

Reply to Comment on “Analysis of Ringdown Overtones in GW150914”

Gregorio Carullo,¹ Roberto Cotesta,² Emanuele Berti,² and Vitor Cardoso^{1,3}

¹*Niels Bohr International Academy, Niels Bohr Institute, Blegdamsvej 17, 2100 Copenhagen, Denmark*

²*Department of Physics and Astronomy, Johns Hopkins University,*

3400 N. Charles Street, Baltimore, Maryland, 21218, USA

³*CENTRA, Departamento de Física, Instituto Superior Técnico – IST,*

Universidade de Lisboa – UL, Avenida Rovisco Pais 1, 1049-001 Lisboa, Portugal

Summary – The Comment [1] discusses the impact on our analysis (“the Article”) [2] of: (i) a spurious shift in the discretized time axis within `pyRing` [3]; (ii) too short of an analysis segment ($T = 0.1$ s instead of $T \geq 0.2$ s).

In this Reply, we repeat the analysis including these two effects, alleviating some of the previous discrepancies observed with respect to Ref. [4]. However, even after accounting for the aforementioned effects, the Bayesian log-evidence for an overtone remains negative, supporting the main findings of the Article. This overall conclusion is in agreement with two independent reanalyses, Refs. [5, 6], and with Fig. 1 of Ref. [4], by the authors of the Comment. The latter shows how at times consistent with the peak ($t \sim 2M$ in the units of that figure), the significance does not reach 2σ . The fact that the significance is much lower than the original 3.6σ estimate [7] in all of these works is a consequence of the uncertainty in t_{start} , not included in Ref. [7]. As we pointed out in the Article (following ideas introduced in Ref. [8]), accounting for this uncertainty is crucial to obtain unbiased results. This has proven to be important also in the recent analysis of Ref. [9].

Different choices of sampling rate (2048 Hz in Ref. [4], 16384 Hz in the Article), and noise estimation methods (in the Article taken to be the same as in the original analysis of Ref. [7]) are behind the remaining differences. As shown in the Comment, the two codes give in fact indistinguishable output, if given identical inputs. The remaining technical differences give rise to appreciably different features in the parameters estimate, highlighting how at current detector sensitivity and when targeting short-lived components, such analyses can be significantly affected by technical details.

Results – Figure 1 of this Reply updates Fig. 2 of the Article. Taking into account the remarks in the Comment, we have increased the analysis duration from $T = 0.1$ s to $T = 0.2$ s, and removed the time axis shift. The posteriors show a small shift in favor of an overtone at the t_{peak} value employed by Ref. [7], confirming an improved agreement with the latter analysis, but significant railing is still observed when the t_{peak} uncertainty is taken into account. The Bayesian log-evidence for an overtone is negative throughout almost the entire time range. Comparing to Fig. 2 in the Comment, agreement is also observed in the injection set: noise (gray line there) frequently increases the overtone evidence, as expected given the current detection significance.

Although differing in the technical implementation and in certain analysis inputs, both our results and the ones

of Ref. [4] broadly agree with the reanalysis of Ref. [5]. Here, the authors employ frequency-domain methods and marginalize over the uncertainty in t_{peak} , reporting negative Bayesian evidence for an overtone and a significance of 1.8σ . Similar conclusions were found in Ref. [6].

Below, we address the specific technical points raised in the Comment.

Time axis shift – In Refs. [2, 4, 7] the starting time of the analysis, t_{start} , is fixed and assumed to agree with t_{peak} , the time at which the gravitational-wave strain has a peak. The internal selection of t_{start} in `pyRing` conservatively selects the time-stamp immediately following the requested t_{start} . The time axis mislabeling pointed out in the Comment implies a misidentification between the closest data time-stamp and t_{start} . At the sampling rate used in the Article, this approximation leads to a shift of only ~ 0.06 ms. The statistical uncertainty in the value of t_{peak} determined from the data is ~ 2.5 ms, implying that the shift (i.e. the “software bug” discussed in the Comment) has a negligible impact on the analysis. This is confirmed by the bottom panel of Fig. 1 in the Comment. The above mislabeling has been removed since `pyRing v2.1.0`. Finally, the 0.24 ms shift in Fig. 2 of the Article, mentioned in the Comment, was applied in post-processing for better comparison with Ref. [7] when using a lower sampling rate. This shift did not enter our analysis. To facilitate comparison with Ref. [4], we do not apply any time shift in Fig. 1 of this Reply.

Analysis duration – While the signal power of the unfiltered GW150914 ringdown signal after 0.1 s is negligible, this may not be the case for the whitened signal entering the time-domain likelihood, which (depending on both the noise and model features) can be significantly stretched due to the overlap of the mode’s spectrum with the detector’s spectral lines. Increasing the analysis duration to $T = 0.2$ s produces a small increase in the inferred amplitude of the searched overtone, as shown in the bottom panel of Fig. 1 of the Comment, and in Fig. 1 of this Reply. However, while this increase is appreciable at t_{start} of Fig. 1 in [1] (corresponding to $-0.72M$ on our time axis), Fig. 1 below shows significant railing against zero when t_{start} is set to nearby values.

Remaining differences – In Ref. [4], the auto-correlation function (ACF) was estimated using a Fourier (FFT) method (which is closer to the standard method used in LVK analyses within the `pyRing` pipeline [8, 10]), while the Article used the same ACF estimate of the original analysis Ref. [7], obtained through a direct correlation

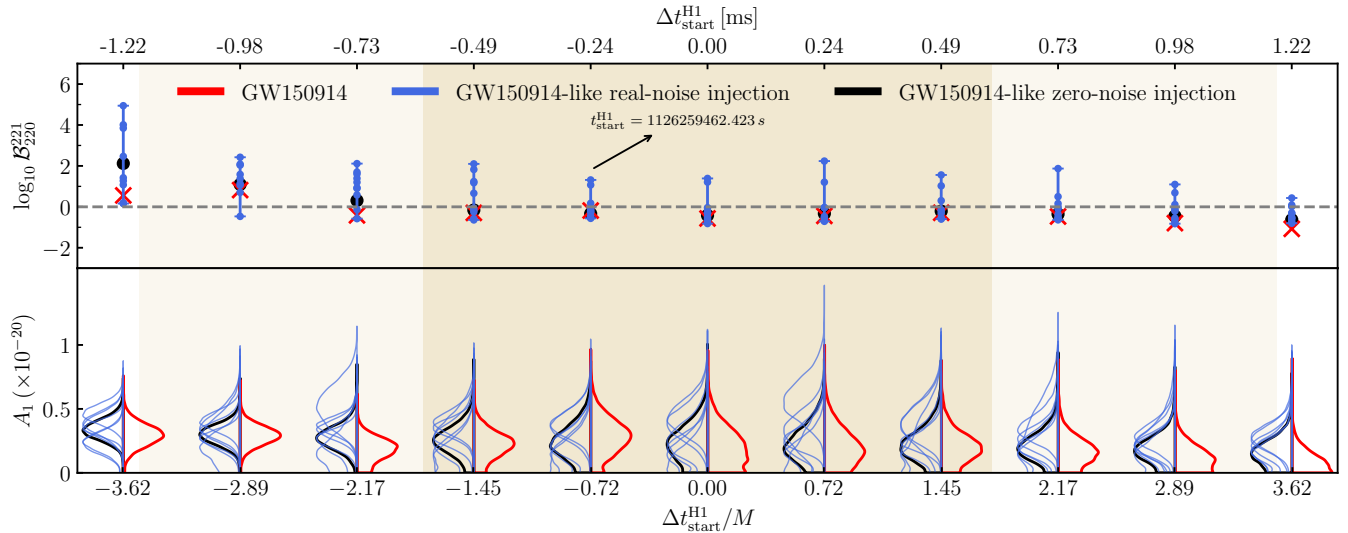


FIG. 1. Same as Fig.2 in the Article, but using the updated settings described above. The results were obtained with the `pyRing` commit 6d7d0e8d.

method. We test the validity of the direct correlation vs FFT estimates by dividing 32 s of data surrounding the event into stretches 200 ms long, whitening each of these. A Kolmogorov-Smirnov test against the gaussian hypothesis reports less than 1% of outliers at 1% significance with both methods for the 32 s surrounding the event, confirming the statistical validity of both noise estimates within this test.

We also test for putative numerical instabilities in our analysis. We evaluate the covariance matrix conditioning number, and compute the time-domain likelihood using supposedly more robust inversion methods (based on the Cholesky decomposition combined with back substitution or the Levinson recursion), instead of a direct inversion. We obtain identical likelihoods up to double precision, implying that no numerical instabilities are present.

When repeating the analysis of Fig. 1 using a sampling rate of 2048 Hz¹ and an FFT ACF estimate, we find results in closer agreement with Ref. [1], confirming that the two analysis softwares (`ringdown` and `pyRing`) achieve close to identical outputs when they are given the same inputs. In this case, the overtone significance is larger, but significant riling is still present for $\Delta t_{\text{start}}^{H1} = 0.72, 1.45 M$, and the Bayesian evidence is always not significant within the t_{peak} uncertainty at 1σ . This is consistent with the substantial riling evident even in the results of [4], as shown in Fig. 4 of the Article. Note that the riling is more apparent when we plot histograms of the samples, instead of the bands shown in Fig. 2 of the Comment.

Additional remarks on the Comment – The Comment still fails to address the key observation of the Article

– namely, that accounting for the statistical uncertainty in t_{peak} (not incorporated in the original analysis of Ref. [7]) is necessary in order to obtain unbiased results. To detect a mode, the evidence should hold throughout the support of the distribution of t_{peak} , not at a single (arbitrary) value. A positive outcome emerged from the present debate is that new analyses by different groups have instead started to take into account this crucial uncertainty [5, 9].

It should also be noted how, for clarity, the time axis of Fig. 2 of the Comment should be centered around the best estimates of the peak time derived by the Comment’s authors in Ref. [4], namely $\sim 2M$ later than the original estimate t_{ref} used in Ref. [7]. At that point, the significance has dropped substantially: as shown in Fig. 7 of Ref. [4], even at this best estimate for t_{peak} , the log-Bayes factor is below unity. As we show with the injection of numerical relativity waveforms (for which we know the true value of t_{peak}), analyzing the signal at times earlier than t_{peak} can only artificially increase the evidence for an overtone, as the model is trying to fit the merger part of the waveform. In summary, even assuming the analysis of the Comment authors’, we find the evidence for a second mode not significant (see also Fig. 4 of the Article).

Theoretical considerations – Beyond considerations of statistical evidence, recent studies of the overtones [11, 12] and of nonlinear modes excitation in numerical simulations [13–15] have clearly shown that the test presented in Ref. [7] is not relevant for the observational investigations of black hole spectra. The claimed “overtone” does not correspond to the physical excitation of a true quasinormal mode close to the peak. It is merely a phenomenological term fitting away nonlinearities near the peak of the radiation. Even if the significance of a subdominant damped exponential would have been higher, as claimed in previous analyses, it would be wrong to identify this second exponential with the first overtone and to

¹ The higher sampling rate in our analysis was chosen to obtain the maximum available precision on the start time.

use it for spectroscopy tests. This argument alone implies

that the analysis of Ref. [7] should not be interpreted as a test of the “no-hair theorem.”

-
- [1] M. Isi and W. M. Farr, *Phys. Rev. Lett.* (2023).
 - [2] R. Cotesta, G. Carullo, E. Berti, and V. Cardoso, *Phys. Rev. Lett.* **129**, 111102 (2022), [arXiv:2201.00822 \[gr-qc\]](#).
 - [3] G. Carullo, W. Del Pozzo, and J. Veitch, **pyRing: a time-domain ringdown analysis python package**, [git.ligo.org/lscsoft/pyring](#) (2023).
 - [4] M. Isi and W. M. Farr, (2022), [arXiv:2202.02941 \[gr-qc\]](#).
 - [5] E. Finch and C. J. Moore, *Phys. Rev. D* **106**, 043005 (2022), [arXiv:2205.07809 \[gr-qc\]](#).
 - [6] M. Crisostomi, K. Dey, E. Barausse, and R. Trotta, *Phys. Rev. D* **108**, 044029 (2023), [arXiv:2305.18528 \[gr-qc\]](#).
 - [7] M. Isi, M. Giesler, W. M. Farr, M. A. Scheel, and S. A. Teukolsky, *Phys. Rev. Lett.* **123**, 111102 (2019), [arXiv:1905.00869 \[gr-qc\]](#).
 - [8] G. Carullo, W. Del Pozzo, and J. Veitch, *Phys. Rev. D* **99**, 123029 (2019), [Erratum: *Phys.Rev.D* 100, 089903 (2019)], [arXiv:1902.07527 \[gr-qc\]](#).
 - [9] H. Siegel, M. Isi, and W. M. Farr, *Phys. Rev. D* **108**, 064008 (2023), [arXiv:2307.11975 \[gr-qc\]](#).
 - [10] R. Abbott *et al.* (LIGO Scientific, Virgo), *Phys. Rev. D* **103**, 122002 (2021), [arXiv:2010.14529 \[gr-qc\]](#).
 - [11] V. Baibhav, M. H.-Y. Cheung, E. Berti, V. Cardoso, G. Carullo, R. Cotesta, W. Del Pozzo, and F. Duque, (2023), [arXiv:2302.03050 \[gr-qc\]](#).
 - [12] P. J. Nee, S. H. Völkel, and H. P. Pfeiffer, (2023), [arXiv:2302.06634 \[gr-qc\]](#).
 - [13] M. H.-Y. Cheung *et al.*, *Phys. Rev. Lett.* **130**, 081401 (2023), [arXiv:2208.07374 \[gr-qc\]](#).
 - [14] K. Mitman *et al.*, *Phys. Rev. Lett.* **130**, 081402 (2023), [arXiv:2208.07380 \[gr-qc\]](#).
 - [15] N. Khera, A. Ribes Metidieri, B. Bonga, X. J. Forteza, B. Krishnan, E. Poisson, D. Pook-Kolb, E. Schnetter, and H. Yang, (2023), [arXiv:2306.11142 \[gr-qc\]](#).