On the detection of ringdown overtones in GW150914

Roberto Cotesta,1 Gregorio Carullo,2,3 Emanuele Berti,1 and Vitor Cardoso4,5

1Department of Physics and Astronomy, Johns Hopkins University, 3400 N. Charles Street, Baltimore, Maryland, 21218, USA
2Dipartimento di Fisica “Enrico Fermi”, Università di Pisa, Pisa I-56127, Italy
3INFN sezione di Pisa, Pisa I-56127, Italy
4CENTRA, Departamento de Física, Instituto Superior Técnico – IST, Universidade de Lisboa – UL, Av. Rovisco Pais 1, 1049-001 Lisboa, Portugal
5Niels Bohr International Academy, Niels Bohr Institute, Blegdamsvej 17, 2100 Copenhagen, Denmark

(Dated: January 5, 2022)

We analyze GW150914 post-merger data to understand if ringdown overtone detection claims are robust. We find no evidence in favor of an overtone in the data after the waveform peak. Around the peak, the log-Bayes factor does not indicate the presence of an overtone, while the support for a non-zero amplitude is sensitive to changes in the starting time much smaller than the overtone damping time. This suggests that claims of an overtone detection are noise-dominated. We perform GW150914-like injections in neighboring segments of the real detector noise, and we show that noise can indeed induce artificial evidence for an overtone.

Introduction. Since the first detection of gravitational waves (GWs) from a binary black hole (BH) merger, GW150914 [1], the LIGO-Virgo-KAGRA (LVK) Collaboration [2–4] reported 90 events with a probability of astrophysical origin $p_{\text{astro}} > 0.5$ during the first three observing runs [5–8]. These GW signals, combined with those detected by independent groups [9–13], have broadened our understanding of cosmology [14], the astrophysics of compact objects [15], matter at supranuclear densities [16], and general relativity (GR) in the strong-field regime [17].

Among the numerous tests of GR proposed over the years, BH spectroscopy with the so-called “ringdown” relaxation phase following the merger presents unique opportunities to characterize the remnant as a Kerr BH. In linearized GR, the two GW polarizations $h_{+,x}$ can be decomposed as $h_{+} - i h_{\times} \equiv \sum_{\ell m} h_{\ell m}(t) - 2 Y_{\ell m}(\iota, \phi)$, where the (spin-weighted) spherical harmonics $-2 Y_{\ell m}(\iota, \phi)$ depend only on two angles that characterize the direction from the source to the observer. Each multipolar component is a superposition of damped exponentials known as quasinormal modes (QNMs):

$$h_{\ell m}(t) \equiv \sum_{n} A_{\ell m n} e^{i\left[\omega_{\ell m n}(t - t_{\text{start}}) + \phi_{\ell m n}\right]} e^{-\left(t - t_{\text{start}}^n\right)/\tau_{\ell m n}},$$

(1)

where we ignored mode-mixing between different $\ell$ modes and counterrotating modes (a valid assumption for GW150914). In GR, the QNM frequencies $\omega_{\ell m n}$ and damping times $\tau_{\ell m n}$ depend only on the remnant BH’s mass $M_\text{f}$ and spin $a_\text{f}$ [18–24]. The QNM amplitudes $A_{\ell m n}$ and phases $\phi_{\ell m n}$ were unknown before the first numerical BH merger simulations, and early work on BH spectroscopy [23] had to rely on educated guesses [25]. We now know that radiation from a binary BH merger is dominated by the $\ell = |m| = 2$ component, while higher multipoles are subdominant [26, 27]. For fixed $(\ell, m)$, the QNMs are sorted by the magnitude of $\tau_{\ell m n}$, the fundamental mode ($n = 0$) has the longest damping time, and the integer $n$ labels the so-called “overtones.”

It has long been known that including overtones improves the agreement between ringdown-only fits and the complete gravitational waveforms from perturbed BHs. This was first shown by direct integration of the perturbation equations sourced by infalling particles or collapsing matter [29–32] and then, more rigorously, using Green’s function techniques [33–37]. Overtones were shown to improve agreement with numerical simulations of collapse [38], head-on collisions [39] and quasicircular mergers [26] leading to BH formation, and their omission leads to significant biases in mass and spin estimates [40, 41]. However, standard QNM tests often relied only on fundamental modes for two main reasons: overtones are short-lived and difficult to confidently identify in the data [42], and it is unclear whether multiple overtones have physical meaning or they just happen to phenomenologically fit the nonlinear part of the merger signal [26, 27].

Recently, Ref. [43] showed that including overtones up to $n = 7$ in the ringdown model improves the agreement with numerical simulations for all times beyond the time $t_{\text{peak}}$ where $|h_{+}^2 + h_{\times}^2|$ has a maximum, claiming that this observation “implies that the spacetime is well described as a linearly perturbed BH with a fixed mass and spin as early as the peak.” Their study’s insistence on an intrinsically linear physical description spurred a sequence of additional investigations, both on the modeling and on the observational side [37, 44–52]. If higher overtones can indeed be measured by starting at the peak, the larger ringdown signal-to-noise ratio (SNR) would open the door to more precise tests of GR. This theoretical argument motivated a reanalysis of GW150914. Ref. [53] fitted the post-peak waveform with a QNM superposition including overtones, and claimed evidence for “at least one overtone [...] with 3.6σ confidence.” The claim seems at odds with Ref. [46] and with the subsequent LVK analysis [17], both reporting weak evidence (with a log-Bayes factor of only $\sim 0.6$ in favor of the “overtone model”) including both $n = 0$ and $n = 1$ (henceforth Kerr$_{221}$) relative to the model including only $n = 0$ (henceforth Kerr$_{220}$).
Methods. We fix the free overtone amplitudes and phases, respectively. Since there is no evidence for misaligned spins in GW150914, we also assume that the waveform amplitudes satisfy $h_{\text{start}} = h_{\text{peak}}$, quoted in units of $M$. All $\Delta h_{\text{start}}$ values used in panels with dark (light) gold backgrounds are consistent with the median of the $h_{\text{peak}}$ distribution at 1 $\sigma$ (2$\sigma$). In each panel, dashed black, solid red, and solid blue contours correspond to 90% credible level in the BH parameters measured using the full IMR [28], Kerr_{221}, and Kerr_{220} models, respectively.

In this paper we ask whether overtone detection claims in GW150914 data are robust. We use geometrical units $G = c = 1$, restoring physical units when needed, and we always quote redshifted BH masses as measured in a geocentric reference frame.

Methods. The $\ell = |m| = 2$ multipole is largely dominant in GW150914 [17, 54], so we can ignore higher multipoles and mode-mixing contributions in the general waveform model (1). The system does not show evidence for antialigned progenitor spins, so counterrotating modes can be safely ignored [17, 55]. We make several assumptions to match as closely as possible the analysis of Ref. [53]. First, we include only one or two QNMs ($n = 0, 1$) and assume that all overtones start at the same time $t_{\text{start}} = t_{\text{peak}}$. We fix ($\ell, \phi) = (1, 0)$ rad, since in our model these parameters are strongly degenerate with the free overtone amplitudes and phases, respectively. Since there is no evidence for misaligned spins in GW150914, we also assume that the waveform amplitudes satisfy $h_{\text{start}} = h_{\text{peak}}$, a good approximation when the progenitor spins are nearly aligned with the orbital angular momentum of the binary. The strain measured by GW detectors is $h_D(t) = F_+ h_+ + F_\times h_\times$, where the detector pattern functions $F_+, \times (\alpha, \delta, \psi)$ depend on the right ascension, declination and polarization angles $\alpha, \delta$ and $\psi$ [56]. Following Ref. [53] we set $(\alpha, \delta, \psi) = (1.95, -1.27, 0.82)$ rad. We fix $t_{\text{start}}$ in the Hanford detector and compute the starting time in the Livingston detector using a fixed time delay determined from the sky position parameters listed above. We assume flat priors on all free parameters in the ranges $M_f \in [20, 200] M_\odot, a_f \in [0, 0.99], A_{22n} \in [0.5 \times 10^{-20}], \phi_{22n} \in [0, 2\pi]$.

We analyze the ringdown signal using the Bayesian parameter estimation package pyRing [54, 57], employed by the LVK collaboration to perform ringdown-only tests of GR. The autocorrelation function of the background noise (appropriately cropped to avoid contaminations from earlier stages of the coalescence [58]) was chosen to be as close as possible to the settings of Ref. [53]. The pyRing package relies on the nested sampling algorithm cpnest [59], that allows us to compare alternative hypotheses by computing their relative Bayes factors. We analyze 0.1 s of publicly available data from GWOSC [60] with a sampling rate of 16384 Hz (the maximum resolution available). This rate, larger than the rate of 2048 Hz used in Ref. [53], was chosen to minimize the impact of the time discretization.

In fact, when investigating the consequences of slightly changing the analysis settings, we found that the choice of $t_{\text{start}}$ (which has been set equal to $t_{\text{peak}}$ according to the theoretical arguments in [43]) has by far the largest impact. The effect of varying $\psi, \iota$ and the segments used to estimate the autocorrelation function is milder, and it will be discussed in a forthcoming paper [61], together with the impact of dropping the symmetry assumption on the amplitudes $h_{\text{start}}$. Ref. [53] assumed $t_{\text{start}} = t_{\text{peak}} = 1126259462.423$ s. However the value of $t_{\text{peak}}$ must be estimated from the data, and as such it is uncertain. Fixing it to a specific value can induce systematic biases. We quantify this uncertainty by reconstructing $t_{\text{peak}}$ using the posterior distributions of the parameters of GW150914 [62] obtained with the IMR waveform model SEOBNRv4 [63] (see the Supplemental Material for details). In the Hanford detector, the resulting posterior distribution has median $t_{\text{peak}} = 1126259462.42323$ s and standard deviation $\sigma = 0.00059$ s. We will vary $t_{\text{start}}$ within the $\pm 2 \sigma$ interval of its posterior distribution.
Mass and spin estimates. In Fig. 1 we show the mass and spin of the GW150914 BH remnant estimated using the Kerr$_{220}$ (blue), Kerr$_{221}$ (red) and full IMR model[28] (dashed black) for 10 selected values of $\Delta t_{\text{start}}^{H1} \equiv t_{\text{start}}^{H1} - \hat{t}_{\text{peak}}^{H1}$. For $\Delta t_{\text{start}}^{H1}/M \geq -1.45$, the IMR posterior overlaps with both the Kerr$_{220}$ and Kerr$_{221}$ models at 90\% credibility, although the Kerr$_{221}$ reconstruction peaks closer to the IMR estimate. The Kerr$_{221}$ model agrees much better than Kerr$_{220}$ with the IMR posterior especially when we start fitting before the peak ($\Delta t_{\text{start}}^{H1}/M \leq -2.17$), where such a fit is not well motivated by the overtone model (see Fig. 1 of [43]). The starting time used in Ref. [53] corresponds to $\Delta t_{\text{start}}^{H1}/M = -0.72$ in Fig. 1. Note that the $(M_f,a_f)$ measurements obtained with the Kerr$_{221}$ model overlap with the GR prediction even when $\Delta t_{\text{start}}^{H1}/M = -3.62$, outside of the 2\(\sigma\) confidence interval on the peak location. This is likely due to a combination of two effects: (i) since $\omega_{221} < \omega_{220}$, any overtone model naturally includes a low-frequency component, thus improving the fit to the low-frequency, pre-merger part of the signal; and (ii) the Kerr$_{221}$ model has a larger number of parameters than the Kerr$_{220}$ model, thus at low signal-to-noise ratios it can still fit the signal with the values of $(M_f,a_f)$ determined by the late-time ringdown behavior. 

Bayes factors. To quantify the evidence for the presence of an overtone in GW150914, we compare the hypotheses that the data can be described by the Kerr$_{221}$ vs. Kerr$_{220}$ models and compute the resulting Bayes factor, $B_{221}^{220}$. In the top panel of Fig. 2 we show $\log_{10} B_{221}^{220}$ (red crosses) for selected values of $\Delta t_{\text{start}}^{H1}$. In the bottom panel we show the posterior of the overtone amplitude $A_1 \equiv A_{221}$ for the Kerr$_{221}$ model (red curves). When $\Delta t_{\text{start}}^{H1}/M \geq -1.45$, there is no evidence for the overtone in the data ($\log_{10} B_{221}^{220} < 0$), and the posterior distributions in the bottom panel have significant support for $A_1 = 0$, hence the Kerr$_{220}$ model is favored with respect to Kerr$_{221}$. We observe significant Bayesian evidence for the presence of the overtone ($\log_{10} B_{221}^{220} > 2$) only for $\Delta t_{\text{start}}^{H1}/M \leq -4.34$, i.e., well outside of the nominal region of validity of the Kerr$_{221}$ model. For $\Delta t_{\text{start}}^{H1}/M = -0.72$, which corresponds to the $t_{\text{peak}}$ value used in Ref. [53], we find that $\log_{10} B_{221}^{220} = -0.60$, while the amplitude has large support for zero. At the peak time $A_1$ is maximum away from zero, but there is still some support for zero amplitude. This may lead us to conclude that the overtone is measurable in this ringdown signal. However, both the Bayes factor and $A_1$ decrease for values of $\Delta t_{\text{start}}^{H1}$, located immediately before and after $\Delta t_{\text{start}}^{H1}/M = 0$. Now, the decay time for the overtone in question is $\tau_{221} \approx 1.3 \text{ ms} \approx 4M$. If the overtone were measurable, we would expect to find evidence for its presence when changing $t_{\text{peak}}$ by only $\sim 0.24 \text{ ms} \approx 0.72 \text{ M}$. Since this is not the case, we must consider the hypothesis that the (weak) evidence in favor of an overtone for $\Delta t_{\text{start}}^{H1}/M = 0$ could be driven by a noise fluctuation.

We test this hypothesis by using a synthetic signal ("injection", in LVK jargon) obtained from a numerical solution of the Einstein equations consistent with the GW150914 signal (see the Supplemental Material for details). In this case, $t_{\text{peak}}$ is known exactly. We analyze the signal using different values of $t_{\text{start}}^{H1}$, such that $\Delta t_{\text{start}}^{H1}$ is consistent with the values used for the real signal. For each selected $\Delta t_{\text{start}}$, we first perform the analysis described above in the case of the real signal, but we now set the
noise realization to zero (“zero-noise” injection). The resulting parameter distributions will thus have an uncertainty consistent with the actual signal, while eliminating a possible shift of the posterior median due to noise fluctuations coincident with the signal. The values of $\log_{10} B_{220}^{221}$ and $A_1$ obtained from this zero-noise injection are shown as black dots and black curves in the upper and lower panels of Fig. 2. When $\Delta t_{\text{start}}^H/M = 0$ there is no evidence for an overtone ($\log_{10} B_{220}^{221} = -0.21 < 0$) and $A_1$ has a large support for zero. For the zero-noise injection, the log Bayes factor is positive only when $\Delta t_{\text{start}}^H/M \leq -1.45$, and it generally increases for lower values of $\Delta t_{\text{start}}^H$, similarly to what happens for the real signal. The inferred amplitude of the overtone is consistent with the behavior observed for the Bayes factor, increasing for large negative values of $\Delta t_{\text{start}}^H/M$.

To assess the impact of the detector noise on the measurement of $\log_{10} B_{220}^{221}$ and $A_1$, for each $\Delta t_{\text{start}}^H$, we repeat the above analysis superposing the simulated signal to 10 different segments of the real detector noise close to the time of coalescence of GW150914 (see the Supplementary Material). The resulting Bayes factors are reported as “error bars” on $\log_{10} B_{220}^{221}$; for each time $\Delta t_{\text{start}}^H$, the upper (lower) error bar corresponds to the largest (smallest) $\log_{10} B_{220}^{221}$ obtained from these injections. The blue curves in the lower panel are posterior distributions of $A_1$ for the configuration with the largest value of $\log_{10} B_{220}^{221}$ at a given $\Delta t_{\text{start}}^H$. These distributions (to be compared with the zero-noise black curves) are estimates of the largest possible overtone amplitude obtainable when accounting for noise fluctuations. For $\Delta t_{\text{start}}^H/M = 0$ and neighboring points, the negative values of $\log_{10} B_{220}^{221}$ measured in the real signal are consistent with the negative values measured in the synthetic signal, if we account for the detector noise. The posterior distribution of $A_1$ shows that a “favorable” realization of the detector noise can lead to a measurement of $A_1$ that peaks away from zero (blue curve) – similarly to the actual signal (red curve) – although $A_1$ is consistent with zero in the case of the zero-noise injection (black curve). We conclude that the mild support for an overtone observed in the amplitude posterior (although never confirmed by the Bayesian evidence) is driven by the detector noise.

**Discussion.** We have performed a Bayesian data analysis of the GW150914 ringdown signal to understand if ringdown overtone detection claims are robust. We found no Bayesian evidence in favor of an overtone, nor a significant overtone amplitude measurement in GW150914 data after the waveform peak, where the inclusion of overtones in the ringdown model is expected to improve the agreement with numerical relativity simulations [41, 43]. There is mild support for a non-zero overtone amplitude in the data at the peak, but such support for $A_1 = 0$ is sensitive to changes in the starting time smaller than the overtone damping time. Most importantly, the log-Bayes factor never favors the detection of an overtone when varying the starting time within the 1σ credible region of the peak time reconstruction. This suggests that the detection is noise-dominated. We verified this hypothesis by performing GW150914-like injections in different segments of the real detector noise. These results differ from Ref. [58], where the impact of the real detector noise and peak time uncertainty were not considered.

For both real and synthetic signals, the evidence for the overtone and the uncertainty on the evidence (as measured by the black “error bars”) generally increase for large negative values of $\Delta t_{\text{start}}^H$. The overtone model is not expected to be valid in this region, but the larger number of degrees of freedom in the model can pick up a larger portion of the low-frequency, pre-merger signal power. At the same time, the evidence uncertainty grows dramatically – spanning up to four orders of magnitude for the earliest times shown in Fig. 2 – because the poorly constrained model can easily pick up noise fluctuations.

Our results reveal an intrinsic instability of the inference based on such a model. The instability may happen even in the absence of noise, because the mass and spin of the remnant extracted from numerical simulations vary significantly close to the peak of the radiation [27, 41, 64], and thus the assumption of a linear superposition of QNMs starting at the peak can lead to conceptual issues [44, 65]. As reported in Table I of Ref. [43], the amplitude of the fundamental mode is stable up to a few parts in $10^3$ under the addition of overtones, but higher overtones have much less stable amplitudes: $A_{221}$ varies by 8%, while $A_{223}$ varies by more than 200%. This is inconsistent with our understanding of ringdown in the linearized regime, where (by definition) the QNM amplitudes should be constant [42, 45, 66, 67]. In the absence of fitting errors for the overtone amplitudes, it is difficult to quantify how much of this variation can be ascribed to the current accuracy of numerical BH merger simulations, rather than being due to a time-evolving background. This instability might also explain the incompatibility of the measurement $A_{221}/A_{220} < 2$ reported in [53, 58], compared to the predicted value $A_{221}/A_{220} \sim 4$ reported in Table I of [43].

A physical parametrization of the overtone amplitudes as a function of the progenitors parameters, similar to the one proposed in Refs. [42, 67] for the fundamental modes, may alleviate this problem. However parametrizations of nonspinning binary BH mergers find that such a “global” fit is not robust under variations of the starting time: see e.g. Figs. 3 and 4 of [45]. Overfitting issues are particularly difficult to address. For example, the accuracy of overtone models constructed using GR QNMs can be matched (or even surpassed) by adding “unphysical” low-frequency components corresponding to non-GR values of the frequency and damping time [44, 48]. Similar “pseudo-QNMs” were introduced in the context of effective-one-body models [68–70].

Our results for the Bayes factors are consistent with previous work. The large number of free parameters in the overtone model introduces an Occam penalty that must be balanced by large SNRs [46]. Even when modeling the overtone amplitudes as functions of the properties...
of the remnant progenitors, measuring several overtone frequencies may still be impractical: Fisher matrix estimates [45] suggest that it will be easier to obtain evidence for multiple modes using higher angular harmonics rather than overtones. These results are in contrast with the predictions of [58], which employed a different detection criterion. In future work we plan to investigate strategies for a robust modeling and measurement of higher overtones, and to revisit the BH spectroscopy horizon estimates of Refs. [71, 72].

Acknowledgments. We are grateful to Max Isi for help in reproducing the analysis settings of Ref. [53], and to Walter Del Pozzo for valuable comments and suggestions. We thank Vishal Baibhav, Swetha Bhagwat, Juan Calderón Bustillo, Will Farr, Xisco Jiménez-Forteza, Danny Laghi, Lionel London, Paolo Pani, Saul Teukolsky and the Testing GR group of the LVK collaboration. We thank Vishal Baibhav, Swetha Bhagwat, to Walter Del Pozzo for valuable comments and suggestions. We thank Vishal Baibhav, Swetha Bhagwat, Juan Calderón Bustillo, Will Farr, Xisco Jiménez-Forteza, Danny Laghi, Lionel London, Paolo Pani, Saul Teukolsky and the Testing GR group of the LVK collaboration. We thank Vishal Baibhav, Swetha Bhagwat, to Walter Del Pozzo for valuable comments and suggestions.

Software. LIGO-Virgo data are interfaced through GWpy [73]. Projections onto detectors are computed through LALSuite [74]. The pyRing package is publicly available at: https://git.ligo.org/lscsoft/pyring. This study made use of the open-software python packages: corner, cython, h5py, matplotlib, numpy, scipy, seaborn [75–81].

[1] B. P. Abbott et al. (LIGO Scientific, Virgo), Phys. Rev. Lett. 116, 061102 (2016), arXiv:1602.03837 [gr-qc].
[2] J. Aasi et al. (LIGO Scientific), Class. Quant. Grav. 32, 074001 (2015), arXiv:1411.4547 [gr-qc].
[3] F. Acernese et al. (VIRGO), Class. Quant. Grav. 32, 024001 (2015), arXiv:1408.3978 [gr-qc].
[4] T. Akutsu et al. (KAGRA), PTEP 2021, 05A101 (2021), arXiv:2005.05574 [physics.ins-det].
[5] B. P. Abbott et al. (LIGO Scientific, Virgo), Phys. Rev. X 9, 031040 (2019), arXiv:1811.12907 [astro-ph.HE].
[6] R. Abbott et al. (LIGO Scientific, Virgo), Phys. Rev. X 11, 021053 (2021), arXiv:2010.14527 [gr-qc].
[7] R. Abbott et al. (LIGO Scientific, Virgo), (2021), arXiv:2108.01045 [gr-qc].
[8] R. Abbott et al. (LIGO Scientific, Virgo, KAGRA), (2021), arXiv:2110.03606 [astro-ph.HE].
[9] A. H. Nitz, C. Capano, A. B. Nielsen, S. Reyes, R. White, D. A. Brown, and B. Krishnan, Astrophys. J. 872, 195 (2019), arXiv:1811.01921 [gr-qc].
[10] A. H. Nitz, T. Dent, G. S. Davies, S. Kumar, C. D. Capano, I. Harry, S. Mozon, L. Nuttall, A. Lundgren, and M. Tápai, Astrophys. J. 891, 123 (2020), arXiv:1910.05331 [astro-ph.HE].
[11] A. H. Nitz, C. D. Capano, S. Kumar, Y.-F. Wang, S. Kastha, M. Schäfer, R. Dhurkunde, and M. Cabero, Astrophys. J. 922, 76 (2021), arXiv:2105.09151 [astro-ph.HE].
[12] T. Vennumadhav, B. Zackay, J. Roulet, L. Dai, and M. Zaldarriaga, Phys. Rev. D 101, 083030 (2020), arXiv:1904.07214 [astro-ph.HE].
[13] B. Zackay, L. Dai, T. Vennumadhav, J. Roulet, and M. Zaldarriaga, Phys. Rev. D 104, 063030 (2021), arXiv:1910.09528 [astro-ph.HE].
[14] R. Abbott et al. (LIGO Scientific, VIRGO, KAGRA), (2021), arXiv:2111.03604 [astro-ph.CO].
[15] R. Abbott et al. (LIGO Scientific, VIRGO, KAGRA), (2021), arXiv:2111.03634 [astro-ph.HE].
[16] K. Chatziioannou, Gen. Rel. Grav. 52, 109 (2020), arXiv:2006.03168 [gr-qc].
[17] R. Abbott et al. (LIGO Scientific, Virgo), Phys. Rev. D 103, 122002 (2021), arXiv:2010.14529 [gr-qc].
[18] W. H. Press, Astrophys. J. Lett. 170, L105 (1971).
[19] S. Chandrasekhar and S. L. Detweiler, Proc. Roy. Soc. Lond. A 344, 441 (1975).
[20] S. L. Detweiler, Astrophys. J. 239, 292 (1980).
[21] K. D. Kokkotas and B. G. Schmidt, Living Rev. Rel. 2, 2 (1999), arXiv:gr-qc/9909058.
[22] O. Dreyer, B. J. Kelly, B. Krishnan, L. S. Finn, D. Garrison, and R. Lopez-Aleman, Class. Quant. Grav. 21, 787 (2004), arXiv:gr-qc/0309007.
[23] E. Berti, V. Cardoso, and C. M. Will, Phys. Rev. D 73, 064030 (2006), arXiv:gr-qc/0512160.
[24] E. Berti, V. Cardoso, and A. O. Starinets, Class. Quant. Grav. 26, 163001 (2009), arXiv:0905.2975 [gr-qc].
[25] E. E. Flanagan and S. A. Hughes, Phys. Rev. D 57, 4535 (1998), arXiv:gr-qc/9701039.
[26] A. Buonanno, G. B. Cook, and F. Pretorius, Phys. Rev. D 75, 124018 (2007), arXiv:gr-qc/0610122.
[27] E. Berti, V. Cardoso, J. A. Gonzalez, U. Sperhake, M. Hannam, S. Husa, and B. Bruegmann, Phys. Rev. D 76, 064034 (2007), arXiv:gr-qc/0703053.
[28] B. P. Abbott et al. (LIGO Scientific, Virgo), Phys. Rev. Lett. 116, 241102 (2016), arXiv:1602.03840 [gr-qc].
[29] M. Davis, R. Ruffini, W. H. Press, and R. H. Price, Phys. Rev. Lett. 27, 1466 (1971).
[30] C. T. Cunningham, R. H. Price, and V. Moncrief, Astrophys. J. 224, 643 (1978).
[31] C. T. Cunningham, R. H. Price, and V. Moncrief, Astro-
R. Abbott

S. Bhagwat, X. J. Forteza, P. Pani, and V. Ferrari, Phys. Rev. D 89, 044018 (2013), arXiv:1305.4306 [gr-qc].

N. Oshita, (2021), arXiv:2109.09757 [gr-qc].

R. F. Stark and T. Piran, Phys. Rev. Lett. 55, 891 (1985).

A. Dhani and B. S. Sathyaprakash, (2021), arXiv:2102.07794 [gr-qc].

M. Okounkova, S. Bhagwat, P. Pani, and V. Ferrari, Phys. Rev. D 101, 044033 (2020), arXiv:1910.07078 [gr-qc].

M. Isi, M. Giesler, W. M. Farr, M. A. Scheel, and S. A. Teukolsky, Phys. Rev. D 102, 044053 (2020), arXiv:2005.03260 [gr-qc].

J. C. Bustillo, P. D. Lasky, and E. Thrane, Phys. Rev. D 103, 024041 (2021), arXiv:2010.01857 [gr-qc].

M. Okounkova, (2020), arXiv:2004.00671 [gr-qc].

M. Isi and W. M. Farr, (2021), arXiv:2110.02156 [gr-qc].

E. Finch and C. J. Moore, Phys. Rev. D 103, 084028 (2021), arXiv:2107.14195 [gr-qc].

E. Berti, A. Sesana, E. Barausse, V. Cardoso, and K. Belczynski, Phys. Rev. Lett. 117, 101102 (2016), arXiv:1605.09286 [gr-qc].

I. Ota and C. Chirenti, (2021), arXiv:2108.01774 [gr-qc].

D. Macleod et al., “gwpy/gwpy: 2.0.3,” (2021).

LIGO Scientific Collaboration, “LIGO Algorithm Library - LALSuite,” free software (GPL) (2018).

D. Foreman-Mackey et al., “dfm/corner.py: corner.py v.2.2.1,” (2021).

S. Behnel, R. Bradshaw, C. Citro, L. Dalcin, D. Seljebotn, and K. Smith, Comput. Sci. Eng. 13, 31 (2011).

A. Collette, Python and HDF5 (O'Reilly, 2013).

J. D. Hunter, Comput. Sci. Eng. 9, 90 (2007).

C. R. Harris et al., Nature (London) 585, 357 (2020).

P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovskiy, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. Jarrod Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. Carey, I. Polat, Y. Feng, E. W. Moore, J. Vand erPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and the contributors, Nature Methods (2020).

M. Wasik et al., “mwasikom/seaborn: v0.11.2 (august 2021),” (2021).

S. Husa, S. Khan, M. Hannam, M. Pürrer, F. Ohme, X. Jiménez Forteza, and A. Böh, Phys. Rev. D 93, 044006 (2016), arXiv:1508.07250 [gr-qc].

A. Böh, L. Haegel, S. Husa, F. Ohme, G. Pratten, and M. Pürrer, Phys. Rev. Lett. 113, 151101 (2014), arXiv:1308.3271 [gr-qc].

A. Hannam, M. Pürrer, A. Böh, F. Ohme, and S. Husa, Phys. Rev. D 93, 044007 (2016), arXiv:1508.07253 [gr-qc].

https://github.com/johnveitch/cpnest.

R. Abbott et al. (LIGO Scientific, Virgo), SoftwareX 13, 100658 (2021), arXiv:1912.11716 [gr-qc].

R. Cotesta, G. Carullo, E. Berti, and V. Cardoso, in preparation.
Details of the peak time reconstruction. The peak time in the Hanford detector is reconstructed by generating the $h_+, h_\times$ waveform polarizations in post-processing, using the LVK posterior samples [62], and computing the maximum of $h_+^2 + h_\times^2$. In the main text we use the peak time $t_{\text{start}}^\text{H1}$ reconstructed from the SEOBNRv4 model [63], but to quantify waveform systematics we have repeated the calculation using also the IMRPhenomPv2 model [82–84]. The resulting distribution has median $\bar{t}_{\text{H1,peak}}^{\text{Pv2}} = 1126259462.42371$ s and standard deviation $\sigma_{\text{Pv2}} = 0.00063$ s, i.e., it is shifted $\sim 0.5$ ms after the time inferred from SEOBNRv4. Thus, using the reconstruction from this alternative model would reinforce our conclusions. This difference also highlights the need to properly marginalize over the peak time when evaluating the robustness of ringdown analyses.

Details of the injection study. In the injection study, we use the numerical relativity simulation SXS:BBH:0305 from the public catalog [85] of the Simulating eXtreme Spacetimes (SXS) collaboration. This simulation was set up to reproduce the GW150914 signal. The BH binary in the numerical waveform has mass ratio $q = 1.22$ and spins aligned with the orbital angular momentum, with dimensionless magnitudes $\chi_1 = 0.33$ and $\chi_2 = -0.44$. For the synthetic signal, we place the system at a luminosity distance of $D_L = 410$ Mpc and we use a redshifted total mass $M = 72M_\odot$, in agreement with the median values estimated by the LVK collaboration [28]. Finally, the simulated signal is superimposed to the real detector noise at times $[-30, -25, -20, -5, 5, 10, 15, 20, 25, 30]$ s with respect to the peak time $t_{\text{peak, inj}}^\text{H1} = 1126259472.423$ s, approximately 10 s after the coalescence time of GW150914. We use the same noise autocorrelation function used in the analysis of the GW150914 event.