Testing General Relativity with Gravitational Wave Signals using Hybrid Sampling

Noah E. Wolfe,[1] Jacob Golomb,[2] and Colm Talbot[2]

[1] *Department of Physics, North Carolina State University, Raleigh NC 27695*
[2] *LIGO Laboratory, California Institute of Technology, Pasadena, CA 91125, US*

(Dated: 24 September 2021)

The population of observed gravitational wave transients continues to grow, and with it, our ability to further constrain deviations from the signals predicted by general relativity. However, our current procedures for computing these constraints will not successfully scale with future transient catalogs. Thus, we leverage modern statistical methods into a new hybrid sampling method, in order to provide more efficient and more complete investigations of the parameter space of deviations from general relativity given gravitational wave observations of binary black hole mergers.

I. INTRODUCTION

General relativity is currently our most successful theory of gravity. However, modern developments in both theoretical and observational physics may provide hints that general relativity may not be a complete theory of gravity. For example, mathematical developments in string theory and quantum gravity have the potential to provide a unified theory of gravitation across length scales. Some of the most popular alternatives to general relativity that address theoretical and observational concerns include scalar-tensor theories [1] and dynamical Cherns-Simons gravity [2].

General relativity has been rigorously tested in the non-dynamical, weak-field regime, through experiments like the Gravity Probe B experiment and time-delay measurements with the Cassini space probe [3]. It has passed these tests with flying colors. Thus, the next gravitational frontier lies in the highly dynamical, strong-field regime of compact object mergers. LIGO has already begun tests in this regime, analyzing both single-events [4, 5] and the burgeoning population of gravitational wave transients [6–8], and to date has not yet identified deviations from general relativity.

The number of observed mergers will only continue to grow, further enhancing our resolution on key parameters describing the strong-field regime. However, this also necessitates that our statistical and computational techniques can support larger and more complex analyses. Current population analyses searching for non-GR effects like [8], based on the procedure of single-event analyses like [4], are approaching limits of reasonable computational efficiency.

Previous work with this approach includes tests with single events (a single BBH [4] and a single neutron star-neutron star merger [5]), as well as tests with observing run O1 [6], the first gravitational wave transient catalog (GWTC-1) [7], and most recently with GWTC-2 [8]. We will briefly focus on the last study, as we intend to improve upon this analysis in particular.

In Section V, subsection A of [8], the authors constrain a set of parameters $\delta\hat{\varphi}_i$ that denote deviations from coefficients in a phenomenological approximation to the waveform during the inspiral phase (see equation 4 of [8]). Using, alternatively, the `SEOBNRv4_ROM` and `IMRPhenomPv2` models, and `LALInference`, [8] varies each parameter $\hat{\varphi}_i$. From the modifications to the predicted waveforms with these variations, they construct posterior distributions on the parameterized violations of general relativity (see Figure 6 of [8]). Ultimately, they find no evidence for violations of general relativity.

However, there are two key limitations in the method employed by [8] that we seek to remedy. First, the method in [8] is computationally expensive. They begin with a set of 15 parameters from general relativity, with an additional variation parameter $\delta\hat{\phi}_i$. So, in total, they must infer a set of 16 parameters for each variation $\delta\hat{\phi}_i$, requiring many evaluations of the signal model in order to explore the parameter space of physically relevant $\delta\hat{\phi}$. Second, their method only varies one parameter $\hat{\varphi}_i$ at a time, due to the computational expensive involved in one parameter variation and an observed loss of information in posterior distributions with multiple parameter variation. Although there is reason to believe that variations in one parameter may be sufficient [8], there remains the possibility that violations of general relativity will only appear in multi-coefficient deviations from the PN expansion.

The goal of this work is to improve this analysis procedure with a new method for parameter estimation: hybrid sampling. This method is more computationally efficient, allowing us to scale our analysis as the population of observed mergers grows, and further constrain deviations in gravitational wave signals predicted by general relativity.

In this paper, we outline work completed in the LIGO SURF 2021 program. In Section II, we provide relevant background and description of the statistical methods used to implement hybrid sampling. In Section IV, we describe the physical context of astrophysical gravitational wave generation and the particular waveform approximation we use, as well as a physically-motivated method for constraining the priors on deviations from GR. In Section V, we describe initial results of gravitational wave parameter estimation with hybrid sampling, and in Section VI we summarize our current work and outline future research directions.

II. PARAMETER ESTIMATION METHODS

A. Bayesian Inference

For our description of Bayesian inference, we will begin by following the description and notation of [9]. In Bayesian inference, we consider a set of parameters θ in the context of the data d. For a physical situation of interest, we will have a model parameterized in terms of θ. For example, in this work, we will have a set of parameters which includes black hole binary properties (like mass and spin), with additional parameters to denote deviations from the predictions of general relativity. Our goal is to constrain these parameters given the data; statistically, we want to construct the posterior distribution $p(\theta|d)$, read as "the probability of getting a particular set of parameter values given observed data."

Bayesian inference allows us to construct the posterior distribution via Bayes' theorem, which relates $p(\theta|d)$ to our observations:

$$p(\theta|d) = \frac{\mathcal{L}(d|\theta)\pi(\theta)}{\mathcal{Z}} \tag{1}$$

where $\mathcal{L}(d|\theta)$ is known as the likelihood, $\pi(\theta)$ is the prior distribution. The normalization factor $\mathcal{Z}$ is known as the evidence, i.e. the probability of observing the data given the parameteric model we choose:

$$\mathcal{Z} \equiv \int d\theta \mathcal{L}(d|\theta)\pi(\theta) \tag{2}$$

The likelihood is a model for our observations under different parameters θ, and includes a model for noise; for gravitational wave observations, we most commonly use a likelihood that assumes Gaussian detector noise. The prior distribution is, effectively, an initial guess for the distribution of parameter values; when making this guess, we need to ensure that we do not bias the posterior distribution, and so may choose, for example, a uniform prior distribution.

We observe that $p(\theta|d)$ provides a distribution on the entire (multi-dimensional) set of parameters θ. To extract information on specific parameters of interest θ_i, we must "marginalize", i.e. integrate, over the rest of the parameters:

$$p(\theta_i|d) = \int \left(\prod_{k\neq i} d\theta_k\right) p(\theta|d) \tag{3}$$

This integration may be difficult to compute through standard numerical methods, especially if we have a high-dimensional parameter space. One common method is to use a Monte Carlo Markov Chain (MCMC) approach to approximate the posterior distribution. In this approach, a series of "walkers" explore the space of parameters θ in such a way that, given enough iterations, their paths will produce a representative sample of the posterior distribution.

Finally, in order to combine observations from multiple events that we suspect are part of a contiguous population (e.g. a population of binary black hole mergers) we will use a hierarchical analysis. In a hierarchical analysis, we assume that the parameters we infer given each observation are drawn from the same underlying distribution. For example, we will see that we assume parameterization of deviations from general relativity are common among all binary black hole mergers, and thus these deviations are drawn from the same posterior *hyper*parameter distribution. The mechanics of hierarchical inference rely on the same mechanics of standard Bayesian inference, wherein we seek to infer hyper-parameters given a population of observations through application of Bayes theorem and numerical techniques like MCMC.

B. Nested Sampling

For the first step of the hybrid sampling method, we use nested sampling. As developed in [10], the primary goal of nested sampling is to evaluate the evidence $\mathcal{Z}$ using an intelligent division of the prior mass.

Conceptually, when we take part of the prior mass over some domain of our parameters, there is an associated part of the likelihood surface associated to it via that same domain. We can parameterize the prior mass in terms of λ, a lower bound on the likelihood surface, which will limit the part of the parameter space, and thus prior mass that we accumulate. Mathematically,

$$X(\lambda) = \int_{\mathcal{L}(\theta)>\lambda} \pi(\theta) d\theta. \tag{4}$$

Therefore, our evidence from Equation (2) becomes

$$\mathcal{Z} = \int_0^1 \mathcal{L}(X) dX. \tag{5}$$

Numerically, we can approximate $\mathcal{Z}$ as a weighted sum of values $\mathcal{L}(X)$, e.g.

$$\mathcal{Z} \approx \sum_{i=1}^{N} w_i \mathcal{L}_i \tag{6}$$

for some number of samples N, with weights w_i that are chosen by some numerical integration rule, and thus dependent on how we divide up the domain of $\mathcal{L}(X)$, X. For example, if we were to use trapezoidal integration, $w_i = 1/2(X_{i-1} - X_{i+1})$. Then, the challenge of any implementation of nested sampling is to properly choose which points X_i, and associated L_i, to use in calculating $\mathcal{Z}$. A "crude" method introduced by [10] is to use $X_i = \exp(-i/N)$, and in effect, draw increasingly tight contours on the likelihood surface, as shown in Figure 1. More advanced implementations may perform sampling within these contour regions, e.g. with monte carlo techniques, but remain conceptually similar. For this work, we use the `dynesty` nested sampler [11].

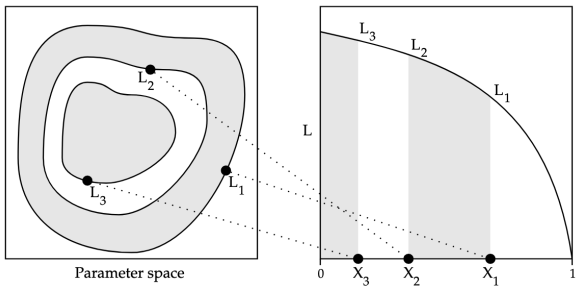

FIG. 1: The above figure is adapted from [10]. As we choose X_i, we associate each with a value of the likelihood $\mathcal{L}_i$ that defines a contour on the likelihood surface. (Note, Skilling uses L_i instead of $\mathcal{L}_i$).

Once the evidence has been calculated, we can trivially generate samples from the posterior distribution by resampling according to $w_i\mathcal{L}_i$ in Equation (6).

The key benefit of nested sampling, especially in the context of a hybrid approach, are to computationally efficient compression of the prior mass into information on the evidence, and thus the posterior. Additionally, nested sampling has been widely used in other LIGO parameter estimation activities, such as in analysis of GWTC-2 [12]. However, while nested sampling can provide ever increasing precision on the evidence by increasing N, the number of posterior samples generated is determined by the number of points as well as the shape of the likelihood surface (although methods to mitigate this have been suggested, e.g. dynamic nested sampling [11]).

C. Parallel-Tempered Ensemble Monte Carlo Markov Chains

In contrast with nested sampling, MCMC methods directly explore the posterior and can be run as long as necessary, continually generating additional samples from the posterior distribution. Ensemble MCMC methods build upon existing MCMC methods by replacing a single walker, as used in traditional approaches like the Metropolis-Hastings algorithm, with an ensemble of walkers that explore the parameter space in parallel [13]. A key feature of this approach is that we can reduce the number of samples we need to generate for our Bayesian inference methods, as at any single step the ensemble of walkers provides us a representation of the target posterior distribution.

Further, ensemble MCMC results can be parallel-tempered to explore isolated modes in the posterior distribution by raising the posterior to a "temperature" β_T, like

$$p_{\beta_T}(\theta|d) = \frac{\mathcal{L}^{\beta_T}(d|\theta)\pi(\theta)}{\mathcal{Z}^{\beta_T}}. \tag{7}$$

A parallel-tempered ensemble MCMC method then uses a number of walkers, in parallel, each exploring a tempered likelihood surface $\mathcal{L}^{\beta_T}$.

Finally, we note that MCMC methods can converge more quickly and accurately when given starting conditions that are close to the true values we are trying to estimate. Thus, we will use the results from another method, like nested sampling, to provide the initial conditions for hybrid sampling, under the assumption that they are near the desired values. As we next show in Section III B, this model is effective even when the model has been misspecified in the first hybrid sampling step. In this work, we use the `ptemcee` implementation of parlleltempered ensemble MCMC [13, 14].

D. Hybrid Sampling

We employ a hybrid sampling algorithm whereby nested sampling generates an initial solution to a simple problem and parallel-tempered ensemble MCMC is used to extend this to a more complex problem. Relative to one another, the simple problem may rely on a low-dimensional, non-degenerate, and well-understood model, while the complex case may have more dimensions, more degenercies, or be less well-understood. In the context of testing GR, the simple case assumes that GR is correct, and the more complex problem includes additional, degenerate, non-GR deviation parameters.

We have implemented this using the `dynesty` nested sampler and `ptemcee` parallel-tempered ensemble MCMC sampler. To properly initialize `ptemcee` on `dynesty` output, we need to create starting points for each tempered chain from the results of nested sampling.

In `dynesty`, a key part of these results are the "posterior weights" p_i that determine how each sample from the joint parameter distribution is weighted in constructing the final posterior. From Equation (36) in [10], the posterior weights are

$$p_i = \frac{\mathcal{L}_i w_i}{\mathcal{Z}}. \tag{8}$$

We construct tempered versions of these weights by raising the likelihood and evidence in p_i to a series of powers β_T; as in Equation (7),

$$p_{i,\beta_T} = \frac{\mathcal{L}_i^{\beta_T} w_i}{\mathcal{Z}^{\beta_T}}. \tag{9}$$

An example of these tempered posterior weights is shown in Figure 2; we note that $\beta_T = 1$ corresponds to the original posterior weights and in the limit $\beta_T = 0$ we recover the prior.

For each β_T, we then perform rejection sampling on the normalized weights $p_{i,\beta_T}/\sum_i p_{i,\beta_T}$ to obtain representative initial positions for parallel tempered walkers in `ptemcee`.

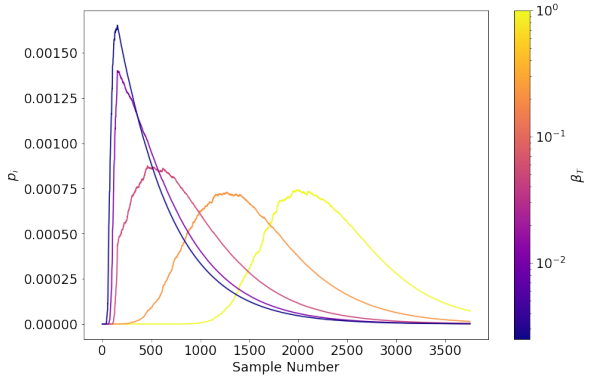

FIG. 2: Tempered posterior weights at five temperatures in the form of Equation (9), chosen by the `ptemcee` `make_ladder` method. Sample number is i in Equation (6) and Figure 1; increasing sample number corresponds to including a smaller region further up the likelihood surface, as pictorially described in Figure 1.

III. HYBRID SAMPLING WITH A GENERALIZED NORMAL DISTRIBUTION

To test our hybrid sampling framework, we consider a generalized normal distribution model with the probability density function

$$P(x) = \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-(|x-\mu|/\alpha)^\beta}. \tag{10}$$

Here α takes a role akin to the standard deviation σ in a normal distribution. We note that, in the limit of $\beta = 2$, this becomes a normal distribution (and we have $\alpha = \sqrt{2}\sigma$), and as $\beta \to \infty$, the generalized normal distribution approaches a tophat function.

For the first stage, we use a normal distribution as the "simple" model and sample with `dynesty`. Following this, we initialize `ptemcee` using the output of the nested sampling algorithm with initial values for β clustered about $\beta = 2$. Then, to check our results, we sample in the more complex model directly using `dynesty`.

A. Well-Specified Models

We demonstrate that this procedure could recover the parameters of normally distributed data, i.e. where both the first and second hybrid sampling steps had properly specified models.

Starting with the nested sampler `dynesty`, we infer the mean μ and standard deviation σ of a normal distribution generated with $\mu = 3$ and $\sigma = 4$. The results of this first step are shown in Figure 3.

Then, we tempered the nested sampling results and provided them as initializations for `ptemcee`, sampling in the more complex generalized normal distribution. The results of this analysis are shown in 4. From this figure, we achieve $\beta = 2$ for a normal distribution, as expected, and approximately recover $\mu = 3$ and $\alpha = \sqrt{(2)}\sigma = 4\sqrt{2} \approx 5.5$.

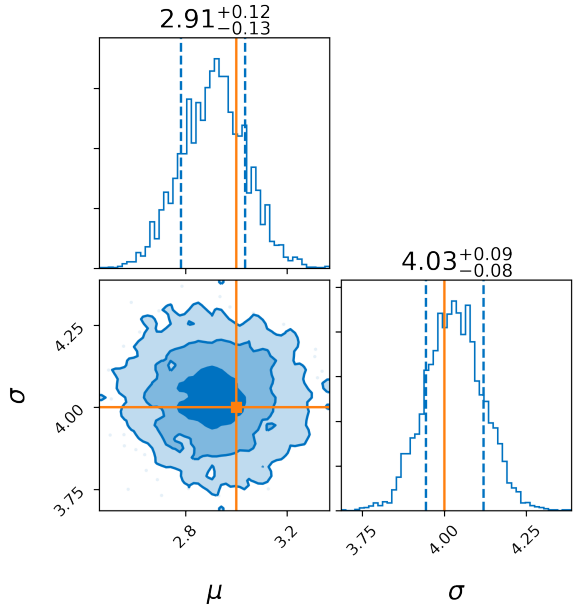

FIG. 3: Marginal and joint distributions on μ and σ inferred with `dynesty`.

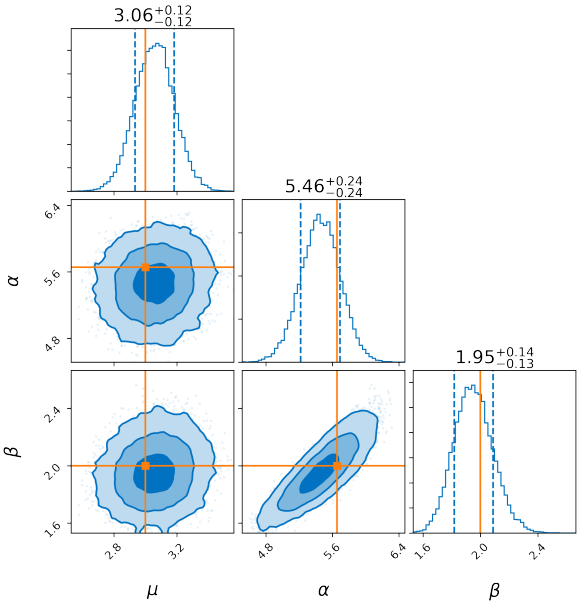

FIG. 4: Marginal and joint distributions on μ, α, and β inferred with `ptemcee`, using tempered `dynesty` input weights.

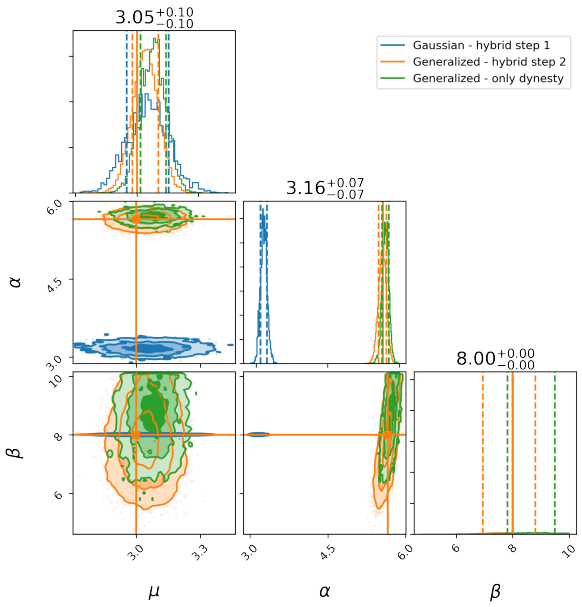

FIG. 5: Marginal and joint distributions on μ, α, and β inferred the hybrid sampling procedure (orange and blue), and with only `dynesty` (green).

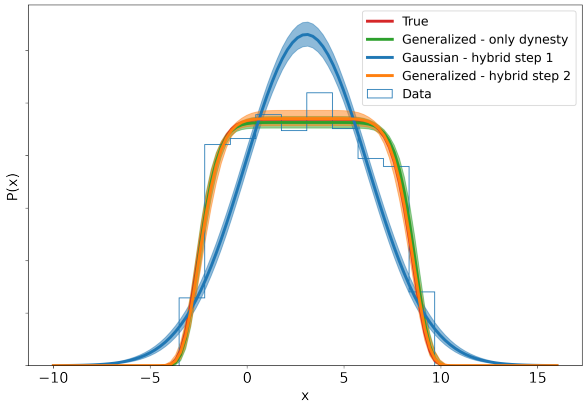

FIG. 6: Posterior predictive distributions, drawn from the distributions shown in Figure 5. Shaded regions are bounded between the 5th and 95th percentiles of the data, and lines are drawn at the mean.

B. Misspecified Models

Next, we demonstrate our method when the model is misspecified in the first step. We use use data drawn from a generalized normal distribution with $\beta = 8$, $\alpha = 4\sqrt{2}$, and $\mu = 3$ as input into the first step.

Hybrid sampling was able to recover these parameters despite assuming only a normal distribution in the first step, as shown in Figures 5 and 6. Also shown in these figures, we performed the same parameter estimation with only `dynesty` sampling in the generalized normal distribution, and achieved the same results, lending confidence to our hybrid sampling method.

IV. GRAVITATIONAL WAVE SIGNAL GENERATION

A. Anatomy of a Binary Black Hole Merger

A BBH merger can be described by 15 intrinsic parameters (such as masses and spins) and extrinsic parameters (characterizing location and orientation); As these are inferred from a GR waveform, we refer to these as "GR parameters". We can divide a BBH merger into three time-domains, in order: the *inspiral*, *merger*, and *ringdown* [8]. The inspiral phase begins when the black holes have formed a binary system. This phase is typically characterized by quasi-circular orbits, and lasts until weak-field approximations like the post-Newtonian expansion breakdown. The point at which inspiral approximations break down depends on the component masses in the system. Following the inspiral is the merger phase, which begins with a "plunge". In the plunge, the quasi-circular orbits suddenly become unstable, and the horizons merge. This phase is only possible to describe through numerical methods. Finally, the ringdown phase is the asymptotic relaxation of the combined black hole to a stable, isolated black hole state. This phase is well-described by analytical quasi-normal modes (QNMs).

Each regime is traditionally modeled through different mathematical and computational methods, owing to their differing physical time and length scales. The inspiral phase can be modeled analytically, through a parameterized post-Newtonian (PN) expansion of the gravitational potential. This formalism was introduced in its modern form by [15], and specified for gravitational wave emission from compact object mergers by [16] and [17]. As initially identified by [15], each term in the PN expansion can vary depending on the underlying theory of gravity and so is one of the key quantitative models for differentiating between alternative theories of gravity. The merger phase can only be fully understood through numerical relativity simulations, as it is described with nonlinear field equations in the chosen theory of gravity; for a recent example, see [18]. Quasi-normal modes can be understood analytically, and are specified by mode frequencies and damping times that depend on the assumed model of gravity [8, 19]. In this work, we focus on testing GR during the inspiral phase, as this provides a purely phenomenological framework without the need to specify a particular theory from the outset.

B. TaylorF2 Waveform Generation

To generate gravitational wave signals, we used the `TaylorF2` waveform approximant. This is a PN expansion of the gravitational wave strain in the frequency domain, broken into expansions of the amplitude and phase

of the signal, as

$$\tilde{h}(f) = A_{\rm PN} e^{-i\psi(f)} \tag{11}$$

where A is the amplitude and ψ is the phase; both may depend on intrinsic or extrinsic parameters of the BBH. The form of the amplitude expansion, in geometric units ($c = G = 1$) is

$$A_{\rm PN} = A_0 \sum_i \mathcal{A}_i (\pi f)^{i/3}. \tag{12}$$

Typically, for `TaylorF2`, only one term in the expansion is used. This yields, with physical units,

$$A_{\rm PN} = \sqrt{\frac{5}{24}\frac{\pi}{\eta}} \frac{m_1 m_2}{d_L} \left(\pi M f \frac{G}{c^3}\right)^{-7/6}, \tag{13}$$

where m_1, m_2 are the component masses of the BBH (by convention, $m_1 \geq m_2$), $M = m_1 + m_2$, $\eta = m_1 m_2 / M^2$ is the symmetric mass ratio, and d_L is the luminosity distance to the BBH.

The phase is calculated with an expansion of the form to 3.5PN order, adapted from [20],

$$\psi(f) = -\frac{\pi}{4} + \frac{3}{128\eta v^5} \sum_i^7 \varphi_i v^i, \tag{14}$$

where $v = \left(\pi M f G c^{-3}\right)^{1/3}$ (we nominally set the time, t_c, and phase, ϕ_c, of coalescence to zero). The coefficients φ_i can be found in Equation 3.18 of [20]; for example,

$$\varphi_1 = v^{-1}, \tag{15}$$

$$\varphi_2 = \frac{20}{9}\left(\frac{743}{336} + \frac{11}{4}\eta\right). \tag{16}$$

Although this approximant is already implemented in `PyCBC`, we manually implemented these waveforms in python, so that we could directly add non-GR corrections to the phase terms, effectively yielding

$$\psi(f) = -\frac{\pi}{4} + \frac{3}{128\eta v^5} \sum_i^7 \left(\varphi_i + \delta\varphi_i\right) v^i \tag{17}$$

with non-GR corrections $\delta\varphi_i$. In future, we could apply similar corrections to the amplitude terms $\mathcal{A}_i$, although LIGO is less sensitive to these deviations.

We verified our implementation of `TaylorF2` by checking that it replicates the amplitude and phase evolution of the implementation found in `PyCBC` [21].

C. Overlap Calculation

As described in Equation (17), the non-GR corrections $\delta\varphi_i$ can take any real value. However, we operate under the assumption that at the very least GR is nearly correct, and so there should be a limit on the magnitude of $\delta\varphi_i$. Therefore, we will ultimately impose a cut on the prior distributions for $\delta\varphi_i$ based on how "similar" a waveform with $\delta\varphi_i \neq 0$ is to a fully-GR waveform.

In practice, we implement this prior cut by calculating the magnitude of the complex overlap $\mathcal{O}$ between two frequency-domain waveforms, $\tilde{h}_1(f), \tilde{h}_2(f)$, and throwing out any waveforms generated during our sampling procedure where $\mathcal{O}$ is less than some limit. In our implementation, one waveform is the pure-GR waveform, and the other includes non-GR corrections. The overlap is calculated as

$$\mathcal{O} = \left| \frac{\left\langle \tilde{h}_1(f), \tilde{h}_2(f) \right\rangle}{\sqrt{\left\langle \tilde{h}_1(f), \tilde{h}_1(f) \right\rangle \left\langle \tilde{h}_2(f), \tilde{h}_2(f) \right\rangle}} \right|, \tag{18}$$

where we take the absolute value to maximize over an arbitrary phase shift; as a phase shift would correspond to a different merger time, our overlap calculation neglects time offsets when comparing GR and non-GR signals. In this calculation, $\langle \cdot, \cdot \rangle$ denotes an inner product of the form

$$\left\langle \tilde{h}_1(f), \tilde{h}_2(f) \right\rangle = 4\Delta f \sum_i^N \frac{\tilde{h}_{1,i} \tilde{h}^*_{2,i}}{S_i}, \tag{19}$$

where we sample each waveform, as well as a detector power spectral density S, at N discrete sampling frequencies, evenly spaced in frequency by Δf.

We performed an initial investigation of how the overlap varies with mass ratio q at fixed values of the total mass M, qualitatively identifying how a minimum overlap of $\mathcal{O} \geq 0.9$ would limit the parameter space. As shown in Figure 7, we observed that as the total mass M increases, so does the spread in $\mathcal{O}$ versus $\delta\varphi_2$, meaning that more values of $\delta\varphi_2$ would be allowed with a cutoff of $\mathcal{O} \geq 0.9$.

V. RESULTS

We demonstrate our hybrid sampling method using simulated gravitational wave signals in two cases; in the first, where there are no deviations from GR, and in the second, where there is a deviation from GR.

A. Recovery of GR Parameters

We first inject a purely-GR signal (with all $\delta\varphi_i = 0$), with the goal of recovering the signal parameters using hybrid sampling, in an effort to test both our sampling procedure and `TaylorF2` implementation. To inject a signal, we model the waveform with certain GR parameters and then simulate data as measured by the LIGO Hanford and Livingston detectors, including the power spectral density corresponding to the Advanced LIGO design sensitivity.

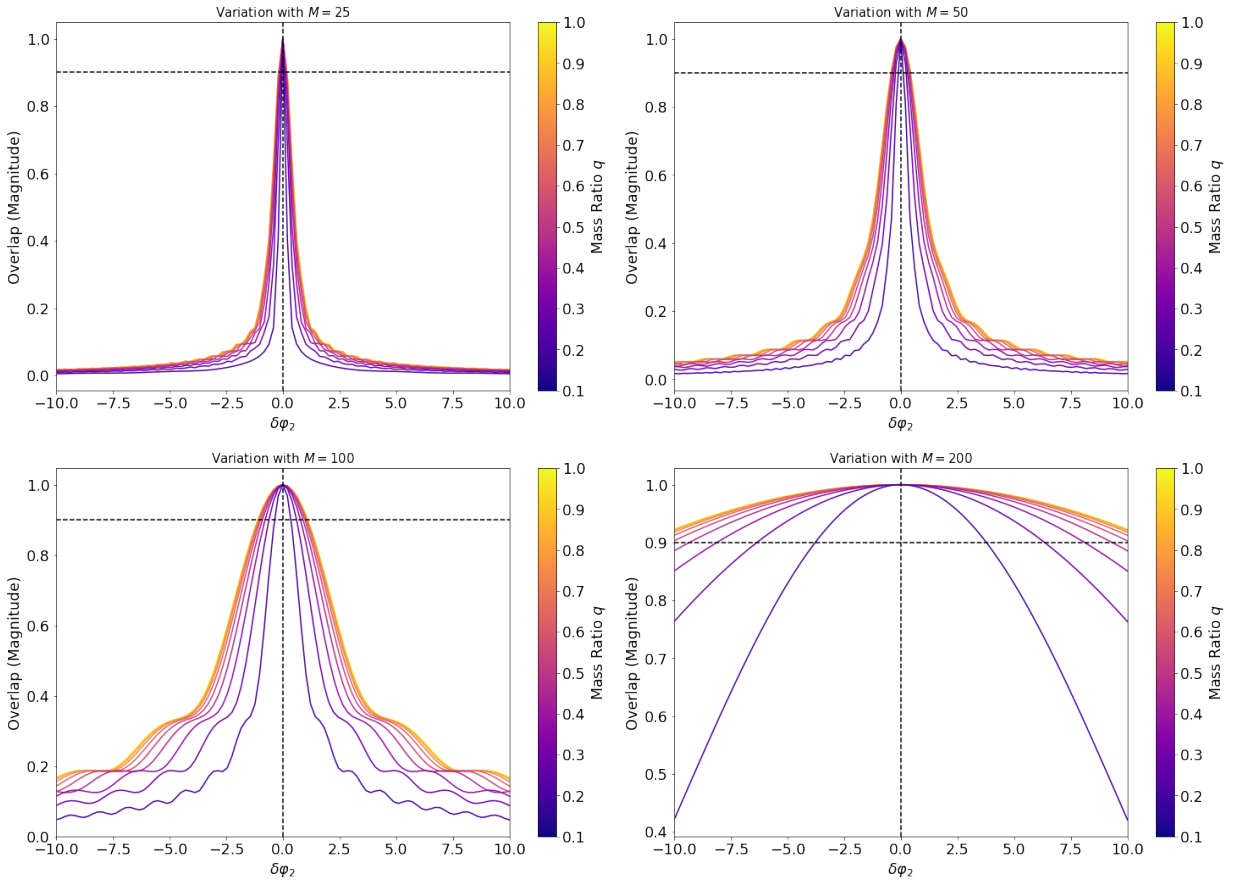

FIG. 7: Overlap with respect to variations in the mass ratio q, at four fixed total masses M. The horizontal dotted line shows the cutoff $\mathcal{O} = 0.9$, and the vertical dotted line shows $\delta\varphi_2 = 0$.

We injected a pure-GR signal with the following parameters: component masses $m_1 = 36M_\odot$, $m_2 = 29M_\odot$ (corresponding to a chirp mass $\mathcal{M} = 28.1M_\odot$ and mass ratio $q = 0.8$), dimensionless black hole spin parameters $a_1 = a_2 = 0$, inclination angles of 0, polarization angle $\psi = 2.659$, angle between the total angular momentum of the system and observer's line of sight $\theta_{jn} = 0.4$, geocenter time of $t_c = 1126259642.413$, right ascension and declination of 1.375 and -1.2108, and a luminosity distance of $d_L = 2000$ Mpc. We performed our initial nested sampling step by varying a subset of the GR parameters, in particular, $\mathcal{M}$, q, d_L, t_c (and marginalizaing over the phase of the signal). Then, in the second step, we also allow $\delta\varphi_2$ to vary, with the goal of recovering $\delta\varphi_2 = 0$. We also performed the full analysis, including variation of $\delta\varphi_2$, with `dynesty` to verify our results.

As shown in the top left panel of Figure 8, we are able to recover the injected waveform parameters of a purely-GR signal.

B. Recovery of non-GR Corrections

Then, we injected non-GR waveforms, with $\delta\varphi_1 = 10^{-5}$, $\mathcal{M} = 8M_\odot$, mass ratio $q = 1$, luminosity distance $d_L = 200$ Mpc, and marginalizing over the geocent time and the phase of the signal. We used an overlap cut on the priors, only allowing waveforms with $\mathcal{O} \geq 0.9$. Conceptually, we can understand this overlap cut as a prior assumption that GR is at least 90% correct in its description of the inspiral phase of BBH mergers.

At first, we found that the posterior distribution on $\delta\varphi_1$ returned the uniform prior, cut by the overlap limit; we effectively had no information on the value of $\delta\varphi_1$, as shown in the top right panel of Figure 8. This signal had a signal-to-noise-ratio (SNR) of 50. Under the hypothesis that this SNR may have been too low to resolve the posterior on $\delta\varphi_1$, we performed subsequent trials with increasing $\mathcal{M}$, at the same d_L, to raise the SNR. With $\mathcal{M} = 15M_\odot$ and $d_L = 200$ Mpc, the injected signal had an SNR of 80, and the resultant posteriors are shown in the bottom right panle of Figure 8. Although the posterior on $\delta\varphi_1$ is no longer strictly the prior, it is still uninformative. Finally, with $\mathcal{M} = 30M_\odot$ and $d_L = 200$ Mpc, the injected signal had an SNR of 140, and we found an informative posterior distribution on $\delta\varphi_1$, as shown in the bottom left panel of Figure 8. Thus, we observe that if we assume GR is correct to a 90% level, we cannot distinguish between GR and non-GR models for the inspiral phase of BBH mergers with an SNR less than 100.

VI. CONCLUSIONS AND FUTURE WORK

In this work, we implemented a hybrid sampling method using nested sampling to solve a simple parameter estimation problem, and then subsequently parallel-tempered ensemble MCMC to solve a more complex parameter estimation problem. We applied this to a toy model using the generalized normal distribution, demonstrating that hybrid sampling is able to return the same results as the well-tested `dynesty` sampler used on its own. In particular, we demonstrated that hybrid sampling performs well with this toy model, even when the model in the first step is misspecified.

Then, we used hybrid sampling to perform parameter estimation on gravitational wave signals generated with the `TaylorF2` waveform approximant. We demonstrated that hybrid sampling is able to infer the a limited set of GR waveform parameters, as well as the non-GR variation parameter $\delta\varphi_1$ accurately, in both pure and non-GR waveforms (i.e. $\delta\varphi_1 = 0$ and $\delta\varphi_1 \neq 0$ cases). However, we observed that in order to return informative posterior distributions on $\delta\varphi_1$, we need to inject a signal with SNR greater than 100.

In our hybrid sampling model, we only need the results of the first step, assuming GR is correct, once; then, from these results, we can conduct many non-GR analyses, varying different sets of deviation parameters. In the future, we will repeat the same analysis on the full set of 15 GR parameters and one non-GR variation parameter; if this analysis is successfully, we will again repeat the analysis with different sets of multiple simultaneous non-GR variations. In addition, we will further investigate the SNR required to gain information on $\delta\varphi_i$, with the aim of identifying lower limits on the signal properties required to identify deviations from GR in the context of current and next-generation gravitational wave detectors.

VII. ACKNOWLEDGEMENTS

I would like to thank my mentors, Dr. Colm Talbot and Jacob Golomb, for their support, encouragement, and guidance; they helped me learn and grow as a scientist throughout the summer. I would like to thank Dr. Alan Weinstein and the rest of the LIGO Laboratory for additional input and guidance. I am grateful to the LIGO Laboratory and Caltech for organizing the LIGO SURF program. I want to acknowledge the support of the National Science Foundation (NSF); LIGO is funded by the NSF and this research was supported by the NSF REU program. Finally, I am grateful for my fellow LIGO SURF participants, especially Celine Wang.

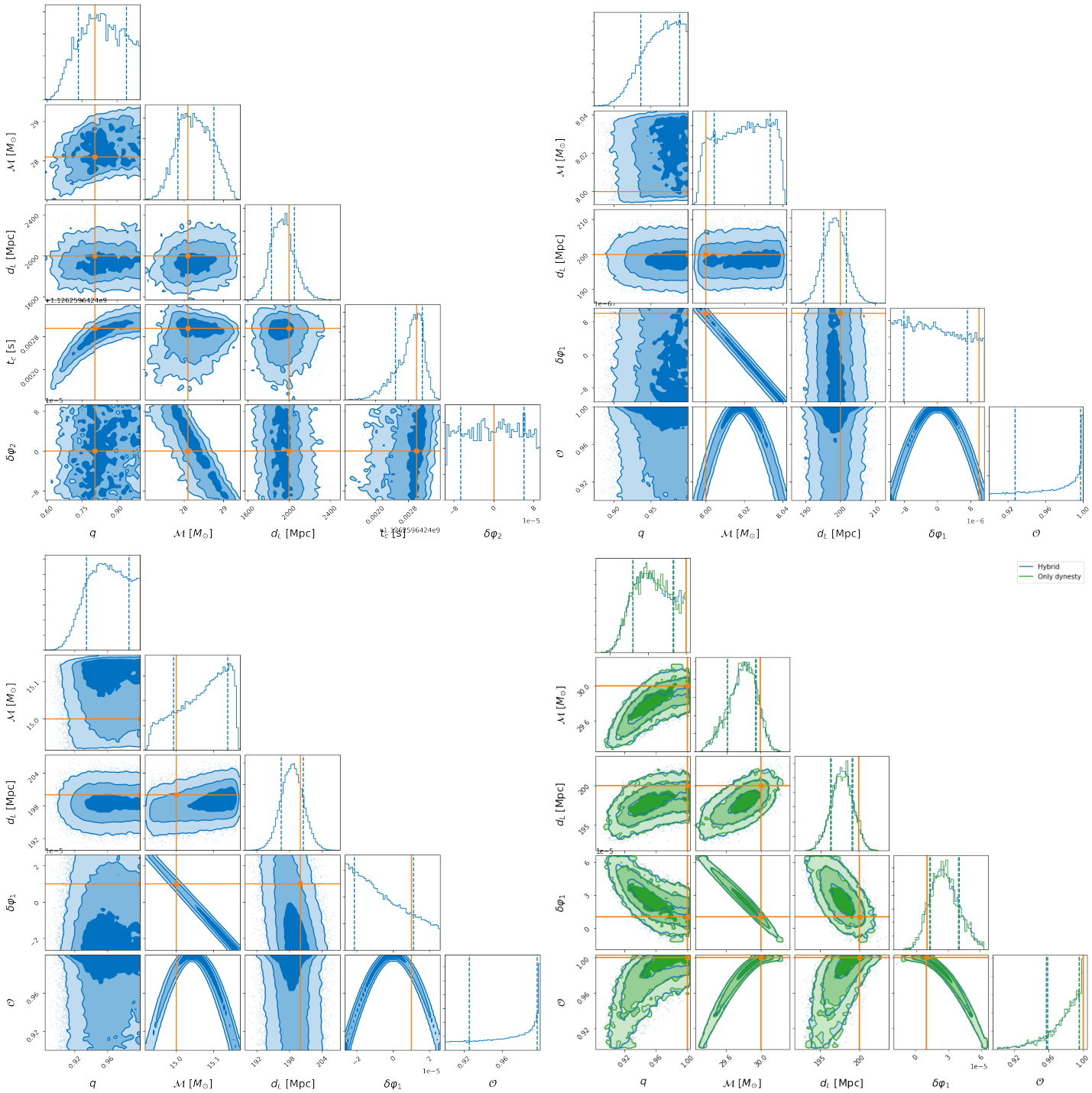

FIG. 8: Posteriors generated via hybrid sampling for parameters of pure-GR and non-GR waveforms. The lines and points in orange denote injected waveform parameters. Clockwise, starting from top left, these plots are for: a pure-GR waveform with injected $\delta\varphi_2 = 0$, a non-GR waveform with an SNR of 50 with injected $\delta\varphi_1 = 10^{-5}$, a non-GR waveform with an SNR of 80 with injected $\delta\varphi_1 = 10^{-5}$, and a non-GR waveform with an SNR of 140 with injected $\delta\varphi_1 = 10^{-5}$. We estimated the geocenter time t_c for a pure-GR waveform in the first panel, and performed time marginalization in the non-GR cases to speed up parameter estimation. We note that our distributions on non-GR parameters are uninformative until we study the signal with the highest SNR, in the last panel at the bottom left. In this panel, we also overlay the same parameter estimation performed with the `dynesty` sampler alone, shown in green, and recover the same distributions.

[1] Israel Quiros. Selected topics in scalar-tensor theories and beyond. *International Journal of Modern Physics D*, 28(7):1930012-156, January 2019. doi: 10.1142/S021827181930012X.

[2] Stephon Alexander and Nicolás Yunes. Chern-Simons modified general relativity. , 480(1-2):1–55, August 2009. doi:10.1016/j.physrep.2009.07.002.

[3] Clifford M. Will. The Confrontation between General Relativity and Experiment. *Living Reviews in Relativity*, 17(1):4, December 2014. doi:10.12942/lrr-2014-4.

[4] The LIGO Scientific Collaboration, the Virgo Collaboration, R. Abbott, et al. Tests of general relativity with gw150914. *Phys. Rev. Lett.*, 116:221101, May 2016. doi: 10.1103/PhysRevLett.116.221101. URL https://link.aps.org/doi/10.1103/PhysRevLett.116.221101.

[5] The LIGO Scientific Collaboration, the Virgo Collaboration, R. Abbott, et al. Tests of general relativity with gw170817. *Phys. Rev. Lett.*, 123:011102, Jul 2019. doi: 10.1103/PhysRevLett.123.011102. URL https://link.aps.org/doi/10.1103/PhysRevLett.123.011102.

[6] The LIGO Scientific Collaboration, the Virgo Collaboration, R. Abbott, et al. Binary black hole mergers in the first advanced ligo observing run. *Phys. Rev. X*, 6:041015, Oct 2016. doi:10.1103/PhysRevX.6.041015. URL https://link.aps.org/doi/10.1103/PhysRevX.6.041015.

[7] The LIGO Scientific Collaboration, the Virgo Collaboration, R. Abbott, et al. Tests of general relativity with the binary black hole signals from the ligo-virgo catalog gwtc-1. *Phys. Rev. D*, 100:104036, Nov 2019. doi:10.1103/PhysRevD.100.104036. URL https://link.aps.org/doi/10.1103/PhysRevD.100.104036.

[8] The LIGO Scientific Collaboration, the Virgo Collaboration, R. Abbott, et al. Tests of General Relativity with Binary Black Holes from the second LIGO-Virgo Gravitational-Wave Transient Catalog. *arXiv e-prints*, art. arXiv:2010.14529, October 2020.

[9] Eric Thrane and Colm Talbot. An introduction to Bayesian inference in gravitational-wave astronomy: Parameter estimation, model selection, and hierarchical models. , 36:e010, March 2019. doi:10.1017/pasa.2019.2.

[10] John Skilling. Nested sampling for general Bayesian computation. *Bayesian Analysis*, 1(4):833 – 859, 2006. doi: 10.1214/06-BA127. URL https://doi.org/10.1214/06-BA127.

[11] Joshua S. Speagle. DYNESTY: a dynamic nested sampling package for estimating Bayesian posteriors and evidences. , 493(3):3132–3158, April 2020. doi: 10.1093/mnras/staa278.

[12] The LIGO Scientific Collaboration, the Virgo Collaboration, R. Abbott, et al. GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run. *arXiv e-prints*, art. arXiv:2010.14527, October 2020.

[13] Daniel Foreman-Mackey, David W. Hogg, Dustin Lang, and Jonathan Goodman. emcee: The MCMC Hammer. , 125(925):306, March 2013. doi:10.1086/670067.

[14] W. D. Vousden, W. M. Farr, and I. Mandel. Dynamic temperature selection for parallel tempering in markov chain monte carlo simulations. *Monthly Notices of the Royal Astronomical Society*, 455(2):1919–1937, Nov 2015. ISSN 1365-2966. doi:10.1093/mnras/stv2422. URL http://dx.doi.org/10.1093/mnras/stv2422.

[15] Clifford M. Will. Theoretical Frameworks for Testing Relativistic Gravity. II. Parametrized Post-Newtonian Hydrodynamics, and the Nordtvedt Effect. *Astrophys. J.*, 163:611, February 1971. doi:10.1086/150804.

[16] Chandra Kant Mishra, K. G. Arun, Bala R. Iyer, and B. S. Sathyaprakash. Parametrized tests of post-Newtonian theory using Advanced LIGO and Einstein Telescope. *Phys. Rev. D*, 82(6):064010, September 2010. doi:10.1103/PhysRevD.82.064010.

[17] T. G. F. Li, W. Del Pozzo, S. Vitale, C. Van Den Broeck, M. Agathos, J. Veitch, K. Grover, T. Sidery, R. Sturani, and A. Vecchio. Towards a generic test of the strong field dynamics of general relativity using compact binary coalescence. *Phys. Rev. D*, 85(8):082003, April 2012. doi: 10.1103/PhysRevD.85.082003.

[18] Maria Okounkova, Leo C. Stein, Mark A. Scheel, and Daniel A. Hemberger. Numerical binary black hole mergers in dynamical Chern-Simons gravity: Scalar field. *Phys. Rev. D*, 96(4):044020, 2017. doi: 10.1103/PhysRevD.96.044020.

[19] Emanuele Berti, Vitor Cardoso, and Andrei O. Starinets. TOPICAL REVIEW: Quasinormal modes of black holes and black branes. *Classical and Quantum Gravity*, 26(16):163001, August 2009. doi:10.1088/0264-9381/26/16/163001.

[20] Alessandra Buonanno, Bala R. Iyer, Evan Ochsner, Yi Pan, and B. S. Sathyaprakash. Comparison of post-Newtonian templates for compact binary inspiral signals in gravitational-wave detectors. *Phys. Rev. D*, 80(8):084043, October 2009. doi: 10.1103/PhysRevD.80.084043.

[21] Alex Nitz, Ian Harry, Duncan Brown, Christopher M. Biwer, Josh Willis, Tito Dal Canton, Collin Capano, Thomas Dent, Larne Pekowsky, Andrew R. Williamson, Gareth S Cabourn Davies, Soumi De, Miriam Cabero, Bernd Machenschalk, Prayush Kumar, Duncan Macleod, Steven Reyes, dfinstad, Francesco Pannarale, Thomas Massinger, Sumit Kumar, Márton Tápai, Leo Singer, Sebastian Khan, Stephen Fairhurst, Alex Nielsen, Shashwat Singh, Koustav Chandra, shasvath, and Bhooshan Uday Varsha Gadre. gwastro/pycbc:, August 2021. URL https://doi.org/10.5281/zenodo.5347736.