Black Hole Spectroscopy with Coherent Mode Stacking

Huan Yang,[1] Kent Yagi,[1] Jonathan Blackman,[2] Luis Lehner,[3,4] Vasileios Paschalidis,[1] Frans Pretorius,[1,4] and Nicolás Yunes[5]
[1]Department of Physics, Princeton University, Princeton, New Jersey 08544, USA
[2]TAPIR, Walter Burke Institute for Theoretical Physics, California Institute of Technology, Pasadena, California 91125, USA
[3]Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada
[4]CIFAR, Cosmology and Gravity Program, Toronto, Ontario M5G 1Z8, Canada
[5]eXtreme Gravity Institute, Department of Physics, Montana State University, Bozeman, Montana 59717, USA
(Received 20 January 2017; published 20 April 2017)

The measurement of multiple ringdown modes in gravitational waves from binary black hole mergers will allow for testing the fundamental properties of black holes in general relativity and to constrain modified theories of gravity. To enhance the ability of Advanced LIGO/Virgo to perform such tasks, we propose a coherent mode stacking method to search for a chosen target mode within a collection of multiple merger events. We first rescale each signal so that the target mode in each of them has the same frequency and then sum the waveforms constructively. A crucial element to realize this coherent superposition is to make use of *a priori* information extracted from the inspiral-merger phase of each event. To illustrate the method, we perform a study with simulated events targeting the $\ell = m = 3$ ringdown mode of the remnant black holes. We show that this method can significantly boost the signal-to-noise ratio of the collective target mode compared to that of the single loudest event. Using current estimates of merger rates, we show that it is likely that advanced-era detectors can measure this collective ringdown mode with one year of coincident data gathered at design sensitivity.

DOI: 10.1103/PhysRevLett.118.161101

Introduction.—The recent detection of gravitational waves (GWs) emitted during the coalescence of binary black holes [1,2] marked the beginning of the era of gravitational wave astronomy, a feat that heralds a boom of scientific discoveries to come. GWs not only provide a new window to our Universe, they also offer a unique opportunity to test general relativity (GR) in the dynamical and highly nonlinear gravitational regime [3–7]. One celebrated prediction of GR is the uniqueness, or "no-hair," property of vacuum black holes (BHs) [8–12]: *All* isolated BHs are described by the Kerr family of solutions, each uniquely characterized by only its mass and spin [13]. This property has many wide-ranging consequences, the two most relevant here being (a) that the spacetime of an isolated binary black hole (BBH) inspiral is uniquely characterized by a small, finite set of parameters identifying the two BHs in the binary and the properties of the orbit and (b) that this same set of parameters uniquely determines the merger remnant and the full spectrum of its quasinormal mode (QNM) ringdown waveform.

This latter point forms the basis of black hole spectroscopy, where measurements of multiple ringdown modes are used to test this no-hair property. The idea is as follows. If the no-hair property holds, a measurement of the (complex) frequency of one QNM can be inverted to find a discrete set of possibilities for the spherical harmonic (ℓ, m) plus the overtone number n of the mode, and the BH mass M and spin parameter $a = |\vec{S}|/M^2$, where $\vec{S}$ is the BH spin angular momentum. However, if we have *a priori* information about the objects that merged to form the perturbed BH, then we also have information about the dominant (ℓ, m, n) QNM, and the measurement of its complex frequency then provides information about the mass and spin of the perturbed object. The measurement of any additional QNM frequencies then overconstrains this mass and spin measurement, providing independent tests of the no-hair property. Naturally, the results of such tests can then be leveraged to place constraints on (or to detect) non-Kerr BHs in modified gravity theories, exotic compact objects, the presence of exotic or unexpected matter fields, etc. (e.g., [15–30]).

In fact, aLIGO has already given us a "zeroth-order" test of the no-hair property from event GW150914: The inspiral-only portion of the signal was matched to a best-fit numerical relativity template, giving an estimate of the mass and spin of the remnant *and* informing that the waveform shortly after the peak amplitude should be dominated by the fundamental harmonic of the $(\ell, m) = (2, 2)$ QNM ("22 mode" for short); this was consistent with the independently measured properties of the postmerger signal [2]. More stringent tests of the no-hair property of the final BH require the observation of subleading QNMs [31]. This is challenging using individual merger events given how weak these subleading modes are relative to the primary mode [33,34]. For example, GW150914 has a ringdown signal-to-noise ratio (SNR) of ≈ 7, but a ringdown SNR upwards of 45 would have been needed to detect the first subleading QNM [18,35]. Thus, the

detection of such modes in individual events will require third-generation GW detectors, as even a loud GW150914-like event at aLIGO's design sensitivity would have a ringdown SNR of ≈20 [33]. On the other hand, many such events are expected after years of operation, leading us to consider how the information from multiple detections could be used to extract faint signals from a population of events.

Here then, we propose a way to coherently combine (or "stack") multiple, high total SNR (low ringdown SNR) binary BH coalescence events, to boost the detectability of a chosen secondary QNM mode. An earlier study in Ref. [36] considered a similar problem, though their approach effectively amounted to an incoherent assembly of ringdown signals, where, all else being equal, one expects $N^{1/4}$ scaling of the SNR for N events, compared to $N^{1/2}$ for a coherent method (see Supplemental Material [37] for more details). Key to achieving coherent stacking is using the information gleaned from the inspiral portion of each event to predict the relative phases and amplitudes of the ringdown modes excited in the remnant.

Signal stacking.—Given a set of BBH coalescence observations, we first select the loudest subset, here taken to consist of the signals with a ringdown SNR in the primary 22 mode alone of $\rho_{22} > 8$. Based on the studies in Refs. [16,34,35,48–50], the 33 mode is typically one of the next loudest ringdown modes. Therefore, we concentrate on the 33 mode as a target for our analysis, although the methodology presented here is generally applicable to other modes, as well as other features common to a population of GW events. Similar to the analysis in Refs. [33,34], we use the two-mode approximation to describe each detected ringdown signal $s_j(t)$:

$$s_j = n_j + h_{22,j} + h_{33,j}, \tag{1}$$

where the subscript j refers to the jth event, n_j is the corresponding detector noise, and $h_{\ell m,j}$ is a ringdown mode of the form (for $t > 0$)

$$h_{\ell m,j}(t) = A_{\ell m,j} e^{-\gamma_{\ell m,j} t} \sin(\omega_{\ell m,j} t - \phi_{\ell m,j}). \tag{2}$$

For each ringdown mode, $(\omega_{\ell m,j} + i\gamma_{\ell m,j})$ is its complex frequency, $A_{\ell m,j}$ its real amplitude, and $\phi_{\ell m,j}$ its constant phase offset.

Next, each entire jth signal is fitted to inspiral-merger-ringdown (IMR) waveform models in GR to accurately extract certain binary parameters that characterize the inspiral (e.g., the individual masses and spins) [51]. Using this, we can compute the QNM frequencies, phase offsets, and amplitudes for all modes as expected in GR [the extrinsic parameters, such as the polarization and inclination angles, do not affect the phase difference between the 22 and $\ell\ell$ modes ($\ell > 2$) [52], as we discuss in Supplemental Material [37]]. This is a key ingredient of our coherent mode stacking, as we need to properly align the phase offsets $\phi_{33,j}$ *and* frequencies $\omega_{33,j}$ of the targeted modes to achieve an optimal improvement in the SNR relative to a single-event analysis.

To perform the alignment, out of the set of N events, we arbitrarily pick one (e.g., the ith one) as the base case and shift or rescale all others to give the same expected secondary mode phase offset $\phi_{33,i} \equiv \phi_{33}$ and frequency $\omega_{33,i} \equiv \omega_{33}$. Specifically, we scale and shift each signal in time via $\mathbf{s}_j(t) \equiv s_j(t/\alpha_j + \Delta_j)$, with $\alpha_j \equiv \omega_{33,j}/\omega_{33}$ and $\Delta_j \equiv (\phi_{33,j} - \phi_{33})/\omega_{33,j}$.

We are now ready to combine the individual signals. For convenience, we work in the frequency domain, denoting the Fourier transform of a function $g(t)$ by $\tilde{g}(f)$. The Fourier transform of Eq. (2) is given by [35]

$$\tilde{h}_{\ell m,j}(f) = A_{\ell m,j} \frac{\omega_{\ell m,j} \cos\phi_{\ell m,j} - (\gamma_{\ell m,j} - i\omega)\sin\phi_{\ell m,j}}{\omega^2_{\ell m,j} - \omega^2 + \gamma^2_{\ell m,j} - 2i\omega\gamma_{\ell m,j}} \tag{3}$$

with $\omega = 2\pi f$ the angular Fourier frequency. In the frequency domain, the secondary mode alignment of Eq. (1) is achieved via $\tilde{\mathbf{s}}_j(f) \equiv \alpha_j e^{i\omega\Delta_j\alpha_j}\tilde{s}_j(\alpha_j f)$. We then sum up these phase- and frequency-aligned signals to obtain our composite signal: $\tilde{\mathbf{s}} = \sum_j c_j \tilde{\mathbf{s}}_j \equiv \tilde{\mathbf{n}} + \tilde{\mathbf{h}}_{22} + \tilde{\mathbf{h}}_{33}$, where the identification of $\tilde{\mathbf{n}}$, $\tilde{\mathbf{h}}_{22}$, and $\tilde{\mathbf{h}}_{33}$ is obvious, and we describe later how to optimize the choice of weight constants c_j. If the frequencies and phase offsets are known exactly, $\tilde{\mathbf{h}}_{33}$ contains a single oscillation frequency ω_{33}, and $\tilde{\mathbf{h}}_{22}$ contains a family of modes with (rescaled) frequencies $\in (0.623, 2/3)\omega_{33}$ as the dimensionless BH spin a ranges from 0 to 1 [16,54].

Parameter uncertainty.—Equation (1) decomposes a measured event into a true underlying signal and detector noise. The rescaling we have just described makes crucial use of parameters of the signal during the IMR phase, which can only be estimated to within some uncertainty, and this will introduce what we call "parameter estimation noise" n_h, that we will add to the composite signal $\tilde{\mathbf{s}}$. We investigate the role of this uncertainty here, leaving detailed derivations of some of the conclusions to Supplemental Material [37].

Parameter uncertainty produces two main sources of parameter estimation noise n_h. The first arises from subtracting an imperfectly estimated $\mathbf{h}_{22}$ from the data. This noise source has frequency components quite close to the scaled frequencies $\omega_{22,j}$, which (in relative terms, when comparing to $\omega_{33,j}$) are far from ω_{33}; the latter is the frequency at which $\tilde{\mathbf{h}}_{33}$ peaks, and, thus, the impact of this noise source on ρ_{33} is small. The second noise source is due to the imperfect scaling and alignment of the 33 mode, which is resonant at frequency ω_{33}.

Let us denote any variable with a prime as the maximum likelihood estimator, i.e., $Y' = Y + \delta Y$, with Y the true

(scaled or not) value and δY the corresponding uncertainty in its estimation. With this, the time domain, estimated composite GW signal is

$$\mathbf{h}_{22}{}' = \mathrm{Im}\left(\sum_j \mathbf{A}'_{22,j} e^{i(\Lambda'_{22,j}t - \Phi'_{22,j})}\right),$$
$$\mathbf{h}_{33}{}' = \mathrm{Im}\left(e^{i(\omega_{33}t-\phi_{33})}\sum_j \mathbf{A}'_{33,j} e^{-\Gamma'_{33,j}t + i(\delta\Omega_{33,j}t - \delta\Phi_{33,j})}\right), \quad (4)$$

where $\Omega_{\ell m,j} + i\Gamma_{\ell m,j} \equiv (\omega_{\ell m,j} + i\gamma_{\ell m,j})/\alpha_j \equiv \Lambda_{\ell m,j}$ and $\Phi_{\ell m,j} \equiv \phi_{\ell m,j} - \Delta_j\omega_{\ell m,j}$ are the scaled frequencies and phase offsets, respectively, and we have absorbed the c_j coefficients into rescaled amplitudes $\mathbf{A}_{\ell m,j}$. The parameter estimation noise for each (ℓ, m) mode is $n_{h_{\ell m}} = \mathbf{h}'_{\ell m} - \mathbf{h}_{\ell m}$, which is approximately given by

$$n_{h_{\ell m}} \approx \mathrm{Im}\Big\{\sum_j [\delta\mathbf{A}_{\ell m,j} e^{i(\Lambda_{\ell m,j}t - \Phi_{\ell m,j})} + \mathbf{A}_{\ell m,j} e^{i(\Lambda_{\ell m,j}t - \Phi_{\ell m,j})}(e^{i(\delta\Lambda_{\ell m,j}t - \delta\Phi_{\ell m,j})} - 1)]\Big\}. \quad (5)$$

In the subsequent analysis, we assume that $\delta\mathbf{A}$, $\delta\Lambda$, and $\delta\Phi$ are independent, normal random variables in the probability space of $\mathbf{n}$ [55].

We are unaware of any closed-form, analytic formula in the literature that describes parameter uncertainties given the SNR of a particular detection, even when the waveform model is known analytically. Let us then assume one characterizes the data with an inspiral-merger-ringdown model, where the ringdown contains the 22 and 33 modes. These ringdown modes depend (of course) on the ringdown parameters and the underlying gravitational theory governing the dynamics, though in our analysis we are assuming GR as the theory and hence they fundamentally depend on the parameters of the inspiral. The uncertainty in the inspiral parameters depends inversely on the total SNR ρ of the observation, as can be shown via a simple Fisher analysis, which then also provides the uncertainty of the ringdown parameters to within a factor given by the propagation of errors from the inspiral to ringdown parameters. Guided by an estimate of this propagation factor as outlined in Supplemental Material [37], together with aLIGO's parameter estimation errors for event GW150914 [1,6,56], we estimate the variance of mode parameter uncertainties as $\sigma_{\Phi_{ii,j}} = 0.3 \times (20/\rho_j)$ rad ($i = 1, 2$) and use the QNM frequency formula and the formula for $A_{ii,j}$ to propagate the mass uncertainty of event GW150914 to obtain estimates for $\sigma_{\Lambda_{ii,j}}$ and $\sigma_{A_{ii,j}}$.

Hypothesis testing.—With the combined signals, we perform a Bayesian hypothesis test [35] to derive the conditions of detectability of the 33 mode. In particular, we want to test the following two nested hypotheses:

$$\mathcal{H}1: \tilde{y} \equiv \tilde{\mathbf{s}} - \tilde{\mathbf{h}}_{22} = \tilde{\mathbf{n}} + A\tilde{\mathbf{h}}_{33},$$
$$\mathcal{H}2: \tilde{y} \equiv \tilde{\mathbf{s}} - \tilde{\mathbf{h}}_{22} = \tilde{\mathbf{n}}. \quad (6)$$

For convenience, we have introduced an overall amplitude factor A such that when $A \neq 0$ the 33 mode is nonzero, and vice versa. The probability that the observed data are consistent with $\mathcal{H}_1$ is

$$P_A \propto \exp\left[-\int_0^\infty df \frac{2|\tilde{y} - A\tilde{\mathbf{h}}_{33}|^2}{S_n}\right], \quad (7)$$

with $S_n = \sum_j c_j^2 S_{n_j}(\alpha_j f)\alpha_j$ the one-sided and shifted noise spectrum (with S_{n_j} the unscaled detector noise spectral density for each detection).

With the above probability function, we can derive the maximum likelihood estimator for A and then perform a generalized likelihood ratio test (GLRT) [35]. As we explain in detail in Supplemental Material [37], parameter uncertainties shift the mean and expand the variance of the distribution of the likelihood ratio between the two hypotheses. The former effectively reduces the 33 mode to

$$\mathbf{H}_{33} = \left[1 + \frac{1}{2}\left(\left\langle \frac{\langle n_{h_{33}}|n_{h_{33}}\rangle}{\langle \mathbf{h}_{33}|\mathbf{h}_{33}\rangle}\right\rangle - \left\langle \frac{\langle \mathbf{h}_{33}|n_{h_{33}}\rangle^2}{\langle \mathbf{h}_{33}|\mathbf{h}_{33}\rangle^2}\right\rangle\right)\right]\langle \mathbf{h}_{33}\rangle \quad (8)$$

(see Supplemental Material [37] for the definition of the inner product $\langle|\rangle$ and explicit form of $\langle \mathbf{h}_{33}\rangle$), while the latter directly reduces the SNR of the 33 mode by $\sqrt{1+\sigma_p^2}$, where σ_p^2 is the variance of $\langle \mathbf{h}_{33}|n_{h_{22}} - n_{h_{33}}\rangle[\langle \mathbf{h}_{33}|\mathbf{h}_{33}\rangle]^{-1/2}$. Thus, the requirement to favor $\mathcal{H}_1$ over $\mathcal{H}_2$ is

$$\rho_{33} \equiv \frac{\sqrt{\langle \mathbf{H}_{33}|\mathbf{H}_{33}\rangle}}{\sqrt{1+\sigma_p^2}} \geq \rho_{\mathrm{crit}}, \quad (9)$$

where ρ_{crit} is related to the false-alarm rate P_f and detection rate P_d in the GLRT. If we choose $P_f = 0.01$ and $P_d = 0.99$, ρ_{crit} would be 4.65, which is close to the threshold 5 set in Ref. [34]. Here we also pick $\rho_{\mathrm{crit}} = 5$.

Assessing observational prospects.—To investigate the detectability of the 33 mode after coherent stacking, we employ a Monte Carlo (MC) sampling of possible events, repeating each sampling 100 times to accumulate statistics. Given the predictions derived from the recent GW detections [6], we assume a uniform merger rate of quasicircular inspirals of 40 $\mathrm{Gpc}^{-3}\ \mathrm{yr}^{-1}$ in comoving volume. For simplicity, we assume the BHs are nonspinning (see Supplemental Material [37] for the effect of BH spins on the relative phase difference between the 22 and 33 modes) with masses uniformly distributed $\in [10\text{–}50]M_\odot$ and employ the empirical fitting formula of Ref. [58] to connect the initial BH masses to the final mass and spin of

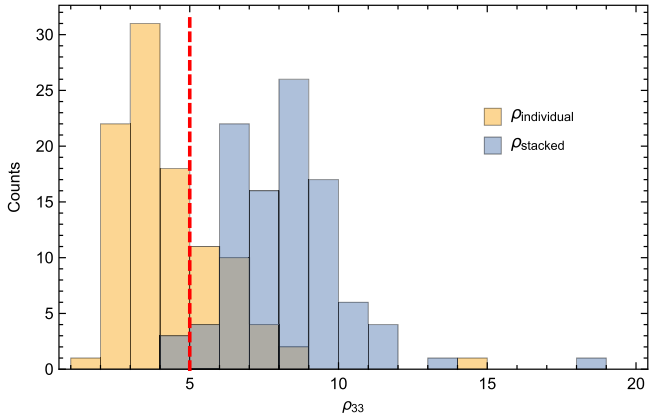

FIG. 1. A histogram of the SNR of the 33 mode, ρ_{33}, from 100 randomly sampled sets of detections, assuming a one-year data acquisition time for aLIGO and a uniform comoving merger rate of 40 Gpc^{-3} yr^{-1} [6]. We present the statistics of the largest ρ_{33} event from each set (orange bins) and those with the stacked SNR using only the 15 largest SNR events from each set (blue bins). The 33 mode is detected if ρ_{33} is above the detection threshold of $\rho_{33} = 5$ (red dashed line). Refer to the main text for more details.

the remnant. We compute the total SNR for each individual event using the sky-averaged IMRPhenomB waveform model [59], choose the amplitude of the primary 22 mode to match the ringdown SNR in Eq. (1) of Ref. [33], and set the amplitude of the 33 mode following the fitting formula for A_{33}/A_{22} in Ref. [48]. We adopt the zero-detuned, high-power noise spectral density of aLIGO at design sensitivity [60] for $S_{n_j}(f)$. For each MC sampling, we randomly distribute merger events within redshift $z = 1$ (with $H_0 = 70$ km s^{-1} Mpc^{-1}, $\Omega_m = 0.3$) over a one-year observation period and, as discussed earlier, select only those with $\rho_{22} > 8$. Each MC sampling contains about 1000–2000 events, giving rise to 40–65 events with $\rho_{22} > 8$, which is roughly 2 times higher than samples taken using population synthesis models [33]. In computing the stacked signal SNR of Eq. (9), it suffices to use a small number (15) of loudest events in each sample [61], and we determine the weight constants c_j in the sum to maximize the SNR using the downhill simplex optimization method [62,63].

The resulting distribution (Fig. 1) indicates that there is roughly a 28% chance for aLIGO to resolve at least one 33 mode from a single event in one year of data at design sensitivity. After stacking, the probability of a collective 33-mode detection increases to 97%. These probabilities, of course, depend on the actual merger rate, as well as additional factors we have not taken into account here, including initial BH spins and precession. For example, if we take the more pessimistic event rate estimate of 13 Gpc^{-3} yr^{-1} [6], the probability for detection with a single event drops to ∼12%, while the collective mode detection probability drops to 50% (still using 15 events).

In theory, all else being equal, coherent stacking should provide a $\sqrt{N}$ scaling of the SNR. Here $N = 15$, so the ideal scenario would see a factor of ∼3.8 improvement in the collective ρ_{33} relative to a single event. In our MC realizations, we achieved improvement factors of between 1.3 and 3.1 relative to the loudest event over the set of 100 realizations (see Supplemental Material [37] for some additional comments and figures about the distribution) [64]. The primary reason for this is simply the nonuniform nature of the sampling, where it is typically the small handful of loudest events that contribute most to the collective SNR. The parameter uncertainty noise has a smaller impact, in particular, because the fainter events that have larger uncertainties are weighted less in the sum.

Discussion.—We have presented a coherent mode stacking method that uses multiple high-quality BBH coalescence detections to obtain better statistics for BH spectroscopy. Crucial to the method's success is the appropriate alignment of the phase and frequency from different signals. For the class of BBH merger events we have targeted here, this is achievable for two primary reasons: (i) the no-hair properties of isolated BHs in GR imply that a binary system is likewise described by a small set of parameters, and (ii) the expected events that aLIGO will detect where the primary ringdown mode is visible will also have an inspiral detectable with a high SNR, and this can be used to estimate the parameters in (i) with enough accuracy to predict the initial phases and amplitudes of subdominant ringdown modes. In this first, proof-of-principle study, we have demonstrated that the detection of a collective secondary BH ringdown mode through stacking is likely with the current "advanced" generation of ground-based GW detectors, even if the corresponding modes are not loud enough to be detected in any single-event analysis.

There are many avenues for future work and extensions of this method, including using merger rates predicted by population synthesis models as done in Ref. [33], using different mass distribution functions for BHs [65], considering other ringdown modes (such as the 44 and 21 modes [33,34], or even the fundamental 22 mode in a population of low SNR events where it is not individually detectable), investigating the (precessional) spin effect to the phase of secondary modes, and also targeting secondary inspiral modes. Furthermore, this method could be adapted to constrain or search for other small-amplitude features that might be shared by a population of events, e.g., common parameterized post-Einsteinian-like [66] corrections to the inspiral phase of the mergers, or common equation-of-state-discriminating frequencies excited in hypermassive remnants of binary neutron star mergers [67–77]. In this latter example, one issue in adapting the coherent stacking method would be achieving phase alignment, due to the challenge in accurately calculating the details of the matter dynamics postmerger. If the phases cannot be aligned, incoherent power stacking could still in theory achieve a $N^{1/4}$ SNR scaling (see Supplemental Material [37] for more details).

H. Y. thanks Haixing Miao for sharing the code for downhill simplex optimization. The authors thank Emanueli Berti, Swetha Bhagwat, Vitor Cardoso, Neil Cornish, Kendrick Smith, Chris Van Den Broeck, and John Veitch for valuable discussions and comments. K. Y. acknowledges support from JSPS Postdoctoral Fellowships for Research Abroad. F. P. and V. P. acknowledge support from NSF Grant No. PHY-1607449 and the Simons Foundation. V. P. also acknowledges support from NASA Grant No. NNX16AR67G (Fermi). N. Y. acknowledges support from NSF CAREER Grant No. PHY-1250636. Computational resources were provided by XSEDE/TACC under Grant No. TG-PHY100053. This research was supported in part by NSERC and in part by the Perimeter Institute for Theoretical Physics. Research at Perimeter Institute is supported by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Research and Innovation.

[1] B. P. Abbott *et al.* (Virgo, LIGO Scientific Collaboration), Phys. Rev. Lett. **116**, 061102 (2016).
[2] B. P. Abbott *et al.* (Virgo, LIGO Scientific Collaboration), Phys. Rev. Lett. **116**, 241103 (2016).
[3] N. Yunes and X. Siemens, Living Rev. Relativ. **16**, 9 (2013).
[4] B. P. Abbott *et al.* (Virgo, LIGO Scientific Collaboration), Phys. Rev. Lett. **116**, 221101 (2016).
[5] N. Yunes, K. Yagi, and F. Pretorius, Phys. Rev. D **94**, 084002 (2016).
[6] B. P. Abbott *et al.* (Virgo, LIGO Scientific Collaboration), Phys. Rev. X **6**, 041015 (2016).
[7] P. D. Lasky, E. Thrane, Y. Levin, J. Blackman, and Y. Chen, Phys. Rev. Lett. **117**, 061102 (2016).
[8] W. Israel, Phys. Rev. **164**, 1776 (1967).
[9] B. Carter, Phys. Rev. Lett. **26**, 331 (1971).
[10] S. W. Hawking, Commun. Math. Phys. **25**, 152 (1972).
[11] D. C. Robinson, Phys. Rev. Lett. **34**, 905 (1975).
[12] V. Cardoso and L. Gualtieri, Classical Quantum Gravity **33**, 174001 (2016).
[13] An astrophysical environment is not a pure vacuum, though it is expected that any ambient matter, radiation, or charge about an aLIGO merger event will have an insignificant effect on the spacetime dynamics and corresponding GW emission. Also see discussions in Ref. [14].
[14] E. Barausse, V. Cardoso, and P. Pani, Phys. Rev. D **89**, 104059 (2014).
[15] O. Dreyer, B. J. Kelly, B. Krishnan, L. S. Finn, D. Garrison, and R. Lopez-Aleman, Classical Quantum Gravity **21**, 787 (2004).
[16] E. Berti, V. Cardoso, and C. M. Will, Phys. Rev. D **73**, 064030 (2006).
[17] I. Kamaretsos, M. Hannam, S. Husa, and B. S. Sathyaprakash, Phys. Rev. D **85**, 024018 (2012).
[18] S. Gossan, J. Veitch, and B. S. Sathyaprakash, Phys. Rev. D **85**, 124056 (2012).
[19] R. Jackiw and S. Y. Pi, Phys. Rev. D **68**, 104012 (2003).
[20] S. Alexander and N. Yunes, Phys. Rep. **480**, 1 (2009).
[21] N. Yunes and F. Pretorius, Phys. Rev. D **79**, 084043 (2009).
[22] K. Yagi, N. Yunes, and T. Tanaka, Phys. Rev. D **86**, 044037 (2012).
[23] B. A. Campbell, N. Kaloper, and K. A. Olive, Phys. Lett. B **285**, 199 (1992).
[24] P. Kanti, N. E. Mavromatos, J. Rizos, K. Tamvakis, and E. Winstanley, Phys. Rev. D **54**, 5049 (1996).
[25] N. Yunes and L. C. Stein, Phys. Rev. D **83**, 104002 (2011).
[26] P. Pani and V. Cardoso, Phys. Rev. D **79**, 084031 (2009).
[27] N. Yunes and L. C. Stein, Phys. Rev. D **83**, 104002 (2011).
[28] P. Pani, C. F. B. Macedo, L. C. B. Crispino, and V. Cardoso, Phys. Rev. D **84**, 087501 (2011).
[29] D. Ayzenberg and N. Yunes, Phys. Rev. D **90**, 044066 (2014); **91**, 069905(E) (2015).
[30] R. Brito, V. Cardoso, and P. Pani, Phys. Rev. D **88**, 064006 (2013).
[31] See, e.g., [32] for a consistency test of GR with the dominant ringdown mode only.
[32] H. Nakano, T. Tanaka, and T. Nakamura, Phys. Rev. D **92**, 064003 (2015).
[33] E. Berti, A. Sesana, E. Barausse, V. Cardoso, and K. Belczynski, Phys. Rev. Lett. **117**, 101102 (2016).
[34] S. Bhagwat, D. A. Brown, and S. W. Ballmer, Phys. Rev. D **94**, 084024 (2016).
[35] E. Berti, J. Cardoso, V. Cardoso, and M. Cavaglia, Phys. Rev. D **76**, 104044 (2007).
[36] J. Meidam, M. Agathos, C. Van Den Broeck, J. Veitch, and B. S. Sathyaprakash, Phys. Rev. D **90**, 064009 (2014).
[37] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.118.161101 for details in deriving the hypothesis test, discussion in determining phase uncertainties and comparison to power stacking method, which includes Ref. [38–47].
[38] M. Shahram and P. Milanfar, IEEE Trans. Signal Process. **53**, 2579 (2005).
[39] J. Blackman, S. E. Field, C. R. Galley, B. Szilagyi, M. A. Scheel, M. Tiglio, and D. A. Hemberger, Phys. Rev. Lett. **115**, 121102 (2015).
[40] J. Blackman, S. E. Field, M. A. Scheel, C. R. Galley, D. A. Hemberger, P. Schmidt, and R. Smith, arXiv:1701.00550.
[41] A. Klein, P. Jetzer, and M. Sereno, Phys. Rev. D **80**, 064027 (2009).
[42] E. Berti, V. Cardoso, and M. Casals, Phys. Rev. D **73**, 024013 (2006); **73**, 109902(E) (2006).
[43] E. Berti and A. Klein, Phys. Rev. D **90**, 064012 (2014).
[44] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. Lett. **116**, 241102 (2016).
[45] P. Kalmus, K. C. Cannon, S. Marka, and B. J. Owen, Phys. Rev. D **80**, 042001 (2009).
[46] Z. B. Etienne, J. A. Faber, Y. T. Liu, S. L. Shapiro, K. Taniguchi, and T. W. Baumgarte, Phys. Rev. D **77**, 084002 (2008).
[47] R. Gold, V. Paschalidis, M. Ruiz, S. L. Shapiro, Z. B. Etienne, and H. P. Pfeiffer, Phys. Rev. D **90**, 104030 (2014).
[48] L. London, D. Shoemaker, and J. Healy, Phys. Rev. D **90**, 124032 (2014); **94**, 069902(E) (2016).
[49] A. Buonanno, G. B. Cook, and F. Pretorius, Phys. Rev. D **75**, 124018 (2007).

[50] E. Berti, V. Cardoso, J. A. Gonzalez, U. Sperhake, M. Hannam, S. Husa, and B. Brügmann, Phys. Rev. D **76**, 064034 (2007).
[51] Parameter uncertainties scale inversely with SNR, and thus they will likely be smaller than uncertainties in parameter extraction with GW150914. Such uncertainties should have a small effect on the final BH mass and spin measurements, as was the case for GW150914 [1].
[52] If one wishes to search for a subdominant ringdown mode with $\ell \neq m$, one needs to take into account uncertainties of the extrinsic parameters in the phase difference between the 22 mode and the ℓm mode. For example, based on extrinsic parameter uncertainties for stellar-mass BH binaries in Fig. 11 of Ref. [53] with a network of three interferometers (and neglecting correlations among parameters), we found that such uncertainties introduce an error on the phase difference of the 22 and 21 mode as $\sim 0.2 \times (20/\rho)$ rad, which is smaller than the error from intrinsic parameter uncertainties that we used in our analysis.
[53] J. Veitch and A. Vecchio, Phys. Rev. D **81**, 062003 (2010).
[54] H. Yang, D. A. Nichols, F. Zhang, A. Zimmerman, Z. Zhang, and Y. Chen, Phys. Rev. D **86**, 104006 (2012).
[55] We neglect correlations between these variables. We have verified that, after normalizing the Fisher matrix of these variables calculated from the propagation of errors of inspiral parameters such that all diagonal components are unity, off-diagonal components are smaller than 44% for a GW150914-like event. We leave it to future work to investigate this more thoroughly.
[56] The calibration error could also contribute to the total phase error. According to the discussion in Ref. [57], the calibration phase errors are 10° or ∼0.17 rad and, hence, smaller than the 0.3 rad standard deviation we considered in our work. Thus, as long as all calibration phase errors are comparable to 0.2 rad, we expect that their impact on our results should be small.
[57] B. P. Abbott *et al.* (L. S. Collaboration), Phys. Rev. D **95**, 042003 (2017).
[58] S. Husa, S. Khan, M. Hannam, M. Purrer, F. Ohme, X. J. Forteza, and A. Bohe, Phys. Rev. D **93**, 044006 (2016).
[59] P. Ajith *et al.*, Phys. Rev. Lett. **106**, 241101 (2011).
[60] P. Ajith, Phys. Rev. D **84**, 084037 (2011).
[61] The choice of the 15 loudest events was to lower the computational cost of the optimization, and future work will investigate the optimal choice of N to balance minimizing the computational cost vs maximizing the SNR.
[62] J. A. Nelder and R. J. Mead, Computer Journal (UK) **7**, 308 (1965).
[63] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes: The Art of Scientific Computing* (Cambridge University Press, Cambridge, England, 2007).
[64] In judging the benefit of coherent stacking, a more equitable comparison may be to compare to a common existing strategy of combining posteriors of confirmed detections (as discussed more in Supplemental Material [37]) rather than the SNR gain relative to the single loudest event. However, only 5% of our MC year-long simulated data sets had two events loud enough for individual 33-mode detections. Therefore, unless we are extremely fortunate with loud events early on or rates are more optimistic than expected, coherent stacking (or an equivalently modified Bayesian approach) will be the only practical method to go after these subleading modes in the near future.
[65] K. Belczynski, M. Dominik, T. Bulik, R. O'Shaughnessy, C. Fryer, and D. E. Holz, Astrophys. J. Lett. **715**, L138 (2010).
[66] N. Yunes and F. Pretorius, Phys. Rev. D **80**, 122003 (2009).
[67] N. Stergioulas, A. Bauswein, K. Zagkouris, and H.-T. Janka, Mon. Not. R. Astron. Soc. **418**, 427 (2011).
[68] K. Takami, L. Rezzolla, and L. Baiotti, Phys. Rev. Lett. **113**, 091104 (2014).
[69] K. Takami, L. Rezzolla, and L. Baiotti, Phys. Rev. D **91**, 064001 (2015).
[70] A. Bauswein, N. Stergioulas, and H.-T. Janka, Eur. Phys. J. A **52**, 56 (2016).
[71] A. Bauswein and N. Stergioulas, Phys. Rev. D **91**, 124056 (2015).
[72] V. Paschalidis, W. E. East, F. Pretorius, and S. L. Shapiro, Phys. Rev. D **92**, 121502 (2015).
[73] W. E. East, V. Paschalidis, F. Pretorius, and S. L. Shapiro, Phys. Rev. D **93**, 024011 (2016).
[74] W. E. East, V. Paschalidis, and F. Pretorius, Classical Quantum Gravity **33**, 244004 (2016).
[75] L. Lehner, S. L. Liebling, C. Palenzuela, O. L. Caballero, E. O'Connor, M. Anderson, and D. Neilsen, Classical Quantum Gravity **33**, 184002 (2016).
[76] D. Radice, S. Bernuzzi, and C. D. Ott, Phys. Rev. D **94**, 064011 (2016).
[77] L. Lehner, S. L. Liebling, C. Palenzuela, and P. M. Motl, Phys. Rev. D **94**, 043003 (2016).