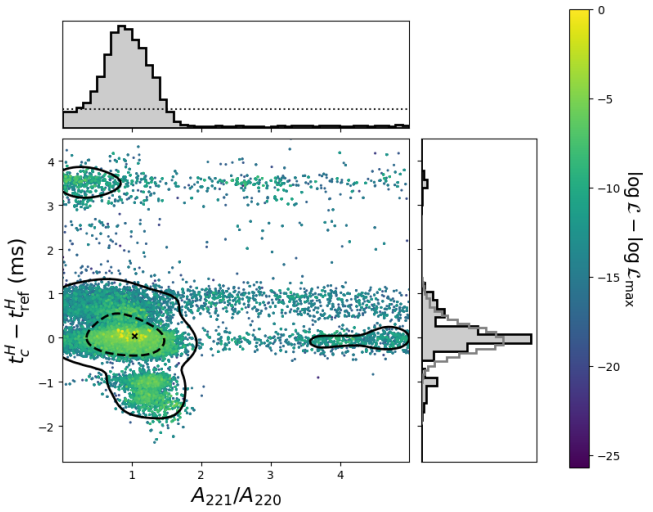


FIG. 2. Marginal posteriors on A_{221}/A_{220} and the coalescence time as measured in the Hanford detector t_c^H; similar results are obtained for the Livingston detector (not shown). The horizontal axis shows the ratio between the $(2,2,1)$ and $(2,2,0)$ mode amplitudes. The vertical axis shows the difference between the measured t_c^H and the reference time $t_{\rm ref}^H$, which is the maximum likelihood t_c^H from Ref. [29]. The filled marginal histograms represent the 1D distributions for each parameter. The prior on A_{221}/A_{220} is overlaid on the corresponding marginal posterior with a dotted black line. The 4-OGC t_c^H posterior is overlaid on the t_c marginal posterior in gray. Each point in the 2D scatterplot is colored according to its log likelihood value relative to the maximum log likelihood. The maximum likelihood point is denoted with a black cross. The 2D credible regions are indicated by the contour lines; shown are the 50th (dashed contour) and 90th (solid contours) percentiles. The Bayes factor for the $(2,2,1)$ mode is estimated from the A_{221}/A_{220} marginal posterior using the Savagey-Dickey ratio. This is the ratio between the posterior density (solid black line) and the prior density (dotted black line) at $A_{221}/A_{220} = 0$.

Gating the pre- and post-merger means that no coupling occurs over the boundary due to whitening. Simultaneously modelling both portions has the added benefit that we use the entire signal to constrain the sky location. We use PyCBC Inference [44] with the dynesty sampler [45] to perform the Bayesian analysis.

RESULTS AND DISCUSSION

The results of our analysis are summarized in Figs. 1 and 2. Figure 1 shows the recovered inspiral-only and ringdown-only templates. In Fig. 2 we plot the marginal posterior on the $(2,2,1)$ amplitude A_{221} and t_c in the Hanford detector. The median and 90% credible interval for the $(2,2,1)$ amplitude is $A_{221}/A_{220} = 0.9^{+1.8}_{-0.7}$; for the coalescence time it is $t_c^H - t_{\rm ref}^H = 0.0^{+1.0}_{-1.1}$ ms in Hanford and $t_c^L - t_{\rm ref}^L = 0.2^{+1.1}_{-1.0}$ ms in Livingston.[46]

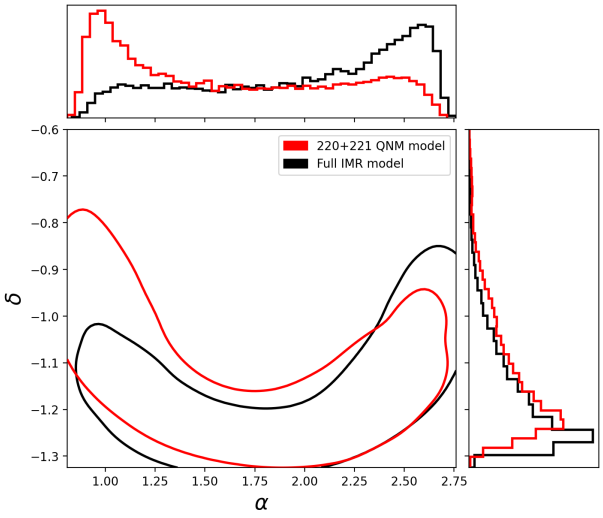


FIG. 3. Sky location posteriors for our analysis depicted in Fig. 2 and a full IMR analysis from 4-OGC [29]. Contours indicate the 90th percentile of points for each run. The orange contour denotes the analysis conducted with sky and time marginalization described above. The blue contour denotes the 4-OGC analysis.

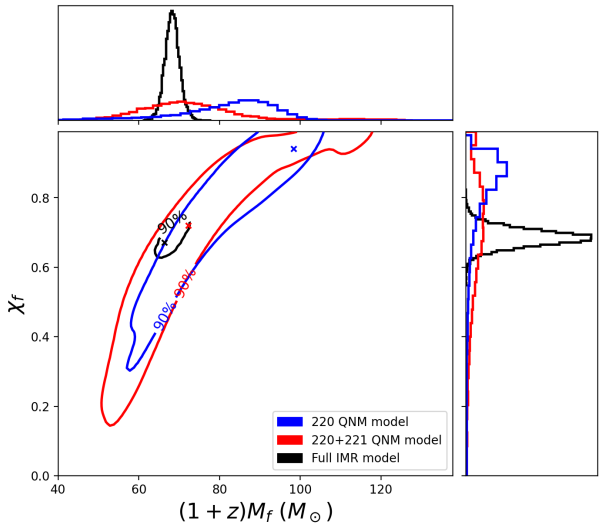


FIG. 4. Final mass and spin posteriors for our analysis depicted in Fig. 2, a $(2,2,0)$-only ringdown analysis, and a full IMR analysis from 4-OGC [29]. Contours indicate the 90th percentile of points for each run. The blue contour denotes our ringdown analysis with both the $(2,2,0)$ and $(2,2,1)$ modes. The red contour denotes an analysis with only the dominant mode present in the ringdown. The black contour denotes the 4-OGC analysis. The maximum likelihood points of the models are indicated by the crosses. The median and 90% credible intervals from the overtone model are $71^{+30}_{-16}\,M_\odot$ and $0.70^{+0.25}_{-0.47}$ for the red-shifted mass and spin, respectively, while for the $(2,2,0)$-only model they are $84^{+13}_{-27}\,M_\odot$ and $0.84^{+0.09}_{-0.55}$.

Fig. 3 compares the sky location posteriors of our analysis with a full IMR analysis from 4-OGC [29]. Our analysis recovers most of the 90th percentile of the 4-OGC posterior. Our marginal posterior on coalescence time is also consistent with the full IMR analysis (see Fig. 2). This indicates that our sky and time marginalization implementation is able to recover nominal sky parameters for GW150914.

We estimate the Bayes factor $\mathcal{B}$ between $Z_{220+221}$ and Z_{220} by calculating the Savage-Dickey ratio [47] at $A_{221}/A_{220} = 0$. Doing so, we obtain $\mathcal{B} = 1.10 \pm 0.03$, indicating low support in favor of the $(2,2,1)$ mode. Similarly, the median of the posterior distribution for A_{221} deviates from 0 by 1.2σ [48]. The reason for the small Bayes factor for the $(2,2,1)$ mode is evident from the marginal posterior plot in Fig. 2. While the maximum likelihood point is at $A_{221}/A_{220} = 1.03$ and $t_c^H - t_{\rm ref}^H = 0.03$ ms, the A_{221}/A_{220} posterior has support at zero over much of the t_c^H posterior (similarly for t_c^L). It is only at earlier times that the posterior on A_{221} is markedly shifted away from zero. This is consistent with previous studies which found the evidence for the $(2,2,1)$ mode increased at earlier times [2, 3, 26]. However, due to the relatively large timing uncertainty – our credible interval on $t_c^{H/L}$ spans roughly $5M$ – we cannot confidently claim detection of the overtone, as reflected by the Bayes factor. A signal with more SNR in the ringdown is needed.

Including the overtone does yield a mass and spin estimate of the final black hole that is more consistent with the full IMR analysis. This can be seen in Fig. 4, which compares the redshifted final mass and final spin parameters of the full IMR analysis to the QNM model that includes the (2,2,1) overtone, and to the QNM model that includes only the (2,2,0) mode [49]. This is consistent with previous analyses of GW150914 [50], which also found that using a $(2,2,0)$-only QNM model at merger yielded biased results. Indeed, a major motivation for including the $(2,2,1)$ mode is that it yields better mass and spin estimates at merger [14].

However, the overtone posterior is also *wider*: the 90% credible intervals from the overtone model span $46\,M_\odot$ and 0.72 for the red-shifted mass and spin, respectively, while for the $(2,2,0)$-only model they span $40\,M_\odot$ and 0.64. It is therefore unclear whether the better fit is simply due to the overtone model's additional freedom. This is borne out by the Bayes factor. A key feature of Bayes factors is they account for model complexity. That it is ~ 1 between the overtone and $(2,2,0)$-only models indicates that the better fit obtained by the former is offset by the Occam penalty from its larger number of parameters.

A detection of the $(3,3,0)$ sub-dominant QNM in the ringdown of GW190521 was claimed in Capano et al. [30][51]. The sky location was also fixed in that analysis, and the Bayes factor was maximized over a grid of possible merger times. Our results here raise the question of whether the evidence for the sub-dominant mode in GW190521 would also decrease if the sky location and

t_c were marginalized over.

A reanalysis of GW190521 is beyond the scope of this paper. However, there are a few differences between the overtone claim in GW150914 and the $(3,3,0)$ claim in GW190521 that indicate that sky/time marginalization will not negatively affect the Bayes factor for the latter, and in fact may increase it. First, the SNR in the ringdown of GW190521 is about twice that of GW150914 [15, 30]. Second, the difference in arrival times between Hanford and Livingston for GW190521 is $\sim 0.1\,\tau_{330}$ [29], where τ_{330} is the damping time of the putative $(3,3,0)$ mode. For comparison, the difference in arrival times in units of the overtone damping time is $\sim 5\,\tau_{221}$ for GW150914. Marginalizing over sky location — which roughly corresponds to marginalizing over uncertainty in the difference in arrival times — is likely to have less of an effect on the evidence for the GW190521 $(3,3,0)$ mode.

Assuming the sky location used in Capano et al., the evidence for the $(3,3,0)$ mode stays elevated across a large range of times after merger [52]. In contrast, the evidence for the $(2,2,1)$ mode for GW150914 drops quickly after merger. A naive marginalization over the discrete times published in Capano et al. — accomplished by summing the log evidence over those times — yields a Bayes factor for the $(3,3,0)$ mode of ~ 5000.[53] A reanalysis that uses a model for the pre- and post-merger signal and marginalizes over time and sky-location uncertainty is necessary to fully evaluate the evidence for the $(3,3,0)$ mode, and any other sub-dominant QNM, in GW190521. Nevertheless, our cursory analysis indicates that the evidence for the $(3,3,0)$ mode in GW190521 may in fact increase with full sky and time marginalization. We plan to revisit this in a future study.

Returning to overtones, our results show that GW150914 is too weak to definitively claim detection of the $(2,2,1)$ mode. Gravitational-wave detectors are continuing to improve in sensitivity [29, 54, 55]. A signal with sufficient SNR to resolve questions about the presence of one or more overtones is likely to eventually be observed.[56] When such an event occurs, it will be necessary to fully marginalize over both time and sky-location uncertainties when performing a QNM analysis of the ringdown, as we have done here.

This research was conducted using PyCBC [57]. Data is available at [58].

We wish to thank Gregorio Carullo and Juan Calderon Bustillo for providing helpful comments. A.C. was supported by funds from the Massachusetts Space Grant Consortium. C.C. acknowledges support from NSF award PHY-2309356. All computations were performed on Unity, a collaborative, multi-institutional high-performance computing cluster managed by UMass Amherst Research Computing and Data.

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. KAGRA 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.

[a] cdcapano@syr.edu

[1] M. Isi, M. Giesler, W. M. Farr, M. A. Scheel, and S. A. Teukolsky, Testing the no-hair theorem with GW150914, Physical Review Letters **123**, 10.1103/physrevlett.123.111102 (2019).

[2] M. Isi and W. M. Farr, Revisiting the ringdown of GW150914 (2022), arXiv:2202.02941 [gr-qc].

[3] R. Cotesta, G. Carullo, E. Berti, and V. Cardoso, Analysis of ringdown overtones in GW150914, Physical Review Letters **129**, 10.1103/physrevlett.129.111102 (2022).

[4] C. V. Vishveshwara, Scattering of Gravitational Radiation by a Schwarzschild Black-hole, Nature **227**, 936 (1970).

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

[6] O. Dreyer, B. Kelly, B. Krishnan, L. S. Finn, D. Garrison, and R. Lopez-Aleman, Black-hole spectroscopy: testing general relativity through gravitational-wave observations, Classical and Quantum Gravity **21**, 787 (2004).

[7] L. T. London, Modeling ringdown. II. Aligned-spin binary black holes, implications for data analysis and fundamental theory, Phys. Rev. D **102**, 084052 (2020), arXiv:1801.08208 [gr-qc].

[8] G. Carullo *et al.*, Empirical tests of the black hole no-hair conjecture using gravitational-wave observations, Phys. Rev. D **98**, 104020 (2018), arXiv:1805.04760 [gr-qc].

[9] A. Buonanno, G. B. Cook, and F. Pretorius, Inspiral, merger and ring-down of equal-mass black-hole binaries, Phys. Rev. D **75**, 124018 (2007), arXiv:gr-qc/0610122.

[10] E. Berti, V. Cardoso, J. A. Gonzalez, U. Sperhake, M. Hannam, S. Husa, and B. Bruegmann, Inspiral, merger and ringdown of unequal mass black hole binaries: A Multi-

polar analysis, Phys. Rev. D **76**, 064034 (2007), arXiv:gr-qc/0703053.

[11] I. Kamaretsos, M. Hannam, S. Husa, and B. S. Sathyaprakash, Black-hole hair loss: learning about binary progenitors from ringdown signals, Phys. Rev. D **85**, 024018 (2012), arXiv:1107.0854 [gr-qc].

[12] L. London, D. Shoemaker, and J. Healy, Modeling ringdown: Beyond the fundamental quasinormal modes, Phys. Rev. D **90**, 124032 (2014), [Erratum: Phys.Rev.D 94, 069902 (2016)], arXiv:1404.3197 [gr-qc].

[13] A. Gupta, B. Krishnan, A. Nielsen, and E. Schnetter, Dynamics of marginally trapped surfaces in a binary black hole merger: Growth and approach to equilibrium, Phys. Rev. D **97**, 084028 (2018), arXiv:1801.07048 [gr-qc].

[14] M. Giesler, M. Isi, M. A. Scheel, and S. A. Teukolsky, Black hole ringdown: The importance of overtones, Phys. Rev. X **9**, 041060 (2019).

[15] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), Observation of gravitational waves from a binary black hole merger, Physical Review Letters **116**, 061102 (2016).

[16] S. Bhagwat, X. J. Forteza, P. Pani, and V. Ferrari, Ringdown overtones, black hole spectroscopy, and no-hair theorem tests, Phys. Rev. D **101**, 044033 (2020), arXiv:1910.08708 [gr-qc].

[17] J. L. Jaramillo, R. Panosso Macedo, and L. Al Sheikh, Pseudospectrum and Black Hole Quasinormal Mode Instability, Phys. Rev. X **11**, 031003 (2021), arXiv:2004.06434 [gr-qc].

[18] S. Ma, K. Mitman, L. Sun, N. Deppe, F. Hébert, L. E. Kidder, J. Moxon, W. Throwe, N. L. Vu, and Y. Chen, Quasinormal-mode filters: A new approach to analyze the gravitational-wave ringdown of binary black-hole mergers, Phys. Rev. D **106**, 084036 (2022), arXiv:2207.10870 [gr-qc].

[19] M. H.-Y. Cheung *et al.*, Nonlinear Effects in Black Hole Ringdown, Phys. Rev. Lett. **130**, 081401 (2023), arXiv:2208.07374 [gr-qc].

[20] K. Mitman *et al.*, Nonlinearities in Black Hole Ringdowns, Phys. Rev. Lett. **130**, 081402 (2023), arXiv:2208.07380 [gr-qc].

[21] P. J. Nee, S. H. Völkel, and H. P. Pfeiffer, Role of black hole quasinormal mode overtones for ringdown analysis, Phys. Rev. D **108**, 044032 (2023), arXiv:2302.06634 [gr-qc].

[22] V. Baibhav, M. H.-Y. Cheung, E. Berti, V. Cardoso, G. Carullo, R. Cotesta, W. Del Pozzo, and F. Duque, Agnostic black hole spectroscopy: Quasinormal mode content of numerical relativity waveforms and limits of validity of linear perturbation theory, Phys. Rev. D **108**, 104020 (2023), arXiv:2302.03050 [gr-qc].

[23] J. Calderón Bustillo, P. D. Lasky, and E. Thrane, Black-hole spectroscopy, the no-hair theorem, and GW150914: Kerr versus Occam, Phys. Rev. D **103**, 024041 (2021), arXiv:2010.01857 [gr-qc].

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

[25] It should be noted that the pipeline [38, 39] used to report overtone Bayes factors by the LVC is the same as used in Cotesta et al., so the agreement between the two is not entirely surprising.

[26] Y.-F. Wang, C. D. Capano, J. Abedi, S. Kastha, B. Krishnan, A. B. Nielsen, A. H. Nitz, and J. Westerweck, A frequency-domain perspective on GW150914 ringdown overtone, (2023), arXiv:2310.19645 [gr-qc].

[27] G. Carullo, R. Cotesta, E. Berti, and V. Cardoso, Reply to Comment on "Analysis of Ringdown Overtones in GW150914", Phys. Rev. Lett. **131**, 169002 (2023), arXiv:2310.20625 [gr-qc].

[28] B. P. Abbott *et al.* (LIGO Scientific, Virgo), Properties of the Binary Black Hole Merger GW150914, Phys. Rev. Lett. **116**, 241102 (2016), arXiv:1602.03840 [gr-qc].

[29] A. H. Nitz, S. Kumar, Y. Wang, S. Kastha, S. Wu, M. Schäfer, R. Dhurkunde, and C. D. Capano, 4-OGC: Catalog of gravitational waves from compact-binary mergers (2022), arXiv:2112.06878 [astro-ph.HE].

[30] C. D. Capano, M. Cabero, J. Westerweck, J. Abedi, S. Kastha, A. H. Nitz, Y.-F. Wang, A. B. Nielsen, and B. Krishnan, Multimode Quasinormal Spectrum from a Perturbed Black Hole, Phys. Rev. Lett. **131**, 221402 (2023), arXiv:2105.05238 [gr-qc].

[31] M. Isi, W. M. Farr, M. Giesler, M. A. Scheel, and S. A. Teukolsky, Testing the black-hole area law with GW150914, Physical Review Letters **127**, 10.1103/physrevlett.127.011103 (2021).

[32] S. Kastha, C. D. Capano, J. Westerweck, M. Cabero, B. Krishnan, and A. B. Nielsen, Model systematics in time domain tests of binary black hole evolution, Physical Review D **105**, 10.1103/physrevd.105.064042 (2022).

[33] C. D. Capano, J. Abedi, S. Kastha, A. H. Nitz, J. Westerweck, Y.-F. Wang, M. Cabero, A. B. Nielsen, and B. Krishnan, Statistical validation of the detection of a sub-dominant quasi-normal mode in GW190521 (2022), arXiv:2209.00640 [gr-qc].

[34] J. Abedi, C. D. Capano, S. Kastha, A. H. Nitz, Y.-F. Wang, J. Westerweck, A. B. Nielsen, and B. Krishnan, Spectroscopy for asymmetric binary black hole mergers, Phys. Rev. D **108**, 104009 (2023), arXiv:2309.03121 [gr-qc].

[35] E. Finch and C. J. Moore, Searching for a ringdown overtone in GW150914, Physical Review D **106**, 10.1103/physrevd.106.043005 (2022).

[36] A. Correia and C. D. Capano, Sky marginalization in black hole spectroscopy and tests of the area theorem, (2023), arXiv:2312.15146 [gr-qc].

[37] M. Isi and W. M. Farr, maxisi/ringdown: Initial ringdown release (2021).

[38] G. Carullo, W. D. Pozzo, and J. Veitch, Observational black hole spectroscopy: A time-domain multimode analysis of GW150914, Physical Review D **99**, 10.1103/physrevd.99.123029 (2019).

[39] G. Carullo, W. Del Pozzo, and J. Veitch, `pyRing`: a time-domain ringdown analysis python package, git.ligo.org/lscsoft/pyring (2023).

[40] B. Zackay, T. Venumadhav, J. Roulet, L. Dai, and M. Zaldarriaga, Detecting gravitational waves in data with non-stationary and non-gaussian noise, Physical Review D **104**, 10.1103/physrevd.104.063034 (2021).

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

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

arXiv:2307.11975 [gr-qc].
[43] G. Pratten, C. García-Quirós, M. Colleoni, A. Ramos-Buades, H. Estellés, M. Mateu-Lucena, R. Jaume, M. Haney, D. Keitel, J. E. Thompson, and S. Husa, Computationally efficient models for the dominant and subdominant harmonic modes of precessing binary black holes, Physical Review D **103**, 10.1103/physrevd.103.104056 (2021).
[44] C. M. Biwer, C. D. Capano, S. De, M. Cabero, D. A. Brown, A. H. Nitz, and V. Raymond, PyCBC Inference: A Python-based parameter estimation toolkit for compact binary coalescence signals, Publ. Astron. Soc. Pac. **131**, 024503 (2019), arXiv:1807.10312 [astro-ph.IM].
[45] J. S. Speagle, dynesty: a dynamic nested sampling package for estimating Bayesian posteriors and evidences, Monthly Notices of the Royal Astronomical Society **493**, 3132 (2020), https://academic.oup.com/mnras/article-pdf/493/3/3132/32890730/staa278.pdf.
[46] We use the maximum likelihood sky location and coalescence time from the 4-OGC analysis [29] to calculate the reference GPS time in each detector. Specifically, we use $t^H_{\rm ref} = 1126259462.423701$ and $t^L_{\rm ref} = 1126259462.416674$.
[47] I. Verdinelli and L. Wasserman, Computing Bayes factors using a generalization of the Savage-Dickey density ratio, Journal of the American Statistical Association **90**, 614 (1993).
[48] The 1.2σ quoted here was computed using the relative amplitude posterior A_{221}/A_{220}. Recalculating using the absolute A_{221} posterior yields the same result.
[49] To obtain the $(2,2,0)$-only posterior on mass and spin, we preform a full Bayesian analysis in which the $(2,2,0)$-only QNM starts at t_c.
[50] B. P. Abbott *et al.* (LIGO Scientific, Virgo), Tests of general relativity with GW150914, Phys. Rev. Lett. **116**, 221101 (2016), [Erratum: Phys.Rev.Lett. 121, 129902 (2018)], arXiv:1602.03841 [gr-qc].
[51] See also Siegel et al. [42], which found evidence for the $(2,1,0)$ and $(3,2,0)$ modes in the same event.
[52] C. Capano, S. Kastha, M. Cabero, J. Westerweck, and Y. Wang, gwastro/BH-Spectroscopy-GW190521: Published data release (2023).
[53] We report the Bayes factor for the $(3,3,0)$ mode as being $Z_{220+330}/\max(Z_{220}, Z_{220+221})$, as was done in Capano et al. [30].
[54] R. Abbott *et al.* (LIGO Scientific, VIRGO, KAGRA), GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run, (2021), arXiv:2111.03606 [gr-qc].
[55] LIGO, Virgo, and KAGRA observering run plans, `https://observing.docs.ligo.org/plan/`, Accessed: 2023-12-05.
[56] Although larger SNR may result in an ambiguous number of overtones needed to fully model the merger, raising questions of overfitting [23].
[57] A. Nitz, I. Harry, D. Brown, C. M. Biwer, J. Willis, T. D. Canton, C. Capano, T. Dent, L. Pekowsky, G. S. C. Davies, S. De, M. Cabero, S. Wu, A. R. Williamson, B. Machenschalk, D. Macleod, F. Pannarale, P. Kumar, S. Reyes, dfinstad, S. Kumar, M. Tápai, L. Singer, P. Kumar, veronica villa, maxtrevor, B. U. V. Gadre, S. Khan, S. Fairhurst, and A. Tolley, gwastro/pycbc: v2.3.3 release of pycbc (2024).
[58] A. Correia and C. Capano, Sky marginalization in black hole spectroscopy (2024).