Comments on:
“Echoes from the abyss: Evidence for Planck-scale structure at black hole horizons”

Gregory Ashton,[1,2] Ofek Birnholtz,[1,2,*] Miriam Cabero,[1,2] Collin Capano,[1,2] Thomas Dent,[1,2] Badri Krishnan,[1,2] Grant David Meadors,[1,2,3] Alex B. Nielsen,[1,2] Alex Nitz,[1,2] and Julian Westerweck[1,2]

[1]*Max-Planck-Institut für Gravitationsphysik, D-30167 Hannover, Germany*
[2]*Leibniz Universität Hannover, D-30167 Hannover, Germany*
[3]*Max-Planck-Institut für Gravitationsphysik, D-14476 Potsdam-Golm, Germany*

Recently, Abedi, Dykaar and Afshordi claimed evidence for a repeating damped echo signal following the binary black hole merger gravitational-wave events recorded in the first observational period of the Advanced LIGO interferometers. We discuss the methods of data analysis and significance estimation leading to this claim, and identify several important shortcomings. We conclude that their analysis does not provide significant observational evidence for the existence of Planck-scale structure at black hole horizons, and suggest renewed analysis correcting for these shortcomings.

The detections by the Advanced LIGO detectors of gravitational wave signals from binary black hole mergers [1–3] has opened up the possibility of new tests of the nature of these objects [3–5]. A recent work [6] has claimed to find evidence of near-horizon Planck-scale structure using data[7] from the three Advanced LIGO events GW150914, LVT151012 and GW151226. In the model of [6] this near-horizon structure gives rise to so-called echoes [8–10]. Their inferred amplitude parameters suggest that a lot of gravitational wave energy was emitted in the echoes: a very rough calculation implies that the amount of energy emitted in the echoes was approximately 0.1 solar masses (for GW150914) and 0.2 solar masses (for LVT151012). This should be compared to the total estimated energy emitted by the original signal of 3 solar masses (for GW150914) and 1.5 solar masses (for LVT151012). The data used is from the LIGO Open Science Center (LOSC) [7] which contains a total of 4096 seconds of strain data from both Advanced LIGO detectors around the three events. Of these data the authors use only 32 seconds centered around each event for their analysis. The authors claim to find such echoes in data following the three events with combined significance of 2.9σ (p-value 3.7×10^{-3}; with the one-sided significance convention used in [1–3], this value corresponds to 2.7σ). If this claim were true, it would force a major re-evaluation of the standard picture of black holes in vacuum Einstein gravity.

Besides the marginal claimed significance, there are a number of aspects of the analysis of [6] that lead us to suspect that the true significance of their detection may be considerably weaker. Here we will not examine the theoretical motivations for the existence of such near-horizon Planck-scale structure, nor the model templates the authors have chosen to search for. Instead we will focus on the data analysis methods as reported and the significance estimates assigned to them. Regarding these we highlight some major data analysis drawbacks, which cast doubt on this aspect of their result.

The first problem arises at how strong the relative signal should be for the three events. The two binary black hole events GW150914 and GW151226 were detected by the Advanced LIGO detectors with significance levels $> 5.3\sigma$ and signal-to-noise ratios of 23.7 and 13.0 respectively[3]. The other event, LVT151012, had a reported significance of only 1.7σ and a signal-to-noise ratio of 9.7 combined between the two Advanced LIGO detectors. However, in Table II of [6] we see that the signal-to-noise ratio of the claimed echo signal is actually largest for LVT151012. The higher SNR cannot be due to the projected number of echoes for LVT151012 over 32 seconds of data (~ 180) being greater than the number of echoes for GW150914 over that duration (~ 110), because late echoes are strongly damped, decreasing to a factor of 10 over ~ 22 echoes. Thus in order for the echoes of LVT15012 to have a higher SNR than the echoes of GW150914, their amplitude must be very high. In fact to account for the reported SNRs, the initial amplitude for the first echo of LVT151012 would have to be about 10% higher than that of GW150914[1], while the original event’s peak is about 2-3 times lower for LVT151012 in comparison to GW150914’s. This would require their parameter A to be about 2-3 times larger for LVT151012 than for GW150914. It would therefore be interesting to see plots and estimated parameters for LVT151012 (and GW151226), similar to those presented in Table I of [6] for GW150914.

A second worrying aspect is the determination of the values for their echo waveform model, Equation 9. The model depends on six parameters: a phase factor, three time parameters Δt_{echo}, t_{echo} and t_0, and two amplitude parameters A and γ. The phase is modeled as a simple sign flip at each reflection[2], A is maximized over analytically, and Δt_{echo} is determined by the parameters of

---

[1] $\frac{\rho_{LVT151012}}{\rho_{GW150914}} = \frac{A_{LVT151012}\sqrt{\sum_{p=1}^{180} \gamma^{2p}}}{A_{GW150914}\sqrt{\sum_{p=1}^{100} \gamma^{2p}}} = \frac{4.52}{4.13} \sim 1.1$, where we have used the nomenclature of [6], and $\gamma = 0.9$.

[2] ignoring the phase accumulated over the travel between the light ring and the near-horizon Planck-scale boundary.

the final black hole given in [3] as given in Equation 6. The three parameters γ, t_0 and $t_{\rm echo}$ are determined by maximization, with γ and t_0 kept fixed between the different events. This maximization is done over a prior range, as displayed in their Table I, and the values resulting from this maximization for γ and t_0 are found to lie very close to the boundary of this prior range, 0.9 and -0.1 respectively. This suggests that there may be support for values of these parameters that lie outside of this range (no error ranges are given). This would be particularly worrisome in the case of γ since a value greater than unity means that each successive echo would have an amplitude greater than the previous echo. Such a result would seem unphysical, and if supported by the analysis method, would cast considerable doubts on the method's validity. Even if values $\gamma \geq 1$ are not supported, the railing of the reported parameter values against their prior range is a sign that these values may not be the best fits to the data; if these values are in fact arbitrary, reflecting the priors rather than the data, they cannot be reliably considered as evidence for a detection claim. It would be both helpful and prudent to show results of the analysis for wider prior ranges.

The third problem relates to how the background is estimated for their result, as displayed in their Fig. 5. For each time $t_{\rm echo}$ in a window covering offsets up to $\pm 5\%$ of $\Delta t_{\rm echo}$ after the merger, the matched filter SNR [11, 12] is maximized over the remaining parameters $\Delta t_{\rm echo}$, t_0, A and γ, either for GW150914 or for the combined events. In both cases the resulting peak of SNR is found to actually lie within 0.54% of $\Delta t_{\rm echo}$. They then estimate in each case how likely this peak would be in random noise by finding how often such a high peak occurs in data away from the merger. However, since they originally allowed the time offset to range over $\pm 5\% \Delta t_{\rm echo}$, they should account for possible, comparable background peaks occurring over that full range not only the restricted range $(0 - 0.54\%)\Delta t_{\rm echo}$. A naïve accounting for this post-hoc reduction in the extent of the parameter range would apply a trials factor of about 20 to the number of higher-SNR background samples, which would reduce the significance below 2σ. A more sophisticated treatment of false positives over the reduced parameter range [13] indicates a trials factor of $\mathcal{O}(10)$, weakly dependent on the number of independent samples in the SNR time series.

It is unclear why this background estimation was performed using a range of $t_{\rm echo}$ values that is only 9 to 38 echo periods away from the merger. If there is indeed an echo signal in the data then this region will not be entirely free of the signal being searched for. At the beginning of the region the amplitude of the echoes would only have dropped by a factor $0.9^9 \sim 0.4$. One therefore expects a contaminated background estimation. Each of the data sets released at [7] consist of 4096 seconds of data. Both GW150914 and LVT151012 are located 2048 seconds into this data, thus for large stretches of the data such contamination would be negligibly small. We expect that use of this relatively uncontaminated data would give a more self-consistent background estimate.

A full analysis of the data is outside the scope of this comment. Without a full analysis it is not possible to say whether the signals contain any true evidence of an echo signal, but as discussed here there are sufficient problems with the data analysis methodology of [6] to cast grave doubt on their claimed significance of a 2.9σ effect. It would be interesting to see the results of the analysis with these problems addressed, regarding both estimated parameters and estimated significance. In conclusion, we find that the evidence as presented in [6] is lacking in several key aspects, such that their current methodology cannot provide observational evidence for or against the existence of near-horizon Planck-scale structure in black holes.

The authors thank Karl Wette, Francesco Salemi, Marco Drago, Andrew Lundgren and Vitor Cardoso for useful discussions, and the authors of [6] for helpful communications.

---

[*] ofek.birnholtz@aei.mpg.de.

[1] LIGO Scientific and Virgo Collaborations, Phys. Rev. Lett. **116** (2016) no.6, 061102 [arXiv:1602.03837].
[2] LIGO Scientific and Virgo Collaborations, Phys. Rev. Lett. **116** (2016) no.24, 241103 [arXiv:1606.04855].
[3] LIGO Scientific and Virgo Collaborations, Phys. Rev. X **6** (2016) no.4, 041015 [arXiv:1606.04856].
[4] LIGO Scientific and Virgo Collaborations, Phys. Rev. Lett. **116** (2016) no.24, 241102 [arXiv:1602.03840].
[5] LIGO Scientific and Virgo Collaborations, Phys. Rev. Lett. **116** (2016) no.22, 221101 [arXiv:1602.03841].
[6] J. Abedi *et al.*, arXiv:1612.00266.
[7] LIGO Scientific Collaboration, "LIGO Open Science Center release of GW150914", 2016, `http://losc.ligo.org`
[8] V. Cardoso *et al.*, Phys. Rev. D **94** (2016) no.8, 084031 [arXiv:1608.08637].
[9] V. Cardoso *et al.*, Phys. Rev. Lett. **116** (2016) no.17, 171101 Erratum: [Phys. Rev. Lett. **117** (2016) no.8, 089902] [arXiv:1602.07309].
[10] B. Holdom *et al.*, [arXiv:1612.04889].
[11] LIGO Scientific Collaboration, "LOSC Event tutorial v1.4", 2016, `http://losc.ligo.org/s/events/GW150914/LOSC_Event_tutorial_GW150914.html`
[12] B. Allen *et al.*, Phys. Rev. D **85**, 122006 (2012) [gr-qc/0509116], Section IV.
[13] V. Connaughton *et al.*, Astrophys. J. **826** (2016) no.1, L6 [arXiv:1602.03920], Appendix B, 'Significance of two-parameter coincidence'.