Marginalizing Over Noise Properties in Parameter Estimation

Cailin Plunkett
Advisors: Katerina Chatziioannou, Sophie Hourihane
(Dated: September 24 2021)

The traditional gravitational wave parameter estimation process relies on sequential estimation of noise properties and binary parameters, which assumes the noise variance is perfectly known. Using new capabilities of the `BayesWave` algorithm and recent developments in noise uncertainty modeling, we simultaneously estimate the noise and compact binary parameters, marginalizing over uncertainty in the noise. We compare the sequential estimation method and the marginalized method on real GW events from GWTC-2 using both the wavelet- and template-based models in `BayesWave`. We find that the recovered signals and posterior parameter distributions agree in median and width. At current sensitivities, PSD uncertainty is a subdominant effect compared to other sources of uncertainty.

I. INTRODUCTION

Gravitational wave (GW) data analysis requires models of both the genuine GW signal and the frequency-dependent noise in the raw data. Accurate estimation of black hole and neutron star properties from compact binary coalescence (CBC) signals depends on the robustness of both of these models [1]. While creating waveform templates by numerically solving Einstein's equations has been the subject of many research operations over the last decades [2], noise models have not been traditionally given the same amount of attention.

The traditional parameter estimation (PE) process uses sequential estimation of noise properties and binary parameters. The random noise in GW detectors is characterized by its power spectral density (PSD), $S_n(f)$, or its frequency-dependent variance. The PSD is modeled first, then given as a fixed input to `LALInference` (LI) or its successor `Bilby`, the primary PE pipelines used by the LIGO and Virgo collaboration [3][4]. LI and `Bilby` are template-based GW searches that require a provided model for the noise to whiten the data prior to evaluating the likelihood function for the CBC parameters.

Such analyses of CBC signals depend on three assumptions about the noise properties: first, that the noise is Gaussian; second, that it is stationary in time; and third, that its frequency-dependent variance is perfectly known. We discuss how these assumptions appear in the likelihood function in Appendix A. Of particular relevance for this study, the third is invoked in the use of a point estimate for the noise as input to the CBC PE pipeline. However, in practice, all three assumptions are invalid to some extent. Recent research in noise modeling has sought to address these assumptions to increase accuracy and prevent bias in PE.

The stationarity assumption, for instance, is addressed in Chatziioannou *et al.* [5], which compares two methods of estimating the noise PSD for use as a point estimate in LI. First, they consider the "off-source" method, which averages the PSD from many data segments around, but not including, the data segment containing the signal [3]. They compare this to the "on-source" method, which uses the `BayesLine` (BL) algorithm to estimate the PSD with the same data segment that is used for PE [6]. BL is integrated into the broader `BayesWave` (BW) algorithm, a variable dimension, parameterized model to separate transient GW signals from noise [1][7]. BL estimates the PSD as a sum of splines and Lorentzians. By using the same data segment for PE and noise modeling, the on-source estimation method mitigates the amount of time that the stationarity assumption must hold. Chatziioannou *et al.* [5] found that the on-source PSD produced whitened data more consistent with a Gaussian likelihood and better recovered injection parameters.

In order to address the assumption that the noise variance is perfectly known, the noise uncertainty must be marginalized over during PE. The BW algorithm is capable of employing BL to simultaneously marginalize over the noise and the signal. Until recently, it could do so only while modeling the GW signal using sine-Gaussian wavelets; there was not yet a template-based algorithm that did not require a fixed estimate of the noise. An augmentation to the BW algorithm, described in detail in Chatziioannou *et al.* [8] and discussed in Section II A, is a CBC template-based GW search. When used with BL, it marginalizes over noise uncertainty in parameter estimation. This method simultaneously mitigates the assumptions of noise stationarity and known variance.

While the impact of the stationarity premise has been thoroughly explored, we now seek to study the known variance assumption. In this work, we expand on the methods in Chatziioannou *et al.* [8], which tested the model on simulated signals, to compare the PE results on real GW events when using a fixed input PSD as opposed to BL. In doing so, we test the effect of marginalizing over the noise properties on real-data PE.

The rest of the paper is organized as follows. In Section II, we describe the BW algorithm and models for the GW signal available within it. In Section III, we discuss our analyses on a set of events from GWTC-2 [9], considering the waveform reconstructions as well as the parameter posteriors. To further understand our real data results, we then explore the impact of using a shifted PSD in PE before mathematically considering the effect of PSD marginalization. Finally, in Section IV we conclude.

II. METHODS

A. Algorithm Description

`BayesWave` is a Markov Chain Monte Carlo algorithm that simultaneously models GW signals, detector noise, and glitches. [1][7]. For this work, we focus on the noise and two GW signal models.

The *noise model* uses the `BayesLine` algorithm, fully integrated into BW, to describe the PSD as the sum of a broadband spline and Lorentzians for spectral lines [6]. The number and parameters of the spline control points and Lorentzians are marginalized over.

The *signal model* finds astrophysical signals by searching for coherent power between the detectors, which it describes as the sum of Morlet-Gabor wavelets. The algorithm is trans-dimensional, meaning the number and morphology of the wavelets are not fixed *a priori*. Its flexible approach enables it to find weakly modeled signals, such as bursts, in addition to well-modeled events like CBCs. While other models for the GW signal are now available in BW, we keep the original names, so "signal model" refers to the wavelet signal model.

The second model we employ is the *CBC model*, which uses general relativity templates to model the signal. The latest development of `BayesWave`, the CBC model is thoroughly described in Chatziioannou *et al.* [8]. The CBC model marginalizing over noise in CBC parameter estimation. For this work, we use the `IMRPhenomD` [10] waveform model for signal recovery.

Lastly, we use the *glitch cleaning phase* to obtain the PSDs that we use for the fixed PSD analyses. The cleaning phase serves a dual purpose: first, it models and removes excess power outside of a given window. Second, it estimates the PSD. When run on its own, the cleaning phase output can be used as the fixed estimate for a traditional PE analysis. When run as a preprocessing step before the CBC or signal model, the output is used as the starting point for the BL sampler.

B. Analyses on Detected GW Events

We perform four primary analyses on each GW event. To conduct the traditional fixed PSD analyses, we first run the cleaning phase to obtain a PSD estimate. We take the median output PSD as the input to signal and CBC model runs, described from here on as "signal + fixed" and "CBC + fixed" analyses. We then run both models using BL, which we call "signal + BL" and "CBC + BL" analyses, to find the signal while marginalizing over noise properties. Finally, we compare the two signal model and two CBC model analyses.

We also check results within each method for consistency when using, versus not using, a heterodyned likelihood, a method that speeds up the computation described in Cornish [11], and compare results when using fixed PSDs offset by a constant.

III. ANALYSES ON DETECTED GW EVENTS

We run each analysis on several events from GWTC-2 [9], chosen to cover the range of masses and SNRs of the events observed in O1-O3a. The chosen events and their published SNRs, masses in the source frame, and distances are listed in Table I. The run parameters that we use for each event are in Table II.

Event	SNR	$\mathcal{M}/M_\odot$	$m_1/M_\odot$	$m_2/M_\odot$	Distance (Mpc)
GW150914	24.2	28.6	35.6	30.6	440
GW151226	13.1	8.9	13.7	7.7	450
GW170104	13	21.4	30.8	20.0	990
GW170814	15.9	24.1	30.6	25.2	600
GW190412	18.9	13.3	30.1	8.3	740
GW190521	14.2	69.2	95.3	69.0	3920

TABLE I: The GW events we analyzed from GWTC-2 and their median parameters, as published in [12] and [9].

Event	$f_{\rm low}$	Segment length (s)	Sampling rate (Hz)
GW150914	16	4	2048
GW151226	16	8	2048
GW170104	16	4	2048
GW170814	16	4	2048
GW190412	16	4	1024
GW190521	16	4	1024

TABLE II: Run settings for each event we analyzed.

A. Recovered Waveforms

Figures 1 and 2 show the results of the four analyses in the time and frequency domains for GW150914. The former compares the signal model results while the latter compares those of the CBC model.

We first observe that the signal and CBC models agree in terms of waveform shape, power, and cutoff frequency. We see this with each event we analyzed, further confirming the established result that GR templates agree with the weakly modeled wavelet method [13]. The signal model analyses have much greater uncertainty, as anticipated.

Within each model, the recovered waveforms for the fixed and BL methods agree in median and uncertainty; despite the BL method adding uncertainty to the analysis, the methods have comparable error. The signal model recovers slightly more power and SNR when used with BL, which can be seen in the time-domain plot. For every GW event we analyze, there is this close agreement between the recovered waveforms for the fixed and BL methods in both models. We show the direct comparison for GW150914 as a representative example.

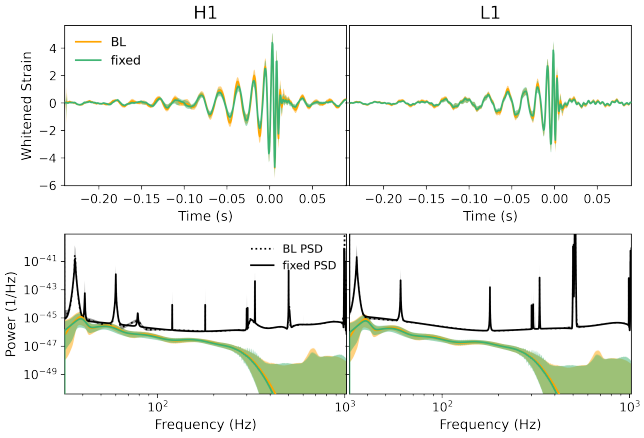

FIG. 1: *Top row:* Time-domain waveform comparison between fixed PSD (green) and BL PSD (orange) methods for GW150914 in Hanford (left) and Livingston (right) using the signal model. Time is with respect to GPS 1126259462.4. *Bottom row:* Frequency-domain waveform and PSD comparison.

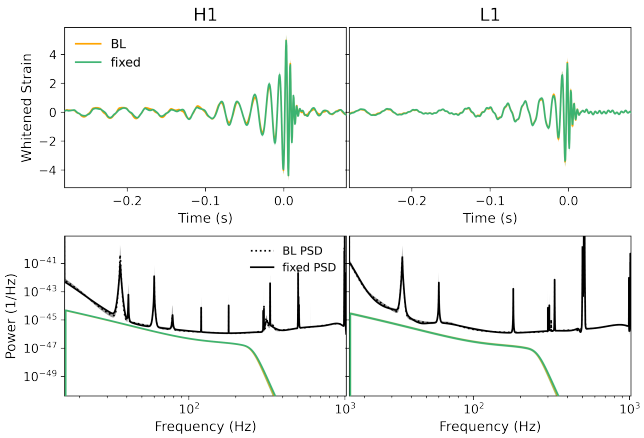

FIG. 2: Same as Fig. 1, but for the CBC model.

B. Parameter Posteriors

We compare the parameter posterior distributions of the two methods for all events. We focus on chirp mass and effective spin for the intrinsic parameters and distance for extrinsic parameters. Figure 3 compares the parameter posteriors for detector-frame chirp mass, effective spin, and luminosity distance for GW150914. The medians are overplotted and labeled with the 90% confidence intervals. Although the CBC+fixed posteriors for chirp mass and spin are slightly shifted and irregular in shape, the medians agree to within uncertainty and the 90% widths are comparable.

Figure 4 shows the detector-frame chirp mass posteriors for all events we analyzed with their medians overplotted. The medians slightly differ for GW150914 and GW170104, but to well within the uncertainty. Additionally, the posterior widths are dissimilar for GW170814, but agree for all other events. Overall, the distributions agree well in shape and spread. We find no systematic difference between the chirp mass posterior medians or uncertainties between the two methods. The spin and distance posteriors show similar trends and we have the same conclusion.

That the posterior widths of the fixed and BL methods agree is non-intuitive; by adding additional uncertainty from the PSD, one would expect the BL posterior to have greater uncertainty. In the case of GW170814, the event for which the CBC+BL posterior is significantly different in width than the CBC+fixed posterior, the PSD uncertainty is not notably higher than the other events, suggesting the observed widening is not attributable to additional PSD error.

C. Shifted PSD

To understand why posterior uncertainty does not increase when we marginalize over PSD uncertainty, we first look at the posterior width when we use a fixed PSD that is shifted. Overestimation of the noise variance leads to underestimation of the SNR, and as a consequence, overestimation of posterior uncertainty. Similarly, if one is falsely confident by underestimating the PSD, the posteriors are incorrectly narrow.

To see this effect in the data, we run CBC+fixed PE for GW159014 with the median PSD from BL—the standard one used—as well as the low and high 90% confidence interval PSDs. Note that these runs are conducted with $f_{\rm low} = 32$ and no heterodyned likelihood, while the other fixed PSD runs presented here used $f_{\rm low} = 16$ and a heterodyned likelihood. The posteriors are plotted in Figure 5 and show the expected posterior broadening with a higher PSD. The shift in the medians is likely sampling error and insignificant. Figure 5 also plots the $\mathcal{M}$ 90% width as a function of ASD height, given as fractional difference between the ASD and the median at 100 Hz. With three points, we can draw limited conclusions from the plot, but can see the increasing trend. $\chi_{\rm eff}$, the second-best measured parameter after $\mathcal{M}$, shows similar widening, but distance does not. This is likely because distance is much less well-measured, with $\sim 100\%$ fractional uncertainty, so the impact of PSD height on posterior width is outweighed by other effects.

D. PSD marginalization

When using BL to marginalize over PSD uncertainty, each MCMC step computes an estimate for the PSD, which may have different spline and Lorentzian parameters, and likely a different height, than the "true" PSD. BW then samples the intrinsic CBC parameters using this PSD estimate. Because the PSD height impacts the posterior width, at each step, one effectively samples

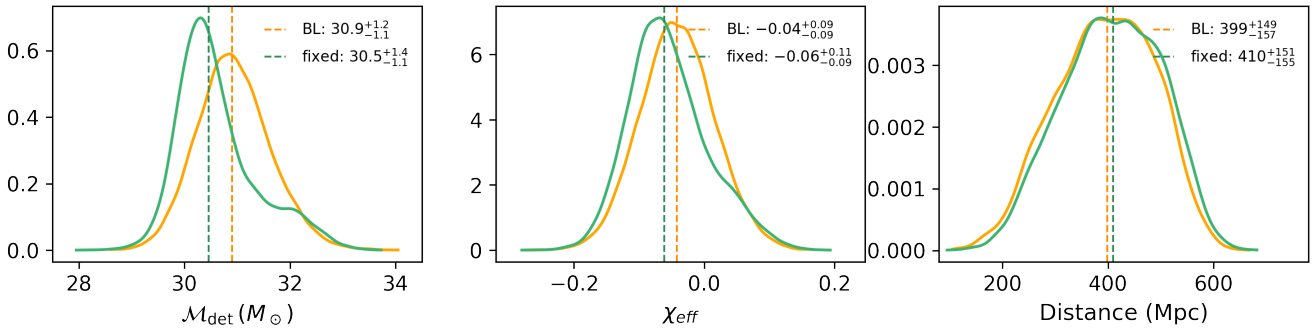

FIG. 3: GW159014 posterior distributions for $\mathcal{M}$, χ, and distance using BL (orange) and a fixed PSD (green). The medians are overplotted. Uncertainties are quoted at the 90% confidence level.

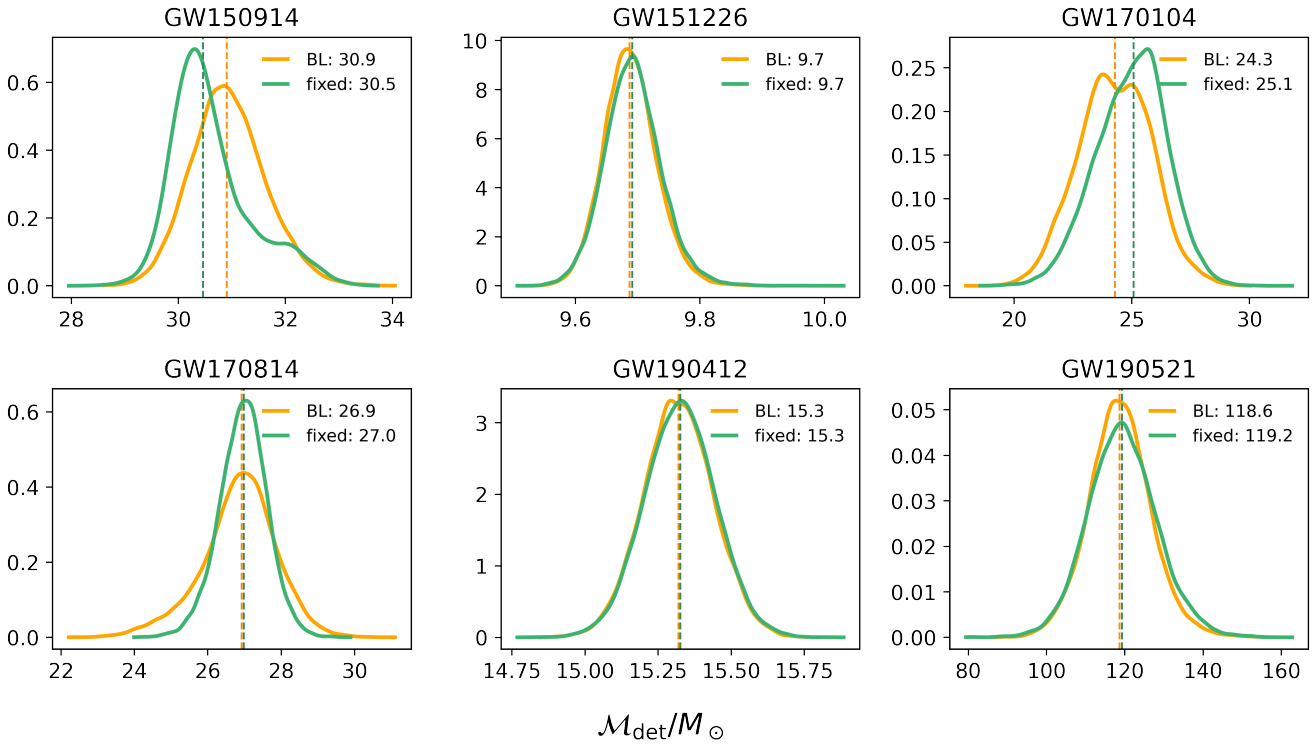

FIG. 4: Detector-frame chirp mass posterior comparison between BL and fixed PSD methods for real events.

from a distribution of a width that is impacted by the PSD height. The posterior that results from marginalizing over PSD parameters will be a collection of samples from posteriors with different standard deviations, which is comparable to a normal distribution integrated over a range of standard deviations. The difference between the posteriors that result from using a fixed PSD versus BL will be similar to the difference between a normal distribution and one integrated over a range:

$$\frac{1}{2\varepsilon}\int_{-\varepsilon}^{\varepsilon}\mathcal{N}\big(\theta,(1+x)\sigma\big)dx-\mathcal{N}(\theta,\sigma) \qquad (1)$$

Figure 6 plots Eqn. 1 for $\theta=0$, $\sigma=1$, and several values of ε. The difference between the two distributions peaks at their mean and goes as ε^4. Because the change in width is proportional to ε^4, for small PSD uncertainty, which is analogous to a small ε value, there will be little difference between the posteriors. As demonstrated in Figure 5, typical BL uncertainty at current sensitivities results in a relatively narrow range of posterior widths. The ε^4 approximation gives that the posterior marginalized over PSD uncertainty will agree with fixed PSD posterior in width to well within sampling error.

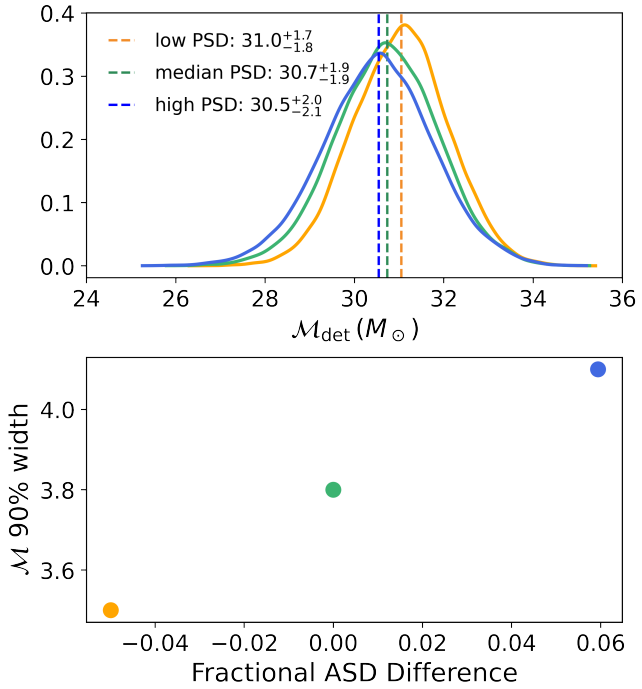

FIG. 5: *Top:* $\mathcal{M}$ posteriors for GW150914 using the median, low 90% CI, and high 90% CI PSDs from the BL output. *Bottom:* Fractional difference between ASD and the median at 100 Hz, vs. $\mathcal{M}$ 90% width.

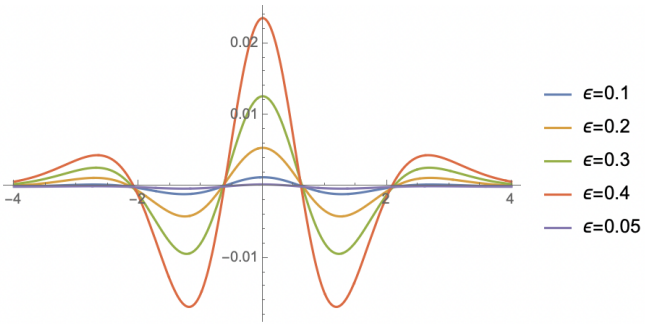

FIG. 6: Difference between $\mathcal{N}(0,1)$ and $\mathcal{N}(0,1+x)$ integrated over x, for various ranges ε.

IV. CONCLUSION

We study the impact of the known noise variance assumption in real-data GW PE by comparing the posteriors when noise properties are marginalized over to those when a point estimate is used for the noise. We find that the BL and fixed PSD methods recover nearly identical waveforms for both the signal and CBC models. For the CBC model analyses, the parameter posterior medians for chirp mass, distance, and effective spin agree to well within uncertainty, confirming the expectation that using a fixed PSD does not systematically bias the posterior. Moreover, the BL method does not produce wider posteriors; the 90% width is comparable to the fixed PSD method in all instances. This is attributable to the relatively low uncertainty in the PSD.

Our results show that current analyses are not underestimating error in binary parameters by not accounting for uncertainty in the PSD. At current sensitivities, PSD uncertainty is a subdominant effect in posterior estimation compared to other sources of uncertainty.

V. ACKNOWLEDGEMENTS

I would like to thank my mentors, Katerina Chatziioannou and Sophie Hourihane, for their guidance and support throughout this project. This research was made possible through the National Science Foundation REU Program, the LIGO Laboratory, and the Caltech LIGO SURF Program.

Appendix A: Likelihood Function

A detailed derivation of the GW PE likelihood function is given in Veitch *et al.* [3]. We summarize the function here to highlight where the assumptions about the noise appear. GW PE assumes that data from GW detectors, d, is the sum of a coherent signal h between detectors and uncorrelated random noise n: $d = h + n$. For an accurate signal model h', the residual data $r = d - h'$ should follow the same distribution as the noise. The likelihood function $\mathcal{L}(d|h')$ for GW PE, the probability density of measuring data d if the true signal is h', can be expressed in terms of these residuals. Assuming Gaussian noise allows for describing the residuals with a normal distribution. Further assuming stationarity over the time period of interest, the noise covariance reduces to a diagonal matrix in the frequency domain, which enables simplifying the likelihood function to

$$\ln \mathcal{L}(d|h') = -2\sum_{i}^{N/2} \frac{\tilde{r}_i \tilde{r}_i{}^*}{T S_n(f_i)} + \text{const},$$

which explicitly depends on $S_n(f)$. GW PE depends on its likelihood function, which in turn depends on the validity of these three assumptions, hence the recent work to measure and mitigate their effects.

[1] N. J. Cornish and T. B. Littenberg, Classical and Quantum Gravity **32**, 135012 (2015).

[2] L. Blanchet, Living Reviews in Relativity **17**, 2 (2014), arXiv:1310.1528 [gr-qc].

[3] J. Veitch, V. Raymond, B. Farr, W. Farr, P. Graff, S. Vitale, B. Aylott, K. Blackburn, N. Christensen, M. Coughlin, W. Del Pozzo, F. Feroz, J. Gair, C. J. Haster, V. Kalogera, T. Littenberg, I. Mandel, R. O'Shaughnessy, M. Pitkin, C. Rodriguez, C. Röver, T. Sidery, R. Smith, M. Van Der Sluys, A. Vecchio, W. Vousden, and L. Wade, Phys. Rev. D **91**, 042003 (2015), arXiv:1409.7215 [gr-qc].

[4] G. Ashton, M. Hübner, P. D. Lasky, C. Talbot, K. Ackley, S. Biscoveanu, Q. Chu, A. Divakarla, P. J. Easter, B. Goncharov, F. Hernandez Vivanco, J. Harms, M. E. Lower, G. D. Meadors, D. Melchor, E. Payne, M. D. Pitkin, J. Powell, N. Sarin, R. J. E. Smith, and E. Thrane, apjs **241**, 27 (2019), arXiv:1811.02042 [astro-ph.IM].

[5] K. Chatziioannou, C.-J. Haster, T. B. Littenberg, W. M. Farr, S. Ghonge, M. Millhouse, J. A. Clark, and N. Cornish, Physical Review D **100** (2019), 10.1103/physrevd.100.104004.

[6] T. B. Littenberg and N. J. Cornish, Physical Review D **91** (2015), 10.1103/physrevd.91.084034.

[7] N. J. Cornish, T. B. Littenberg, B. Bécsy, K. Chatziioannou, J. A. Clark, S. Ghonge, and M. Millhouse, Phys. Rev. D **103**, 044006 (2021), arXiv:2011.09494 [gr-qc].

[8] K. Chatziioannou, N. Cornish, M. Wijngaarden, and T. B. Littenberg, Phys. Rev. D **103**, 044013 (2021), arXiv:2101.01200 [gr-qc].

[9] R. Abbott, T. D. Abbott, and S. e. a. Abraham, arXiv e-prints , arXiv:2010.14527 (2020), arXiv:2010.14527 [gr-qc].

[10] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. J. Forteza, and A. Bohé, Phys. Rev. D **93**, 044007 (2016).

[11] N. J. Cornish, arXiv e-prints , arXiv:1007.4820 (2010), arXiv:1007.4820 [gr-qc].

[12] B. Abbott, R. Abbott, and T. e. a. Abbott, Physical Review X **9** (2019), 10.1103/physrevx.9.031040.

[13] B. P. Abbott and et al., Phys. Rev. D **93**, 122004 (2016), arXiv:1602.03843 [gr-qc].