Gravitational wave spectroscopy of binary neutron star merger remnants with mode stacking

Huan Yang, Vasileios Paschalidis, Kent Yagi, Luis Lehner, Frans Pretorius, and Nicolás Yunes

Gravitational wave spectroscopy of binary neutron star merger remnants with mode stacking

Huan Yang,[1] Vasileios Paschalidis,[1, 2] Kent Yagi,[1] Luis Lehner,[3, 4] Frans Pretorius,[1, 4] and Nicolás Yunes[5]

[1] *Department of Physics, Princeton University, Princeton, New Jersey 08544, USA.*
[2] *Theoretical Astrophysics Program, Departments of Astronomy and Physics,*
University of Arizona, Tucson, AZ 85721, USA.
[3] *Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada*
[4] *CIFAR, Cosmology & Gravity Program, Toronto, ON M5G 1Z8, Canada*
[5] *eXtreme Gravity Institute, Department of Physics,*
Montana State University, Bozeman, Montana 59717, USA

A binary neutron star coalescence event has recently been observed for the first time in gravitational waves, and many more detections are expected once current ground-based detectors begin operating at design sensitivity. As in the case of binary black holes, gravitational waves generated by binary neutron stars consist of inspiral, merger, and post-merger components. Detecting the latter is important because it encodes information about the nuclear equation of state in a regime that cannot be probed prior to merger. The post-merger signal, however, can only be expected to be measurable by current detectors for events closer than roughly ten megaparsecs, which given merger rate estimates implies a low probability of observation within the expected lifetime of these detectors. We carry out Monte-Carlo simulations showing that the dominant post-merger signal (the $\ell = m = 2$ mode) from individual binary neutron star mergers may not have a good chance of observation even with the most sensitive future ground-based gravitational-wave detectors proposed so far (the Einstein Telescope and Cosmic Explorer, for certain equations of state, assuming a full year of operation, the latest merger rates, and a detection threshold corresponding to a signal-to-noise ratio of 5). For this reason, we propose two methods that stack the post-merger signal from multiple binary neutron star observations to boost the post-merger detection probability. The first method follows a commonly-used practice of multiplying the Bayes factors of individual events. The second method relies on an assumption that the mode phase can be determined from the inspiral waveform, so that coherent mode stacking of the data from different events becomes possible. We find that both methods significantly improve the chances of detecting the dominant post-merger signal, making a detection very likely after a year of observation with Cosmic Explorer for certain equations of state. We also show that in terms of detection, coherent stacking is more efficient in accumulating confidence for the presence of post-merger oscillations in a signal than the first method. Moreover, assuming the post-merger signal is detected with Cosmic Explorer via stacking, we estimate through a Fisher analysis that the peak frequency can be measured to a statistical error of $\sim$ 4–20 Hz for certain equations of state. Such an error corresponds to a neutron star radius measurement to within $\sim$ 15-56 m, a fractional relative error $\sim$ 4%, suggesting that systematic errors from theoretical modeling ($\gtrsim$ 100 m) may dominate the error budget.

I. INTRODUCTION

The LIGO/Virgo collaboration recently announced the first detection of a gravitational wave (GW) signal consistent with the inspiral and merger of a binary neutron star (BNS) system [1], corroborated by numerous observations of electromagnetic counterparts across the spectrum, from radio to gamma rays [2]. This one event has already provided a wealth of new information: highlights include the establishment of a connection between NS mergers and (at least a class of) short gamma ray bursts, evidence that a significant fraction of the universe's r-process elements are born in NS mergers, an upper-bound constraint on the tidal deformability of neutron stars, a measurement of the Hubble constant independent of the cosmic distance ladder, and a stringent constraint that the speed of gravitational waves equals the speed of light.

As loud as GW170817 was in gravitational waves with a network signal-to-noise (SNR) ratio of 32, this still all came from the inspiral phase of the event, and no detectable merger/post-merger signal was reported by the LIGO/Virgo collaboration. This is not surprising, as regardless of what the outcome of the merger may have been—prompt or delayed collapse to a black hole, or a stable high mass NS remnant— the corresponding GW emission is not expected to be loud enough to allow extraction from the noise at the relevant frequencies > 1kHz. Thus, at present only informed guesses as to the nature of the remnant can be made, based on the consistency of models of post-merger electromagnetic (EM) emission processes with observations (insofar as the emission depends on properties of the remnant; see e.g. [3]). Given how crude existing models of the post-merger central engine of the EM counterparts are, it would be ideal to instead measure properties of the remnant in GWs, and use that to inform interpretation of the counterpart emission.

Beyond helping to decipher the EM data, the merger/post-merger GW signal can contain much information of intrinsic value in understanding the physics of

the remnant. As mentioned, following the merger, the BNS remnant may either promptly collapse to a black hole (BH), form a supramassive neutron star or form a hypermassive neutron star (HMNS) that will ultimately undergo delayed collapse to a BH (see [4] for a recent review). The latter two scenarios lead to a remnant that spins rapidly and undergoes non-axisymmetric oscillations, emitting GWs in the process. More than a decade of simulations of BNS mergers have revealed that the post-merger GW spectrum is rich, with several distinct peaks that can be used to probe the merger remnant through spectroscopy (see e.g. [5–17] for some historical and recent work, [18–21] for related work on BH spectroscopy, and [22, 23] for recent reviews). In the first 10-20 ms after merger, the dominant component of a post-merger GW is the $\ell = m = 2$ mode (which we call here the 22 mode for short). For BNS merger remnants that may survive for longer times, a one-arm mode ($\ell = 2, m = 1$, or 21 mode) can dominate the GW emission [24–28].

Extracting post-merger information from GWs is also crucial for obtaining a full understanding of the physics of nuclear matter. Individual NSs (with mass M_{NS}) in inspiralling binaries are described by cold nuclear matter, whereas BNS merger remnants (with mass $\sim 2M_{NS}$) are described by hot nuclear matter. Therefore, GWs from BNS merger remnants encode the physics of dense nuclear matter in a regime that is not accessible in the inspiral phase. In addition, measured post-merger GWs could reduce the uncertainties in information drawn from the inspiral phase, just as with binary BH mergers (e.g. [29, 30]). Moreover, these waves will provide further insight to help disentangle degeneracies between modulations due to tidal effects from those induced by deviations from General Relativity (e.g. [31–34]). We can also anticipate that information on the interior composition of cold neutron stars will be available through independent electromagnetic observations, for example with the recently launched NICER [35] (under suitable assumptions, NICER may determine neutron star radii to $\simeq 5\%$ accuracy which, in turn, will help constrain the cold nuclear EOS). One can then envision either employing such knowledge to further constrain GW predictions, or use the GW observations independently and crosscheck for consistency with results from EM observations.

GWs from BNS post-merger oscillations are challenging to detect, as indicated by previous studies [36, 37], and evidenced by a lack of detection of any with GW170817 [38]. With certain binary parameters, EOS, etc., an event not too much closer than GW170817 could produce a detectable post-merger signal with aLIGO sensitivities. However, even with the more optimistic BNS merger rates of $1540(+3200 - 1220)/\text{Gpc}^3/\text{yr}$ implied by GW170817 [1], a similarly loud merger is roughly a once-per-decade event. On the other hand, these merger rates suggest several events (including GW170817) could be expected in the aLIGO era, and even more in the era of third-generation ground-based detectors, that are all

within a factor of a few in SNR of having individually detectable post-merger signals. If there is a common post-merger signal in these anticipated events, we can therefore attempt to go after the common component by combining, or stacking, the data from multiple events appropriately. The data may also be combined through unmodelled algorithms (e.g., [39]) for parameter estimation purpose.

To simplify the analysis in our first study of this idea, we only include the 22 mode of the post-merger signal. We model it as an exponentially decaying sinusoidal function, which is consistent with the leading order behavior identified in the principal component analysis of BNS post-merger waveforms in [37]. We propose two methods to stack this data from different detections. With the first method, we treat all events as independent and combine the Bayes factor, following a similar approach as discussed in [40] (referred to as "power stacking" in this work). In the second method, we assume that the theoretical uncertainties in future numerical BNS simulations can be significantly reduced, such that the inspiral waveform can be used to predict the phase of post-merger modes. In this case, the dominant modes from different post-merger signals can be coherently stacked together, as shown in the black hole ringdown scenario [41]. Essentially, coherent mode stacking is the shifting and rescaling of N signals to align their phase using information from the inspiral in order to construct a weighted, linear superposition that boosts the post-merger SNR. Both methods are able to boost the detectability of post-merger oscillations. The coherent stacking approach outperforms the first method by taking advantage of the additional phase information.

There are several important issues to note in this work. First, the coherent stacking procedure presented here is similar to the treatment we developed in [41], which was designed to boost particularly relevant features in signals from binary black hole mergers, for instance the SNR of secondary modes in BH ringdowns. However, there are important differences between the stacking approach developed in this paper and in [41]. In particular, the inspiral-merger-ringdown waveform of binary BHs is known from numerical relativity simulations sufficiently accurately that it can be used to predict the phase of secondary modes, which in turn set the basis to align the secondary modes from different detected events. By contrast, current numerical relativity simulations of BNS mergers cannot reliably determine the phase of post-merger oscillations, partly because there are important pieces of physics (such as turbulent magnetohydrodynamics, microphysical effects, NS spin effects etc.) that are not fully resolved or accounted for. Despite the significant progress in our understanding of BNS post-merger physics (see [22, 23] for recent reviews), there remain obstacles both in the computational aspect and the physical understanding of the problem that must be overcome before reliable GWs from numerical relativity simulations can be used to construct GW templates. Therefore, in

this work we emphasize the application of power stacking, and also generalize the hypothesis test formalism (Generalized Likelihood Ratio Test or GLRT) of [41] to signals with unknown phase.

Second, although for a given EOS the frequency (unlike the phase) of the 22 mode can be robustly determined by numerical simulations [12], the true underlying EOS is unknown. Thus, in order to perform the hypothesis test for detection, we assume an underlying EOS to compute the mode frequencies for each event. Picking an incorrect EOS would in principle generate frequency mismatch which would degrade the SNR of the signal. On the other hand, one can perform a model-selection study to compare different EOSs for their relative consistency with the data. We investigate this issue here as well.

Third, as mentioned earlier, the BNS merger remnant can undergo collapse to a BH promptly after the merger, in which case there is no GW signal from a HMNS to stack. According to the work of [8] there exists a threshold total binary mass that determines whether prompt collapse will take place, independently of the mass ratio. Therefore, we only consider events with total mass below this threshold in our Monte-Carlo (MC) simulations before stacking. Since in this study we focus on finite-temperature, realistic nuclear EOSs we use the threshold masses for prompt collapse determined in [42]. We show how detection/non-detection of a stacked signal from a suitable population of events can provide a direct test of this collapse hypothesis, and further be used to place constraints on the nuclear EOS.

Finally, the starting time of coalescence may be subject to systematic uncertainties in modeling the tidal effects of binary NSs in the inspiral stage using post-Newtonian methods. This does not significantly affect the calculations in this paper, as we mainly focus on the properties of the 22 (peak) mode, which radiates waves with frequency above 2kHz (well above the merger frequency). In addition, an accurate numerical waveform would naturally take into account all tidal effects.

A. Executive summary

We now summarize the main results of this paper. Unless otherwise specified, for the sake of presentation our calculations will focus on the Cosmic Explorer (CE) experiment as the representative third-generation instrument; we expect similar conclusions to hold for both CE and the Einstein Telescope (ET). Based on the MC simulations we have performed, given an EOS (TM1 [43] for reference) and with the adopted BNS merger rate, the chances of detecting a single post-merger event after one year of observations with third generation detectors are good, but not certain (for simplicity, here and henceforth we use the word "event" to only refer to the post-merger signal). By stacking the loudest events, a detection becomes almost certain after a year of observations with CE. For example, if the SNR threshold for detection is

set to 5, a one-year observation with CE has a $\sim 79\%$ chance of detecting a post-merger oscillation signal in a single event, while the chances increase to $\sim 100\%$ after power stacking the top 5 loudest events.

Apart from power stacking that simply multiplies the Bayes factor of each event [44–46] [1], we also investigate combining signals if the phase of the post-merger modes can be predicted using simulations informed by source parameters measured from the inspiral waveform. This is reasonable to expect by the era of third generation gravitational wave observatories, as future numerical modelling of binary neutron star mergers is anticipated to become sufficiently accurate by then. We compare these two methods and find that coherent stacking is more efficient at enhancing the SNR of BNS post-merger signals than power stacking (see Sec. III C and Fig. 4). This is partially because the coherent stacking method we propose requires extra phase information.

We also carry out a Bayesian model selection analysis to see how well one can distinguish between two different EOS models. For example, the TM1 EOS can be well distinguished from the DD2 EOS [43], with the average log-Bayes factor in the range 20–100 using the single loudest event (and $130 - 300$ for power-stacked signals). We further perform a parameter estimation study to derive how accurately one can measure the peak frequency of post-merger oscillations. We convert such a statistical error on the peak frequency to a statistical error on the NS radius of a $1.6 M_\odot$ NS using a universal relation between these quantities [12]. We find that with the power-stacked signal and using CE, the statistical error on the NS radius ranges from 15 m to 56 m, depending on the underlying EOS, which constitutes a fractional relative error of $\sim 4\%$. Such a measurement would thus compete with NICER measurement of the mass-radius relation of isolated NSs [49]. However, at this time systematic error in the universal relations between post-merger oscillation frequency and binary total mass, as well as in the template construction, dominate over the statistical error; this may be reduced in the future through, e.g., better modeling of NSs and more accurate BNS merger simulations.

A recent study by Bose et al. [50] also proposes to perform stacking of multiple BNS post-merger events with a focus on parameter estimation. Apart from considering only second-generation gravitational-wave detectors, and using results from simulations that do not employ finite temperature EOSs, our results are distinct from this work in at least two additional, significant aspects.

———

[1] Calling this "power stacking" is a slight abuse of historic notation, as this term has mostly been used to refer to analysis strategies that add excess power in select tiles in a time-frequency decomposition of multiple signals; see e.g. [47, 48]. These methods also give a composite SNR that scales as $N^{1/4}$ for N identical events each with low individual SNR, as multiplication of Bayes factors does [41], which is why we have borrowed this nomenclature.

First, we focus on the dominant 22 mode modeled as a single damped sinusoid, while in [50] a several parameter fit to the entire post-merger signal is used, with additional assumptions on the phases of the modes used in the fit. This alters the parameter estimation because the target templates are different. Second, here we perform a more detailed investigation and in-depth study to assess the detectability of GWs from BNS merger remnants by applying a Generalized-Likelihood-Ratio-Test (for hypothesis testing). In particular, instead of focusing on parameter estimation alone, we also discuss in detail the performance of stacking methods in making a detection of the dominant 22 mode, as well as a comparison to the power stacking method. This discussion is necessary because strictly speaking the interpretation of results of parameter estimation using modes from a set of events is only valid after a statistically significant confirmation of the existence of the modes have been made.

B. Organization

This paper is organized as follows. In Sec. II we develop the hypothesis test (GLRT) formalism for signals with unknown phase and perform a MC study to probe the detectability of post-merger oscillations from individual BNS remnants, assuming several different EOSs[2] and focusing on third-generation GW detectors. In Sec. III we apply the hypothesis test (GLRT) formalism for signals with unknown phase to stacked signals. We use the individual signals from the MC study of Sec. II to demonstrate that these stacking methods significantly amplify the SNR of BNS post-merger GWs and their detectability. Moreover, we show that coherent stacking works more efficiently than power stacking. In Sec. IV we discuss the possibility of distinguishing different EOSs using the stacked signal by carrying out a Bayesian model selection study. We also perform a parameter estimation study to derive the measurement accuracy of the post-merger peak frequency, and in turn, that of the NS radius. We conclude in Sec. V and discuss possible directions for future work.

II. SINGLE EVENT DETECTION

In this section we present the GLRT formalism we develop for single events, and perform a MC study to assess the detectability of BNS post-merger oscillations from individual events using third-generation ground based GW interferometers.

A. Hypothesis testing with unknown phase

Let us begin by describing how we generalize the Bayesian hypothesis test formalism of [41, 51, 52] such that it is applicable to coherent stacking of signals without prior phase information. In this section, we extend the formalism to the case of individual signals with unknown phase offset, which is suitable for finding oscillations of BNS merger remnants. In Sec. III B, we describe the procedure of coherently stacking a set of events, which involves frequency rescaling and phase alignment.

As we mentioned in the introduction, in the first $10-20$ ms following a typical BNS merger the 22 mode is the dominant one. Thus, instead of trying to model full signatures of post-merger waveforms, we focus on the dominant peak of the 22 mode component (see also the Principal Component Template in [37]). Here, we model the 22 mode oscillation as a damped sinusoid

$$h(t) = A'A_r \sin(2\pi f_{\text{peak}}t - \phi^0)e^{-\pi f_{\text{peak}}t/Q}\Theta(t), \quad (1)$$

where A' is the amplitude, f_{peak} is the 22 mode peak oscillation frequency, while we label the time coordinate in a way that the waveform starts at $t = 0$ (hence the Heaviside step function $\Theta(t)$) and ϕ^0 is a constant phase offset. The factor A_r denotes the reduction of the wave amplitude arising from source inclination and the response of the detector. Finally, Q is the quality factor of the mode.

We now explain the GLRT formalism and its extension. The one used in [41, 51, 52] assumes that all the parameters in the waveform are known *a priori* except for the amplitude. In our context, we assume that f_{peak}, Q and A_r are known from the inspiral information together with a given underlying EOS. One can then repeat the analysis with a different choice of EOS and carry out a Bayesian model selection study to see which one is preferred (see Sec. IV A). On the other hand, the phase ϕ^0 is unknown for BNS post-merger GWs, which requires one to extend the GLRT formalism. We begin by rewriting Eq. (1) as

$$h(t) = \left[A'_s \sin(2\pi f_{\text{peak}}t) + A'_c \cos(2\pi f_{\text{peak}}t)\right] e^{-\frac{\pi f_{\text{peak}}t}{Q}}\Theta(t)$$
$$= A_s h_s(t) + A_c h_c(t), \quad (2)$$

with $A'A_r = \sqrt{A'^2_c + A'^2_s}$ and $\tan\phi^0 = -A'_c/A'_s$. Here, h_c (h_s) is proportional to the above cosine (sine) function with an arbitrary normalization constant. Therefore, testing for a signal with unknown phase can be phrased as a test between the following two hypotheses:

$$\mathcal{H}_1 : \tilde{y}(f) = A_c\tilde{h}_c(f) + A_s\tilde{h}_s(f) + \tilde{n}(f), \quad (3)$$
$$\mathcal{H}_2 : \tilde{y}(f) = \tilde{n}(f), \quad (4)$$

with $A_c^2 + A_s^2 > 0$ in $\mathcal{H}_1$. Here $\tilde{h}_c$ and $\tilde{h}_s$ are two frequency-domain bases of the waveform which are nearly orthogonal to each other (this is generally true if $Q \gg 1$, i.e. there are enough cycles in the relevant waveform), so

that $\langle h_c | h_s \rangle \approx 0$, where the inner product is defined as

$$\langle \chi | \xi \rangle \equiv 2 \int_0^\infty \frac{\tilde{\chi}^*(f)\tilde{\xi}(f) + \tilde{\chi}(f)\tilde{\xi}^*(f)}{S_n} df \,, \quad ||\xi||^2 \equiv \langle \xi | \xi \rangle \tag{5}$$

with respect to the one-sided spectral density of detector noise $\langle \tilde{n}(f)\tilde{n}^*(f') \rangle = [S_n(f)/2]\,\delta(f - f')$.

The posterior probability of a hypothesis $\mathcal{H}$ being correct given some data y is given by Bayes' theorem [53, 54]

$$P(\mathcal{H}|y) = \frac{P(\mathcal{H})\,P(y|\mathcal{H})}{P(y)} \,, \tag{6}$$

where $P(\mathcal{H})$ is the prior belief in $\mathcal{H}$, while $P(y)$ is the probability of the data, which serves as an irrelevant normalization constant. The evidence $P(y|\mathcal{H})$ is given by

$$P(y|\mathcal{H}) \equiv \int d\vartheta\, P(\vartheta|\mathcal{H})\,P(y|\vartheta\mathcal{H}) \,, \tag{7}$$

where $P(\vartheta|\mathcal{H})$ is the prior on the model parameters ϑ, while $P(y|\vartheta\mathcal{H})$ is the likelihood function.

The likelihood of Hypothesis 1 is given by

$$P(y|\vartheta^i\mathcal{H}_1) \propto \prod_{f>0} \exp\left(-\frac{2|\tilde{y} - A_s\tilde{h}_s - A_c\tilde{h}_c|^2}{S_n} \right)$$
$$\propto \exp\left(-\frac{||y - A_s h_s - A_c h_c||^2}{2} \right) \,, \tag{8}$$

where $\vartheta^i = \{A_c, A_s\}$. For a uniform prior on A_c and A_s, the marginalization over ϑ in Eq. (7) corresponds to maximizing the above likelihood over A_c and A_s. The maximum likelihood estimator, using the shorthand notation $c = h_c$ and $s = h_s$, is then given by

$$\hat{A}_c = \frac{\langle c|y \rangle}{\langle c|c \rangle} \,, \quad \hat{A}_s = \frac{\langle s|y \rangle}{\langle s|s \rangle} \,. \tag{9}$$

Thus, according to Eqs. (7), (8) and the discussion above Eq. (9):

$$P(y|\mathcal{H}_1) \propto \exp\left(-\frac{||y - \hat{A}_s s - \hat{A}_c c||^2}{2} \right) \,, \tag{10}$$

and consequently

$$P(\mathcal{H}_1|y) \propto P(\mathcal{H}_1)\,P(y|\mathcal{H}_1)$$
$$\propto P(\mathcal{H}_1) \exp\left(-\frac{||y - \hat{A}_s s - \hat{A}_c c||^2}{2} \right) \,. \tag{11}$$

Repeating these steps for Hypothesis 2 ($A_s = A_c = 0$) gives

$$P(y|\mathcal{H}_2) \propto \exp\left(-\frac{||y||^2}{2} \right) \,, \tag{12}$$

$$P(\mathcal{H}_2|y) \propto P(\mathcal{H}_2) \exp\left(-\frac{||y||^2}{2} \right) \,. \tag{13}$$

The betting odds of $\mathcal{H}_1$ over $\mathcal{H}_2$, known as the odds ratio, is given by

$$O_{12} \equiv \frac{P(\mathcal{H}_1|y)}{P(\mathcal{H}_2|y)} = \frac{P(\mathcal{H}_1)}{P(\mathcal{H}_2)} B_{12} \,, \tag{14}$$

where

$$B_{12} \equiv \frac{P(y|\mathcal{H}_1)}{P(y|\mathcal{H}_2)} \tag{15}$$

is the Bayes factor. We focus on using B_{12} throughout this paper, though it agrees with O_{12} in the case of equal priors $P(\mathcal{H}_1) = P(\mathcal{H}_2)$ or in the case of uninformative priors $P(\mathcal{H}_1) = 0.5$ and $P(\mathcal{H}_2) = 0.5$.

We next compute the log of the Bayes factor, which using Eqs. (9), (10) and (12) is

$$\hat{T}_{\text{single}} \equiv \log\left[\frac{P(y|\mathcal{H}_1)}{P(y|\mathcal{H}_2)} \right]_{A_{c,s} \to \hat{A}_{c,s}} = \frac{\langle c|y \rangle^2}{2\langle c|c \rangle} + \frac{\langle s|y \rangle^2}{2\langle s|s \rangle}$$
$$= \frac{\langle c|c \rangle}{2}(A_c^2 + A_s^2) \tag{16}$$
$$+ \frac{\langle c|n \rangle^2 + \langle s|n \rangle^2}{2\langle c|c \rangle} + \frac{A_s\langle s|n \rangle + A_c\langle c|n \rangle}{\langle c|c \rangle}$$
$$= s_T + n_T \,, \tag{17}$$

where s_T is defined as the first term in equation (16) i.e., $s_T = \langle c|c \rangle(A_c^2 + A_s^2)/2\,(= \rho^2/2)$, and the remaining terms are defined as n_T. Note that in going from the first line to the second and third lines in the above we replaced the data y on the right-hand-side of the equality in the first line by Eq. (3). Here, we have chosen the normalization of h_c and h_s such that $\langle c|c \rangle = \langle s|s \rangle$. If we are to take the log of the odds ratio (defined in Eq. (14)) instead of the log of the Bayes factor and $P(\mathcal{H}_1) \neq P(\mathcal{H}_2)$, s_T needs to be shifted by $\log[P(\mathcal{H}_1)/P(\mathcal{H}_2)]$.

The evidence to favor (or disfavor) $\mathcal{H}_1$ over $\mathcal{H}_2$ depends on the signal part s_T, and the distribution of the noise part n_T. The GLRT ratio variable $\hat{T}_{\text{single}}$ can be intuitively thought of as an approximate spectral power of y near the central frequency f_{peak}. The distribution of n_T is in general non-Gaussian, but when $A_{c,s} = 0$, it becomes χ_2^2 (chi-squared with 2 degrees of freedom). Here and throughout we assume that $\langle c|n \rangle$ and $\langle s|n \rangle$ are normally distributed. If we denote the right-tail probability function of n_T (whose probability distribution is P_{n_T}) as

$$R(x) = \int_x^\infty P_{n_T}(z)dz \,, \tag{18}$$

(with $R_{A_{c,s}=0}$ corresponding to that of the χ_2^2 distribution) and the false-alarm probability is P_f, the criteria for rejecting hypothesis $\mathcal{H}_2$ with $\hat{T}_{\text{single}}$ computed from observation data is

$$R_{A_{c,s}=0}(\hat{T}_{\text{single}}) \leq P_f \,, \tag{19}$$

or

$$\hat{T}_{\text{single}} \geq R_{A_{c,s}=0}^{-1}(P_f) \,. \tag{20}$$

Now, notice that under $\mathcal{H}_1$, $\hat{T}_{\text{single}} = s_T + n_T$ is a random variable depending on the underlying signal and detector noise. Based on its distribution, one can infer the probability that the above inequality is satisfied, giving the target detection rate (probability) P_d:

$$P_d \geq R(R_{A_{c,s}=0}^{-1}(P_f) - s_T)\,. \qquad (21)$$

The amplitude of signal $A_{c,s}$ required to satisfy the above bound is then

$$\frac{1}{2}\langle c|c\rangle(A_c^2 + A_s^2) \geq R_{A_{c,s}=0}^{-1}(P_f) - R^{-1}(P_d)\,, \qquad (22)$$

or equivalently, the SNR required to satisfy the bound is given by

$$\rho \geq \rho_{\text{thres}} \equiv \sqrt{2[R_{A_{c,s}=0}^{-1}(P_f) - R^{-1}(P_d)]}\,, \qquad (23)$$

where we used the relation $s_T = \rho^2/2$.

The probability distribution function $P_{n_T}(z)$ inside the integral of the right-tail probability function R in Eq. (18) is obtained as follows. For simplicity, we choose the normalization such that $\langle c|c\rangle = 1 = \langle s|s\rangle$, and denote $X \equiv \langle c|n\rangle^2/2 + A_c\langle c|n\rangle$, $Y \equiv \langle s|n\rangle^2/2 + A_s\langle s|n\rangle$. Given that X and Y are independent random variables, the probability distribution of the random variable $Z = X + Y$ is given by

$$P_Z(z) = \int_{-\infty}^{\infty} P_X(x')P_Y(z - x')dx'\,. \qquad (24)$$

However, notice that by definition $X = (\langle c|n\rangle + A_c)^2/2 - A_c^2/2 \geq -A_c^2/2$, and similarly $Y \geq -A_s^2/2$. Therefore, Eq. (24) becomes

$$P_{n_T}(z) = \int_{-A_c^2/2}^{z+A_s^2/2} P_X(x')P_Y(z - x')dx'\,. \qquad (25)$$

Here

$$P_X(x') = \sqrt{\frac{2}{\pi}} \frac{e^{-A_c^2 - x'}\cosh[A_c\sqrt{A_c^2 + 2x'}]}{\sqrt{A_c^2 + 2x'}}\,, \qquad (26)$$

which is obtained from a non-central χ^2 distribution with an appropriate change of variable. In fact, P_{n_T} can also be obtained from a non-central χ_2^2 distribution with an appropriate change of variable. As expected, P_X reduces to a Gaussian distribution in the large A_c limit, and reduces to the χ^2 distribution with one degree of freedom for $A_c \to 0$. The distribution P_Y follows similarly, with $c \to s$. For completeness, we show the variance of n_T in Appendix B. We also show the signal-to-noise level of $\hat{T}$ that one can use instead of ρ to discuss the detection criterion.

There are two important facts regarding R. First, the distribution R depends on $A_{c,s}$ only through $A_c^2 + A_s^2$.

This can be seen by writing n_T as

$$n_T = A_s\langle s|n\rangle + A_c\langle c|n\rangle + \frac{1}{2(A_c^2 + A_s^2)}(A_c\langle c|n\rangle + A_s\langle s|n\rangle)^2$$
$$+ \frac{1}{2(A_s^2 + A_s^2)}(A_s\langle c|n\rangle - A_c\langle s|n\rangle)^2\,, \qquad (27)$$

and noting that $A_c\langle c|n\rangle + A_s\langle s|n\rangle$ and $A_s\langle c|n\rangle - A_c\langle s|n\rangle$ are independent Gaussian random variables with the same variance, $A_c^2 + A_s^2$. Second, if $A_{c,s} = A_{c,s}^m$ is the marginal solution that satisfies the equality in Eq. (22), and if we further scale the detector noise such that $n \to Cn$ without changing the definition of the inner product $\langle|\rangle$ so that c and s do not have to be renormalized, it is straightforward to see that $A_{c,s} = CA_{c,s}^m$ still satisfy the equality in Eq. (22) with the rescaled noise. Such a property is important as it means there is a one-to-one mapping between the threshold event SNR (schematically $\sim \sqrt{A_c^2 + A_s^2}/n$) and P_f and P_d. Such a property also carries over to the stacked signal we consider in Sec. III B. Following the convention in [37], we shall set the threshold SNR to $\rho_{\text{thres}} = 5$, which is consistent with setting $P_f = 0.01$ and $P_d = 0.982$.

B. MC study

The detectability of post-merger oscillations from BNS remnants is discussed in [37], assuming optimal sky orientation and source inclination. The results indicate that post-merger oscillations from individual sources are detectable only by third-generation GW detectors. Here we extend the analysis, but with two important modifications that make the analysis more realistic (although unfortunately greatly reducing detectability):

1. Instead of assuming the optimal sky location and source inclination that maximize the SNR, we randomly sample sources in sky direction, orbit inclination angle and polarization angle. According to [55, 56], the sky-averaged amplitude for a given type of source receives a $2/5$ reduction factor compared to the optimized configuration (assuming an "L"-shaped GW detector). In addition, the opening angle between arm cavities in the design for ET is $60°$, leading to an overall $\sqrt{3}/2$ reduction in signal amplitude comparing to an "L"-shaped interferometer with the same arm length. In our MC simulations, we have a different reduction factor for each source based on its parameters, although on average it recovers the $2/5$ factor for "L"-shaped detectors. To obtain the reduction factor for each source, we assume the "L"-shape antenna pattern

function [55] for CE:

$$F_+ = \frac{1}{2}(1 + \cos^2\theta)\cos 2\phi \cos 2\psi$$
$$- \cos\theta \sin 2\phi \sin 2\psi \,, \tag{28}$$

$$F_\times = \frac{1}{2}(1 + \cos^2\theta)\cos 2\phi \sin 2\psi$$
$$+ \cos\theta \sin 2\phi \cos 2\psi \,, \tag{29}$$

and the single-detector antenna pattern function for ET [56]:

$$F_+ = -\frac{\sqrt{3}}{4}\left[(1 + \cos^2\theta)\sin 2\phi \cos 2\psi\right.$$
$$\left. +2\cos\theta \cos 2\phi \sin 2\psi\right] \,, \tag{30}$$

$$F_\times = \frac{\sqrt{3}}{4}\left[(1 + \cos^2\theta)\sin 2\phi \sin 2\psi\right.$$
$$\left. -2\cos\theta \cos 2\phi \cos 2\psi\right] \,. \tag{31}$$

Here θ and ϕ are the angular coordinates of the source in the detector frame and ψ is the polarization angle. The amplitude fraction in each polarization can be computed by

$$\mathcal{A}_+ = F_+\frac{1 + \cos^2\varsigma}{2}\,, \quad \mathcal{A}_\times = F_\times \cos\varsigma\,, \tag{32}$$

where ς is the inclination angle of the BNS orbit with respect to the line of sight. The overall amplitude reduction factor with respect to the optimal configuration that enters in Eq. (1) is then given by $A_r = \sqrt{\mathcal{A}_+^2 + \mathcal{A}_\times^2}$ for an "L"-shape detector and $A_r = (2/\sqrt{3})\sqrt{\mathcal{A}_+^2 + \mathcal{A}_\times^2}$ for a single detector following the ET design. If we allow three detectors placed in a triangle geometry as explained in [56], the corresponding factor is $A_r = 2/\sqrt{3}\sqrt{\sum_{i,j}\mathcal{A}_i^2(\theta, \phi + 2\pi j/3, \psi, \varsigma)}$ with the summation over $i = (+, \times)$ and $j = (-1, 0, 1)$. The total SNR receives a factor of $\sqrt{3}$ boost on average compared to the single detector case. In fact, if there are N_d identical detectors, the total SNR is a factor of $\sqrt{N_d}$ larger than the single detector SNR.

2. We adopt a more up-to-date estimate of the BNS merger rate from [1] based on the observed BNS merger. Such a rate ($R_{BNS} = 1.54$ Mpc^{-3} Myr^{-1}) is 1.5 times higher than the "realistic" rate of [57], which was adopted in [37]. Naturally, as a result, we predict more detections over a one-year observation period. Our conclusions can be easily modified if the true rate turns out to be different than this number. An argument about the relevant scaling goes as follows. Considering the case where for a given rate R_{BNS} only one event is above the detection threshold within a volume of space V after $T_{\rm obs} = 1$ yr of observations, we have

$R_{BNS} \times V \times T_{\rm obs} = 1$. But, $V \propto d^3 \propto 1/\rho^3$, with d the distance. Thus, the SNR should scale with the merger rate as $\rho \propto R_{BNS}^{1/3}T_{\rm obs}^{1/3}$. In reality different merger events are not identical, their source parameters and sky locations all affect their SNR, and for sufficiently high redshift V is not simply proportional to d^3. Nevertheless, the above simple expression can be used to approximately scale the SNR that we present in our study below for different merger rates or observation periods.

Our analysis is based on the waveform model of Eq. (1), which depends on the peak frequency $f_{\rm peak}$, the quality factor Q and the 22 mode amplitude A', the angle-dependent amplitude factor A_r and phase offset ϕ^0. We estimate the 22 mode frequency ($f_{\rm peak}$) using the fit of [12] (see also [11, 14, 58–61] for other fits)

$$\frac{f_{\rm peak}}{1{\rm kHz}} = \frac{m_1 + m_2}{M_\odot}\left[a_2\left(\frac{R_{1.6M_\odot}}{1{\rm km}}\right)^2 + a_1\frac{R_{1.6M_\odot}}{1{\rm km}} + a_0\right]\,, \tag{33}$$

where $a_0 = 5.503$, $a_1 = -0.5495$ and $a_2 = 0.0157$ are EOS-independent parameters; $R_{1.6M_\odot}$ is the radius of a non-rotating NS with gravitational mass $1.6M_\odot$, and this parameter therefore encodes the EOS dependence. We choose the masses by independently sampling the Gaussian distribution [62]

$$P(M_{\rm NS}; M_0, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(M_{\rm NS} - M_0)^2}{2\sigma^2}\right] \tag{34}$$

with $M_0 = 1.33M_\odot$ and $\sigma = 0.09M_\odot$.

The quality factor and 22 mode amplitude in Eq. (1) should also depend on the NS EOS, the mass ratio and mass of the binary, but the detailed dependence is currently unknown. In order to enable comparison to the results in [37], we set A' and Q such that the peak value of the characteristic strain and the SNR of Eq. (1) match the peak characteristic strain and SNR of the dominant 22 mode component in Fig. 11 of [37], which corresponds to the post-merger signal arising from a $1.35M_\odot + 1.35M_\odot$ BNS with optimal extrinsic parameters (sky location and inclination angle), at luminosity distance $d = 50$Mpc with the Hempel et al. EOS (TM1) [43]. The matching process yields $Q = 34$, $A' = 2.5 \times 10^{-22}$. For a binary obeying the TM1 EOS, but with different component masses and luminosity distance we still set A' based on a $1.35M_\odot + 1.35M_\odot$ BNS, i.e.,

$$A' = 2.5 \times 10^{-22} \times \frac{50{\rm Mpc}}{d}\,. \tag{35}$$

While choosing A' based on results from $1.35M_\odot + 1.35M_\odot$ BNSs is not ideal, it should provide a reasonable approximation if Eq. (34) is valid for merging BNSs, because it is narrowly peaked around $1.33M_\odot$ and hence the majority of BNSs are near equal mass binaries with total mass $\sim 2.7M_\odot$. Nevertheless, such a prescription
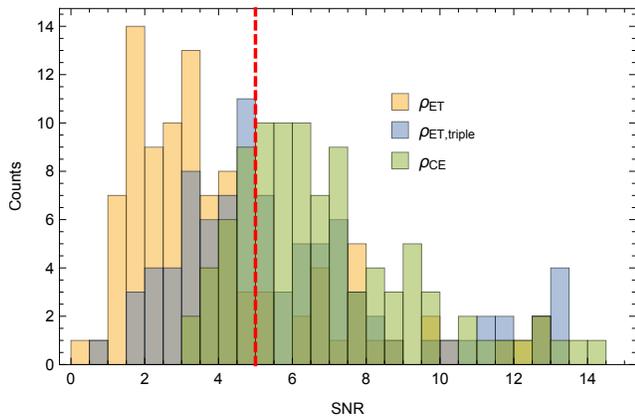
FIG. 1. Histogram for single event SNR for 100 realizations in the MC simulation. Orange bins represent the SNR with respect to the sensitivity of the ET (single detector). Blue bins are associated with the triple-detector, triangle design of the ET [56]. Green bins represent the SNR with respect to the sensitivity of the CE (wide-band configuration). The detection threshold ($\rho = 5$) is indicated by the red, dashed line. The TM1 EOS and one year observation is assumed, and the binary merger rate is taken to be $R_{BNS} = 1.54 \text{Mpc}^{-3}\text{Myr}^{-1}$.
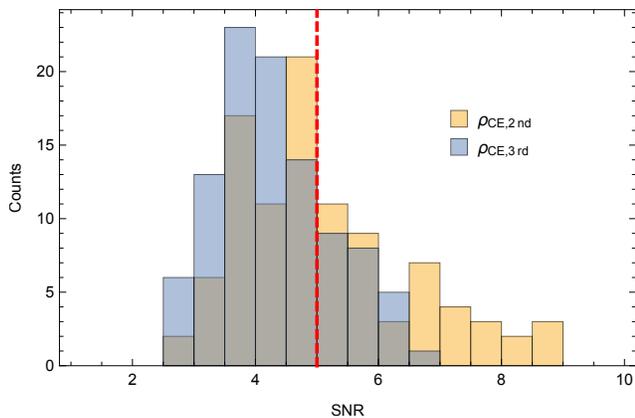
FIG. 2. Histograms of the second and third loudest events with the CE sensitivity, from the same MC runs as in Fig. 1. Orange bins represent the SNR of the *second loudest* event, while blue bins represent the SNR of the *third loudest* event.

needs to be revised once we gain more systematic (and accurate) understanding of the functional dependence of A' and Q on the binary intrinsic parameters from future numerical simulations of BNS mergers, complemented by actual observations. Since in this section we are only interested in SNRs of individual events, we will set the phase offset to zero.

We ran 100 MC realizations each covering one-year of observations to calculate SNRs. In each realization, we reject binaries with total mass exceeding the prompt-collapse threshold mass M_{thres} based on the results of [42]. We note that, strictly speaking, a BNS with total mass just below M_{thres} in general cannot survive for more than a few ms following merger, and hence it

cannot exhibit any loud post-merger oscillations. The simulations of [8] suggest that the threshold mass for rejecting binaries from our MC realizations should be $\sim 0.95 \times M_{\text{thres}}$. However, this effect has a small contribution to our results because the total mass distribution for BNSs derived from Eq. (34) is also Gaussian and is given by Eq. (34) with $M_0 = 2.66 M_\odot$ and $\sigma = 0.1273 M_\odot$. As a result, the fraction of binaries with mass above $0.95 \times M_{\text{thres}}$ is only 1% (for the TM1 EOS[3]).

For the single-event study we consider both the ET and CE third-generation ground-based GW observatories. The ET sensitivity is obtained from [63] and the CE sensitivity from [64]. For the CE, we choose the wide-band configuration, because it has better sensitivity than the "standard configuration" above 1kHz. For the ET configuration, we consider both a single interferometer and a triangular arrangement with three interferometers.

In each MC realization there are about 40-70 events with $\rho > 1$. In Fig. 1 we present the SNR of the loudest event in each of the 100 MC realizations with the TM1 EOS. Our MC simulations show that for the ET (with single interferometer) and CE sensitivities, there is a 25% and 79% chance respectively to have a single loud event passing the detection threshold after acquiring data for a full year (for the triple-detector ET case it is a 56% chance). In Fig. 2 we present the SNR of the second and third loudest event in each of 100 MC realizations for CE. The plot shows that after a year of observations there is about a 43% chance to have a second and a 23% chance to have a third event above detection threshold. The even lower chance of detecting secondary and tertiary events above the detection threshold of 5 implies that there is little room for stacking if one insists on using only signals above this threshold in order to obtain a much stronger signal that will reduce the statistical uncertainties in parameter estimation.

Given that the ET and CE are the most sensitive ground-based GW detectors proposed so far, our results indicate that (for the currently envisioned configurations) unless the true BNS merger rate turns out to be substantially higher than the rate we adopt in our study, or nature supplies a more "favorable" EOS as discussed in the next paragraphs, over the next few decades the prospect to directly probe the dominant peak of BNS merger remnant oscillations from *individual* events does not appear very promising. Of course, this observation is a consequence of adopting $\rho_{\text{thres}} = 5$. Since post-merger oscillations will be an example of a *triggered* search, it is conceivable that a lower threshold could be targeted.

All of the above results were obtained with the TM1 EOS, but we also studied several other popular and realistic, finite temperature nuclear EOSs. In particular,

---

[3] For other EOSs we introduce later, this fraction is practically the same for the LS220 EOS, even lower for the DD2 and Shen EOSs, and a bit larger for the SFHo EOS (in the latter case, however, we will see that the post-merger GWs are difficult to detect in the first place).

we considered the Steiner et al. EOS SFHo [65], the Lattimer Swesty EOS [66] with compressibility parameter $K = 220$ MeV (LS220), the Hempel et al. EOS DD2 [43], and the Shen et al. EOS [67]. These EOSs were chosen because they all have a maximum mass above $2.0 M_\odot$ [68, 69], they cover a range of stiffness, and because they take into consideration finite temperature effects self-consistently. The parameters for performing the MC simulations with these different EOSs are listed in Table I. The strain amplitude A' and the quality factor Q are chosen such that the peak value of the characteristic strain and SNR of Eq. (1) match the peak value of the characteristic strain and SNR of the post-merger dominant 22 component reported in the BNS merger simulations of [13, 70, 71]. In Appendix A we show how well the Lorentzian profile of Eq. (1) approximates the post-merger spectra in the vicinity of the dominant post-merger peak found in numerical relativity simulations.

Assuming $R_{BNS} = 1.54 \text{Mpc}^{-3}\text{Myr}^{-1}$, the results of the MC realizations with different EOSs are presented in Fig. 3, which shows that, among the EOSs that we study, both SFHo and DD2 EOS have small detection rates ($\sim$ 13% and 30% respectively, with 76% for LS220 and 100% for Shen) for post-merger oscillations after a full year of observations with CE. However, it should be stressed that these results should be considered only as approximate, with the detailed numbers subject to change with more accurate modeling of NS mergers in the coming years.

Given the richness and importance of the physics encoded in the post-merger signal, there is strong motivation to improve its detectability by exploiting the information we can anticipate from the current/planned generation of detectors, and informing designs for future GW detectors to maximize their sensitivity to this phase of BNS mergers. In this work we are focusing on the former approach, and in next section show that stacking signals from multiple detections can significantly enhance the sharpness of the post-merger signal. We describe the details of the power and coherent mode stacking methods we propose in the next section.

TABLE I. Parameters for different EOS

EOS	$R_{1.6M_\odot}$	f_{peak}(kHz)	$\frac{M_\odot}{m_1 + m_2}$	$\frac{A'(50\text{Mpc})}{10^{-22}}$	Q	$\frac{M_{\text{thres}}}{M_\odot}$
SFHo	11.77	1.21		2.7	25.7	2.95
LS220	12.5	1.09		4.3	25.7	3.05
DD2	13.26	0.98		2.8	12.7	3.35
Shen	14.42	0.84		5.0	23.3	3.45
TM1	14.36	0.85		2.5	34.2	3.1

III. MULTIPLE EVENT DETECTION

In this section we present the GLRT formalism for stacking multiple events, and we assess the detectability of BNS post-merger oscillations through such a stacking
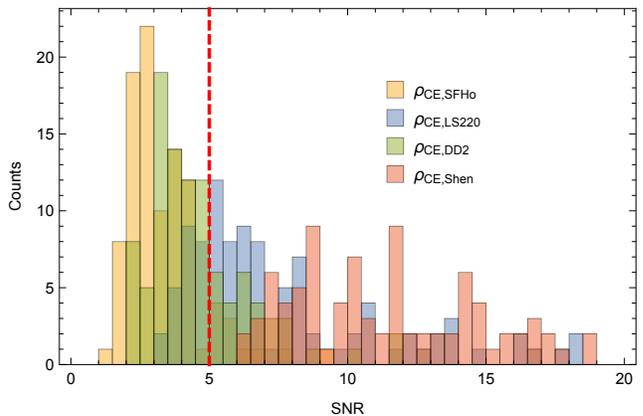

FIG. 3. The same setting as in Fig. 1 but with different EOSs and with respect to the CE sensitivity alone. Orange bins represent the SNR for the SFHo EOS, blue bins for the LS220 EOS, green bins for the DD2 EOS and red bins for the Shen EOS.

analysis using third-generation ground based GW interferometers.

The GLRT formalism described in Section II A justifies the rationale of claiming detection from a single event with unknown phase, within the Bayesian framework. When it is applied to single-event detections, it should give consistent results with previous studies [37]. However, as discussed in Sec. II B, these single events are not likely to allow a direct detection of post-merger oscillations even when using the most sensitive ground-based GW detectors proposed so far, unless event rates are higher than expected or the EOS is Shen-like. In the few cases where we manage to beat the odds and have a loud event, the chance of also having a second sufficiently loud event is even slimmer (Fig. 2). Nevertheless, regardless of whether there is a loud event passing the detection threshold, our MC studies indicate that there will likely be several tens of events with modest SNR (e.g. $1 \leq \rho \leq 5$), which we exploit in this work to increase the chances of detection and improve the accuracy of parameter estimation. Notice that although the post-merger SNRs of these events are low, their inspiral SNRs will be significantly higher and easily detectable. This point will be further discussed in Sec. III B.

A. Hypothesis testing with power stacking

In this section, we apply the Bayesian model selection approach discussed in [40] for multiple events, which we refer to as power stacking. With a group of events y_i ($i = 1, .., N$), the combined Bayes factor is [40]

$$B_{12} = \prod_{i=1}^{N} \frac{P(y_i|\mathcal{H}_1)}{P(y_i|\mathcal{H}_2)}, \quad (36)$$

where the combined $\hat{T}$ variable is

$$
\begin{aligned}
2\hat{T}_{\text{power}} \equiv &\, 2\log\prod_{i=1}^{N}\frac{P(y_i|\mathcal{H}_1)}{P(y_i|\mathcal{H}_2)} \\
= &\sum_{i=1}^{N}\left\{\frac{(\langle c_i|y_i\rangle)^2}{\langle c_i|c_i\rangle}+\frac{(\langle s_i|y_i\rangle)^2}{\langle s_i|s_i\rangle}\right\} \\
= &\sum_{i=1}^{N}\langle c_i|c_i\rangle(A_{c,i}^2+A_{s,i}^2)+\sum_{i=1}^{N}\frac{\langle c_i|n_i\rangle^2+\langle s_i|n_i\rangle^2}{\langle c_i|c_i\rangle} \\
&+2\sum_{i=1}^{N}\frac{A_{s,i}\langle s_i|n_i\rangle+A_{c,i}\langle c_i|n_i\rangle}{\langle c_i|c_i\rangle} \\
= &\, 2\mathbf{s}_{Tp}+2\mathbf{n}_{Tp}\,.
\end{aligned}
\tag{37}
$$

Note that in writing down such a combination of Bayes factors, we have implicitly assumed that all events follow the same hypothesis. This assertion relies on the assumption that post-merger oscillations should exist if the mass of the remnants is below the (EOS-dependent) threshold. This GLRT variable does not have the same type of noise distribution as that analyzed in Sec. II A. Instead, its distribution is obtained from a non-central χ_{2N}^2 distribution with an appropriate change of variable. For sufficiently large N, due to the central limit theorem, the distribution of $\mathbf{n}_{Tp}$ is a Gaussian with mean N and variance (all c_i, s_i are normalized such that $\langle c_i|c_i\rangle = \langle s_i|s_i\rangle = 1$)

$$
\text{Var}[\mathbf{n}_{Tp}] = N + \sum_{i=1}^{N}(A_{c,i}^2+A_{s,i}^2) \equiv \sigma_{\mathbf{n}_{Tp}}^2\,.
\tag{38}
$$

The distribution of the noise $(2\mathbf{n}_{Tp})$ associated with the null hypothesis $A_{c,s}=0$ is just a χ_{2N}^2 distribution, which also asymptotes to a Gaussian distribution in the large N limit with mean $2N$ and variance $4N$. Let us denote

$$
Q_\sigma(x) \equiv \frac{1}{\sqrt{2\pi}\sigma}\int_x^\infty dy\, e^{-y^2/(2\sigma^2)}\,,
\tag{39}
$$

and

$$
U_{2N}(x) \equiv \int_x^\infty dy\, P_{\chi_{2N}^2}(y)\,,
\tag{40}
$$

so that the requirement to reject the null hypothesis with significance level P_f, and the change of success (detection rate) is P_d, is

$$
\begin{aligned}
\sum_{i=1}^{N}(A_{c,i}^2+A_{s,i}^2) \geq &\, U_{2N}^{-1}(P_f) - R_{2N}^{-1}(P_d) \\
&\approx U_{2N}^{-1}(P_f) - 2N - 2Q_{\sigma_{\mathbf{n}_{Tp}}}^{-1}(P_d)\,,
\end{aligned}
\tag{41}
$$

where R_{2N} is the right-tail probability function for the random variable $2\mathbf{n}_{Tp}$. In practice, P_d being around 0.99 is already a decent detection rate.

B. Hypothesis testing with coherent stacking

The coherent stacking approach developed in [41] for black hole ringdowns relies on extra information to align the phase between modes in different events. For BNS mergers such accuracy in theoretical modelling (even given the EOS) is unavailable at present. However, with improvements in numerical simulations expected in the future, it is possible that the inspiral part of BNS waveforms could be used to predict the phase of post-merger modes (again, given the EOS). The investigation in this section relies on the above assumption.

We develop a method of coherent stacking that relies on the existence of a dominant post-merger peak frequency, universally related to the component masses and the radius via Eq. (33). The component masses and the time of merger can be determined from the inspiral part of the waveform to high accuracy. For example, a Fisher analysis suggests that the mass of each individual component in a $1.35M_\odot + 1.35M_\odot$ NS binary located at 300Mpc away from Earth can be measured to a precision better than $\sim 0.1\%$, assuming optimal sky location and orientation of the source using CE. The relation in Eq. (33) itself is not exact, but we expect the theoretical understanding leading to it to improve over time with more accurate numerical simulations and better constraints on the EOS. Of course, one should note that even if Eq. (33) were exact for the set of candidate EOSs studied, there could still be a systematic error if none of these EOSs are close enough to the true finite temperature NS EOS. If this is the case, f_{peak} will be erroneously predicted, degrading the efficiency of the coherent stacking process. We refer to Sec. IV A for more discussions on comparing different EOSs.

In our stacking approach we pick events with modest SNR ($\rho \geq 1$) and assume that the phase ϕ^0 can be determined by the inspiral waveform within an uncertainty

$$
\sigma_{\phi^0} \approx \frac{C}{\rho_{\text{inspiral}}}\,,
\tag{42}
$$

where C is a constant to be determined by future simulations. For the Monte-Carlo investigation in Sec. III C we choose $C = 2\pi$.

Different events will in general have different remnant masses and hence different 22 mode frequencies; therefore, we need to rescale the data before stacking so that all 22 modes have the same frequency. Such a procedure has been described in detail in [41] to reprocess data before coherent stacking of BH ringdown modes, and in [37] for constructing a "universal" template bank.

With this at hand, let us now proceed with stacking. Suppose we have a set of *rescaled* data from N different events with $\rho > 1$

$$
\tilde{y}_i(f) = \tilde{g}_i(f) + \tilde{n}_i(f)\,,
\tag{43}
$$

where $i = 1, \ldots, N$ labels different events with $\tilde{g}_i \equiv A_{c,i}\tilde{h}_{c,i} + A_{s,i}\tilde{h}_{s,i}$. We further assume that $\phi_i = \phi_i^0 + \delta\phi_i$

is the estimator for the phase of each event, where ϕ_i^0 are the unknown, true underlying phases while $\delta\phi_i$ is the measurement uncertainty of ϕ_i. We then align the phases and coherently sum up the data with different weights via ($0 < w_i \leq 1$):

$$
\tilde{\mathbf{y}} = \sum_i w_i e^{i\phi_i} (\tilde{g}_i + \tilde{n}_i)
$$
$$
= \tilde{\mathbf{g}}_y + \tilde{\mathbf{n}}_y \,, \tag{44}
$$

where $\tilde{\mathbf{g}}_y$ ($\tilde{\mathbf{n}}_y$) is the signal (noise) part of the stacked data $\tilde{\mathbf{y}}$.

The stacked data can now be used to construct the log Bayes factor for the hypothesis that a signal is present versus one where no signal is present. Based on the discussion in Sec. II A, we then need to evaluate the quantity $\langle \mathbf{y} | \mathcal{I} | \mathbf{y} \rangle$, with $\mathcal{I} \equiv |c\rangle\langle c| + |s\rangle\langle s|$ and the brackets $\langle | \rangle$ here are defined with respect to the spectrum of n_y. The quantities c and s are again defined as in Sec. II A but with individually rescaled frequencies. It is also straightforward to verify that $\mathcal{I}|g_i\rangle = |g_i\rangle$, because we assume that the frequency uncertainty with a *known* EOS is negligible. Using this property, $\hat{T}_{\text{coherent}}$ is given by

$$
2\hat{T}_{\text{coherent}} \equiv \langle \mathbf{y} | \mathcal{I} | \mathbf{y} \rangle
$$
$$
= \sum_i w_i^2 \langle g_i | g_i \rangle + \sum_{i \neq j} w_i w_j \langle g_i e^{i\phi_i} | g_j e^{-i\phi_j} \rangle
$$
$$
+ \sum_{ij} w_i w_j \langle n_i e^{i\phi_i} | \mathcal{I} | n_j e^{-i\phi_j} \rangle
$$
$$
+ \sum_{ij} w_i w_j [\langle g_i e^{i\phi_i} | n_j e^{-i\phi_j} \rangle + \langle n_i e^{i\phi_i} | g_j e^{-i\phi_j} \rangle]
$$
$$
= 2\mathbf{s}_{Ty} + 2\mathbf{n}_{Ty} \,, \tag{45}
$$

where $2\mathbf{s}_{Ty}$ is used to designate the term appearing on the first line and $2\mathbf{n}_{Ty}$ all remaining terms. We refer to $\mathbf{s}_{Ty}$ ($\mathbf{n}_{Ty}$) as the signal (noise) part of $\hat{T}_{\text{coherent}}$.

The signal part of the stacked data can now be used to determine a detection criterion. We begin by evaluating $\mathbf{s}_{Ty}$ with an ensemble average over the phase uncertainties (using $\langle e^X \rangle = e^{-\langle X^2 \rangle / 2}$ for any Gaussian random variable X with zero mean)

$$
2\langle \mathbf{s}_{Ty} \rangle = \sum_i w_i^2 \langle g_i | g_i \rangle
$$
$$
+ \sum_{i \neq j} w_i w_j \langle g_i e^{i\phi_i^0} | g_j e^{-i\phi_j^0} \rangle e^{-\frac{\sigma_{\phi_i}^2}{2} - \frac{\sigma_{\phi_j}^2}{2}} \,, \tag{46}
$$

which corresponds to the stacked SNR squared, and where $\sigma_{\phi_i}^2 = \langle \delta\phi_i^2 \rangle$. Based on the discussion in Sec. II A, the SNR of y has to be larger than 5 to pass the detection threshold. Then, the detection criteria for the stacked signal is just

$$
\sqrt{2\langle \mathbf{s}_{Ty} \rangle} \geq 5 \,. \tag{47}
$$

The weight coefficients w_i are chosen such that $\langle \mathbf{s}_{Ty} \rangle$ is maximized, and in this work it is achieved using the

downhill simplex optimization method [72, 73]. Similar to the single event case, we present the variance of $\mathbf{n}_{Ty}$ in Appendix B, together with the signal-to-noise level of $\hat{T}_{\text{coherent}}$.

The performance of stacking is discussed in Sec. III C, but let us make an immediate observation. If there are N events under stacking and all of them have comparable SNR, this coherent stacking method would produce an $\mathcal{O}(N^{1/2})$ boost in $\mathbf{s}_{Ty}$[4]. In reality, there is always a small group of events with high SNR, while the remaining events have low SNR. Thus, in practice the improvement factor over the event with *best* SNR can never achieve $N^{1/2}$-type scaling. The same observation was made when coherently stacking ringdown modes from BH coalescences [41].

C. MC study

In this section, we show how stacking enhances the chance of detecting BNS post-merger signals by using the results of our MC simulations. We first compare the results for power stacking against single event detection. We next compare coherent stacking against power stacking and show that the former works more efficiently than the latter.

We note that it is difficult to define a SNR for a combined set of events, because the statistical distributions of $\hat{T}$ for the true and null hypotheses ($\mathcal{H}_{1,2}$) are different from those of a single event (see Eq. (16)). As a result, we define a new quantity α, which is the universal scale factor that the SNR of all events should be divided by (or detector noise should be multiplied by): $A_i \to A_i/\alpha$, in order to exactly satisfy the detection bound in Eq. (41) or Eq. (47). The larger this detection-threshold-matching factor α is, the more efficiently an analysis method performs. We shall apply this α to characterize the performance of stacking in this Section.

1. Power stacking versus single event detection

In each MC realization performed in Sec. II B, we pick the top N events to construct the Bayes factor in Eq. (37). In Fig. 4, for illustration purposes we choose N to be 5 and 30, although N can be any positive integer less than or equal to the total number of events detected in general.

In Fig. 4, α for any single event is equal to its SNR/5. Before applying power stacking with the TM1 EOS, there is roughly a 79% chance to detect a post-merger 22 mode with CE operating for a year. Power stacking leads to a decrease in the false alarm rate. For $N = 5$, this decrease is enough to allow the stacked signals to pass the

———

[4] Notice that S_n in the definition of the inner product scales linearly with N.
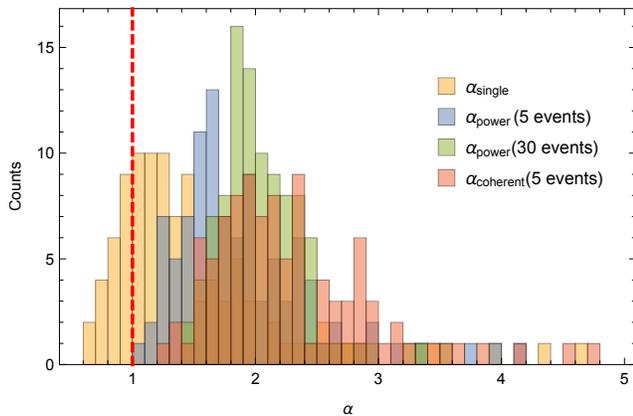
FIG. 4. Histogram for power-stacked, coherently stacked and single-event α (a proxy for SNR, with $\alpha = 1$ being the detection threshold), in each realization for the TM1 EOS. Orange bins represent the top event in each MC realization without stacking, with 79/100 passing the detection threshold. Blue bins represent power stacking using the loudest 5 events, and demonstrate that all realizations pass the detection threshold. Green bins represent power stacking using the loudest 30 events, showing an improvement compared to power stacking 5 events. Pink bins represent coherently stacking the top 5 events with $C = 2\pi$ in Eq. (42); all cases pass the detection threshold, and the skew of the distribution toward larger values of α indicates that coherently stacking the top 5 events is more efficient than power stacking the top 30 events.
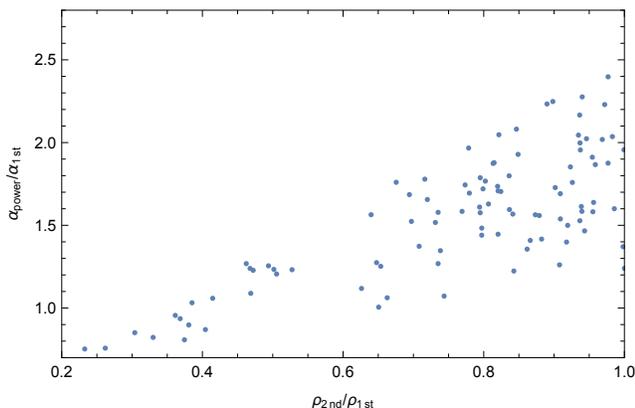
FIG. 5. Vertical axis: the improvement factor of effective SNR for the power stacked signal ($N = 30$) over the best single event in each MC realization with the TM1 EOS. Horizontal axis: the ratio of ρ between the second best event and the best event in each MC realization.

detection threshold in all MC realizations. For $N = 30$, all stacked events are able to pass the detection threshold, and the improvement factor in α is roughly a factor of 1.5. In principle, events with low SNR could contribute more noise fluctuations than signal improvement in $\hat{T}_{\text{power}}$, which means that adding more events does not necessarily lead to better statistics for detection. This is shown in Fig. 5 because some of the MC realizations have $\alpha_{1\text{st}} > \alpha_{\text{power}}$. Here $\alpha_{1\text{st}}$ stands for the α factor of

the top event in each MC realization.

Intuitively we can interpret $5 \times \alpha_{\text{power}}$ as the "effective SNR" of the power stacked signals, as $5 \times \alpha_{\text{single}}$ is the SNR of a single event. In this sense, $\alpha_{\text{power}}/\alpha_{1\text{st}}$ just characterizes the improvement in effective SNR by power stacking. As shown in Fig. 5, this effective SNR improvement ranges between 0.7 and 2.5, with median value at 1.57. When the SNR of the best event is much higher than the rest, so that $\rho_{2\text{nd}}/\rho_{1\text{st}}$ is small, the effective SNR improvement tends to be smaller. Therefore the power stacking approach works better for events with more uniformly distributed SNRs.

We conclude this subsection with a short discussion of how our results would change if we had chosen a different EOS. Assuming that the enhancement in SNR (a factor of ~ 1.5) due to power stacking relative to a single event does not depend strongly on the EOS, one can roughly estimate the distribution of ρ after stacking for various EOSs by shifting the histograms in Fig. 3 to larger SNR by a factor of ~ 1.5. Doing so, one finds that it is very likely that the stacked signal can be detected for all EOS we consider here except for the case of the SFHo EOS. This clearly shows that the detectability of the post-merger signal is sensitive to the underlying EOS.

2. Coherent stacking versus power stacking

We now compare the SNR improvement between coherent and power stacking. In Fig. 4 we show α for a coherently stacked signal using the top 5 events in each of the 100 MC realizations discussed earlier, and C in (42) is assumed to be 2π. Because the inspiral SNRs of these events are much greater than the post-merger SNR, the effect due to phase and frequency uncertainties of modes is negligible. The weight w_i is basically proportional to post-merger SNR_i, and the SNR of the coherently stacked signal to close to $\sqrt{\sum_i \text{SNR}_i^2}$. Figure 4 demonstrates that the α distribution for coherent stacking of *five events* is skewed toward larger values than power stacking of *thirty events*. Thus, coherent mode stacking outperforms power stacking in this setting.

Another way to compare the two stacking approaches is to consider a simple scenario in which all N events have identical SNR, and ask how many events are needed to satisfy the detection threshold for each stacking method. For power stacking, this gives the following equation for the threshold number N in terms of the individual event SNR ρ (for simplicity, we ignore the fact that N has to be integer)

$$\frac{\sqrt{N}}{2}\rho^2 = \frac{1}{2\sqrt{N}}[U_{2N}^{-1}(P_f) - 2N] - \sqrt{1+\rho^2}Q_1^{-1}(P_d)$$
$$\approx Q_1^{-1}(P_f) - \sqrt{1+\rho^2}Q_1^{-1}(P_d). \quad (48)$$

One immediate observation is that unlike the coherent stacking case discussed in Sec. III B, the $N - \rho$ relation

is not a single power-law. For example, if $P_d = 0.5$, the second term in Eq. (48) vanishes and we can see that the threshold SNR satisfies $\rho \propto N^{-1/4}$. On the other hand, if the second term dominates over the first term in Eq. (48) and $\rho \gg 1$, the threshold SNR satisfies $\rho \propto N^{-1/2}$.

To compare the performance between coherent mode stacking and power stacking, assuming all events have the same single SNR ρ, we compute the number of identical-SNR events N needed to satisfy the equality in Eq. (47) for coherent stacking, and the equality in Eq. (41) or Eq. (48) for power stacking. In the coherent stacking case, this can be computed exactly:

$$\sqrt{N}\rho = 5, \tag{49}$$

without considering phase uncertainty and

$$\rho^2[1 + (N-1)e^{-\sigma_{\phi^0}^2}] = 25, \tag{50}$$

where for illustration purpose we also include a case with phase uncertainty $\delta\phi^0 \approx 1/\rho$ [5]. In the power stacking case, one must carry out the calculation numerically for a given P_f and P_d, as shown in Fig. 6. In this idealized scenario, coherent stacking always outperforms power stacking as it requires fewer events to pass the detection threshold for the same $P_f = 0.01$ and $P_d = 0.982$. The Gaussian distribution approximation (the second line in Eq. (48)) corresponds to the blue dashed line, which underestimates the performance of power stacking (blue solid line) for small N, but agrees better with the exact expression in the first line of Eq. (48) for larger N, as expected. We also find that the phase uncertainty in the coherent stacking case becomes more important in the low-ρ regime – the red solid line departs more from the red dashed line – as expected from Eq. (50).

IV. MODEL SELECTION AND PARAMETER ESTIMATION

We now discuss how well one can distinguish two different EOS models, mainly focusing on power stacking. We also show how accurately one can measure the peak frequency (and in turn the NS radius) with the stacked events.

A. Model selection for EOSs

In order to compare the likelihood of different EOSs based on the measured data, and in particular based on the stacked signal, we perform a Bayesian model selection method to evaluate the relative performance between
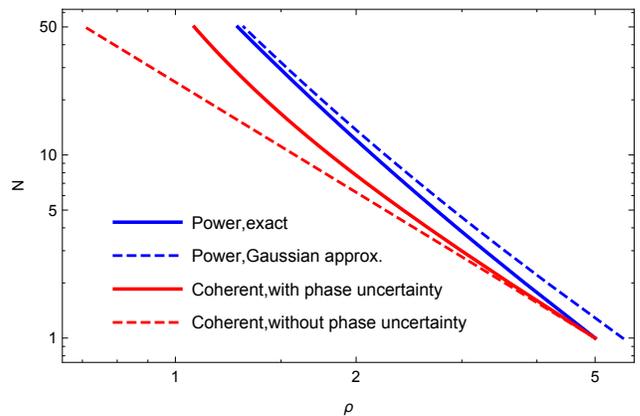

FIG. 6. Number of identical events needed to satisfy the detection threshold, as a function of single event SNR ρ, with the false alarm rate $P_f = 0.01$ and the detection probability $P_d = 0.982$. The blue solid (dashed) line represents the requirement for power stacking with (without) the Gaussian approximation. The red solid (dashed) line represents the requirement for coherent mode stacking with (without) phase alignment uncertainty considered. Note that coherent stacking is always more efficient than power stacking (fewer events are needed to cross above the detection threshold).

different models. In general, for a given data set y and two possible models $\mathcal{H}_1$, $\mathcal{H}_2$, one can evaluate the Bayes factor given in Eq. (15)[6].

Given that we can perform the analysis we present in Sec. III B for multiple EOSs, we can determine the EOS which gives the best SNR, which here we call model 1, and then perform a model selection test for other EOSs by using the data set y corresponding to model 1. Then within the GLRT framework, we evaluate the following Bayes factor:

$$\mathcal{B}_{1|2} \equiv \frac{P(y|\mathcal{H}_1)}{P(y|\mathcal{H}_2)}. \tag{51}$$

We shall denote the two basis functions in model 1 as $c^{(1)}$ and $s^{(1)}$ and the basis functions in model 2 as $c^{(2)}$ and $s^{(2)}$. According to Eq. (8), we then have

$$\hat{\mathcal{T}}_{1|2} = \log \mathcal{B}_{1|2}$$
$$= -\frac{||y - \hat{A}_{1c}c^{(1)} - \hat{A}_{1s}s^{(1)}||^2}{2}$$
$$+ \frac{||y - \hat{A}_{2c}c^{(2)} - \hat{A}_{2s}s^{(2)}||^2}{2}. \tag{52}$$

Inserting the expressions for the maximum likelihood es-

---

[5] For a general event we expect ρ to be proportional to $\rho_{\rm inspiral}$, so that $\delta\phi^0 \propto 1/\rho$ according to Eq. (42). We arbitrarily picked a coefficient 1 to illustrate the effect of phase uncertainty.

---

[6] For coherently stacking multiple events, we generically obtain different stacked signals for different EOSs, and thus different data sets that one has to compare to perform model selection. This introduces a subtlety that we discuss in Appendix C, but for simplicity we ignore it in the following analysis.

timators (c.f. Eq. (9)) in the above equation, we obtain

$$\hat{\mathcal{T}}_{1|2} = \frac{(\langle c^{(1)}|y\rangle)^2}{2\langle c^{(1)}|c^{(1)}\rangle} + \frac{(\langle s^{(1)}|y\rangle)^2}{2\langle s^{(1)}|s^{(1)}\rangle}$$
$$- \frac{(\langle c^{(2)}|y\rangle)^2}{2\langle c^{(2)}|c^{(2)}\rangle} - \frac{(\langle s^{(2)}|y\rangle)^2}{2\langle s^{(2)}|s^{(2)}\rangle}$$
$$= \Delta\mathbf{s}_{Ty} + \Delta\mathbf{n}_{Ty} , \qquad (53)$$

where $\Delta\mathbf{s}_{Ty}$ and $\Delta\mathbf{n}_{Ty}$ are the signal and noise part of $\hat{\mathcal{T}}_{1|2}$ respectively. For multiple events under the power stacking framework, we simply multiply all the posterior distributions, and the total Bayes factor is

$$\mathcal{T}_{1|2,\text{power}} = \log\prod_{i=1}^{N}\mathcal{B}_{1|2,i}$$
$$= -\sum_{i=1}^{N}\frac{||y_i - \hat{A}_{1c,i}c_i^{(1)} - \hat{A}_{1s,i}s_i^{(1)}||^2}{2}$$
$$+ \sum_{i=1}^{N}\frac{||y_i - \hat{A}_{2c,i}c_i^{(2)} - \hat{A}_{2s,i}s_i^{(2)}||^2}{2}$$
$$= \Delta\mathbf{s}_{Tp} + \Delta\mathbf{n}_{Tp} . \qquad (54)$$

In this case, the expectation of $T_{1|2}$ is

$$\langle\Delta\mathbf{s}_{Tp}\rangle = \sum_{i=1}^{N}\langle g_i|\mathcal{I}_{1,i} - \mathcal{I}_{2,i}|g_i\rangle$$
$$:= \langle\log\mathcal{B}_{1|2}\rangle_{,\text{power}} . \qquad (55)$$

On the other hand, we could coherently stack data from different events, if the assumption made in Sec. III B is met. By assuming that model 1 is the true EOS, we then find

$$2\langle\Delta\mathbf{s}_{Ty}\rangle = \sum_i w_i^2(\langle g_i|g_i\rangle - \langle g_i|\mathcal{I}_2|g_i\rangle)$$
$$+ \sum_{i\neq j} w_i w_j\langle g_i e^{i\phi_i^0}|\mathcal{I}_1 - \mathcal{I}_2|g_j e^{-i\phi_j^0}\rangle e^{-\sigma_{\phi_i}^2/2 - \sigma_{\phi_j}^2/2}$$
$$:= \langle\log\mathcal{B}_{1|2}\rangle_{,\text{coherent}} , \qquad (56)$$

where we use $\langle\Delta\mathbf{s}_{Ty}\rangle$ as the expectation of $\log\mathcal{B}_{1|2}$ for coherent stacking. Notice that $2\langle\Delta\mathbf{s}_{Ty}\rangle$ above reduces to $2\langle\mathbf{s}_{Ty}\rangle$ in Eq. (46) when $\mathcal{I}_2 = 0$ (i.e. when $c^{(2)} = 0 = s^{(2)}$).

One can use the *Jeffreys scale of interpretation of Bayes Factor* [74] to determine how significant a Bayes factor is. If $\mathcal{B}_{1|2}$ is between $[1,3]$, the statistical significance is barely worth mentioning; if $3 < \mathcal{B}_{1|2} < 10$ the evidence is strong; if $10 < \mathcal{B}_{1|2} < 100$ the evidence is very strong and beyond 100 it is decisive.

While events under detection threshold can be used to accumulate statistics via stacking in Eq. (16), the interpretation of the results of model selection (and also parameter estimation to be discussed in Sec. IV B) should be used with caution. This is because if the combined
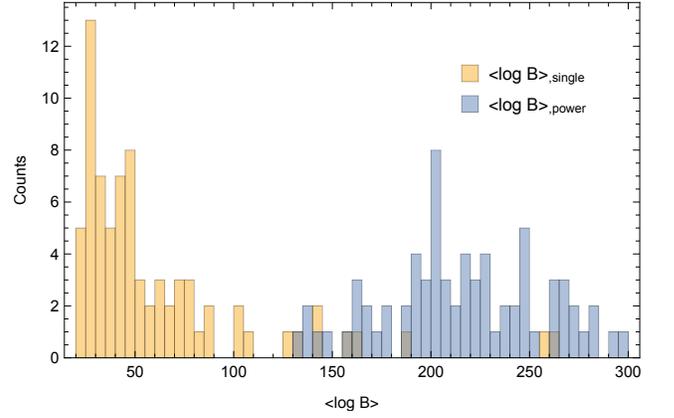

FIG. 7. Average (log of) Bayes factor for different MC realizations, testing EOS TM1 against DD2. Blue bins represent the stacked signals with the loudest 30 events according to Eq. (55). There are 79 realizations where their best events individually pass the detection threshold, and their corresponding Bayes factors are shown in orange bins.

statistics of a set of events does not pass the detection threshold, the existence of a 22 mode in any of these events is not confirmed. For simplicity we only present the distribution of $\langle\log\mathcal{B}_{1|2}\rangle_{,\text{power}}$ (using the 30 loudest events) versus the distribution of $\langle\log\mathcal{B}_{1|2}\rangle$ for single events; repeating the analysis with coherent stacking will introduce the additional complication of dealing with the effect of phase uncertainty, which we leave to future studies.

As an application, we assume TM1 to be the underlying EOS (model 1), and test it against the EOS DD2 (model 2). As discussed in Sec. III, there are 79 out of 100 MC realizations with at least one event passing the detection threshold, and all 100 MC realizations that pass the detection threshold when we apply power stacking for the top 30 events. For each MC realization that can claim a detection, we compute the corresponding Bayes factor, which is shown in blue bins in the histogram in Fig. 7. Notice that these average Bayes factors can only be used to rank the models in a semi-quantitative manner, as it is non-trivial to convert them to probability measures. A full analysis would require one to generate statistical distributions of these Bayes factors for each underlying EOS. In other words, Fig. 7 should be interpreted as the scattering of the average Bayes factor due to the astrophysical distribution of sources. Based on the results of comparing a single pair of EOSs, we conjecture that as long as a single event has passed the detection threshold, it can be used to distinguish between different EOSs with "very strong evidence". However, single events are less likely to be detected than the stacked events, and thus, the latter have more chance of distinguishing different EOSs than the former.

B. Parameter estimation for the peak frequency

Given a signal, we can also study the degree to which we can estimate its peak frequency, which we do here via a Fisher analysis. This can also serve as an alternative approach to distinguish between different EOSs, as they generally predict different peak frequencies. As a simple example, we assume TM1 as the best-fit EOS and construct the stacked signal accordingly. Next we promote the waveform (Eq. (1)) parameter vector to four-dimensions:

$$\lambda^i = \left(A, \phi^0, f_{\text{peak}}, Q \right) \qquad (57)$$

with $A = A' A_r$, and maximize Eq. (8) to obtain maximum likelihood estimators of these parameters.

In the Fisher approximation, the uncertainty in λ^i can be evaluated through the (Fisher) information matrix

$$\Gamma_{ij} = \langle \partial_i h | \partial_j h \rangle , \qquad (58)$$

where $\partial_i \equiv \partial / \partial \lambda^i$ and the inner product is defined with respect to the spectrum of $\tilde{\mathbf{n}}_y$ in Eq. (44). The measurement uncertainty of f_{peak} is simply

$$\delta f_{\text{peak}} \geq \sqrt{(\Gamma^{-1})_{f_{\text{peak}} f_{\text{peak}}}} , \qquad (59)$$

where the right hand side corresponds to the square root of the $(f_{\text{peak}}, f_{\text{peak}})$ element of the variance-covariance matrix. The inequality in the above equation comes about because of the Cramer-Rao bound, which guarantees a best-case measurement for a set of parameters in the high SNR limit [75]. We will use a Fisher analysis here only as a rough estimate of the accuracy to which parameters can be measured; a more complete analysis would construct the posterior probability distribution for each parameter through a detailed mapping of the likelihood surface, but this is beyond the scope of this paper.

Using these arguments and the approximations in [37], we assume that the off-diagonal terms of the Γ matrix are small, so that

$$\delta f_{\text{peak}} \approx \left(\Gamma_{f_{\text{peak}} f_{\text{peak}}} \right)^{-1/2} = \langle \partial_{f_{\text{peak}}} h | \partial_{f_{\text{peak}}} h \rangle^{-1/2}$$
$$\approx 0.7 \frac{f_{\text{peak}}}{Q \rho} , \qquad (60)$$

where in the last approximate equality we used the fact that the Fourier transform of Eq. (1) satisfies $\partial_{f_{\text{peak}}} \tilde{h} \sim Q \tilde{h} / f_{\text{peak}}$. The factor of 0.7 comes from a numerical fit to our set of data using the TM1 EOS, which is also expected by computing $\delta f_{\text{peak}} / f_{\text{peak}}$ for a universal Lorentzian-Type waveform. This shows that for $Q = 34$ (corresponding to the TM1 EOS) with a signal of $\rho \sim 6.5$, f_{peak} can be measured to $\sim 0.3\%$ accuracy at best.

We now look at the effect of the off-diagonal terms in the Fisher matrix. If we include ϕ^0 in addition to f_{peak} into the Fisher analysis, as they both enter the argument of the phase factor in the waveform, we find that the
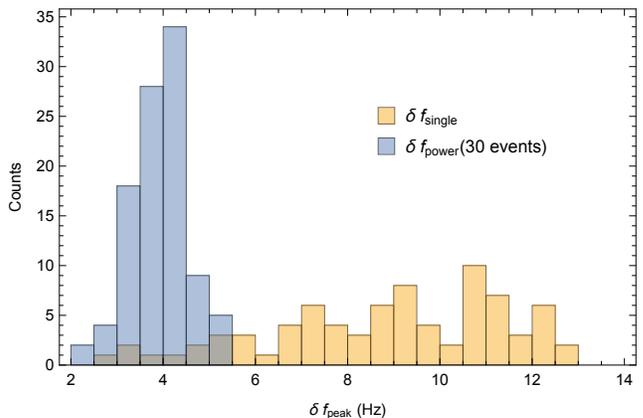

FIG. 8. Histogram of δf_{peak} for different MC realizations, assuming the TM1 EOS and a $1.35 + 1.35 M_\odot$ BNS merger remnant with corresponding $f_{\text{peak}} = 2.3$ kHz. The blue bins correspond to the power-stacked signal using the loudest 30 events from each of the 100 realizations that pass the detection threshold, while the orange bins correspond to the 79 individual events that pass the detection threshold. A $\delta f_{\text{peak}} \sim 8$ Hz measurement roughly corresponds to a ~ 26 m statistical error in the determination of the radius of a NS with mass $1.6 M_\odot$. However, systematic errors in the universal $f_{\text{peak}} - R_{1.6 M_\odot}$ relation used in the mapping could be larger than ~ 100 m, so these errors would dominate over the statistical measurement error.

statistical frequency uncertainty increases by a factor of 1.4. Consequently, the statistical radius uncertainty also increases by 1.4. Therefore, our Fisher analysis for TM1 including ϕ^0 predicts $\delta f_{\text{peak}} \sim 8$ Hz. For all the realizations with the best signal passing the detection threshold, we evaluate the uncertainties in δf_{peak} numerically, shown in Fig. 8.

Let us now map the statistical error in the peak frequency δf_{peak} to that of the radius of a NS with mass $1.6 M_\odot$, $\delta R_{1.6 M_\odot}$. From Eq. (33), one finds the following relation:

$$\frac{\delta R_{1.6 M_\odot}}{1 \text{km}} = \frac{\delta f_{\text{peak}}}{1 \text{kHz}} \frac{M_\odot}{m_1 + m_2} \left[2 a_2 \left(\frac{R_{1.6 M_\odot}}{1 \text{km}} \right) + a_1 \right]^{-1} , \qquad (61)$$

where we have neglected the error in the estimation of the component masses, as this is negligible for third generation detectors. Thus, a ~ 8 Hz statistical uncertainty in the peak frequency roughly corresponds to a ~ 30 m (0.3%, "TM1" EOS) uncertainty in radius (for a NS with mass $1.6 M_\odot$). The total error is the root of the sum of the squares of the statistical and all systematic errors. One source for the latter, as discussed in [37], comes from the $f_{\text{peak}} - R_{1.6 M_\odot}$ relation, and currently is above 100 m for $R_{1.6 M_\odot}$. Therefore, for now, the error budget is dominated by systematic error and not statistical when considering third generation detectors. Of course, as discussed earlier, we expect this systematic uncertainty to be considerably lowered by the time third generation detectors come online, as more accurate understanding and

modeling of BNS merger remnants is developed.

Notice that the uncertainty in frequency, with mean at about 8 Hz in Fig. 8 is significantly smaller than the value of order 50Hz presented in [37]. In fact, we notice that $\delta f_{\rm peak}/f_{\rm peak}$ in [37] does not follow the $0.7/(Q\rho)$ relation derived here. In [37] for the TM1 EOS, $\delta f_{\rm peak}/f_{\rm peak} \sim 50/2300$ with the post-merger SNR being 5. Based on Fig. 11 of [37], the SNR of the dominant 22 component is roughly half of the post-merger total SNR, i.e., $\rho \sim 2.3$. This converts to $\delta f_{\rm peak}/f_{\rm peak}(Q\rho/0.7) \sim 2.4$. We suspect this factor of 2.4 comes from the fact that we are using different waveform templates for parameter estimation.

Let us end this section by commenting on how the measurement accuracy of the NS radius changes if the correct EOS in nature is not TM1. Among the 5 EOSs considered in this paper, there is a very good chance of detecting the post-merger signal after power stacking for the DD2, TM1, LS220 and Shen EOSs, as discussed at the end of Sec. III C 1. The Shen EOS is quite similar to TM1, so let us consider LS220 here. The approximation in Eq. (60) shows that $\delta f_{\rm peak}/f_{\rm peak}$ depends only on Q and ρ. From Table I and comparing Figs. 1 and 3, one sees that $Q\rho$ for the DD2 EOS is roughly a factor of 2 smaller than $Q\rho$ for the TM1 EOS, which leads to a $\delta f_{\rm peak}$ that is approximately a factor of two larger, given the difference in $f_{\rm peak}$. Furthermore, using Eq. (61) one finds that DD2 has a $\delta R_{1.6M_\odot}$ that is roughly a factor of 1.1 times smaller. This means that if the post-merger signal is detected with CE via power stacking, depending on the underlying EOS we expect that $\delta f_{\rm peak}$ and $\delta R_{1.6M_\odot}$ to lie in the range $\sim 4-20$ Hz and $15-56$m, respectively ($\sim 4\%$ accuracy). Thus, systematic errors seem to always dominate statistical errors on the NS radius measurement irrespective of the EOS for third generation detectors (for stacked signals that pass the detection threshold).

V. DISCUSSION AND CONCLUSION

In this work we have studied the possibility of detecting the GWs generated by the oscillations of hypermassive NSs formed following BNS mergers with future ground-based GW detectors. Based on the latest estimates of the BNS merger rate and fitting formulas for the oscillation peak frequency from state-of-the-art BNS merger simulations, we found that the chance of detecting such oscillations from individual sources could be low even for third generation GW detectors, depending on the EOS. However, we point out that detectability of individual events could potentially improve if one considers all components/peaks that arise in the post-merger waveform, and not only the dominant peak, as we do here. Nevertheless, it is not currently clear whether sub-leading modes can persist and contribute substantially to the SNR or whether their frequencies might drift, hence making their detection challenging.

In order to increase the detection rate of the dominant post-merger component, we multiplied the Bayes factor of each event to derive the Bayes factor of combined events. We refer to this approach as power stacking. Such an analysis was used to propose a test of General Relativity in [44], to probe the BH no-hair property in [45] and to explore EOS properties in [46]. We have shown that this method can significantly boost the statistical chance of detection (shown in Fig. 4) as compared to single events. The stacked signal can also be used to distinguish between different NS EOSs. We formulated a Bayesian model selection framework, and illustrated its application by comparing EOS model TM1 vs DD2, assuming the former is the true EOS. In practice, such a model selection method only suggests *relative* preference between the two selected models, both of which do not have to be the true EOS. Thus, the results of model selection should be combined with the signal-to-noise level of the stacked signal, assuming different EOSs, to obtain an overall sense of the "true" EOS.

The power stacking methods can be naturally applied to other post-merger oscillation modes both in isolation or in combination with other modes. For instance, one could apply this to the 21 mode, which can become strong if a one-arm instability develops in the BNS remnant [24–27]. In this case, there is a tight correlation between the frequencies of the 21 and 22 modes which can be exploited to further enhance the achievable stacked SNR. This method can also be used to stack other post-merger GW templates, such as the Principal Component basis developed in [37]. In an even broader context, this approach could also be exploited to help identify decaying modes in cold atom data [76, 77] and their connection with possible BH duals through holographic arguments [78, 79].

If in the future the theoretical uncertainty in modelling gets down to a level that the initial phase of post-merger modes can be estimated given the binary parameters (including EOS), we can make use of this phase information to coherently stack a set of post merger events to further boost the collective SNR. We demonstrated that such coherent stacking could significantly increase the detection probability of the BNS post-merger dominant 22 mode. We explicitly showed that if we require the same level of statistical significance, then coherent stacking is more efficient at increasing the Bayes factor than the current way power stacking is performed (at least in the Bayesian framework we adopt); related comparisons are shown in Figs. 4 and 6. It would be interesting to find a Bayesian formulation that mimics the behavior of coherent stacking for low-SNR events, for example by introducing different weights for different events.

The main limitation of the coherent stacking method is that it requires small phase uncertainty in constructing the coherently stacked signal. If the phase uncertainties are large, the coherent part of the stacked signal (the second line of Eq. (46)) will be reduced dramatically. This could be alleviated if the initial phase of the 22 mode can accurately be estimated using the inferred parameters of the inspiral together with numerical simulations of the

merger event. Though producing full templates of the post-merger signals incorporating all the correct microphysics may not be practical within the next few years, it may not be unreasonable to expect that simulations can at least provide an accurate prediction of the initial phase of the 22 mode, as this will be fixed within the first few ms post-merger[7].

Another limitation of the framework used in this paper is that one needs to assume all the parameters are known except for A_c and A_s (or the amplitude and the phase offset). It would be interesting to extend the framework further to the case with unknown f_{peak} and Q. Then, one does not need to assume the underlying EOS a priori, and one can reformulate the Bayesian hypothesis test problem by taking into account the prior distribution of f_{peak} and Q.

When the post-merger SNR is above unity, the inspiral SNR will be large and one can likely extract nuclear physics information from the measurement of NS tidal deformations that occur in this phase. Thus, it would be interesting to study how the post-merger detection via stacking helps in probing nuclear physics by further including the inspiral measurement. Universal relations between the post-merger oscillation peak frequency and the leading tidal parameter in the inspiral waveform [60, 80] may help in addressing this question. Alternatively, an independent measurement of the tidal deformability and the post-merger peak frequency may allow one to confirm such universal relations from observations. If such relations are altered from the GR prediction in modified theories of gravity, one can use such a measurement to probe strong-field gravity. A similar proposal was already made and demonstrated regarding the universal relation between the tidal deformability and moment of inertia [81–83]. Also, as mentioned, complementary information from electromagnetic observations – coupled with refined numerical studies to connect the behavior of cold and finite temperature nuclear EOS – could be exploited to inform suitable priors for the analysis described here.

---

[7] The instantaneous 22 mode phase may drift with time due to non-linear effects, but that is beyond the scope of this model.
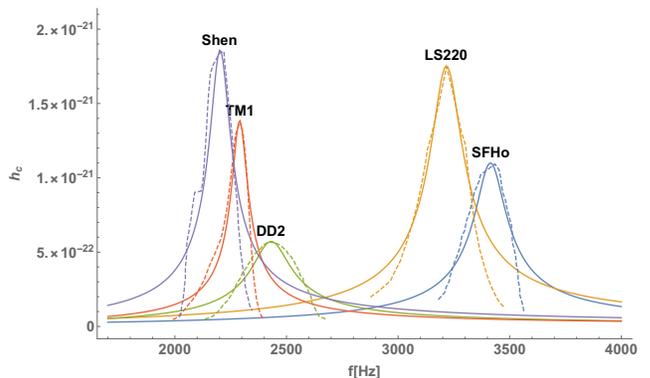

FIG. 9. Characteristic strain h_c vs frequency for the dominant peak in BNS post-merger gravitational wave spectra and for various equations of state considered in this work. The luminosity distance to the source is set to 50 Mpc. Solid lines correspond to Eq. (1) with the values for the amplitude and quality factor listed in Table I. Dashed lines correspond to the post-merger spectra obtained from numerical relativity simulations of [13, 70, 71].

Appendix A: Fitting the post-merger dominant gravitational-wave mode

In this appendix, we show how well Eq. (1) approximates the dominant peak of the post-merger 22 mode obtained from the numerical simulations of [13, 70, 71]. As discussed in the main text the strain amplitude A' and the quality factor Q in Eq. (1) are chosen for each equation of state such that the peak value of the characteristic strain and SNR with Eq. (1) match the peak value of the characteristic strain and SNR of the dominant post-merger 22 component found in the corresponding BNS merger simulations. Figure 9 shows that the Lorentzian profile (1) provides a reasonable approximation of numerical relativity post-merger spectra around the dominant peak.

Appendix B: Variance and Signal-to-noise Level of $\hat{T}$

In this appendix, we explain the variance and the signal-to-noise level of $\hat{T}$, which is the log of the Bayes factor between the hypotheses $\mathcal{H}_1$ and $\mathcal{H}_2$. Such a signal-to-noise level can be used, instead of ρ, to discuss the detection criterion of the post-merger GW signals.

Let us begin with the single event case. The variance of n_T is given by

$$\begin{aligned} \mathrm{Var}[n_T] &= 1 + A_c^2 + A_s^2 \\ &= 1 + \rho^2 \,, \end{aligned} \tag{B1}$$

(recall that we chose $\langle c|c \rangle = 1 = \langle s|s \rangle$). We can then define the ratio

$$\frac{2s_T}{\sqrt{\mathrm{Var}(n_T)}} = \frac{\rho^2}{\sqrt{1+\rho^2}}, \tag{B2}$$

which intuitively measures the signal-to-noise level in the GLRT variable $\hat{T}_{\mathrm{single}}$. In the limit that the detection SNR $= \rho \gg 1$, it is straightforward to see that the above ratio is approximately ρ.

We now explain the coherent stacking case. The noise part of $\hat{T}_{\mathrm{coherent}}$, i.e., $\mathbf{n}_{Ty}$, follows a distribution similar to Eq. (25), with $A_c \to \langle g_y|c \rangle$ and $A_s \to \langle g_y|s \rangle$. Its variance is given by

$$\mathrm{Var}[\mathbf{n}_{Ty}] = \left(\sum_i w_i^2 \int df \frac{4S_{n_i}(f)|\tilde{h}_c(f)|^2}{S_{n_y}(f)^2} \right)^2$$
$$+ \sum_k w_k^2 \sum_{ij} w_i w_j \int df \frac{2(\tilde{g}_i \tilde{g}_j^* e^{i(\phi_i^0 - \phi_j^0)} + c.c.)S_{n_k}}{S_{n_y}(f)^2} . \tag{B3}$$

With Eq. (46) and Eq. (B3), one can compute $\langle \mathbf{s}_{Ty} \rangle / \sqrt{\mathrm{Var}[\mathbf{n}_{Ty}]}$ for the stacked signal $\mathbf{y}$. As before, this quantity is a measure of the signal-to-noise level in the variable $\hat{T}_{\mathrm{coherent}}$. Indeed, if all individual events have the same SNR, then in the SNR $= \rho \gg 1$ limit one finds $\langle \mathbf{s}_{Ty} \rangle / \sqrt{\mathrm{Var}[\mathbf{n}_{Ty}]} \approx N^{1/2}\rho$.

Appendix C: Model selection with different data sets

For different EOSs we generically obtain different stacked signals, thus one is faced with the problem of performing model selection using different data sets as discussed in Sec. IV A. In this appendix, we explain how one can construct appropriate Bayes factors for such a model selection study.

We begin by generalizing further Eq. (15) to allow different data sets:

$$\mathcal{B}_{12} \equiv \frac{P(\mathbf{y}_1|\mathcal{H}_1)}{P(\mathbf{y}_2|\mathcal{H}_2)}, \tag{C1}$$

where $\mathbf{y}_1$ ($\mathbf{y}_2$) is the data $\mathbf{y}$ stacked using the frequency scaling of EOS 1 (2). Within the GLRT framework, one may consider the expectation value and noise distribution of the random variable

$$\hat{\mathcal{T}}_{12} \equiv \log \mathcal{B}_{12} \tag{C2}$$

to do model selection using the Jeffreys criteria. Notice that $\hat{\mathcal{T}}_{12} = -\hat{\mathcal{T}}_{21}$. Let us assume that model 1 represents the true underlying EOS. One interesting feature implied by Eq. (33) is that even if we make an assumption that the EOS is model 2 where the true underlying EOS follows model 1, the frequency rescaling factors depend only on the total mass for each event, and of course the measured mass is EOS independent. Therefore, the main consequence of assuming an "incorrect" EOS is that the data from different events are not coherently stacked onto each other due to the frequency mismatch between the predicted signal and the actual signal. Such a mismatch also brings systematic errors on the phase measurement, making the signals further incoherent. Incomplete coherent stacking may greatly degrade the signal part of GLRT variable.

Alternatively, if model 2 is so incorrect that the ρ of the stacked signal is well below that of model 1, the phase error in constructing the stacked signal is large and the weights obtained by assuming an "incorrect" EOS are far from their optimal values, then $\mathbf{y}_1$ seems to be a convincingly better set of data than $\mathbf{y}_2$. Therefore, it is more appropriate to evaluate the following Bayes factor:

$$\mathcal{B}_{1|2} \equiv \frac{P(\mathbf{y}_1|\mathcal{H}_1)}{P(\mathbf{y}_1|\mathcal{H}_2)} . \tag{C3}$$

Intuitively, $\mathcal{B}_{12}$ may work better at distinguishing two close EOSs, whereas $\mathcal{B}_{1|2}$ is expected to have wider applicability and the associated analysis is more straightforward. In this paper, we chose to study the statistical behavior of $\mathcal{B}_{1|2}$, with model 1 being the one with better SNR from the study in Sec. III B. The extension of the analysis presented here to the case dealing with the random variable $\mathcal{B}_{12}$ goes beyond the scope of the current paper.

[1] B. Abbott *et al.* (Virgo, LIGO Scientific), Phys. Rev. Lett. **119**, 161101 (2017), arXiv:1710.05832 [gr-qc].
[2] B. P. Abbott *et al.* (GROND, SALT Group, Oz-Grav, DFN, INTEGRAL, Virgo, Insight-Hxmt, MAXI Team, Fermi-LAT, J-GEM, RATIR, IceCube, CAAS-TRO, LWA, ePESSTO, GRAWITA, RIMAS, SKA South Africa/MeerKAT, H.E.S.S., 1M2H Team, IKI-GW Follow-up, Fermi GBM, Pi of Sky, DWF (Deeper Wider Faster Program), Dark Energy Survey, MASTER, AstroSat Cadmium Zinc Telluride Imager Team, Swift, Pierre Auger, ASKAP, VINROUGE, JAGWAR, Chandra Team at McGill University, TTU-NRAO, GROWTH, AGILE Team, MWA, ATCA, AST3, TOROS, Pan-STARRS, NuSTAR, ATLAS Telescopes, BOOTES, CaltechNRAO, LIGO Scientific, High Time Resolution Universe Survey, Nordic Optical Telescope, Las Cumbres

Observatory Group, TZAC Consortium, LOFAR, IPN, DLT40, Texas Tech University, HAWC, ANTARES, KU, Dark Energy Camera GW-EM, CALET, Euro VLBI Team, ALMA), Astrophys. J. **848**, L12 (2017), arXiv:1710.05833 [astro-ph.HE].

[3] B. D. Metzger, (2017), arXiv:1710.05931 [astro-ph.HE].

[4] V. Paschalidis, Class. Quant. Grav. **34**, 084002 (2017), arXiv:1611.01519 [astro-ph.HE].

[5] R. Oechslin, S. Rosswog, and F.-K. Thielemann, Phys. Rev. D. **65**, 103005 (2002), gr-qc/0111005.

[6] M. Shibata and K. Uryū, Progress of Theoretical Physics **107**, 265 (2002), gr-qc/0203037.

[7] M. Shibata, K. Taniguchi, and K. Uryū, Phys. Rev. D **71**, 084021 (2005), gr-qc/0503119.

[8] M. Shibata and K. Taniguchi, Phys. Rev. **D73**, 064027 (2006), arXiv:astro-ph/0603145 [astro-ph].

[9] K. Kiuchi, Y. Sekiguchi, M. Shibata, and K. Taniguchi, Phys. Rev. D **80**, 064037 (2009), arXiv:0904.4551 [gr-qc].

[10] A. Bauswein and H. T. Janka, Phys. Rev. Lett. **108**, 011101 (2012), arXiv:1106.1616 [astro-ph.SR].

[11] K. Takami, L. Rezzolla, and L. Baiotti, ArXiv e-prints (2014), arXiv:1403.5672 [gr-qc].

[12] A. Bauswein, N. Stergioulas, and H.-T. Janka, Eur. Phys. J. **A52**, 56 (2016), arXiv:1508.05493 [astro-ph.HE].

[13] C. Palenzuela, S. L. Liebling, D. Neilsen, L. Lehner, O. L. Caballero, E. O'Connor, and M. Anderson, Phys. Rev. **D92**, 044045 (2015), arXiv:1505.01607 [gr-qc].

[14] L. Lehner, S. L. Liebling, C. Palenzuela, O. L. Caballero, E. O'Connor, M. Anderson, and D. Neilsen, (2016), arXiv:1603.00501 [gr-qc].

[15] L. Rezzolla and K. Takami, Phys. Rev. **D93**, 124051 (2016), arXiv:1604.00246 [gr-qc].

[16] T. Dietrich, S. Bernuzzi, M. Ujevic, and W. Tichy, (2016), arXiv:1611.07367 [gr-qc].

[17] D. Radice, S. Bernuzzi, W. Del Pozzo, L. F. Roberts, and C. D. Ott, (2016), arXiv:1612.06429 [astro-ph.HE].

[18] H. Yang, D. A. Nichols, F. Zhang, A. Zimmerman, Z. Zhang, and Y. Chen, Phys. Rev. D **86**, 104006 (2012).

[19] H. Yang, A. Zimmerman, A. i. e. i. f. Zenginoğlu, F. Zhang, E. Berti, and Y. Chen, Phys. Rev. D **88**, 044047 (2013).

[20] H. Yang, F. Zhang, A. Zimmerman, D. A. Nichols, E. Berti, and Y. Chen, Phys. Rev. D **87**, 041502 (2013).

[21] H. Yang, A. Zimmerman, and L. Lehner, Phys. Rev. Lett. **114**, 081101 (2015).

[22] L. Baiotti and L. Rezzolla, (2016), 10.1088/1361-6633/aa67bb, arXiv:1607.03540 [gr-qc].

[23] V. Paschalidis and N. Stergioulas, (2016), arXiv:1612.03050 [astro-ph.HE].

[24] V. Paschalidis, W. E. East, F. Pretorius, and S. L. Shapiro, Phys. Rev. **D92**, 121502 (2015), arXiv:1510.03432 [astro-ph.HE].

[25] W. E. East, V. Paschalidis, F. Pretorius, and S. L. Shapiro, Phys. Rev. **D93**, 024011 (2016), arXiv:1511.01093 [astro-ph.HE].

[26] W. E. East, V. Paschalidis, and F. Pretorius, (2016), arXiv:1609.00725 [astro-ph.HE].

[27] L. Lehner, S. L. Liebling, C. Palenzuela, and P. M. Motl, Phys. Rev. **D94**, 043003 (2016), arXiv:1605.02369 [gr-qc].

[28] D. Radice, S. Bernuzzi, and C. D. Ott, Phys. Rev. **D94**, 064011 (2016), arXiv:1603.05726 [gr-qc].

[29] B. P. Abbott *et al.* (Virgo, LIGO Scientific), Phys. Rev. Lett. **116**, 241102 (2016), arXiv:1602.03840 [gr-qc].

[30] B. P. Abbott *et al.* (Virgo, LIGO Scientific), Phys. Rev. Lett. **116**, 221101 (2016), arXiv:1602.03841 [gr-qc].

[31] L. Sampson, N. Yunes, N. Cornish, M. Ponce, E. Barausse, A. Klein, C. Palenzuela, and L. Lehner, Phys. Rev. **D90**, 124091 (2014), arXiv:1407.7038 [gr-qc].

[32] E. Barausse, C. Palenzuela, M. Ponce, and L. Lehner, Phys.Rev. **D87**, 081506 (2013), arXiv:1212.5053 [gr-qc].

[33] M. Shibata, K. Taniguchi, H. Okawa, and A. Buonanno, Phys. Rev. **D89**, 084005 (2014), arXiv:1310.0627 [gr-qc].

[34] C. Palenzuela, E. Barausse, M. Ponce, and L. Lehner, Phys. Rev. **D89**, 044024 (2014), arXiv:1310.4481 [gr-qc].

[35] K. C. Gendreau *et al.*, "The neutron star interior composition explorer (nicer): design and development," (2016).

[36] J. Clark, A. Bauswein, L. Cadonati, H. T. Janka, C. Pankow, and N. Stergioulas, Phys. Rev. **D90**, 062004 (2014), arXiv:1406.5444 [astro-ph.HE].

[37] J. A. Clark, A. Bauswein, N. Stergioulas, and D. Shoemaker, Classical and Quantum Gravity **33**, 085003 (2016).

[38] B. Abbott, R. Abbott, T. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. Adhikari, V. Adya, *et al.*, The Astrophysical Journal Letters **851**, L16 (2017).

[39] K. Chatziioannou, J. A. Clark, A. Bauswein, M. Millhouse, T. B. Littenberg, and N. Cornish, Physical Review D **96**, 124035 (2017).

[40] J. Meidam, M. Agathos, C. Van Den Broeck, J. Veitch, and B. S. Sathyaprakash, Physical Review D **90**, 064009 (2014).

[41] H. Yang, K. Yagi, J. Blackman, L. Lehner, V. Paschalidis, F. Pretorius, and N. Yunes, Phys. Rev. Lett. **118**, 161101 (2017), arXiv:1701.05808 [gr-qc].

[42] A. Bauswein, T. Baumgarte, and H.-T. Janka, Physical review letters **111**, 131101 (2013).

[43] M. Hempel, T. Fischer, J. Schaffner-Bielich, and M. Liebendorfer, Astrophys. J. **748**, 70 (2012), arXiv:1108.0848 [astro-ph.HE].

[44] M. Agathos, W. Del Pozzo, T. G. F. Li, C. Van Den Broeck, J. Veitch, and S. Vitale, Phys. Rev. **D89**, 082001 (2014), arXiv:1311.0420 [gr-qc].

[45] J. Meidam, M. Agathos, C. Van Den Broeck, J. Veitch, and B. S. Sathyaprakash, Phys. Rev. D **90**, 064009 (2014).

[46] M. Agathos, J. Meidam, W. Del Pozzo, T. G. F. Li, M. Tompitak, J. Veitch, S. Vitale, and C. Van Den Broeck, Phys. Rev. **D92**, 023012 (2015), arXiv:1503.05405 [gr-qc].

[47] P. Kalmus, K. C. Cannon, S. Marka, and B. J. Owen, Phys. Rev. **D80**, 042001 (2009), arXiv:0904.4906 [astro-ph.HE].

[48] K. S. Tai, S. T. McWilliams, and F. Pretorius, Phys. Rev. **D90**, 103001 (2014), arXiv:1403.7754 [gr-qc].

[49] K. C. Gendreau, Z. Arzoumanian, and T. Okajima, in *Space Telescopes and Instrumentation 2012: Ultraviolet to Gamma Ray*, Proc. SPIE, Vol. 8443 (2012) p. 844313.

[50] S. Bose, K. Chakravarti, L. Rezzolla, B. Sathyaprakash, and K. Takami, arXiv preprint arXiv:1705.10850 (2017).

[51] E. Berti, J. Cardoso, V. Cardoso, and M. Cavaglia, Phys. Rev. **D76**, 104044 (2007), arXiv:0707.1202 [gr-qc].

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

[53] N. J. Cornish and T. B. Littenberg, Phys.Rev. **D76**, 083006 (2007), arXiv:0704.1808 [gr-qc].

[54] T. B. Littenberg and N. J. Cornish, Phys. Rev. D80, 063007 (2009), arXiv:0902.0368 [gr-qc].

[55] B. S. Sathyaprakash and B. F. Schutz, Living Reviews in Relativity 12, 2 (2009).

[56] T. Regimbau, T. Dent, W. Del Pozzo, S. Giampanis, T. G. Li, C. Robinson, C. Van Den Broeck, D. Meacher, C. Rodriguez, B. S. Sathyaprakash, et al., Physical Review D 86, 122001 (2012).

[57] J. Abadie, B. Abbott, R. Abbott, M. Abernathy, T. Accadia, F. Acernese, C. Adams, R. Adhikari, P. Ajith, B. Allen, et al., Classical and Quantum Gravity 27, 173001 (2010).

[58] A. Bauswein, H. T. Janka, K. Hebeler, and A. Schwenk, Phys. Rev. D86, 063001 (2012), arXiv:1204.1888 [astro-ph.SR].

[59] A. Bauswein, N. Stergioulas, and H. T. Janka, Phys. Rev. D90, 023002 (2014), arXiv:1403.5301 [astro-ph.SR].

[60] K. Takami, L. Rezzolla, and L. Baiotti, Phys. Rev. D91, 064001 (2015), arXiv:1412.3240 [gr-qc].

[61] A. Bauswein and N. Stergioulas, Phys. Rev. D91, 124056 (2015), arXiv:1502.03176 [astro-ph.SR].

[62] F. Özel and P. Freire, Annual Review of Astronomy and Astrophysics 54, 401 (2016).

[63] P. Amaro-Seoane, N. Andersson, K. Arun, S. Bose, L. Bosi, J. Clark, T. Dent, J. Gair, K. Glampedakis, M. Hannam, et al., Codified Document ET-030-09 (2009).

[64] B. Abbott, R. Abbott, T. Abbott, M. Abernathy, K. Ackley, C. Adams, P. Addesso, R. Adhikari, V. Adya, C. Affeldt, et al., Classical and Quantum Gravity 34, 044001 (2017).

[65] A. W. Steiner, M. Hempel, and T. Fischer, Astrophys. J. 774, 17 (2013), arXiv:1207.2184 [astro-ph.SR].

[66] J. M. Lattimer and F. Douglas Swesty, Nuclear Physics A 535, 331 (1991).

[67] H. Shen, H. Toki, K. Oyamatsu, and K. Sumiyoshi, Astrophys. J. Suppl. 197, 20 (2011), arXiv:1105.1666 [astro-ph.HE].

[68] P. B. Demorest, T. Pennucci, S. M. Ransom, M. S. E. Roberts, and J. W. T. Hessels, nature 467, 1081 (2010), arXiv:1010.5788 [astro-ph.HE].

[69] J. Antoniadis, P. C. C. Freire, N. Wex, T. M. Tauris, R. S. Lynch, M. H. van Kerkwijk, M. Kramer, C. Bassa, V. S. Dhillon, T. Driebe, J. W. T. Hessels, V. M. Kaspi, V. I. Kondratiev, N. Langer, T. R. Marsh, M. A. McLaughlin, T. T. Pennucci, S. M. Ransom, I. H. Stairs, J. van Leeuwen, J. P. W. Verbiest, and D. G. Whelan, Science 340, 448 (2013), arXiv:1304.6875 [astro-ph.HE].

[70] Y. Sekiguchi, K. Kiuchi, K. Kyutoku, and M. Shibata, Phys. Rev. Lett. 107, 211101 (2011), arXiv:1110.4442 [astro-ph.HE].

[71] N. Stergioulas, A. Bauswein, K. Zagkouris, and H.-T. Janka, Mon. Not. Roy. Astron. Soc. 418, 427 (2011), arXiv:1105.0368 [gr-qc].

[72] J. A. Nelder and R. J. Mead, The Computer Journal 7, 308 (1965).

[73] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, Numerical Recipes: The Art of Scientific Computing (Cambridge University Press, Cambridge (UK) and New York, 2007).

[74] H. Jeffreys, Theory of probability (Clarendon Press, Oxford, 1961).

[75] M. Vallisneri, Phys. Rev. D77, 042001 (2008), arXiv:gr-qc/0703086 [GR-QC].

[76] J. Kinast, A. Turlapov, and J. E. Thomas, Phys. Rev. Lett. 94, 170404 (2005).

[77] C. Cao, E. Elliott, J. Joseph, H. Wu, J. Petricka, T. Schfer, and J. E. Thomas, Science 331, 58 (2011), arXiv:1007.2625 [cond-mat.quant-gas].

[78] J. Brewer and P. Romatschke, Phys. Rev. Lett. 115, 190404 (2015), arXiv:1508.01199 [hep-th].

[79] H. Bantilan, J. T. Brewer, T. Ishii, W. E. Lewis, and P. Romatschke, Phys. Rev. A94, 033621 (2016), arXiv:1605.00014 [cond-mat.quant-gas].

[80] S. Bernuzzi, T. Dietrich, and A. Nagar, Phys. Rev. Lett. 115, 091101 (2015), arXiv:1504.01764 [gr-qc].

[81] K. Yagi and N. Yunes, Science 341, 365 (2013), arXiv:1302.4499 [gr-qc].

[82] K. Yagi and N. Yunes, Phys. Rev. D88, 023009 (2013), arXiv:1303.1528 [gr-qc].

[83] K. Yagi and N. Yunes, Phys. Rept. 681, 1 (2017), arXiv:1608.02582 [gr-qc].