

# Accuracy limitations of existing numerical relativity waveforms on the data analysis of current and future ground-based detectors

Aasim Jan<sup>1</sup>,<sup>✉</sup> Deborah Ferguson<sup>1,2</sup>,<sup>✉</sup> Jacob Lange,<sup>1</sup> Deirdre Shoemaker<sup>1</sup>,<sup>✉</sup> and Aaron Zimmerman<sup>1</sup>

<sup>1</sup>*Center for Gravitational Physics, The University of Texas at Austin, Austin, Texas 78712, USA*

<sup>2</sup>*Illinois Center for Advanced Studies of the Universe, The University of Illinois Urbana-Champaign, Urbana, Illinois 61801, USA*



(Received 21 December 2023; accepted 30 May 2024; published 9 July 2024)

As gravitational wave detectors improve in sensitivity, signal-to-noise ratios of compact binary coalescences will dramatically increase, reaching values in the hundreds and potentially thousands. Such strong signals offer both exciting scientific opportunities and pose formidable challenges to the template waveforms used for interpretation. Current waveform models are informed by calibrating or fitting to numerical relativity waveforms and such strong signals may unveil computational errors in generating these waveforms. In this paper, we isolate a single source of computational error, that of the finite grid resolution, and investigate its impact on parameter estimation for aLIGO and Cosmic Explorer. We demonstrate that increasing the inclination angle or decreasing the mass ratio  $q$  ( $q \leq 1$ ) raises the resolution required for unbiased parameter estimation. We quantify the error associated with the highest-resolution waveform utilized in our study using an extrapolation procedure on the median of recovered posteriors and confirm the accuracy of current waveforms for the synthetic sources. We introduce a measure to predict the necessary numerical resolution for unbiased parameter estimation and use it to predict that current waveforms are suitable for equal and moderately unequal mass binaries for both detectors. However, current waveforms fail to meet accuracy requirements for high signal-to-noise ratio signals from highly unequal mass ratio binaries ( $q \lesssim 1/6$ ), for all inclinations in Cosmic Explorer, and for high inclinations in future updates to LIGO. Given that the resolution requirement becomes more stringent with more unequal mass ratios, current waveforms may lack the necessary accuracy, even at median signal-to-noise ratios for future detectors.

DOI: [10.1103/PhysRevD.110.024023](https://doi.org/10.1103/PhysRevD.110.024023)

## I. INTRODUCTION

The detection of gravitational waves (GW) [1–4] by the Advanced Laser Interferometer Gravitational-wave Observatory (LIGO) [5] and Advanced Virgo [6] has resulted in an explosive growth in the field of GW astronomy and this growth is expected to continue due to plans for substantial upgrades to current detectors [7,8] and the development of next-generation ground-based [9–13] and space-based [14–16] detectors over the next two decades. These advancements will enable the detection of GW events at significantly higher signal-to-noise ratios (SNRs) than observed to date. Detection of such strong signals not only promises unprecedented science but also raises a question: Are current waveforms accurate enough to yield unbiased parameter estimation (PE) of GW sources? Increased sensitivities necessitate a stricter accuracy requirement on model waveforms and the numerical relativity (NR) waveforms used in their construction. The increase in the accuracy requirement stems from the fact that as the SNR of observed signals increases, the statistical uncertainty in the recovered parameters decreases, amplifying the impact

of underlying systematic uncertainties. As such, to fully extract the wealth of information contained in such strong signals, it is imperative that the waveforms meet these enhanced accuracy requirements.

A coalescing compact binary system is described by Einstein’s field equations, which can be solved using NR [17–19], particularly for the late inspiral, merger, and ringdown. Despite being the most accurate way of obtaining a gravitational waveform of merger, NR waveforms are typically not used directly for GW data analysis as they are computationally expensive to generate, are short in length, and have sparse and uneven parameter space coverage. While some studies have used NR waveforms directly for PE [20,21] including a LIGO-Virgo-KAGRA analysis of GW150914 [22], these analyses are limited by the number of available simulations. As such, to create comprehensive inspiral-merger-ringdown waveform models that cover a continuous parameter space, various modeling strategies have been developed. These include: effective-one-body formalism [23–26], phenomenological modeling [27–29], and NR surrogates [30]. Each of these modeling strategies relies on NR waveforms. Effective-one-body models use

NR waveforms to calibrate the free parameters, phenomenological models use hybridized NR and post-Newtonian waveforms [31] for a multiparameter fit and NR surrogates generate interpolants based on NR waveforms. These waveform models play a critical role in the search and interpretation of GW events, and given that their accuracy is bounded by the accuracy of NR waveforms, it is imperative that the NR waveforms used in their construction meet the required level of accuracy.

Both NR and model waveforms have inherent uncertainties that can affect the interpretation of GW signals. One of the sources of systematic uncertainty in NR waveforms is inaccuracy from finite grid resolution. The uncertainties in model waveforms arise from the approximations made during their construction, omission of certain physics and their reliance on finite-resolution NR waveforms. However, it is worth noting that it has been shown that currently available NR waveforms demonstrate an adequate level of accuracy for the signals detected to date [32].

While multiple studies [33–37] have been done to investigate the impact of model waveform systematics on PE, an area that remains relatively unexplored is the effect of bias introduced by finite-resolution NR waveforms on full PE. In a previous study [20], one-dimensional PE of an NR injection was conducted to study the impact of numerical resolutions on the final posterior distribution, with the recovery being done using NR waveforms. Apart from being restricted to a single dimension, the study was carried out at an SNR of 25, limiting the applicability of such a study to future, more sensitive detectors. In another study [32], a criterion was proposed that relates the minimum resolution required for producing an NR waveform that is indistinguishable [38] from a true signal to the SNR. This criterion is conservative, as indistinguishability itself is an unnecessarily strict requirement [34,39] and significant parameter bias may not arise even for technically distinguishable waveforms. Consequently, its application can become impractical, especially in scenarios where higher modes play a crucial role as the criterion sets a considerably higher resolution requirement than actually needed, placing a burden on computational resources.

In this work, we study the impact of using finite-resolution NR waveforms on PE by performing PE on synthetic signals (injections) produced by NR simulations run at multiple resolutions. We start by comparing the PE outputs of equal mass ratio binary injections differing only in resolution, and investigate the impact of NR truncation errors on the recovered binary parameters. We demonstrate that using a resolution higher than what is required for unbiased PE does not yield additional information for PE. We confirm that the resolution requirement based on waveform indistinguishability [40] is stricter than required for unbiased PE. We show the accuracy requirement for NR waveforms is dependent on the detector, as the impact of truncation errors varies according to the sensitivity curve of

the detector. We show that for systems with a greater higher mode content in their GWs, higher resolution is needed to achieve unbiased PE than for those without. Moreover, we introduce and employ an extrapolation procedure to estimate errors associated with the highest resolution waveforms used in our work. Finally, we utilize our results, combined with a waveform criterion, to make predictions for future NR codes, predicting the minimum resolution necessary for unbiased PE of signals expected to be observed by upcoming detectors. It should be noted that the results of this study apply to NR codes that use finite differencing methods to solve Einstein’s equations.

The rest of the paper is organized as follows: in Sec. II we review the NR waveforms used in the study, the indistinguishability criterion for NR waveforms generated using finite-differencing codes, and the PE code RIFT. In Sec. III we study the impact of numerical biases on PE across a variety of systems and for two detectors, LIGO Hanford (H1) and the proposed third generation detector Cosmic Explorer (CE) [12]. In Sec. IV, we employ an extrapolation procedure to determine bias in PE due to the highest resolution waveforms used in this study. In Sec. V, we predict the SNR at which we will see significant biases and compare them with what we obtained from full PE. We also assess the accuracy of current waveforms and determine if they are sufficiently accurate for current and future detectors. In Sec. VI we summarize our results.

## II. METHODS

In this section, we introduce the methods we use to investigate the effect of NR truncation errors on PE and assess the accuracy limitations of existing NR waveforms in the context of PE. We start by discussing the NR waveforms used as injections in our PE studies in Sec. II A. We then discuss the indistinguishability criterion for NR waveforms, as proposed in [32], in Sec. II B. We then review the PE algorithm RIFT in Sec. II C.

### A. Numerical relativity waveforms

For our PE investigations, we use NR waveforms generated using the MAYA code [41–44] as injections. MAYA is a branch of the Einstein Toolkit, and is built upon the Cactus framework, incorporating Carpet [45] for mesh refinement. It employs the BSSN [46] formulation to derive the initial constraints and evolution equations from Einstein’s field equations. In its calculations, MAYA utilizes sixth-order spatial finite-differencing and fourth-order Runge-Kutta for time evolution.

Our NR waveform injections were extracted from two sets of quasicircular MAYA NR simulations. Each had identical initial conditions and parameters, differing only in grid resolution. These simulations are parametrized, up to an arbitrary total mass  $M$  (in code units), by the intrinsic parameters  $\lambda$  of the binary. These intrinsic parameters

TABLE I. Synthetic sources: Parameters of the three sets of synthetic sources used in our study. For these sources, we have chosen right ascension  $\text{ra} = 0.57$ , polarization angle  $\psi = 0$ , declination  $\text{dec} = 0.1$ , and coalescence phase  $\phi_c = 0$  (all in radians). The GPS time at the geocenter was set to  $10^9$ s. Each set of injections was recovered at multiple SNRs, achieved by changing the luminosity distance  $D_L$ .

| $q$ | $m_1/M_\odot$ | $m_2/M_\odot$ | $\mathcal{M}_c/M_\odot$ | $\chi_1$        | $\chi_2$        | $\iota$ (radians) | $\Delta$                     |
|-----|---------------|---------------|-------------------------|-----------------|-----------------|-------------------|------------------------------|
| 1   | 50            | 50            | 43.5                    | (0.0, 0.0, 0.6) | (0.0, 0.0, 0.6) | 0                 | $M/80, M/120, M/140, M/200$  |
| 1   | 50            | 50            | 43.5                    | (0.0, 0.0, 0.6) | (0.0, 0.0, 0.6) | $\pi/6$           | $M/80, M/120, M/140, M/200$  |
| 1/3 | 112.5         | 37.5          | 54.9                    | (0.0, 0.0, 0.0) | (0.0, 0.0, 0.0) | 0                 | $M/100, M/120, M/140, M/180$ |

include primary mass  $m_1$ , secondary mass  $m_2$ , and spin vectors  $\mathbf{S}_1$  and  $\mathbf{S}_2$ . Since the simulations are scale-invariant with respect to total mass, it is conventional to parametrize them in terms of their mass ratio  $q = m_2/m_1 \leq 1$ . Further, we define the dimensionless spin parameters  $\chi_1 = \mathbf{S}_1/m_1^2$  and  $\chi_2 = \mathbf{S}_2/m_2^2$ .

The first set of simulations had  $q = 1$  and spins aligned with the orbital angular momentum, with  $\chi_{1z} = \chi_{2z} = 0.6$ . The other set was nonspinning, with  $q = 1/3$ . These parameters were computed at the beginning of the simulation, but there is evidence that nonphysical junk radiation does not significantly affect their values [47].

The initial separations and grid spacing or resolution of our simulations are expressed in terms of  $M$ . For the  $q = 1$  systems, the initial separation was  $12M$  while for the  $q = 1/3$  systems, the initial separation was  $9M$ . Both sets consist of four differently resolved simulations. The simulations were performed on a grid with 10 refinement levels with the largest grid radii being  $409.6M$  and the smallest being  $0.2M$ . The resolutions are specified by the grid spacing  $\Delta_i$  of the finest grid in each simulation and varied from  $\Delta_i = M/200$  ( $M/180$ ) for our highest resolution  $q = 1$  ( $q = 1/3$ ) simulation to  $\Delta_i = M/80$  ( $M/100$ ) for our lowest. The lowest resolution in each set is lower than typically used for GW data analysis and was chosen to illustrate the impact of resolution at moderate SNRs. The typical resolution of  $q = 1$  and  $q = 1/3$  simulations in the MAYA catalog [48] is  $M/200$  and  $M/370$  respectively.

The waveforms were computed from the Weyl scalar  $\Psi_4$  extracted at a finite radius of  $75M$  from the binary system. We avoided extrapolating the waveforms to infinity in order to isolate the impact of finite resolution.  $\Psi_4$  is related to the gravitational wave polarizations as follows:

$$\Psi_4(t) = \ddot{h}_+(t) - i\ddot{h}_\times(t). \quad (1)$$

Further decomposition of GW polarizations involves expressing them as a sum of spherical harmonics  $_{-2}Y_{lm}$  and GW modes  $h_{lm}$ , given by

$$h_+(t) - ih_\times(t) = \sum_{l,m} h_{lm}(t) {}_{-2}Y_{lm}. \quad (2)$$

In our injections, we only used  $\ell \leq 4$  modes.

In order to compute the GW strain  $h(t)$  as measured at a detector, we must specify the detector frame total mass  $M_{\text{tot}}$  so that the component masses and dimensionful scales are determined. Additionally, the extrinsic parameters of the binary must be defined. These extrinsic parameters determine the space-time location and orientation relative to the detector and include luminosity distance  $D_L$ , right ascension  $\text{ra}$ , declination  $\text{dec}$ , polarization  $\psi$ , inclination  $\iota$ , orbital phase  $\phi_c$ , and coalescence time  $t_c$ . The intrinsic and extrinsic parameters chosen for our NR waveforms, along with the grid spacing used in the finest grid, are provided in Table I. There and throughout this study we quote the detector-frame masses for the system. Two sets of these waveforms are assumed to be observed face-on, with  $\iota = 0$ , while the third has a modest inclination  $\iota = \pi/6$ . Regardless of the detector sensitivity used in this study, we assume the detector is located at the site of the Hanford GW detector and aligned with it. Figure 1 depicts the time-domain strains imprinted on the detector for the face-on cases. These waveforms are aligned at their peaks, and we can see a clear dephasing with differing grid resolutions.

## B. Criterion for assessing waveform accuracy

A gravitational waveform  $h_i$  extracted from an NR simulation can be expressed in terms of the exact solution  $h$  as  $h_i = h + \delta h_i$ , where  $\delta h_i$  is the error in the extracted waveform. If the code used to generate the waveform uses finite-differencing to solve partial differential equations, then the truncation error can be expressed as  $\delta h_i = c(\Delta_i)^\alpha$ , where  $\alpha$  is the convergence rate of the code and  $c$  depends on the derivatives of  $h$ . Recall that since Carpet uses adaptive mesh refinement, different sections of the grid have different spacings, and  $\Delta_i$  refers to the spacing of the finest grid.

Given two waveforms  $h_1$  and  $h_2$ , the overlap between them is defined as

$$\mathcal{O}[h_1, h_2] = \frac{\langle h_1 | h_2 \rangle}{\sqrt{\langle h_1 | h_1 \rangle \langle h_2 | h_2 \rangle}}, \quad (3)$$

where the noise-weighted inner product is defined as

$$\langle h_1 | h_2 \rangle = 4\text{Re} \int_{f_{\min}}^{f_{\max}} \frac{h_1^*(f) h_2(f)}{S_n(f)} df, \quad (4)$$

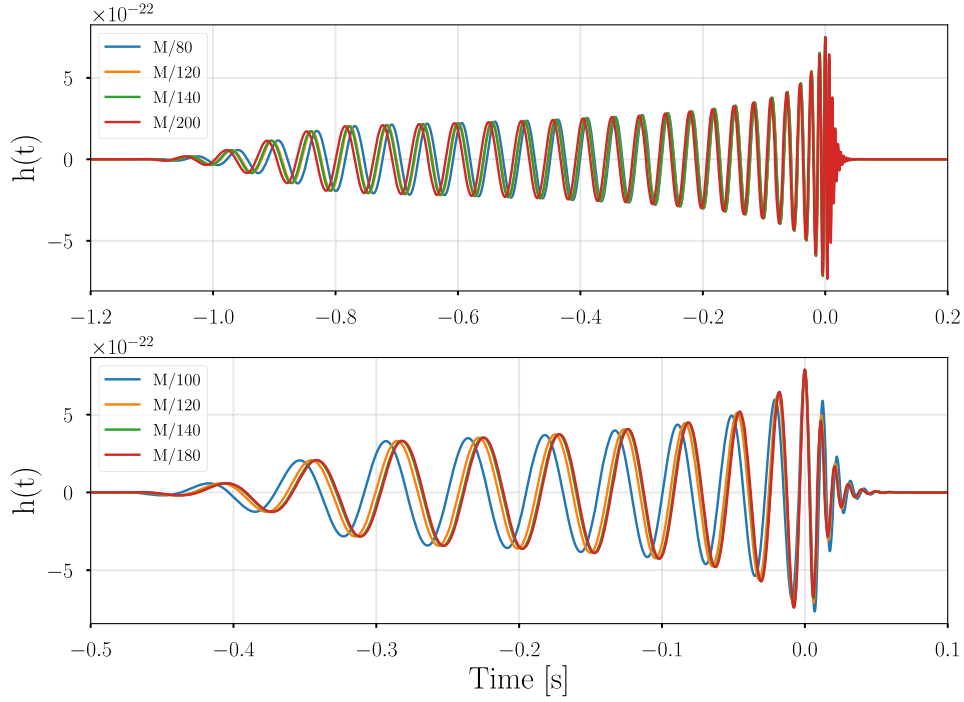


FIG. 1. Comparison between the different resolutions of  $q = 1$  and  $q = 1/3$  waveforms: Top: strain evaluated from the four distinct resolution  $q = 1$  waveforms for a source with  $M_{\text{tot}} = 100M_{\odot}$ ,  $\iota = 0$ . Bottom: strain evaluated from the four different resolution  $q = 1/3$  waveforms for a source with  $M_{\text{tot}} = 150M_{\odot}$ ,  $\iota = 0$ . In both panels, the waveforms are aligned at the peak, evaluated at  $D_L = 1000$  Mpc and the colors represent different resolutions.

with  $S_n(f)$  being the one-sided power spectral density of the detector,  $f_{\text{min}}$  a low-frequency cutoff,  $f_{\text{max}}$  a high-frequency cutoff, and  $*$  denoting the complex conjugate. Here, we have used the fact that  $h_i(t)$  is a real-time series, and as such its Fourier transform satisfies  $h_i(f) = h_i^*(-f)$ , allowing us to define the inner product as an integral over positive frequencies.

Using the overlap, we can define mismatch  $\epsilon[h_1, h_2]$ , which several previous investigations have argued relates to systematic biases in PE [34,38,40,49–52], as

$$\epsilon[h_1, h_2] = 1 - \max_{t_c, \phi_c} \mathcal{O}[h_1, h_2]. \quad (5)$$

We calculate mismatch between two NR waveforms, both having identical parameters but differing in the numerical resolution of simulation grid. We use this mismatch to compute  $\beta$  which is a parameter that only depends on the parameters of the binary system and can be computed as

$$\epsilon[h_1, h_2] = \frac{\beta^2}{2} (\Delta_2^\alpha - \Delta_1^\alpha)^2. \quad (6)$$

Using  $\beta$ , and the convergence rate  $\alpha$  of the finite-differencing code, we can then estimate the minimum resolution necessary for producing NR waveforms indistinguishable from a true signal using the criterion [32]

$$\Delta < (\rho\beta)^{-1/\alpha}, \quad (7)$$

where  $\rho$  is the SNR. Ideally,  $\alpha$  and  $\beta$  would be independent of the resolution of the simulations used for their calculation and this would be true if the waveforms were extracted from simulations that consisted of a single grid, used the same order for all finite-differencing, and the strain was a grid variable. However, given the complicated mesh refinements and boundary conditions as well as the interpolation necessary to extract  $\Psi_4$  and compute strain, we have a less well-defined convergence order. This means the values of  $\alpha$  and  $\beta$  differ slightly between different pairs of resolutions. This introduces an uncertainty of up to  $10M^{-1}$  in our estimates of the minimum resolution necessary for indistinguishability.

### C. RIFT

A GW from a quasicircular binary black hole system undergoing merger can be completely determined by 15 parameters. These parameters are classified into the two groups, *intrinsic* ( $\lambda$ ) and *extrinsic* ( $\theta$ ), described previously. In discussing the PE results, we also use chirp mass  $\mathcal{M}_c = (m_1 m_2)^{3/5} / (m_1 + m_2)^{1/5}$ , which is paired with the mass ratio  $q$  to represent the mass parameters. We also define another parameter called effective spin, which is a



mass-weighted combination of individual spins and is defined as

$$\chi_{\text{eff}} = \frac{m_1 \chi_{1z} + m_2 \chi_{2z}}{m_1 + m_2}. \quad (8)$$

After a GW is detected, the data  $d$  is analyzed to infer the parameters of the radiating system using a PE algorithm, and RIFT [53] is one such algorithm. It is a highly parallelizable, grid-based, iterative algorithm consisting of two core iterative stages. In the first stage, a “grid” of intrinsic parameters is put forth, and for each point  $\lambda_\alpha$  from the proposed grid, RIFT integrates over the extrinsic variables to compute the marginal likelihood

$$\mathcal{L}_{\text{marg}}(d|\lambda) \equiv \int \mathcal{L}_{\text{full}}(d|\lambda, \theta) p(\theta) d\theta \quad (9)$$

from the likelihood  $\mathcal{L}_{\text{full}}(d|\lambda, \theta)$  of the GW signal, accounting for detector response. Here,  $p(\theta)$  are the priors on the extrinsic parameters. The integration is made possible by factorizing the dependence of  $\mathcal{L}_{\text{full}}$  on the extrinsic parameters, which is partially made possible by expressing the GW polarizations in terms of GW  $\ell, m$  modes; see [54] for a more detailed specification.

Once the marginalized likelihood is evaluated for points on the grid  $(\lambda_\alpha, \mathcal{L}_\alpha)$ , RIFT interpolates this discrete grid of marginalized likelihood points to generate the continuous likelihood distribution  $\mathcal{L}_{\text{marg}}(d|\lambda)$ . With the knowledge of the continuous marginalized likelihood distribution  $\mathcal{L}_{\text{marg}}(d|\lambda)$  and intrinsic prior  $p(\lambda)$ , RIFT constructs the marginalized posterior via Bayes’ theorem:

$$p_{\text{post}}(\lambda|d) = \frac{\mathcal{L}_{\text{marg}}(d|\lambda) p(\lambda)}{\int d\lambda \mathcal{L}_{\text{marg}}(d|\lambda) p(\lambda)}. \quad (10)$$

The integral in the denominator is calculated by performing a Monte Carlo integral: the evaluation points and weights in that integral are weighted posterior samples, which are fairly resampled to generate conventional independent, identically distributed posterior samples.

The grid for the following iteration is generated using a subset of posterior samples from the previous iteration, with an additional expansion of the grid to ensure that regions of high likelihood that might have been missed can be explored. The iterations continue until two successive iterations converge, as determined by examining the Jensen-Shannon (JS) divergence [55] between the one-dimensional marginal posteriors for each intrinsic parameter. In our study, we have taken measures to guarantee convergence by ensuring that the JS divergence between the last two iterations is less than or equal to  $10^{-3}$  for all parameters. Further details and a justification of this choice is given in Appendix A. For further details on RIFT’s

technical underpinnings, performance, and comparison to other PE codes see [53,54,56–58].

### III. PARAMETER ESTIMATION RESULTS

In this section, we present the results of our PE study. Our goal is to investigate the impact of finite-resolution errors in NR waveforms on PE recovery. Ideally, to achieve this goal, we would inject an infinite-resolution NR waveform followed by recovery using identically resolved NR waveforms as template waveforms. This process would be iterated multiple times, with each iteration utilizing template waveforms obtained from NR simulations of differing numerical resolution. Ultimately, we would compare the results of each iteration and assess the impact of truncation errors on the posterior distributions across a range of SNRs. While the grid-based structure of RIFT allows it to perform PE using NR waveforms [20], the available catalog of NR waveforms are all of varying resolution, preventing us from isolating the effect of finite-resolution. As such we must take a different approach, which we describe in Sec. III A. We then describe the setup of our PE study in Sec. III B, and present results for our three sets of NR injections (Table I) in Secs. III C–III E.

#### A. Justification for PE approach

Given the impracticality of utilizing NR waveforms as our model waveform for PE, we approach the problem differently. We consider a sequence  $h_i$  of NR simulations with varying resolution but fixed binary parameters  $\mu_*$ , which we inject into a zero-noise realization for use as our synthetic data  $d_i$ . Next we use a high-accuracy waveform model, NRHybSur3dq8 [ $h_{\text{sur}}(\mu)$ ] for parameter recovery, one which can be evaluated at arbitrary parameters  $\mu$ . The result is a sequence of posteriors  $p(\mu|d_i, h_{\text{sur}})$  under our model hypothesis. We repeat this procedure with a variety of SNR values for our injected waveform, and compare the resulting posteriors across resolution values in order to understand the effect of finite resolution on PE results.

Figure 2 illustrates how we can use these results to explore this question when we recover our posteriors with the model  $h_{\text{sur}}$ . The figure illustrates the space of possible signals. All signals that can be realized by the model  $h_{\text{sur}}$  when evaluated over the relevant range of parameter values  $\mu$  forms a submanifold  $\Sigma$  of the overall signal space. The numerical simulations  $h_i$  do not lie on  $\Sigma$ , but using the match as a distance measure allows us to identify points  $\mu_i$  such that  $h_{\text{sur}}(\mu_i)$  is the best match to  $h_i$ . The distance of each  $h_i$  away from  $h_{\text{sur}}(\mu_i)$  is a measure of the SNR loss due to mismodeling the NR waveform with  $h_{\text{sur}}$ . The difference between the best fit  $\mu_i$  value and  $\mu_*$  is the parameter bias that we are interested in. The sequence of NR waveforms converges to some extrapolated waveform  $h_\infty$ , which differs from  $h_{\text{sur}}(\mu_*)$ .

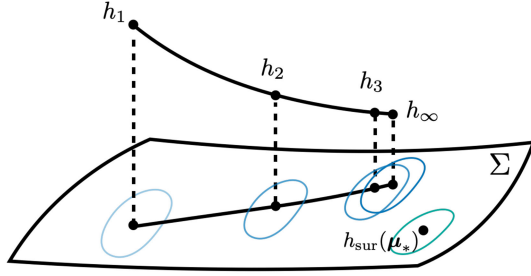


FIG. 2. Illustration of our data analysis strategy, visualizing the waveforms as points in signal space. We perform PE on a sequence of NR waveforms  $h_i$  at fixed binary parameters  $\mu_*$  and recover with a waveform model  $h_{\text{sur}}$ , which produces waveforms on a submanifold  $\Sigma$ . As we increase the resolution, the waveforms converge to an extrapolated signal  $h_\infty$ , and the sequence of posteriors also converges; the manner in which these posteriors converge as a function of injected SNRs allows us to determine resolution requirements for unbiased PE. Meanwhile,  $h_{\text{sur}}(\mu_*)$  is a separate signal which lies off this sequence in general.

From the perspective of PE, the best fit points  $\mu_i$  provide the maximum likelihood values for a zero-noise recovery with the waveform model  $h_{\text{sur}}$ .  $h_{\text{sur}}(\mu_i)$  provides the relevant maximum likelihood waveform. Meanwhile, the shape of the posteriors is determined primarily as if  $h_{\text{sur}}(\mu_i)$  were the injected data, modulo the effects of SNR loss due to the projection of  $h_i$  onto  $\Sigma$ . Thus we can study the sequence of recovered posteriors as we approach  $h_\infty$  and the corresponding recovered posteriors, and ask at what SNR and for which pairs of resolutions the PE results are indistinguishable. This allows us to state what approximate resolution is required at a given SNR for unbiased parameter recovery. Potentially the most important complicating factor relative to an idealized study is the variable SNR loss as we move along the sequence  $h_i$ ; if this SNR loss is not too severe, this approach provides the desired information about the PE bias from finite resolution.

### B. Setup of PE study

To investigate the impact of finite-resolution on PE, we first injected four different resolutions of  $q = 1$  gravitational waveforms (refer to Table I for all injected parameters) at a range of SNRs. Instead of directly selecting SNR values, we first selected eight different resolutions and then found the corresponding critical SNR using Eq. (7). Going forward, when referring to critical SNR and resolution, we mean the SNR and resolution values at which the inequality transitions to equality. The choice for resolution levels was made strategically such that for the first four SNR values, two would correspond to cases where  $M/80$  waveform is predicted to be indistinguishable from true signal and two to cases where it is predicted to be distinguishable from true signal. We applied the same approach for  $M/120$  to select the remaining four SNR values. The selected resolutions and corresponding critical SNRs are given in the first two

TABLE II. Normalized bias in the marginalized  $\mathcal{M}_c$  posterior distributions of  $q = 1$ ,  $\iota = 0$  injections: Bias observed in the lower resolution posteriors, calculated with respect to  $M/200$ , for H1 (top) and CE (bottom). Normalized bias were calculated using Eq. (11).

| SNR | $\Delta_{\text{critical}}$ | $M/80$ | $M/120$ | $M/140$ |
|-----|----------------------------|--------|---------|---------|
| 9   | $M/70$                     | -0.28  | -0.04   | -0.02   |
| 12  | $M/75$                     | -0.39  | -0.07   | -0.03   |
| 20  | $M/85$                     | -0.66  | -0.12   | -0.06   |
| 31  | $M/95$                     | -1.09  | -0.17   | -0.09   |
| 47  | $M/105$                    | -1.70  | -0.27   | -0.14   |
| 67  | $M/115$                    | -3.10  | -0.49   | -0.22   |
| 94  | $M/125$                    | -7.89  | -0.92   | -0.25   |
| 128 | $M/135$                    | -9.59  | -1.22   | -0.78   |

| SNR | $\Delta_{\text{critical}}$ | $M/80$ | $M/120$ | $M/140$ |
|-----|----------------------------|--------|---------|---------|
| 5   | $M/70$                     | -0.10  | -0.02   | -0.01   |
| 7   | $M/75$                     | -0.28  | -0.05   | -0.03   |
| 11  | $M/85$                     | -0.55  | -0.11   | -0.05   |
| 16  | $M/95$                     | -0.89  | -0.17   | -0.08   |
| 22  | $M/105$                    | -1.21  | -0.24   | -0.09   |
| 31  | $M/115$                    | -1.74  | -0.32   | -0.11   |
| 44  | $M/125$                    | -2.53  | -0.52   | -0.22   |
| 60  | $M/135$                    | -3.58  | -0.69   | -0.28   |

columns of Table II. Subsequently, we analyzed and compared the resulting posterior distributions after recovery with the hybridized NR surrogate waveform model NRHybSur3dq8 [59] using all available  $\ell \leq 4$  modes, and the RIFT PE code, quantifying the impact of truncation errors as a function of SNR.

It has been shown that waveforms involving significant contributions from higher modes [32], require a higher resolution to satisfy accuracy requirements in comparison to waveforms that lack significant contributions from higher modes. As such we also injected the same  $q = 1$  NR waveforms but at  $\iota = \pi/6$ . At this inclination, the detected GW is shaped not just by the modes observed at  $\iota = 0$  but also by  $(2, -2)$  and  $(4, 4)$ . To quantify the impact of truncation errors as a function of SNR, the SNR selection was done the same way as for the face-on case. For both of these cases, we use  $M_{\text{tot}} = 100M_\odot$  to ensure the majority of the injected waveform falls in band.

We also extended our study to include  $q = 1/3$  waveform injections which were generated at  $M_{\text{tot}} = 150M_\odot$  and  $\iota = 0$ . We adjusted our SNR selection approach slightly, opting for six SNRs instead of eight. This selection involved choosing six resolution levels in such a way that, for the first three levels, two corresponded to scenarios where  $M/100$  is predicted to be indistinguishable, and one where it is predicted to be distinguishable. We applied the same methodology to determine the next three SNRs, this time focusing on  $M/120$ . Since the  $q = 1/3$  injections were generated at a different total mass, to ensure the

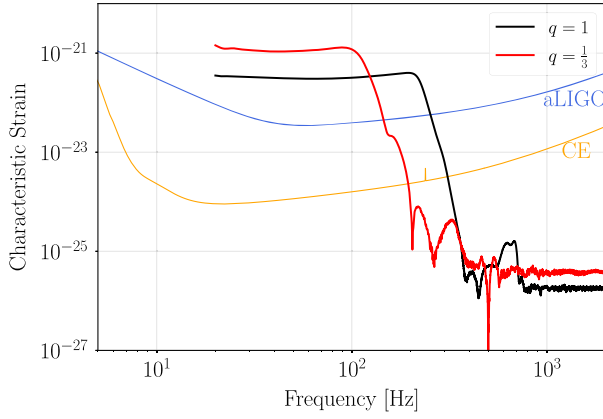


FIG. 3. Characteristic strains and noise amplitudes: Characteristic strain evaluated from the highest resolution  $q = 1$  (black) and  $q = 1/3$  waveform (red). Black curve is from a source with parameters  $M_{\text{tot}} = 100M_{\odot}$ ,  $\iota = 0$ , SNR = 12, and red curve is from a source with parameters  $M_{\text{tot}} = 150M_{\odot}$ ,  $\iota = 0$ , SNR = 37. Also plotted are the noise amplitudes for design aLIGO and CE. The extrinsic parameters for both waveforms are as given in Table I.

injected waveform starts in band, it would be unreasonable to directly compare the results of  $q = 1$  and  $q = 1/3$  injections. To enable a direct comparison, the same total mass must be injected for both  $q = 1$  and  $q = 1/3$  cases. In such a comparison, we would expect that in the  $q = 1/3$  case, the same resolution introduces a significantly higher parameter bias than in the  $q = 1$  case, at a given SNR. To emphasize this point, we injected a set of  $q = 1$  waveforms at  $M_{\text{tot}} = 150M_{\odot}$ , detailed in Sec. III E.

Furthermore, it is important to note that parameter biases are also influenced by the shape of the noise power spectral density (PSD) of the detector. Therefore, our study encompassed two detectors, namely design aLIGO [60,61] and CE. Figure 3 depicts the characteristic strain [62] for the face-on  $q = 1$  and  $q = 1/3$  waveforms, alongside the noise amplitude spectral densities for comparison.

For each set of NR injections, differing only in resolution and recovered at a sequence of SNRs, we kept our settings the same so any differences in the posterior can be attributed to the difference in NR resolution. However, despite our best efforts, there are other errors that can potentially impact the final posterior distribution, with some of them being Monte Carlo integration errors when evaluating marginalized likelihood and the fit to the finite grid of intrinsic parameters.

For our PE runs, in the first iterative step of RIFT, where the marginalized likelihood is evaluated, we marginalized over all the extrinsic parameters, except for  $\text{ra}$  and  $\text{dec}$  which remained fixed for all runs. Thus we marginalized over  $t_c$ ,  $\phi_c$ ,  $\psi$ ,  $D_L$ , and  $\iota$  with the likelihood integration starting from 20 Hz and ending at 2048 Hz, for both H1 and CE. We chose to maintain the same frequency range, for both H1 and CE, to ensure a direct comparison of PE

results. In the second iterative step, we approximated the likelihood using random forests, and the Monte Carlo sampling was carried out using Gaussian mixture models. To ensure our choice of likelihood approximation and sampling method used within RIFT does not impact our final results, we reanalyzed the highest SNR injection for each synthetic source using Gaussian process regression as the approximation method and adaptive cartesian as the sampling method (refer to [20,58] for details on the different approximation and sampling method implemented in RIFT). We found our results remained consistent regardless of our choice of likelihood approximation and sampling method.

To quantify the impact of truncation errors on PE, we calculate normalized bias for all parameters. The normalized bias for a parameter  $\lambda$  is calculated as:

$$\text{Normalized bias} = \Delta\tilde{\lambda}/\sigma \quad (11)$$

where  $\tilde{\lambda}$  is the median of the marginalized  $\lambda$  posterior distribution,  $\Delta\tilde{\lambda}$  is the shift in the median of marginalized  $\lambda$  posterior distribution relative to the highest resolution posterior and  $\sigma$  is the standard deviation of the highest resolution posterior. In Fig. 4, we plot normalized bias in multiple parameters when the  $q = 1$ ,  $\iota = 0$ ,  $\Delta = M/80$  waveform is used as injection. We observe that  $\mathcal{M}_c$  is the first parameter to display an absolute normalized bias of unity, which we will consider as a significant parameter bias in our work. Therefore, moving forward, we place primary emphasis on  $\mathcal{M}_c$ .

### C. $q = 1$

For the  $q = 1$  system, we generated injections from each of the four differently resolved waveforms at eight different SNRs. We started from SNR 9, where  $M/70$  is predicted to

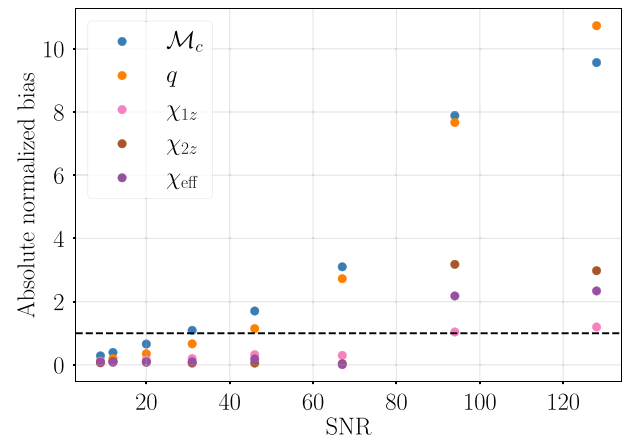


FIG. 4. Absolute normalized bias in multiple parameters: Plot showing absolute normalized bias in 1D marginal posteriors for selected source parameters when  $q = 1$ ,  $\Delta = M/80$  is used as an injection. The dashed horizontal line marks an absolute normalized bias of unity.

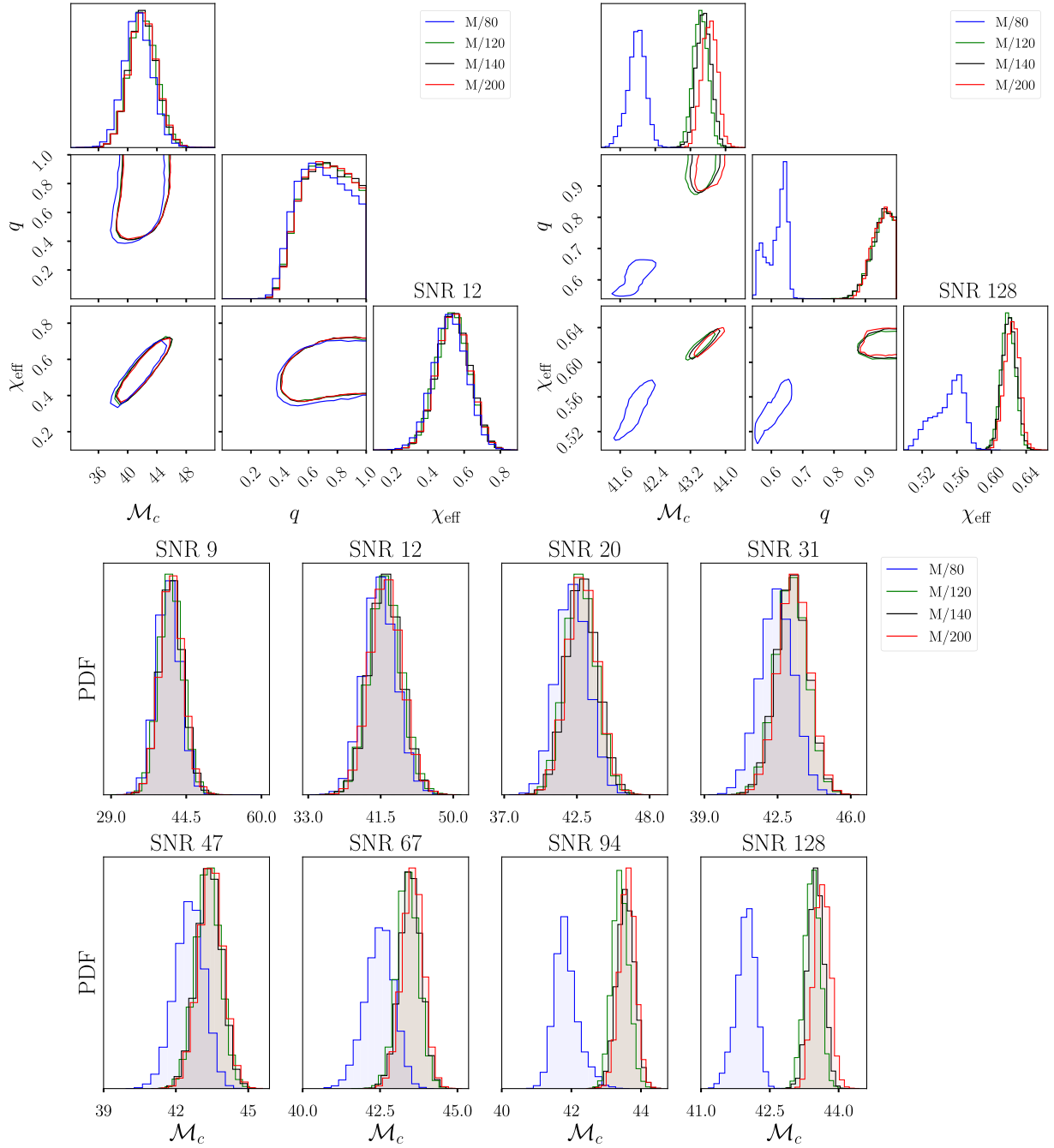


FIG. 5. PE results for  $q = 1, t = 0$  injections (H1): Top left: one- and two-dimensional marginal posterior distributions for  $\mathcal{M}_c$ ,  $q$ , and  $\chi_{\text{eff}}$ . Diagonal panels show the one-dimensional marginal posterior distribution, while contours in the off-diagonal panels show the 90% credible intervals for the two-dimensional marginal posterior distribution. Different colored curves correspond to different resolutions. Injections had an SNR of 12 and the minimum resolution for indistinguishability at that SNR is predicted to be  $(M/75)^{-1}$ . Top right: corner plot produced after performing PE at an SNR of 128, where the minimum resolution for indistinguishability is predicted to be  $(M/135)^{-1}$ . Bottom: one-dimensional marginalized posterior distributions for  $\mathcal{M}_c$  are presented here. PE was conducted at a sequence of SNRs, with all parameters held constant except for  $D_L$ . Each panel illustrates the outcomes for a specific SNR, and distinct colored curves represent different resolutions. With increasing SNR, both  $M/80$  and  $M/120$  posteriors gradually separate from the others.

be the critical grid spacing, and went all the way up to SNR 128, where  $M/135$  is predicted to be the critical grid spacing. All eight sets of injections had identical parameters except  $D_L$ .

The top panel of Fig. 5 shows selected 1-D and 2-D marginals of our recovered posteriors for SNR 12 and SNR 128, where  $M/75$  and  $M/135$  are the predicted critical grid spacings respectively. At SNR 12, the 1-D and 2-D



posterior distributions for all four resolutions lie on top of each other, whereas at SNR 128 the  $M/80$  posteriors peel away from the rest, introducing notable biases in all three parameters displayed. Consequently, at SNR 12, all four resolutions yield equivalent results for PE, rendering the use of higher-resolution waveforms unnecessary. However, at SNR 128, the use of  $M/80$  leads to biased PE. Additionally, we observe truncation errors cause a downward shift in the marginalized posteriors of the three parameters relative to the highest resolution posterior.

The bottom panel of Fig. 5 shows the detector-frame  $\mathcal{M}_c$  posterior distribution for all eight SNRs. The posteriors beyond SNR 12, where the influence of prior becomes negligible, peak at the same point indicating that the shift in the lower resolution  $\mathcal{M}_c$  posteriors relative to the highest resolution posterior remains consistent; only its significance increases with rising SNR. From Table II, we can see that  $M/80$  produces an absolute normalized bias greater than one at SNR 29, suggesting that  $M/80$  might be accurate for SNRs lower than 29 but beyond this SNR, it is no longer sufficiently accurate for unbiased PE. Also, even though  $M/80$  becomes distinguishable at SNR 15.7, according to the estimate of Eq. (7), it does not produce a normalized bias of one, underscoring the fact that the criterion of indistinguishability arising from matches does not imply significant parameter biases in PE.

We then repeated our analysis for CE, which has a different PSD shape than that of H1. Examining Fig. 3, we notice that H1 and CE exhibit a similar shape at higher frequencies. However, at lower frequencies, CE surpasses H1 in terms of sensitivity, which leads to greater mismatches and subsequently reduces the critical SNR for a given resolution. Similar to our H1 analysis, we selected eight SNRs, ranging from 5 to 60, where the predicted critical grid spacings are  $M/70$  and  $M/135$  respectively. The results of PE are shown in Fig. 6; the upper part displays the 1-D and 2-D marginalized posteriors for SNR 7 and SNR 60, corresponding to critical grid spacings of  $M/75$  and  $M/135$ . At SNR 7, the 1-D and 2-D posterior distributions for all four resolutions closely overlap. However, at SNR 60 the  $M/80$  posteriors deviate notably from the others, introducing significant biases in all three parameters under consideration. The lower part of Fig. 6 presents the  $\mathcal{M}_c$  posterior distribution for all eight SNRs. We observe that  $M/80$  introduces significant biases at an SNR of 18.3, which is considerably lower than what we observed for H1. This is due to the differences in the shape of the noise PSD of H1 and CE. Therefore, it is necessary to reevaluate the accuracy of NR waveforms when the shape of the PSD undergoes alterations.

We summarize these results in Fig. 7, which shows the absolute normalized bias in  $\mathcal{M}_c$  as a function of SNR across injection resolutions and detector PSDs. It is evident that only in the case of the lowest grid spacing,  $M/80$ , the median of  $\mathcal{M}_c$  exhibits a bias surpassing the statistical

uncertainty in parameter recovery for both PSDs, and this occurs only at SNRs above approximately 18. Additionally, at the highest explored SNR,  $M/120$  exhibits a significant bias in H1.

#### D. $q = 1, \iota = \pi/6$

Higher GW modes require more resolution to be fully resolved than the dominant (2, 2) mode and as such the same resolution will produce more parameter bias at non zero inclination, due to greater contribution from higher modes, than it will for zero inclination. To illustrate this, we injected the same  $q = 1$  waveforms but at a modest inclination of  $\pi/6$ . Even at such a small inclination, two additional GW modes, (2, -2) and (4, 4), significantly contribute to the detected GW, thereby increasing the mismatch and consequently decreasing the critical SNR for the same resolution. To estimate the impact of bias as a function of SNR, we repeat the analysis the same way we did for the face-on case. We started from SNR 6 (4), where  $M/70$  is the predicted critical grid spacing, and went all the way up to SNR 79 (55), where  $M/135$  is the predicted critical grid spacing for H1 (CE). The results for this case are broadly the same as for the  $q = 1, \iota = 0$  case. Only the lowest resolution ( $M/80$ )<sup>-1</sup> injection displays significant bias in its recovered parameters, and truncation errors continue to cause a downward shift in the marginalized posteriors of the three parameters relative to the highest resolution posterior. We provide our posteriors and absolute normalized bias values in Appendix B 1.

We observe an absolute normalized bias of unity due to  $M/80$  at an SNR of 28.5 (18.1) for H1 (CE), lower than  $\iota = 0$  case, showing that the accuracy requirements change as the observed inclination changes. For a more direct comparison, we also recovered inclined injections at SNR 128 (60) for H1 (CE). At these SNRs, we found that  $M/80$  produces an absolute normalized bias of 10.72 (3.63), surpassing the bias observed at the same SNR for the face-on case which was 9.59 (3.58) for H1 (CE).

#### E. $q = 1/3, \iota = 0$

A resolution that is considered adequate for unbiased PE for  $q = 1$  systems may prove insufficient for systems with unequal mass ratios. This is due to the fact that as the mass ratio decreases, a higher resolution is needed to accurately resolve the smaller black hole. Additionally, the inherent asymmetry in the system induces the excitation of higher modes, which require more resolution to be sufficiently resolved. The combination of these factors leads to an increase in the parameter bias introduced by a resolution as the mass ratio decreases. To illustrate this, we extended our analysis to include  $q = 1/3$  waveforms. The waveforms were extracted from simulations that had an initial separation of  $9M$ , compared to the initial separation of  $12M$  for  $q = 1$  simulations, resulting in shorter waveforms.

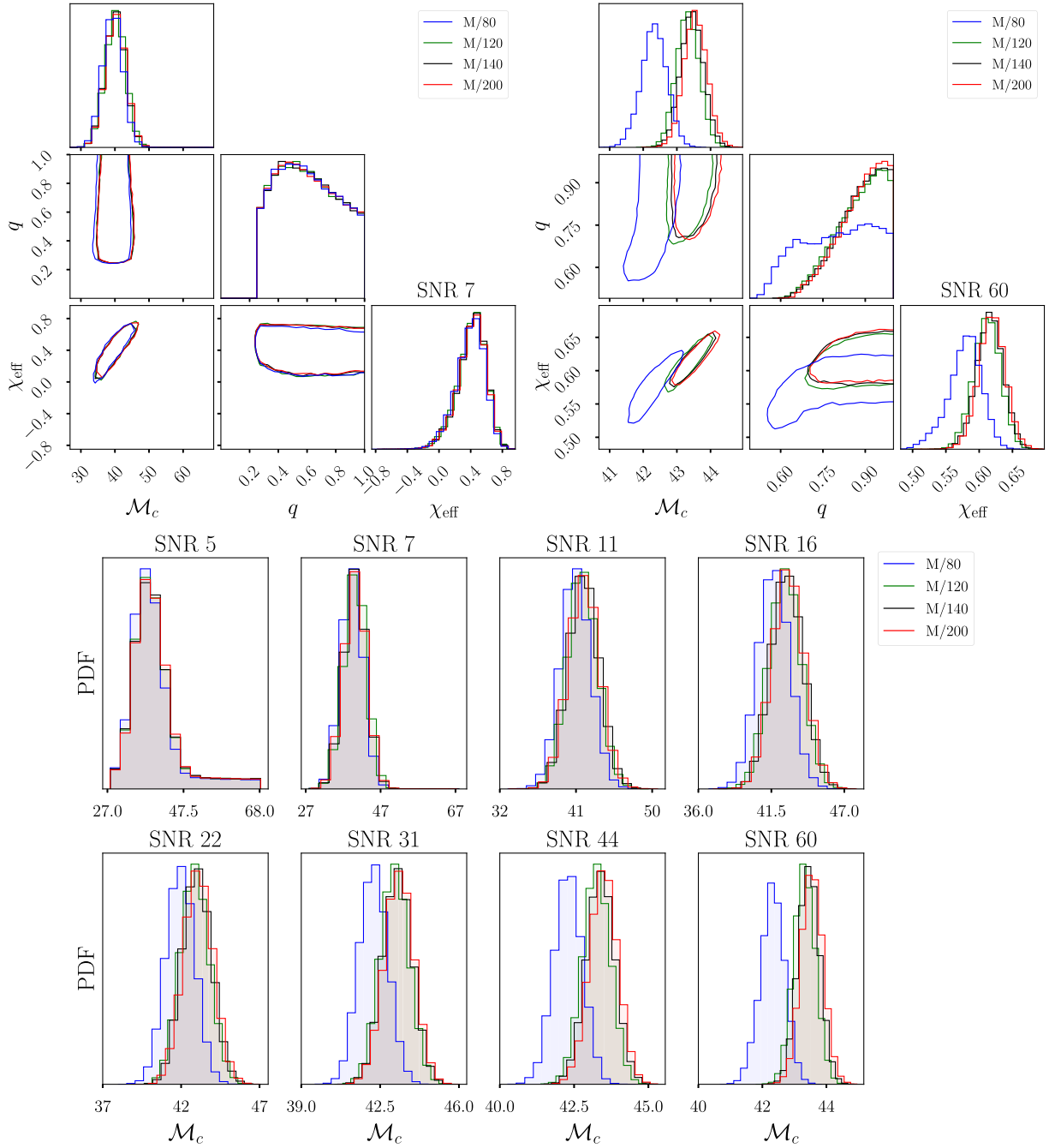


FIG. 6. PE results for  $q = 1, \iota = 0$  injections (CE): Top left: one- and two-dimensional marginal posterior distributions for  $\mathcal{M}_c$ ,  $q$ , and  $\chi_{\text{eff}}$ . Diagonal panels show the one-dimensional marginal posterior distribution, while contours in the off-diagonal panels show the 90% credible intervals for the two-dimensional marginal posterior distribution. Different colored curves correspond to different resolutions. Injections had an SNR of 7 and the minimum resolution for indistinguishability at that SNR is predicted to be  $(M/75)^{-1}$ . Top right: corner plot produced after performing PE at an SNR of 60, where the minimum resolution for indistinguishability is predicted to be  $(M/135)^{-1}$ . Bottom: one-dimensional marginalized posterior distributions for  $\mathcal{M}_c$  are presented here. PE was conducted at a sequence of SNRs, with all parameters held constant except for  $D_L$ . Each panel illustrates the outcomes for a specific SNR, and distinct colored curves represent different resolutions. With increasing SNR, the  $M/80$  posterior gradually separates from the others.

To ensure that the waveforms do not begin in the frequency band of interest, we increased  $M_{\text{tot}}$  from  $100M_{\odot}$  to  $150M_{\odot}$ . As such, we cannot directly compare the normalized bias values for these injections and for  $q = 1$  injections.

The analysis was carried out the same way as we did for  $q = 1$  injections. We started at SNR 37 (21) and went all the way to SNR 203 (122) for H1 (CE), to study the impact of bias as a function of SNR. We observe that  $M/100$

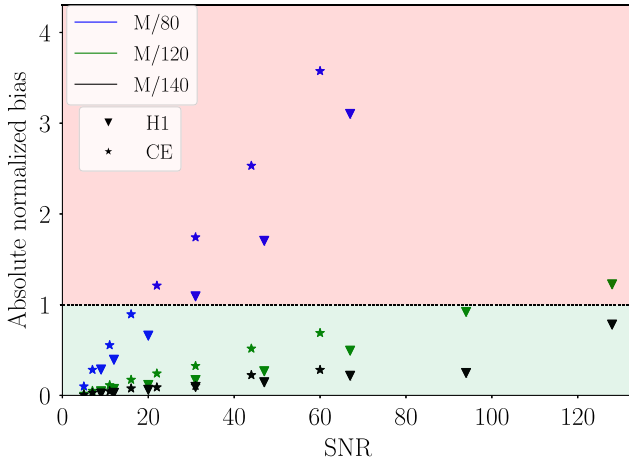


FIG. 7. Absolute normalized bias in the marginalized  $\mathcal{M}_c$  posterior distributions of  $q = 1$ ,  $\iota = 0$  injections as a function of SNR: The dashed horizontal line marks an absolute normalized bias of unity. We have omitted normalized bias values exceeding five to concentrate on the region where the bias is around one. The bias values were calculated with respect to  $M/200$ .

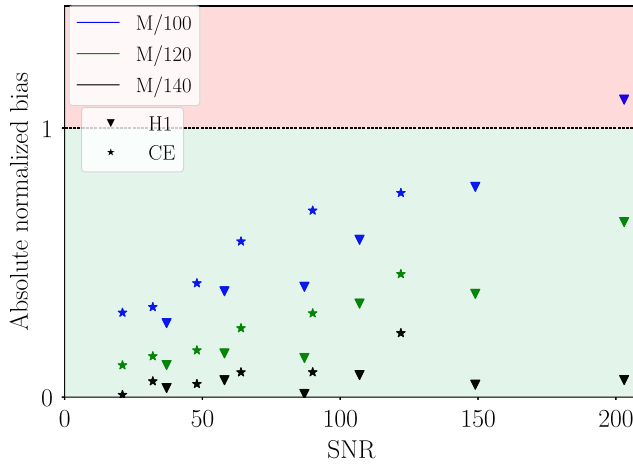


FIG. 8. Absolute normalized bias in the marginalized  $\mathcal{M}_c$  posterior distributions of  $q = 1/3$ ,  $\iota = 0$  injections as a function of SNR: The dashed horizontal line marks an absolute normalized bias of unity. The bias values were calculated with respect to  $M/180$ .

produces a normalized bias of unity at 185.4 (117.1) for H1 (CE). The PE results for this system are summarized in Fig. 8, reflecting similar qualitative trends observed for the previous two synthetic sources. The posteriors and absolute normalized bias values are provided in Appendix B 2.

To emphasize the crucial point that the parameter bias introduced by a resolution increases as the mass ratio decreases, we injected  $q = 1$  waveforms at  $M_{\text{tot}} = 150M_{\odot}$  at SNR 203 (122) for H1 (CE). We find that  $(M/120)^{-1}$  resolution introduces an absolute normalized bias of 0.31 (0.16) for  $q = 1$  and 0.65 (0.46) for  $q = 1/3$  for H1 (CE). This shows that at a given SNR and NR resolution,

waveforms with lower mass ratios introduce greater bias compared to those with higher mass ratios, necessitating a more stringent accuracy requirement.

#### IV. EXTRAPOLATION

In Sec. III, we assessed the errors introduced in PE when finite resolution NR waveforms are used. We accomplished this by comparing the PE recovery of low-resolution NR injections with those of the highest-resolution NR injection, for each of the three synthetic sources. However, it is important to acknowledge that any finite-resolution waveform can introduce bias into PE, and this bias persists even when we use the highest-resolution waveform. As such, in this section, we employ an extrapolation procedure to predict the median of a marginalized posterior distribution for parameter  $\lambda$ , recovered for a hypothetical  $M/\infty$  injection. We then use this extrapolated value to calculate the bias in PE attributable to the use of the highest-resolution waveforms.

We utilize the understanding that the bias in NR waveforms is proportional to  $\Delta^{\alpha}$ . Thus, if all aspects of PE remain consistent across the recovery of each differently resolved injection, the bias in the final posterior due to finite resolution should also be proportional to  $\Delta^{\alpha}$ . Keeping this in mind, we fit the following function to the posteriors recovered for all four injections:

$$\lambda = a\Delta^{\alpha} + b. \quad (12)$$

Here  $a$  and  $b$  are fitting constants, and  $\alpha$  takes a value of 4 [32] for the NR waveforms used in this study. We then fit this function to the  $\mathcal{M}_c$  posteriors, for all three systems and for both detectors at the SNRs mentioned in Table III. Analyzing the posteriors at the SNRs in Table III, we find the standard deviation tends to remain approximately the same for all the different resolution  $\mathcal{M}_c$  posteriors. Using this observation, alongside the median of  $\mathcal{M}_c$  posterior for  $M/\infty$  injection, we can determine the normalized bias in the highest resolution  $\mathcal{M}_c$  posterior, which are listed in Table III. Our findings show that the highest-resolution waveform utilized for each synthetic source meets the accuracy requirements for the injected SNRs, validating the comparison of low-resolution posteriors with the highest-resolution posterior.

TABLE III. Absolute normalized bias in the highest resolution posteriors: Numbers in parenthesis are the SNRs at which these biases are evaluated.

| System                 | $\Delta$ | H1         | CE          |
|------------------------|----------|------------|-------------|
| $q = 1$                | $M/200$  | 0.59 (128) | 0.082 (60)  |
| $q = 1, \iota = \pi/6$ | $M/200$  | 0.87 (128) | 0.092 (60)  |
| $q = 1/3$              | $M/180$  | 0.15 (203) | 0.005 (122) |

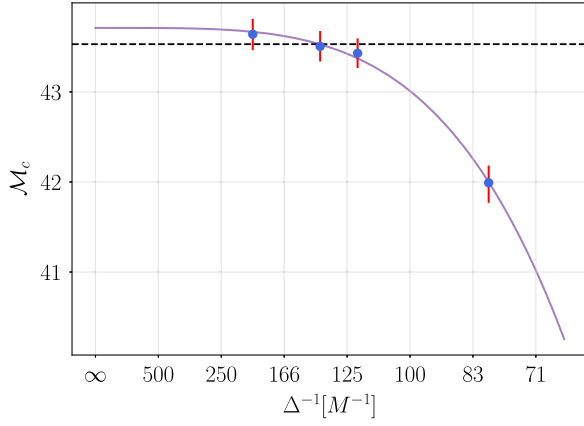


FIG. 9. Extrapolation procedure: This figure shows our extrapolation procedure applied to the recovered marginalized  $\mathcal{M}_c$  posterior for all four H1  $q = 1$ ,  $\iota = 0$  injections at SNR 128. We employ the fitted curve to determine the recovered  $\mathcal{M}_c$  for a hypothetical  $M/\infty$  injection and use it to calculate the bias in our highest resolution injection posterior. The data points represent the median values of recovered  $\mathcal{M}_c$  distributions, the error bars represent one standard deviation and the indigo curve illustrates the injected value. The horizontal dashed line represents the injected value.

Figure 9 demonstrates the application of the extrapolation procedure. It is clear from the figure that the  $\mathcal{M}_c$  median at  $M/\infty$  does not align with the injected value. This discrepancy is expected, due to differences between the extrapolated NR waveforms  $h_\infty$  and the recovery model NRHybSur3dq8  $h_{\text{sur}}$ , as discussed in Sec. III A. We find that the differences between  $h_\infty$  and  $h_{\text{sur}}$  is ultimately due to mismatches between the MAYA waveforms we use for our injections, and the SpEC [63] waveforms which underlie  $h_{\text{sur}}$ . For the parameters shown in Table I, the mismatches between SpEC and  $h_{\text{sur}}$  are on the order of  $10^{-4}$ . Meanwhile, for the same parameters, the mismatches between MAYA and  $h_{\text{sur}}$  and between MAYA and SpEC waveforms are both on the order of  $10^{-3}$ , suggesting that the observed parameter discrepancy can be ascribed to differences between SpEC and MAYA waveforms. These differences can be possibly due to the fact that the MAYA waveforms used in our study were extracted at finite radius whereas the SpEC

waveforms used in the construction of the recovery model were extrapolated to infinity. Furthermore, the two NR codes employ different methods to solve partial differential equations, which could also explain the discrepancy.

## V. PREDICTION

In this section, we aim to determine the SNR at which a finite-resolution NR waveform will produce an absolute normalized bias of unity in at least one of the parameters, without having to do full Bayesian inference. While current PE codes have undergone significant speedups [64,65], full Bayesian inference can still be computationally intensive, especially for signals with higher SNRs. Furthermore, when determining the minimum resolution required for an NR waveform to avoid significant bias in GW interpretation for a particular detector, full PE may be excessive. Therefore we employ the following criterion [34,39] to give an approximate estimation of critical SNR:

$$\epsilon[h_1, h_2] < D/2\rho^2. \quad (13)$$

Here, the prefactor  $D$  is the number of intrinsic parameters affected by waveform inaccuracy. Our study was conducted on aligned systems, and as such we have set  $D$  equal to 4. To get a more accurate assessment of critical SNR, one would need to tune  $D$  by finding the exact SNR at which statistical error becomes equal to systematic error. However, such a refinement typically requires the application of PE and as such we rely on an approximate value of 4 for  $D$ . Examining Table IV, we observe a reasonably good agreement between the critical SNR values obtained through full PE and those determined by the criterion of Eq. (13) for both  $q = 1$  and  $q = 1/3$  systems. The disparity between the two SNR values is approximately 11% for  $q = 1$  systems and roughly 16% for  $q = 1/3$  systems. It is worth noting that, with the exception of  $q = 1$  systems as observed by H1, Eq. (13) tends to provide conservative estimates for the critical SNR.

While Eq. (13) is relatively straightforward to apply, it has a few notable limitations when it comes to its application to finite-resolution NR waveforms. To determine the

TABLE IV. Comparing the critical SNR from full PE, Eqs. (7), (13), and (14): The critical SNRs predicted using the modified criterion [Eq. (14)] tend to be on the conservative side.

| System                 | $\Delta$ | Detector | Critical SNR<br>from PE | Critical SNR<br>[Eq. (7)] | Critical SNR<br>[Eq. (13)] | Critical SNR<br>[Eq. (14)] |
|------------------------|----------|----------|-------------------------|---------------------------|----------------------------|----------------------------|
| $q = 1$                | $M/80$   | H1       | 29.0                    | 15.7                      | 32.3                       | 31.5                       |
|                        |          | CE       | 18.3                    | 8.3                       | 17.0                       | 16.6                       |
| $q = 1, \iota = \pi/6$ | $M/80$   | H1       | 28.5                    | 15.5                      | 31.9                       | 31.0                       |
|                        |          | CE       | 18.1                    | 8.2                       | 17.0                       | 16.5                       |
| $q = 1/3$              | $M/100$  | H1       | 185.4                   | 71.2                      | 157.4                      | 142.3                      |
|                        |          | CE       | 117.1                   | 39.9                      | 88.0                       | 79.7                       |



minimum resolution needed for unbiased PE for a detector, one would generate NR waveforms at various resolutions and then identify the resolution with a critical SNR exceeding the SNR of interest for that detector. Furthermore, mismatches would need to be calculated using an infinite-resolution waveform, which is not feasible. To address these limitations, we adjust Eq. (7) by introducing the pre-factor  $D$ . This modification renders the equation more realistic in its estimation of the required SNR as it takes into account the dimensionality of the search parameter space. By utilizing this adjusted criterion, one can determine the minimum resolution necessary to ensure that PE remains largely unaffected by finite resolution waveforms, all without the need to generate waveforms for a sequence of resolutions or an infinite resolution waveform. One would just need to generate waveforms at two resolutions, compute  $\beta$ , and then find the resolution needed using the modified criterion. The modified criterion is

$$(\beta\Delta^\alpha)^2 < D/\rho^2. \quad (14)$$

Here  $\beta$  is calculated using Eq. (6),  $\alpha$  is the convergence rate and  $\Delta^{-1}$  is the resolution. From Table IV, we can see that

the predictions obtained through the modified criterion closely align with those from Eq. (13). The difference in the critical SNR values by full PE and predictions by modified criterion is approximately 9% for  $q = 1$  systems and roughly 25% for  $q = 1/3$  systems. With the exception of  $q = 1$  systems as observed by H1, Eq. (14) tends to provide conservative estimates for SNR.

Realistic predictions for the minimum resolution required for unbiased PE for CE requires waveforms that span the entire CE frequency range. However, we are limited by the length of the NR waveforms. While it is possible to extend the length of NR waveforms through hybridization, doing so introduces errors in addition to the truncation errors this study is isolating. Thus, we employ power-weighted mismatches [66,67], instead of conventional mismatches, to calculate  $\beta$  in Eq. (6) and subsequently calculate estimates for the minimum resolution required for unbiased PE for CE. Power-weighted mismatches allow us to approximately account for the absent power in the 5–20 Hz range and can be computed as follows:

$$\epsilon_{\text{pow}}[h_1, h_2] = \frac{\langle h|h \rangle_{(20,2048)}}{\langle h|h \rangle_{(5,2048)}} \epsilon[h_1, h_2]. \quad (15)$$

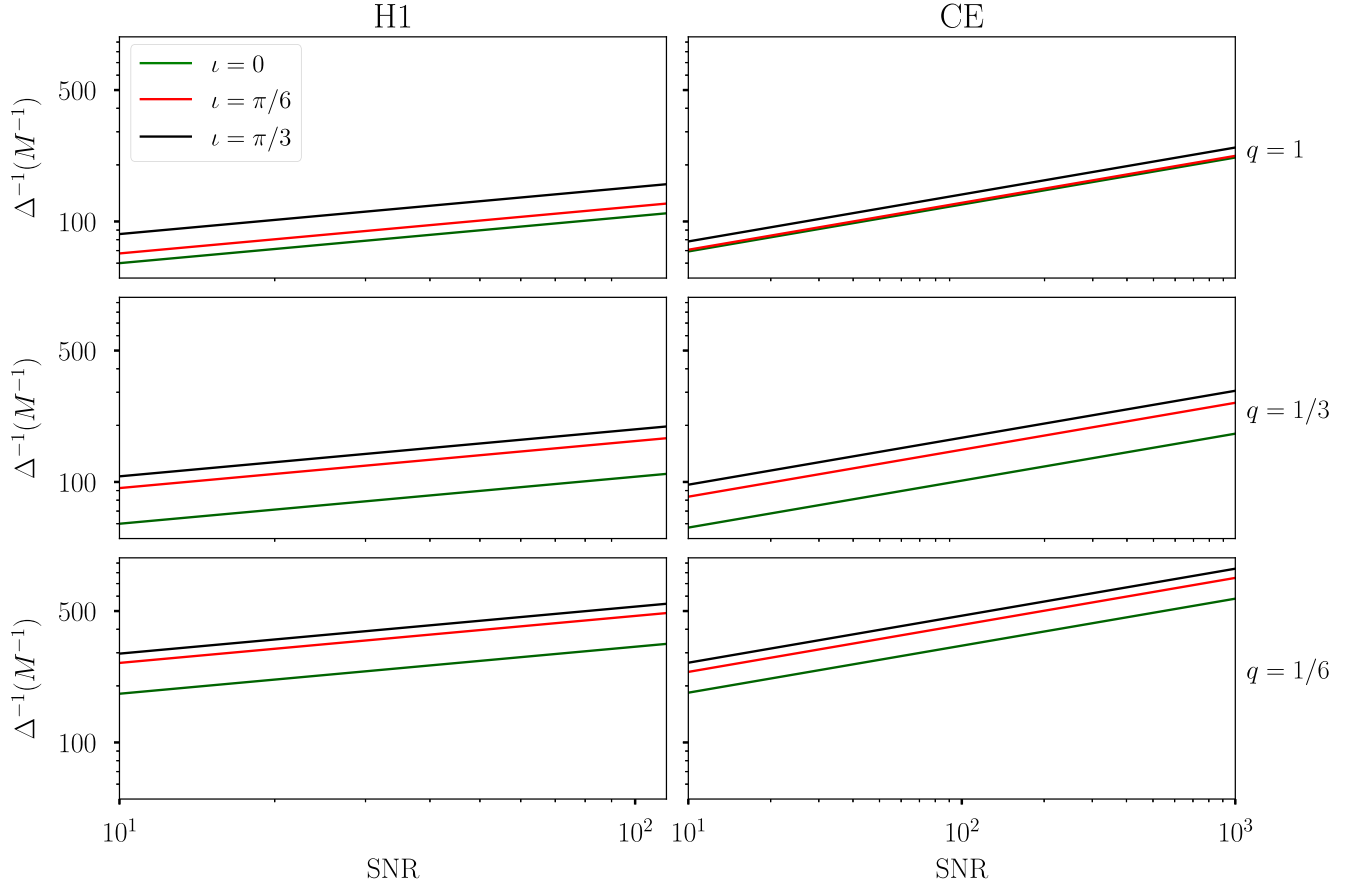


FIG. 10. Applying Eq. (14) to predict accuracy requirements: The predictions for H1 are presented in the left column, while the right column displays predictions for CE. Each row corresponds to predictions for one of the three mass ratios. The colors represent the three inclinations.

Here,  $h_1, h_2$  are two NR waveforms differing only in simulation grid resolution,  $\langle h|h \rangle_{(20,2048)}$  is the noise-weighted inner product integrated from 20–2048 Hz of any reasonably accurate waveform model  $h$  assumed to approximate the inspiral phase, and similarly for  $\langle h|h \rangle_{(5,2048)}$ . The ratio of the power in the reduced bandwidth to the full bandwidth accounts for the early inspiral absent from our analysis, which is optimistically expected to have a perfect match between  $h_1$  and  $h_2$  in the missing frequency band. Here, we use NRHybSur3dq8 for  $h$ . Since,  $\epsilon_{\text{pow}} \leq \epsilon$ , our  $\beta$  values are lower bounds to the actual values computed with  $\epsilon$ . This means that our predicted  $\Delta$  values are upper bounds on the required grid spacing, hence serving as lower bounds on the required  $\Delta^{-1}$  for CE. We find that when using power-weighted mismatches, the resolution prediction for the synthetic sources in Table I for CE is approximately 95% of what we obtain when using normal mismatches calculated in the reduced bandwidth.

We now apply this modified criterion to  $q = 1$  and  $q = 1/3$  waveforms observed at three different inclinations, 0,  $\pi/6$  and  $\pi/3$ . We also apply this criterion to aligned  $q = 1/6$  MAYA waveforms, which have  $\chi_{1z} = 0.2$  and  $\chi_{2z} = 0.0$ . We have set  $M_{\text{tot}} = 100M_{\odot}$  for these waveforms. Figure 10 provides a visual representation of the results when this modified criterion is applied to these three different mass ratio systems observed at three distinct inclinations and for both H1 and CE. The figure illustrates that our current waveforms suffice in terms of accuracy for systems with  $q = 1$  and  $q = 1/3$ , regardless of whether we are dealing with H1 or CE. For systems with  $q = 1/6$ , we find that our current resolutions in the MAYA catalog will be accurate for the median SNR of 10 for aLIGO [4] and 20 for CE [12]. But considering that LIGO Voyager will be at least four times more sensitive than aLIGO [68] and CE will observe signals with SNRs above 600, we find that current resolutions are not sufficient for such high SNR signals for inclinations greater than  $\pi/3$  for H1 and for all possible inclinations for CE.

It is important to note that the accuracy prediction is dependent on the total mass since reducing  $M_{\text{tot}}$  places more of the waveform within the detectors' sensitive bands, causing more truncation error to accumulate and consequently lowering the SNR at which a given resolution introduces significant parameter bias.

## VI. CONCLUSIONS

In this work, we have assessed the impact of NR truncation errors on PE. We accomplished this by performing PE, across a range of SNRs, on simulated GW signals generated at different NR resolutions. We found that for SNRs where a specific resolution ensures unbiased PE, employing a higher resolution does not yield additional scientific insights. Additionally, we have shown that the

resolution required for unbiased PE increases as the mass ratio decreases and/or the observed inclination angle increases. The accuracy requirements are also influenced by the total mass of the system; as the total mass decreases, more accurate waveforms are necessary to achieve an equivalent level of PE accuracy at a given SNR. Furthermore, the shape of the noise curve of a detector is a key factor in defining accuracy requirements. The accuracy demands differ between detectors such as CE and aLIGO, particularly due to the heightened sensitivity of CE at lower frequencies.

We have provided a measure for determining the SNR at which a resolution will produce significant parameter bias. By comparing the critical SNR predictions from this measure with those from full PE for aLIGO and CE, we have shown that the measure provides reasonably accurate estimates of the critical SNR, with most predictions being conservative. To make predictions for the resolution requirements for future NR codes, we applied this measure to three different mass ratio NR waveforms observed at various inclinations and for both detectors. From this application we predict, for equal and moderately unequal mass ratio, our current NR waveforms will be sufficiently accurate, even when observed at high inclinations. For mass ratios around 1/6, our current resolutions will be accurate for the median SNRs for both detectors. However, they will introduce significant parameter bias in the PE for high SNR signals, at all inclinations for CE, and at high inclinations for LIGO Voyager. Considering that, at a given SNR, the resolution required for unbiased PE increases as the mass ratio decreases, current resolutions will produce significant parameter bias in the PE of much lower mass ratio signals even at median SNRs.

In order to attain unbiased PE of signals from binary systems with low mass ratios, we need to generate more accurate NR waveforms. However, the generation of accurate waveforms is a time-consuming process requiring finite-differencing codes to efficiently scale with an increased number of computational nodes. Consequently, significant efforts are underway to improve the parallelization and scalability of NR codes [69,70]. It is imperative to achieve these improvements to maximize the scientific outcomes of GW observations.

In the future, we plan to study the impact of other sources of NR errors, such as extraction radius and finite-differencing order, on PE. Additionally, we aim to explore the impact of truncation error on the recovery of parameters like  $\mathcal{M}_c$  and  $\chi_{\text{eff}}$  over a wide range of injection parameters. We also plan to extend this analysis to the Laser Interferometer Space Antenna (LISA). Given LISA's capability to detect signals with high SNRs and across a wider range of parameters than those examined in this study, our emphasis will be directed toward studying such systems.

## ACKNOWLEDGMENTS

The work presented in this paper is possible due to NSF Grants No. PHY-2207780 and No. PHY-2114581. A. Z. was supported by NSF Grants No. PHY-2207594 and No. PHY-2308833. D. F. was supported by NSF Grant No. OAC-2004879. The computing resources necessary to perform the NR simulations were provided XSEDE PHY120016 and TACC PHY20039. The authors are grateful for the computational resources used for the PE runs provided by the LIGO Laboratories at CIT, LHO, and LLO supported by NSF Grants No. PHY-0757058 and No. PHY-0823459. The authors are also grateful for the computational resources provided on the Nemo cluster by the Leonard E Parker Center for Gravitation, Cosmology, and Astrophysics at the University of Wisconsin-Milwaukee supported by NSF Grants No. PHY-1626190 and No. PHY-1700765. Finally, the authors acknowledge the resources used from the Sarathi cluster by the Inter-University Center for Astronomy & Astrophysics (IUCAA), Pune, India. This material is based upon work supported by NSF's LIGO Laboratory which is a major facility fully funded by the National Science Foundation. The work was done by members of the Weinberg Institute and has an identifier of UT-WI-44-2023.

## APPENDIX A: JS DIVERGENCE

The JS divergence [55], is a statistical tool employed to quantify the dissimilarity between two posterior distributions. It is convenient to use when assessing the agreement between these distributions as it is symmetric in nature, meaning  $D_{\text{JS}}(a|b) = D_{\text{JS}}(b|a)$  and its output is bounded between 0 and 1 bit, where 0 bit means the two distributions are identical and 1 bit means maximal divergence [71].

Given two discrete probability distributions  $a(x)$  and  $b(x)$ , the JS divergence between them is defined as:

$$D_{\text{JS}}(a|b) = \frac{1}{2} \int \left[ a(x) \log_2 \left( \frac{a(x)}{m(x)} \right) + b(x) \log_2 \left( \frac{b(x)}{m(x)} \right) \right] dx \quad (\text{A1})$$

where  $m(x) = [a(x) + b(x)]/2$ . To evaluate the convergence of the RIFT algorithm, we generate kernel density estimators (KDEs) for the marginal probability density functions (PDFs) for each intrinsic parameter. These estimators are constructed using the samples obtained from two consecutive iterations and are subsequently utilized as  $a$  and  $b$  in Eq. (A1) to calculate the JS divergence between them. Our chosen convergence criterion entails achieving a JS divergence of less than  $10^{-3}$  between two successive iterations. This threshold is determined while taking into account the anticipated JS divergence resulting solely from sampling errors. Consider the following experiment: we draw two sets of independent samples from standard

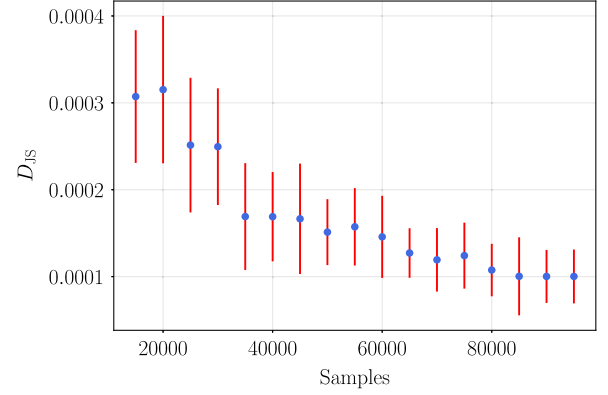


FIG. 11. Sampling errors: JS divergence computed between KDEs built from two sets of samples from standard normal distributions as a function of the number of samples, illustrating the effect of sampling errors. The points are medians, with error bars denoting one standard deviation. These values come from 1000 JS divergence calculations at each  $x$ -axis point.

normal distributions. We then carry out our JS procedure with these samples, forming KDEs for  $a$  and  $b$  from the two sets and computing  $D_{\text{JS}}(a|b)$ . In Fig. 11 we illustrate the median and standard deviation of  $D_{\text{JS}}$  from 1000 iterations of this test as we vary the number of samples. JS divergence values of  $\sim 10^{-4}$  are thus expected due to sampling errors, even for identical distributions. In our study, we typically use 85,000 samples from RIFT to test convergence, to ensure that there is no impact from sampling variance.

## APPENDIX B: ADDITIONAL RESULTS

### 1. $q = 1, \iota = \pi/6$

Figure 12 provides a comprehensive summary of our findings for this system, mirroring the observations made in

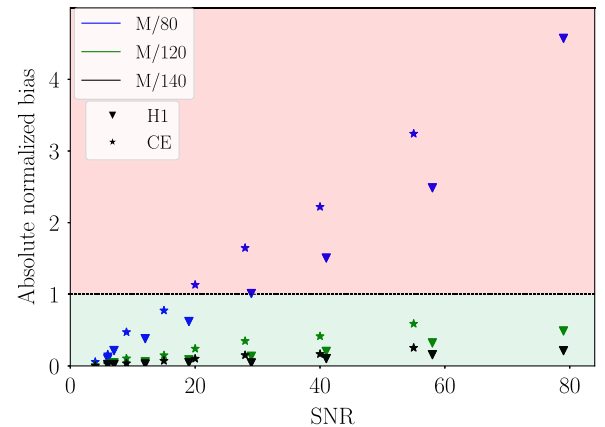


FIG. 12. Absolute normalized bias in the marginalized  $\mathcal{M}_c$  posterior distributions of  $q = 1, \iota = \pi/6$  injections as a function of SNR: The dashed horizontal line marks an absolute normalized bias of unity. The bias values were calculated with respect to  $M/200$ .

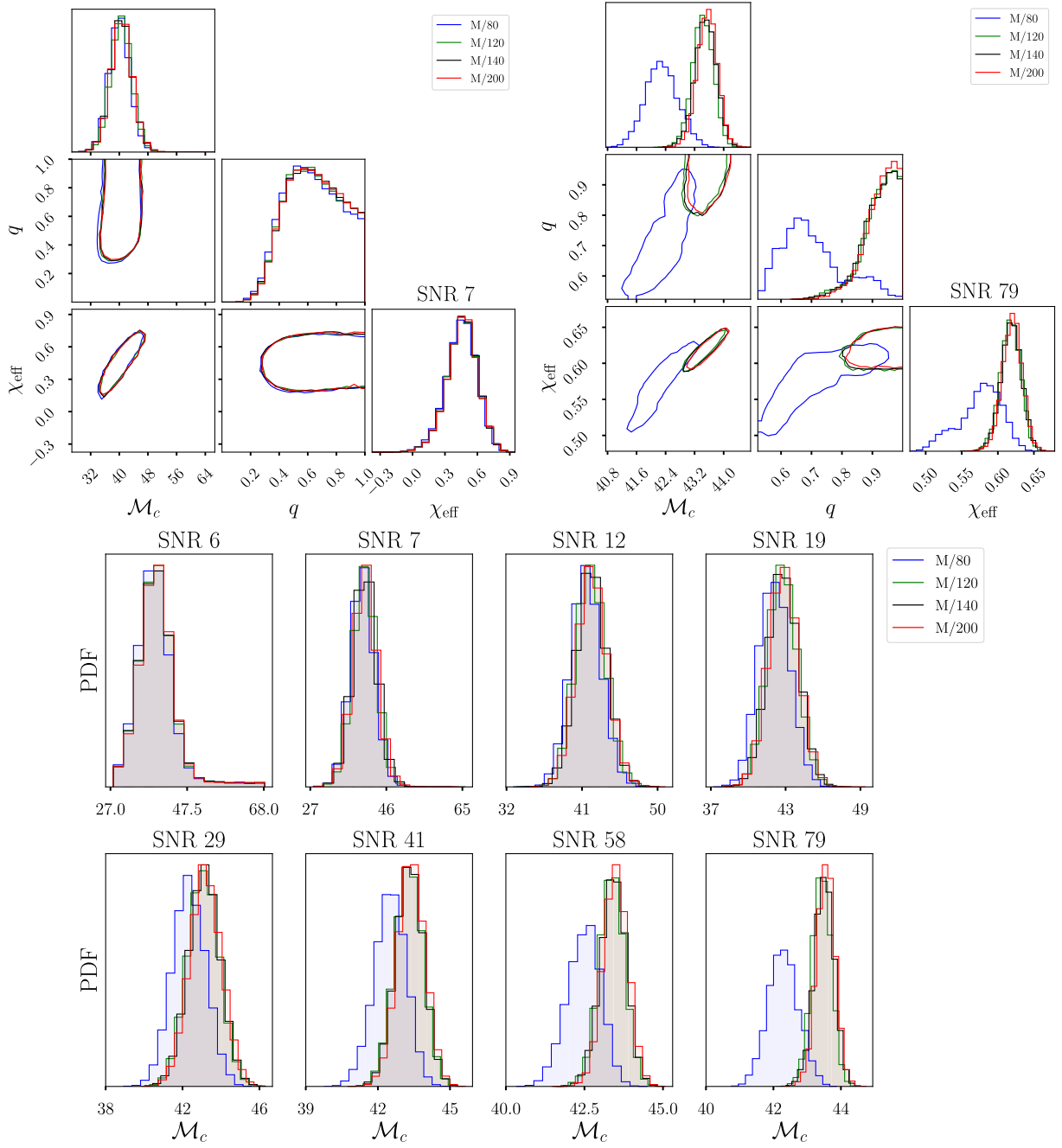


FIG. 13. PE results for  $q = 1, \iota = \pi/6$  injections (H1): Top left: one- and two-dimensional marginal posterior distributions for  $\mathcal{M}_c$ ,  $q$ , and  $\chi_{\text{eff}}$ . Diagonal panels show the one-dimensional marginal posterior distribution, while contours in the off-diagonal panels show the 90% credible intervals for the two-dimensional marginal posterior distribution. Different colored curves correspond to different resolutions. Injections had an SNR of 7 and the minimum resolution for indistinguishability at that SNR is predicted to be  $(M/75)^{-1}$ . Top right: corner plot produced after performing PE at an SNR of 79, where the minimum resolution for indistinguishability is predicted to be  $(M/135)^{-1}$ . Bottom: one-dimensional marginalized posterior distributions for  $\mathcal{M}_c$  are presented here. PE was conducted at a sequence of SNRs, with all parameters held constant except for  $D_L$ . Each panel illustrates the outcomes for a specific SNR, and distinct colored curves represent different resolutions. With increasing SNR, the  $M/80$  posterior gradually separates from the others.

the face-on case. Additionally, the PE results are provided in Fig. 13 (Fig. 14) for H1 (CE). The top panel shows the 1-D and 2-D histogram plots for SNR 7 (6) and SNR 79 (55), where  $M/75$  and  $M/135$  are the predicted critical grid

spacings respectively. The bottom panel shows the detector-frame  $\mathcal{M}_c$  posterior distribution for all eight SNRs and Table V provides the normalized bias values as a function of SNR.



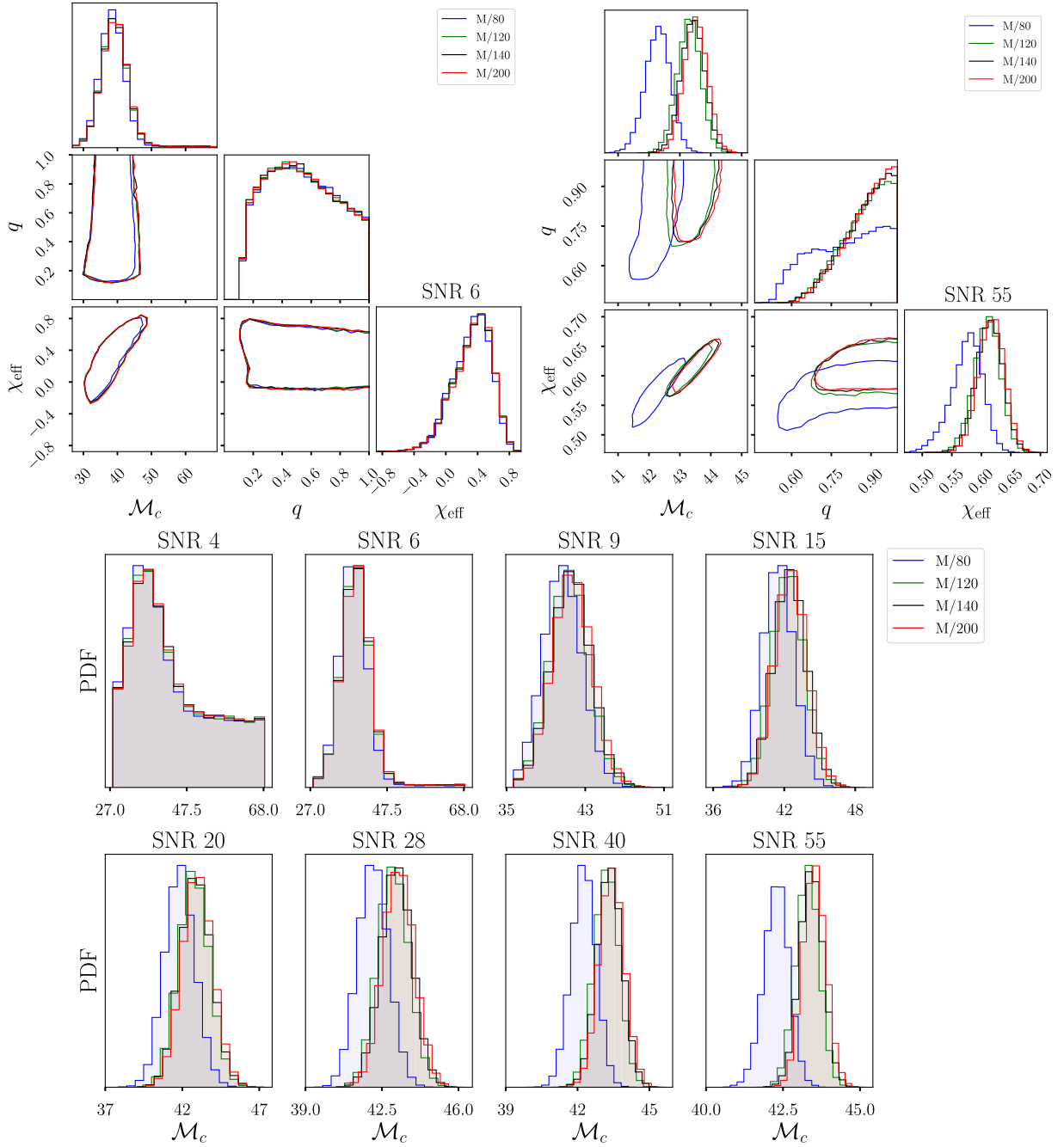


FIG. 14. PE results for  $q = 1, \iota = \pi/6$  injections (CE): Top left: one- and two-dimensional marginal posterior distributions for  $\mathcal{M}_c$ ,  $q$ , and  $\chi_{\text{eff}}$ . Diagonal panels show the one-dimensional marginal posterior distribution, while contours in the off-diagonal panels show the 90% credible intervals for the two-dimensional marginal posterior distribution. Different colored curves correspond to different resolutions. Injections had an SNR of 6 and the minimum resolution for indistinguishability at that SNR is predicted to be  $(M/75)^{-1}$ . Top right: corner plot produced after performing PE at an SNR of 55, where the minimum resolution for indistinguishability is predicted to be  $(M/135)^{-1}$ . Bottom: one-dimensional marginalized posterior distributions for  $\mathcal{M}_c$  are presented here. PE was conducted at a sequence of SNRs, with all parameters held constant except for  $D_L$ . Each panel illustrates the outcomes for a specific SNR, and distinct colored curves represent different resolutions. With increasing SNR, the  $M/80$  posterior gradually separates from the others.

TABLE V. Normalized bias in the marginalized  $\mathcal{M}_c$  posterior distributions of  $q = 1$ ,  $\iota = \pi/6$  injections: Bias observed in the lower resolution posteriors, calculated with respect to  $M/200$ , for H1 (top) and CE (bottom).

| SNR | $\Delta_{\text{critical}}$ | $M/80$ | $M/120$ | $M/140$ |
|-----|----------------------------|--------|---------|---------|
| 6   | $M/70$                     | -0.10  | -0.02   | -0.02   |
| 7   | $M/75$                     | -0.21  | -0.04   | -0.02   |
| 12  | $M/85$                     | -0.38  | -0.06   | -0.03   |
| 19  | $M/95$                     | -0.62  | -0.08   | -0.05   |
| 29  | $M/105$                    | -1.01  | -0.13   | -0.04   |
| 41  | $M/115$                    | -1.51  | -0.20   | -0.10   |
| 58  | $M/125$                    | -2.49  | -0.32   | -0.16   |
| 79  | $M/135$                    | -4.57  | -0.49   | -0.21   |

| SNR | $\Delta_{\text{critical}}$ | $M/80$ | $M/120$ | $M/140$ |
|-----|----------------------------|--------|---------|---------|
| 4   | $M/70$                     | -0.06  | -0.01   | -0.00   |
| 6   | $M/75$                     | -0.16  | -0.03   | -0.02   |
| 9   | $M/85$                     | -0.47  | -0.11   | -0.04   |
| 15  | $M/95$                     | -0.77  | -0.15   | -0.07   |
| 20  | $M/105$                    | -1.13  | -0.24   | -0.10   |
| 28  | $M/115$                    | -1.65  | -0.35   | -0.15   |
| 40  | $M/125$                    | -2.22  | -0.41   | -0.17   |
| 55  | $M/135$                    | -3.24  | -0.59   | -0.25   |

## 2. $q=1/3$ , $\iota=0$

PE results are shown in Fig. 15 for H1 and Fig. 16 for CE. In both figures, the top panels display the 1-D and 2-D histogram plots at SNR 37 (21) and SNR 203 (122) for H1 (CE), where  $M/85$  and  $M/135$  are the predicted critical grid spacings respectively. While the  $M/100$  posterior shows deviation from the  $M/180$  posterior, the deviation is not as pronounced as was for  $q = 1$  injections. This is due

to the resolutions being close to each other, minimizing the disparity. The bottom panel in both figures shows the marginalized  $\mathcal{M}_c$  posteriors for the sequence of SNRs. In this panel, a similar observation is made, the lowest resolution posterior distribution deviates from the highest resolution posterior distribution, although not as prominently as observed for  $q = 1$  injections. Table VI provides the normalized bias values as a function of SNR for both detectors.

TABLE VI. Normalized bias in the marginalized  $\mathcal{M}_c$  posterior distributions of  $q = 1/3$ ,  $\iota = 0$  injections: Bias observed in the lower resolution posteriors, calculated with respect to  $M/180$ , for H1 (top) and CE (bottom).

| SNR | $\Delta_{\text{critical}}$ | $M/100$ | $M/120$ | $M/140$ |
|-----|----------------------------|---------|---------|---------|
| 37  | $M/85$                     | -0.28   | -0.12   | -0.03   |
| 58  | $M/95$                     | -0.39   | -0.16   | -0.06   |
| 87  | $M/105$                    | -0.41   | -0.15   | -0.01   |
| 107 | $M/115$                    | -0.49   | -0.35   | -0.08   |
| 149 | $M/125$                    | -0.78   | -0.38   | -0.05   |
| 203 | $M/135$                    | -1.11   | -0.65   | -0.06   |

| SNR | $\Delta_{\text{critical}}$ | $M/100$ | $M/120$ | $M/140$ |
|-----|----------------------------|---------|---------|---------|
| 21  | $M/85$                     | -0.31   | -0.12   | -0.01   |
| 32  | $M/95$                     | -0.34   | -0.15   | -0.06   |
| 48  | $M/105$                    | -0.42   | -0.17   | -0.05   |
| 64  | $M/115$                    | -0.58   | -0.26   | -0.09   |
| 90  | $M/125$                    | -0.69   | -0.31   | -0.09   |
| 122 | $M/135$                    | -0.76   | -0.46   | -0.24   |

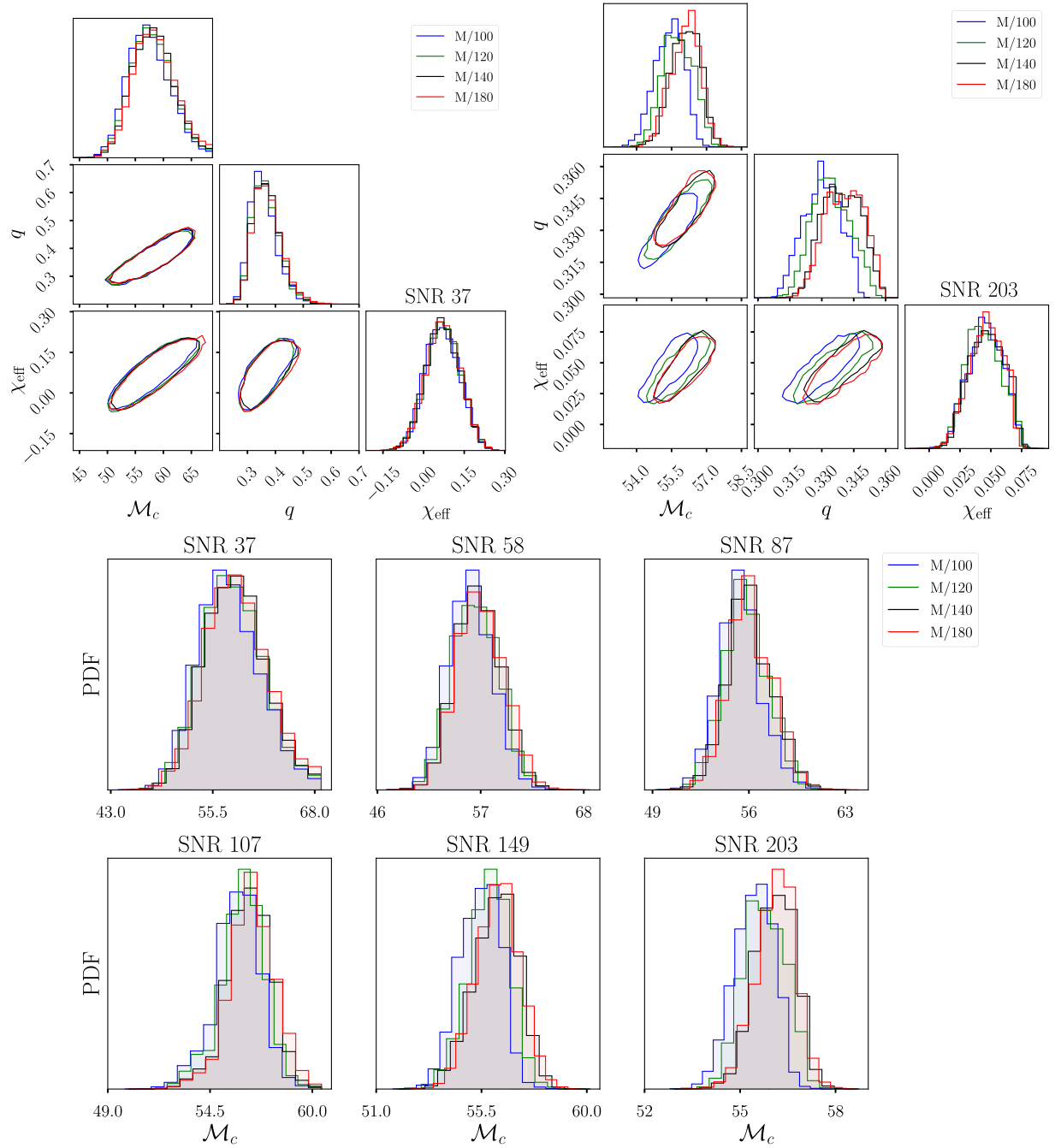


FIG. 15. PE results for  $q = 1/3, \iota = 0$  injections (H1): Top left: one- and two-dimensional marginal posterior distributions for  $\mathcal{M}_c$ ,  $q$ , and  $\chi_{\text{eff}}$ . Diagonal panels show the one-dimensional marginal posterior distribution, while contours in the off-diagonal panels show the 90% credible intervals for the two-dimensional marginal posterior distribution. Different colored curves correspond to different resolutions. Injections had an SNR of 37 and the minimum resolution for indistinguishability at that SNR is predicted to be  $(M/85)^{-1}$ . Top right: corner plot produced after performing PE at an SNR of 203, where the minimum resolution for indistinguishability is predicted to be  $(M/135)^{-1}$ . Bottom: one-dimensional marginalized posterior distributions for  $\mathcal{M}_c$  are presented here. PE was conducted at a sequence of SNRs, with all parameters held constant except for  $D_L$ . Each panel illustrates the outcomes for a specific SNR, and distinct colored curves represent different resolutions.

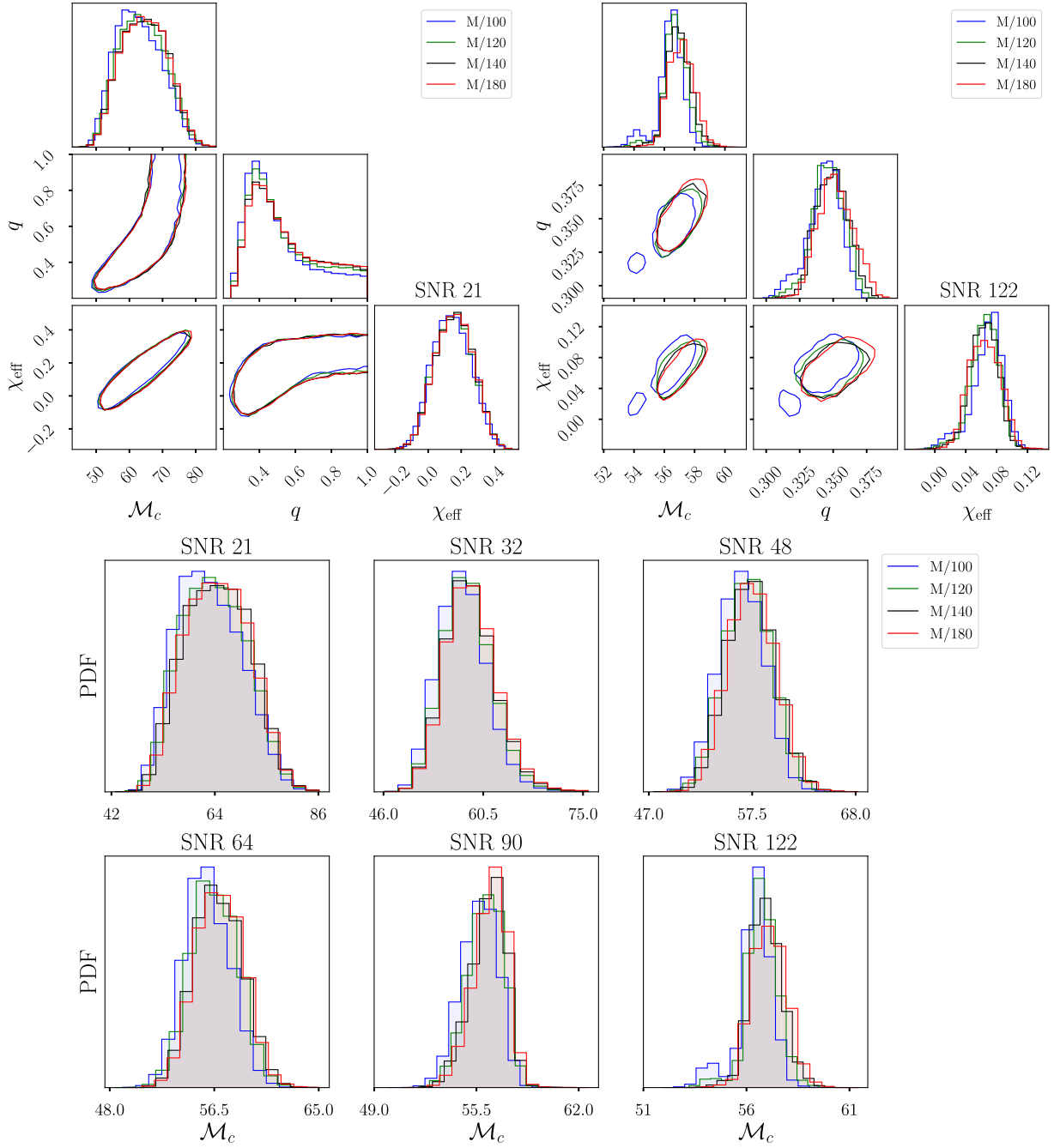


FIG. 16. PE results for  $q = 1/3$ ,  $\iota = 0$  injections (CE): Top left: one- and two-dimensional marginal posterior distributions for  $\mathcal{M}_c$ ,  $q$ , and  $\chi_{\text{eff}}$ . Diagonal panels show the one-dimensional marginal posterior distribution, while contours in the off-diagonal panels show the 90% credible intervals for the two-dimensional marginal posterior distribution. Different colored curves correspond to different resolutions. Injections had an SNR of 21 and the minimum resolution for indistinguishability at that SNR is predicted to be  $(M/85)^{-1}$ . Top right: corner plot produced after performing PE at an SNR of 122, where the minimum resolution for indistinguishability is predicted to be  $(M/135)^{-1}$ . Bottom: one-dimensional marginalized posterior distributions for  $\mathcal{M}_c$  are presented here. PE was conducted at a sequence of SNRs, with all parameters held constant except for  $D_L$ . Each panel illustrates the outcomes for a specific SNR, and distinct colored curves represent different resolutions.



- [1] B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. X. Adhikari *et al.* (The LIGO Scientific and the Virgo Collaborations), *Phys. Rev. X* **9**, 031040 (2019).
- [2] R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, A. Adams, C. Adams, R. X. Adhikari *et al.* (The LIGO Scientific and the Virgo Collaborations), *Phys. Rev. X* **11**, 021053 (2021).
- [3] R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, N. Adhikari, R. X. Adhikari, V. B. Adya *et al.* (The LIGO Scientific and the Virgo Collaborations), *Phys. Rev. D* **109**, 022001 (2024).
- [4] R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, N. Adhikari, R. X. Adhikari *et al.* (The LIGO Scientific, the Virgo, and the KAGRA Collaborations), *Phys. Rev. X* **13**, 041039 (2023).
- [5] J. Aasi, B. P. Abbott, R. Abbott, T. Abbott, M. R. Abernathy, K. Ackley, C. Adams, T. Adams, P. Addesso *et al.* (The LIGO Scientific Collaboration), *Classical Quantum Gravity* **32**, 074001 (2015).
- [6] F. Acernese, M. Agathos, K. Agatsuma, D. Aisa, N. Allemandou, A. Allocca, J. Amarni, P. Astone, G. Balestri, G. Ballardin *et al.*, *Classical Quantum Gravity* **32**, 024001 (2015).
- [7] B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, V. B. Adya *et al.* (The LIGO Scientific and the Virgo Collaborations), *Living Rev. Relativity* **23**, 3 (2020).
- [8] The LIGO Scientific Collaboration, [https://dcc.ligo.org/public/0185/T2200384/002/Instrument\\_Science\\_White\\_Paper\\_2022.pdf](https://dcc.ligo.org/public/0185/T2200384/002/Instrument_Science_White_Paper_2022.pdf) (2022).
- [9] M. Punturo, M. Abernathy, F. Acernese, B. Allen, N. Andersson, K. Arun, F. Barone, B. Barr, M. Barsuglia, M. Beker *et al.*, *Classical Quantum Gravity* **27**, 194002 (2010).
- [10] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, K. Ackley, C. Adams, P. Addesso, R. X. Adhikari, V. B. Adya *et al.* (The LIGO Scientific Collaboration), *Classical Quantum Gravity* **34**, 044001 (2017).
- [11] D. Reitze, R. Abbott, C. Adams, R. Adhikari, N. Aggarwal, S. Anand, A. Ananyeva, S. Anderson, S. Appert, K. Arai *et al.*, *Bull. Am. Astron. Soc.* **51**, 141 (2019).
- [12] D. Reitze, R. X. Adhikari, S. Ballmer, B. Barish, L. Barsotti, G. Billingsley, D. A. Brown, Y. Chen, D. Coyne, R. Eisenstein *et al.*, *Bull. Am. Astron. Soc.* **51**, 035 (2019).
- [13] M. Evans, R. X. Adhikari, C. Afle, S. W. Ballmer, S. Biscoveanu, S. Borhanian, D. A. Brown, Y. Chen, R. Eisenstein, A. Gruson *et al.*, [arXiv:2109.09882](https://arxiv.org/abs/2109.09882).
- [14] J. Luo, L.-S. Chen, H.-Z. Duan, Y.-G. Gong, S. Hu, J. Ji, Q. Liu, J. Mei, V. Milyukov, M. Sazhin *et al.*, *Classical Quantum Gravity* **33**, 035010 (2016).
- [15] P. Amaro-Seoane, H. Audley, S. Babak, J. Baker, E. Barausse, P. Bender, E. Berti, P. Binetruy, M. Born, D. Bortoluzzi *et al.*, [arXiv:1702.00786](https://arxiv.org/abs/1702.00786).
- [16] S. Babak, M. Hewitson, and A. Petiteau, [arXiv:2108.01167](https://arxiv.org/abs/2108.01167).
- [17] F. Pretorius, *Classical Quantum Gravity* **22**, 425 (2005).
- [18] J. G. Baker, J. Centrella, D.-I. Choi, M. Koppitz, and J. van Meter, *Phys. Rev. Lett.* **96**, 111102 (2006).
- [19] M. Campanelli, C. O. Lousto, P. Marronetti, and Y. Zlochower, *Phys. Rev. Lett.* **96**, 111101 (2006).
- [20] J. Lange, R. O'Shaughnessy, M. Boyle, J. C. Bustillo, M. Campanelli, T. Chu, J. Clark, N. Demos, H. Fong, J. Healy *et al.*, *Phys. Rev. D* **96**, 104041 (2017).
- [21] J. C. Bustillo, N. Sanchis-Gual, A. Torres-Forné, J. A. Font, A. Vajpeyi, R. Smith, C. Herdeiro, E. Radu, and S. H. W. Leong, *Phys. Rev. Lett.* **126**, 081101 (2021).
- [22] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams *et al.* (The LIGO Scientific and the Virgo Collaborations), *Phys. Rev. D* **94**, 064035 (2016).
- [23] A. Buonanno and T. Damour, *Phys. Rev. D* **59**, 084006 (1999).
- [24] A. Buonanno and T. Damour, *Phys. Rev. D* **62**, 064015 (2000).
- [25] T. Damour, P. Jaranowski, and G. Schäfer, *Phys. Rev. D* **62**, 084011 (2000).
- [26] T. Damour, *Phys. Rev. D* **64**, 124013 (2001).
- [27] P. Ajith, S. Babak, Y. Chen, M. Hewitson, B. Krishnan, J. T. Whelan, B. Brügmann, P. Diener, J. Gonzalez, M. Hannam *et al.*, *Classical Quantum Gravity* **24**, S689 (2007).
- [28] P. Ajith, S. Babak, Y. Chen, M. Hewitson, B. Krishnan, A. M. Sintes, J. T. Whelan, B. Brügmann, P. Diener, N. Dorband *et al.*, *Phys. Rev. D* **77**, 104017 (2008).
- [29] P. Ajith, M. Hannam, S. Husa, Y. Chen, B. Brügmann, N. Dorband, D. Müller, F. Ohme, D. Pollney, C. Reisswig *et al.*, *Phys. Rev. Lett.* **106**, 241101 (2011).
- [30] J. Blackman, S. E. Field, C. R. Galley, B. Szilágyi, M. A. Scheel, M. Tiglio, and D. A. Hemberger, *Phys. Rev. Lett.* **115**, 121102 (2015).
- [31] L. Blanchet, *Living Rev. Relativity* **9**, 4 (2006).
- [32] D. Ferguson, K. Jani, P. Laguna, and D. Shoemaker, *Phys. Rev. D* **104**, 044037 (2021).
- [33] A. R. Williamson, J. Lange, R. O'Shaughnessy, J. A. Clark, P. Kumar, J. Calderón Bustillo, and J. Veitch, *Phys. Rev. D* **96**, 124041 (2017).
- [34] M. Pürrer and C.-J. Haster, *Phys. Rev. Res.* **2**, 023151 (2020).
- [35] A. Z. Jan, A. B. Yelikar, J. Lange, and R. O'Shaughnessy, *Phys. Rev. D* **102**, 124069 (2020).
- [36] Q. Hu and J. Veitch, *Phys. Rev. D* **106**, 044042 (2022).
- [37] C. B. Owen, C.-J. Haster, S. Perkins, N. J. Cornish, and N. Yunes, *Phys. Rev. D* **108**, 044018 (2023).
- [38] L. Lindblom, J. G. Baker, and B. J. Owen, *Phys. Rev. D* **82**, 084020 (2010).
- [39] K. Chatziioannou, A. Klein, N. Yunes, and N. Cornish, *Phys. Rev. D* **95**, 104004 (2017).
- [40] L. Lindblom, B. J. Owen, and D. A. Brown, *Phys. Rev. D* **78**, 124020 (2008).
- [41] F. Herrmann, I. Hinder, D. Shoemaker, and P. Laguna, *Classical Quantum Gravity* **24**, S33 (2007).
- [42] B. Vaishnav, I. Hinder, F. Herrmann, and D. Shoemaker, *Phys. Rev. D* **76**, 084020 (2007).
- [43] J. Healy, J. Levin, and D. Shoemaker, *Phys. Rev. Lett.* **103**, 131101 (2009).
- [44] L. Pekowsky, R. O'Shaughnessy, J. Healy, and D. Shoemaker, *Phys. Rev. D* **88**, 024040 (2013).
- [45] E. Schnetter, S. H. Hawley, and I. Hawke, *Classical Quantum Gravity* **21**, 1465 (2004).
- [46] T. W. Baumgarte and S. L. Shapiro, *Phys. Rev. D* **59**, 024007 (1998).

- [47] K. Higginbotham, B. Khamesra, J. P. McInerney, K. Jani, D. M. Shoemaker, and P. Laguna, *Phys. Rev. D* **100**, 081501 (2019).
- [48] D. Ferguson, E. Allsup, S. Anne, G. Bouyer, M. Gracia-Linares, H. Iglesias, A. Jan, P. Laguna, J. Lange, E. Martinez *et al.*, [arXiv:2309.00262](https://arxiv.org/abs/2309.00262).
- [49] J. S. Read, C. Markakis, M. Shibata, K. Uryū, J. D. E. Creighton, and J. L. Friedman, *Phys. Rev. D* **79**, 124033 (2009).
- [50] H.-S. Cho, E. Ochsner, R. O’Shaughnessy, C. Kim, and C.-H. Lee, *Phys. Rev. D* **87**, 024004 (2013).
- [51] M. Hannam, S. Husa, F. Ohme, and P. Ajith, *Phys. Rev. D* **82**, 124052 (2010).
- [52] P. Kumar, T. Chu, H. Fong, H. P. Pfeiffer, M. Boyle, D. A. Hemberger, L. E. Kidder, M. A. Scheel, and B. Szilagy, *Phys. Rev. D* **93**, 104050 (2016).
- [53] J. Lange, R. O’Shaughnessy, and M. Rizzo, [arXiv:1805.10457](https://arxiv.org/abs/1805.10457).
- [54] C. Pankow, P. Brady, E. Ochsner, and R. O’Shaughnessy, *Phys. Rev. D* **92**, 023002 (2015).
- [55] J. Lin, *IEEE Trans. Inf. Theory* **37**, 145 (1991).
- [56] D. Wysocki, R. O’Shaughnessy, J. Lange, and Y.-L. L. Fang, *Phys. Rev. D* **99**, 084026 (2019).
- [57] J. Lange, RIFT’ing the wave: Developing and applying an algorithm to infer properties gravitational wave sources, <https://dcc.ligo.org/LIGO-P2000268> (2020).
- [58] J. Wofford, A. B. Yelkar, H. Gallagher, E. Champion, D. Wysocki, V. Delfavero, J. Lange, C. Rose, V. Valsan, S. Morisaki *et al.*, *Phys. Rev. D* **107**, 024040 (2023).
- [59] V. Varma, S. E. Field, M. A. Scheel, J. Blackman, L. E. Kidder, and H. P. Pfeiffer, *Phys. Rev. D* **99**, 064045 (2019).
- [60] The LIGO Scientific Collaboration, <https://dcc.ligo.org/LIGO-T070247-x0/public> (2008).
- [61] The LIGO Scientific Collaboration, <https://dcc.ligo.org/LIGO-T0900288/public> (2009).
- [62] C. J. Moore, R. H. Cole, and C. P. L. Berry, *Classical Quantum Gravity* **32**, 015014 (2015).
- [63] M. Boyle, D. Hemberger, D. A. B. Iozzo, G. Lovelace, S. Ossokine, H. P. Pfeiffer, M. A. Scheel, L. C. Stein, C. J. Woodford, A. B. Zimmerman *et al.*, *Classical Quantum Gravity* **36**, 195006 (2019).
- [64] D. Wysocki, R. O’Shaughnessy, J. Lange, and Y.-L. L. Fang, *Phys. Rev. D* **99** (2019).
- [65] R. J. E. Smith, G. Ashton, A. Vajpeyi, and C. Talbot, *Mon. Not. R. Astron. Soc.* **498**, 4492 (2020).
- [66] F. Ohme, M. Hannam, and S. Husa, *Phys. Rev. D* **84**, 064029 (2011).
- [67] E. Hamilton, E. Fauchon-Jones, M. Hannam, C. Hoy, C. Kalaghatgi, L. London, J. E. Thompson, D. Yeeles, S. Ghosh, S. Khan *et al.*, *Phys. Rev. D* **109**, 044032 (2024).
- [68] The LIGO Scientific Collaboration, <https://docs.ligo.org/voyager/voyagerwhitepaper/main.pdf> (2022).
- [69] F. Foucart, P. Laguna, G. Lovelace, D. Radice, and H. Witek, in *Snowmass2021 Cosmic Frontier White Paper: Numerical relativity for next-generation gravitational-wave probes of fundamental physics* (2022), [arXiv:2203.08139](https://arxiv.org/abs/2203.08139).
- [70] N. Afshordi, S. Akçay, P. A. Seoane, A. Antonelli, J. C. Aurekkoetxea, L. Barack, E. Barausse, R. Benkel, L. Bernard *et al.* (LISA Consortium Waveform Working Group), [arXiv:2311.01300](https://arxiv.org/abs/2311.01300).
- [71] I. M. Romero-Shaw, C. Talbot, S. Biscoveanu, V. D’Emilio, G. Ashton, C. P. L. Berry, S. Coughlin, S. Galaudage, C. Hoy, M. Hübner *et al.*, *Mon. Not. R. Astron. Soc.* **499**, 3295 (2020).