

GPU-accelerated massive black hole binary parameter estimation with LISA

Michael L. Katz*

*Department of Physics and Astronomy, Northwestern University, Evanston, Illinois 60201, USA
and Center for Interdisciplinary Exploration and Research in Astrophysics, Evanston, Illinois 60201, USA*

Sylvain Marsat

Laboratoire Astroparticule et Cosmologie, 10 rue Alice Domon et Léonie Duquet, 75013 Paris, France

Alvin J. K. Chua

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, USA

Stanislav Babak

Laboratoire Astroparticule et Cosmologie, 10 rue Alice Domon et Léonie Duquet, 75013 Paris, France and Moscow Institute of Physics and Technology, Dolgoprudny, Moscow Region, Russia

Shane L. Larson

*Department of Physics and Astronomy, Northwestern University, Evanston, Illinois 60201, USA
and Center for Interdisciplinary Exploration and Research in Astrophysics, Evanston, Illinois 60201, USA*



(Received 4 May 2020; accepted 10 July 2020; published 27 July 2020)

The Laser Interferometer Space Antenna (LISA) is slated for launch in the early 2030s. A main target of the mission is massive black hole binaries that have an expected detection rate of $\sim 20 \text{ yr}^{-1}$. We present a parameter estimation analysis for a variety of massive black hole binaries. This analysis is performed with a graphics processing unit (GPU) implementation comprising the PhenomHM waveform with higher-order harmonic modes and aligned spins; a fast frequency-domain LISA detector response function; and a GPU-native likelihood computation. The computational performance achieved with the GPU is shown to be 500 times greater than with a similar CPU implementation, which allows us to analyze full noise-infused injections at a realistic Fourier bin width for the LISA mission in a tractable and efficient amount of time. With these fast likelihood computations, we study the effect of adding aligned spins to an analysis with higher-order modes by testing different configurations of spins in the injection, as well as the effect of varied and fixed spins during sampling. Within these tests, we examine three different binaries with varying mass ratios, redshifts, sky locations, and detector-frame total masses ranging over 3 orders of magnitude. We discuss varied correlations between the total masses and mass ratios; unique spin posteriors for the larger mass binaries; and the constraints on parameters when fixing spins during sampling, allowing us to compare to previous analyses that did not include aligned spins.

DOI: [10.1103/PhysRevD.102.023033](https://doi.org/10.1103/PhysRevD.102.023033)

I. INTRODUCTION

The Laser Interferometer Space Antenna (LISA), an ESA-led mission, is officially slated for launch in the early 2030s [1]. LISA will add the milliHertz regime of the gravitational wave spectrum to the high-frequency observations by the LIGO-Virgo-KAGRA network [2,3]. A primary source of interest for LISA is the inspiral and merger of massive black hole (MBH) binaries. Recent work has shown that LISA is expected to detect ~ 1 – 20 MBH binaries per year [4–8]. These sources are expected to occur

at high signal-to-noise ratios (SNRs) of ~ 100 – 1000 out to high redshifts ($z \sim 10$ – 20) [1,9,10]. LISA will be sensitive to binaries of $\sim 10^3$ – $10^9 M_{\odot}$ if they exist [1,4,11].

The high SNR of MBH signals will help astronomers understand the formation and evolution of these objects over cosmic time [1,9,10]. Theories of MBH formation channels usually fall into three categories: high-mass seeds of $\sim 10^4$ – $10^6 M_{\odot}$ from pregalactic halo collapse at $z \sim 10$ – 20 [12–17]; intermediate seeds of $\sim 10^3$ – $10^4 M_{\odot}$ from runaway cluster collapse at $z \sim 10$ – 20 [18–21]; and smaller seeds of $\sim 10^2 M_{\odot}$ from the collapse of Population III stars at $z \sim 20$ – 50 [22–27]. The loud signals from MBH binaries can also be used for tests of fundamental physics and

*mikekatz04@gmail.com

independent measurements of cosmological parameters [10,28,29]. In addition, determining MBH binary information such as the sky location and masses is crucial in the search for electromagnetic counterparts. This search requires efficient and reliable measurement of such parameters prior to the actual merger of the two MBHs. With electromagnetic counterparts to MBH mergers, further questions about cosmology can be answered [30,31], as well as questions related to accretion processes and active galactic nuclei [32–35].

Traditionally, Bayesian inference techniques have been suggested for MBH binary searches and statistical analysis (e.g., [36–40]). Prior to the use of modern Bayesian sampling techniques [41], the Fisher information matrix (FIM) approach was an accepted method for determining the uncertainties of measurable quantities; however, it has been shown that FIM analyses do not encompass the necessary information to make strong statements about parameter estimation with MBH binaries (e.g., [39,42,43]). Similarly, it is important to include more advanced waveform models that better describe the physics of binary black hole coalescence. Aspects such as higher-order harmonic modes help to further constrain parameters, to remove systematic bias, and to further understand the characteristics of actual data analysis when the LISA mission is launched. Any difference between the waveform models used to create our templates and the true signals that will be observed by LISA will create systematic error [44]. However, in this work, we will inject the same waveform model as the model used for the templates, therefore bypassing this issue.

In this paper, we examine MBH parameter estimation under the new LISA sensitivity [1,45] by using modern, computationally enhanced inspiral-merger-ringdown waveforms that include higher-order harmonics and aligned spins. A variety of LISA-related parameter estimation studies have previously been performed for MBH binaries. The work in [46–50] addressed this problem with FIM estimates, the low-frequency approximation to the LISA response function, and inspiral-only Newtonian waveforms. The Mock LISA Data Challenges [51] expanded on these studies by developing Markov chain Monte Carlo (MCMC) methods toward the source search problem. Improving upon these analyses further, Bayesian inference techniques were developed in [36,38,39,52–58]. However, these works still employed inspiral-only waveforms.

The effect of using higher-order harmonic corrections in studies of MBH binaries was examined in [42,59–61]. Analyses of waveforms including merger and ringdown were examined with the FIM method in [61–64] and with a Bayesian approach in Babak *et al.* [65]. Babak *et al.* [66] used the FIM method to analyze parameter constraints with a variety of LISA designs. This study used inspiral waveforms with a reweighting scheme to account for the contribution of the merger and ringdown. Baibhav *et al.*

[67] recently studied the ability to constrain parameters using only ringdown signals with higher-order harmonics. This study used the FIM method, no noise, and an analytical description of the ringdown.

Marsat *et al.* [40] was the first study to use a full Bayesian analysis on waveforms that include inspiral, merger, and ringdown. This analysis examined Schwarzschild MBHs and featured a new prescription for the LISA response function by Marsat and Baker [68]. They used MCMC methods to characterize only extrinsic parameters, by employing a noiseless injection that included higher-order modes. In this work, we extend the findings of Marsat *et al.* [40] to include MBH signals with aligned spins, injected into a data stream that includes realistic LISA noise. We also analyze the full parameter space of MBH binaries, including both intrinsic and extrinsic quantities.

Many of the analyses mentioned above were performed using the original LISA sensitivity curve [69]. The changes to the LISA sensitivity curve in the most recent ESA-approved proposal [1] are expected to severely affect our ability to measure and characterize MBH binaries [11]. First, the SNR for a general MBH binary source will be slightly lower due to higher noise at the low-frequency end and a smaller arm length (2.5 Gm versus 5 Gm). Additionally, it can be shown that we will observe MBH mergers for less time with the new LISA sensitivity curve [1], compared to previous predictions under the classic LISA sensitivity curve [69]. For some mergers, we will observe their signal for only a few days, instead of the originally estimated weeks or months.

Figure 1 shows the SNR over time for the binaries (described in Sec. IV) that are studied in this work. For this plot, we employ a simplified calculation of the sky-location-, polarization-, and inclination-averaged SNR for the $l = m = 2$ mode, with the GWSNR CALC package [11] that uses PHenomD [70,71], Eq. (3), and an averaging factor of 16/5 [72]. It can be seen that, for the binaries tested, we will observe them for days, whereas with the classic LISA sensitivity, we can observe them for months. As a result of this difference in observation time, sky localization is not expected to be satisfactory until close to or after the merger. Early, accurate notifications for astronomers are critical for detecting electromagnetic counterparts to MBH binary mergers. Developing low-latency algorithms and computational infrastructure improvements for faster analysis is essential for this process.

In Sec. II, we discuss Bayesian inference in the context of MBH signals and the LISA mission. This includes our graphics processing unit (GPU)-accelerated implementations of the waveform, the LISA response, and the Bayesian likelihood, which prove to be very beneficial when examining a more complete version of the LISA analysis problem that includes a full noise realization. We detail our sampling methods in Sec. III. In Sec. IV, we discuss the sample datasets we analyzed, our ability to infer the proper parameters, and information gained about LISA

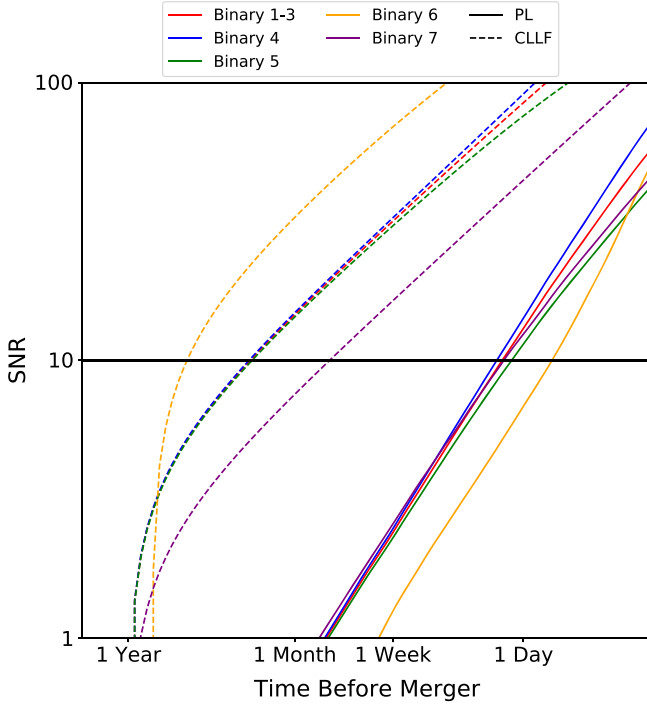


FIG. 1. SNR versus time before merger is shown above. The focus here is when sources become detectable at an SNR of ~ 10 . This calculation was performed using only the dominant $l = m = 2$ mode from PHenomD [70,71] as a part of the GWSNR CALC package by Katz and Larson [11]. Solid lines show the SNR computed using the proposed LISA sensitivity curve (PL; [73]). Calculations for the classic LISA sensitivity curve (CLLF; [69]) are shown with dashed lines. CLLF employs a more realistic low-frequency shape compared to the original classic curve by Larson *et al.* [69]; for more information on this construction, see Katz and Larson [11]. The binaries shown are those we analyze in this work; see Sec. IV for a detailed description of these sources. Binaries 1–3 are grouped into one example because their parameters are identical.

parameter estimation through running our tests. For our analysis, we use units with $G = c = 1$.

II. BAYESIAN INFERENCE WITH GRAVITATIONAL WAVES FROM MASSIVE BLACK HOLES

In the field of gravitational waves, the data stream, $d(t)$, consists of a possible signal, $s(t)$, as well as noise, $n(t)$ [$d(t) = s(t) + n(t)$]. In our theoretical study, we will generate the dataset by injecting a signal both with and without noise. We will then estimate the parameters of the source in the fabricated dataset using Bayesian inference techniques. Bayesian inference methods are based on Bayes rule given by

$$p(\vec{\Theta}|d, \Lambda) = \frac{p(d|\vec{\Theta}, \Lambda)\pi(\vec{\Theta}|\Lambda)}{p(d|\Lambda)}, \quad (1)$$

where Λ and $\vec{\Theta}$ represent the underlying model and model parameters, respectively. The probability on the left-hand side of this equation is the posterior probability, which is the main value we are concerned with determining. The posterior probability represents the probability for the model parameters given the observed data and assumed model. In our study, the parameter set exists in \mathbb{R}^D , where $D = 11$ is the number of dimensions we test. The parameter set we examine is $\{M_T, q, a_1, a_2, D_L, \phi_{\text{ref}}, i, \lambda, \beta, \psi, t_{\text{ref}}\}$, which represents, in order, the total mass, mass ratio, dimensionless spin of the larger black hole, dimensionless spin of the smaller black hole, luminosity distance, reference phase, inclination, ecliptic longitude, ecliptic latitude, polarization angle, and reference time. As discussed in Sec. II A, the spins of the two MBHs are assumed to be aligned to the orbital angular momentum vector, therefore, reducing our parameter space from $D = 15$ to $D = 11$ since the spins are represented only as magnitudes without vector angles. For our underlying model, Λ , we will only examine one waveform model. Therefore, we will drop the Λ throughout the rest of this paper.

The prior probability, $\pi(\vec{\Theta})$, is based on prior knowledge of the potential true parameters representing the injection. For our analysis, we assume limited prior knowledge by employing uniform priors across our parameters throughout an encompassing volume. Information on the priors used can be found in Table I.

The key element of Bayesian analysis involves the calculation of the likelihood, $\mathcal{L}(\vec{\Theta}) = p(d|\vec{\Theta})$. In the case of gravitational waves, \mathcal{L} is the probability for the observed data given a set of parameters and assumed underlying model. The negative log of the likelihood (NLL) is given by

TABLE I. Our priors used in our Bayesian inference are shown below. For each parameter we assume a uniform distribution between the lower and upper bounds. The lower (upper) bound on the distance is based on a redshift of 0.01 (100.0). The inclination (i) and ecliptic latitude (β) are listed in terms of the cosine and sine, respectively, of these values because these quantities represent uniform distributions for isotropically distributed angles on the sphere. The upper bound on t_{ref} represents 10 years in seconds (the quoted maximum lifetime of the mission).

Parameter	Lower bound	Upper bound
$\ln M_T$	$\ln(10^3)$	$\ln(10^9)$
q	0.05	1.0
a_1	-0.99	0.99
a_2	-0.99	0.99
$\ln D_L$	$\ln(0.044)$	$\ln(1298.00)$
ϕ_{ref}	0.0	2π
$\cos i$	-1.0	1.0
λ	0.0	2π
$\sin \beta$	-1.0	1.0
ψ	0.0	π
$\ln t_{\text{ref}}$	$\ln(1)$	$\ln(3.15576 \times 10^8)$

$$-\log \mathcal{L} = \frac{1}{2} \langle d - h | d - h \rangle = \frac{1}{2} (\langle d | d \rangle + \langle h | h \rangle - 2 \langle d | h \rangle), \quad (2)$$

where h is the template waveform, d is the data stream, and $\langle a | b \rangle$ is the noise-weighted inner product of the Fourier transforms of two time series $a(t)$ and $b(t)$. The Fourier transform of $a(t)$ is represented as $\tilde{a}(f) = \mathcal{F}\{a(t)\}$. The inner product is given by

$$\langle a | b \rangle = 2 \int_0^\infty \frac{\tilde{a}(f) \tilde{b}(f)^* + \tilde{a}(f)^* \tilde{b}(f)}{S_n(f)} df, \quad (3)$$

where $S_n(f)$ is the one-sided power spectral density (PSD) of the noise. The template, h , is built from the same model as the injected signal, s ; however, we use s to represent the true signal to avoid confusion. The optimal SNR attainable by a template h is $\langle h | h \rangle^{1/2}$. Here, we treat $S_n(f)$ as constant over the observation duration for simplicity. In reality, the noise is expected to vary slowly with time allowing for noise estimation on the order of \sim week when the instantaneous gravitational wave amplitude is well below the noise. For reference, the error in noise estimation is a second-order effect on the statistics illuminated by parameter estimation [74–76].

The denominator on the right-hand side of Eq. (1) is referred to as the evidence, $Z = p(d)$. The evidence is the marginalization of the likelihood over the parameter space,

$$Z = \int_{\vec{\theta}} \mathcal{L}(\vec{\theta}) \pi(\vec{\theta}) d\vec{\theta}. \quad (4)$$

The evidence, generally, helps estimate the fidelity of a model as well as compare models with one another. Computing the evidence directly in the gravitational wave setting is intractable. However, in the MCMC analysis described in Sec. III, the evidence enters only as a multiplicative constant. Therefore, we do not need to compute the evidence for our chosen analysis method. However, as we will describe in Sec. III, we employ a parallel tempering version of MCMC. Within the parallel tempering method, the evidence can be estimated using techniques such as thermodynamic integration [77,78] or stepping-stone sampling [79].

To determine the signals and templates, we use the frequency domain waveform model for binary black hole coalescence PhenomHM [70,71,80]. We refer the interested reader to Kalaghatgi *et al.* [81] for questions related to systematic errors in this waveform model. PhenomHM includes aligned spins, higher-order spherical harmonic modes, and all three stages of binary black hole coalescence: inspiral, merger, and ringdown. This is the first study of its kind for LISA MBH analysis to include all three of these properties simultaneously. In this work, we analyze the parameter estimation portion of LISA data analysis.

We do not perform the initial search for the source; instead, we assume a source has already been identified in the data stream. Additionally, we also ignore any effects from the superposition of signals from other sources simultaneously evolving in the LISA data.

The set of inputs to the waveform model are $\{M_T, q, a_1, a_2, D_L\}$. In our implementation of the waveform, we receive the amplitude, $A(f)$, and the phase, $\phi(f)$, of each input spherical harmonic mode (l, m) , given by

$$h_{lm} = A_{lm}(f) e^{i\phi_{lm}(f)}. \quad (5)$$

There are six spherical harmonic modes within the PhenomHM model: $(l, m) \in \{(2, 2), (3, 3), (4, 4), (2, 1), (3, 2), (4, 3)\}$. The (2,2) mode is the dominant mode. The other modes are referred to as higher harmonics or higher modes. An example of the PhenomHM amplitudes for each harmonic mode is shown in the left plot of Fig. 2.

The frequency bounds of the waveform are determined from the merger time of the signal. We choose a coalescence time, $t_{\text{coal}} \sim 1$ yr into the LISA observing window. In our methodology, for our true signal we set $t_{\text{coal}} = t_{\text{ref}} + t_0$. t_0 is on the order of a year and defines a temporal marker around which we analyze our signal. t_{ref} represents the fine-tuning of t_0 to the actual coalescence time; therefore, t_{ref} is on the scale of seconds to months. t_{coal} represents the time at which the waveform reaches f_{max} for the (2,2) mode, which is determined internally within PhenomHM to be the frequency at which $f^2 A(f)$ reaches a maximum (see [80] and [71] for more information). Due to the complicated nature of the time-frequency correspondence across multiple higher-order modes, we determine the time evolution of the system using the derivative of the phase and its relation to t_{coal} given by

$$t_{lm}(f) = t_{\text{coal}} - \frac{1}{2\pi} \frac{d\phi_{lm}(f)}{df}. \quad (6)$$

We make a cut in frequency when $t_{lm}(f)$ falls below zero for each individual mode, indicating the initial point of the waveform in the detector when LISA initially begins taking data.

We then apply the transfer function projecting the waveform onto the detector. The LISA response involves a complex time- and frequency-dependent transfer function [46,68,69,85]. We apply the fast Fourier domain response from Marsat and Baker [68]. Therefore, the signal in the LISA detector is represented by

$$h_{lm}^{\text{A,E,T}}(f, t_{lm}(f)) = \mathcal{T}^{\text{A,E,T}}(f, t_{lm}(f)) h_{lm}(f), \quad (7)$$

where $\mathcal{T}(f, t_{lm}(f))$ is the response transfer function and $\{A, E, T\}$ are the time-delay interferometry (TDI) channel indicators. TDI is necessary for LISA to suppress laser noise [86–90]. The A, E, and T channels are

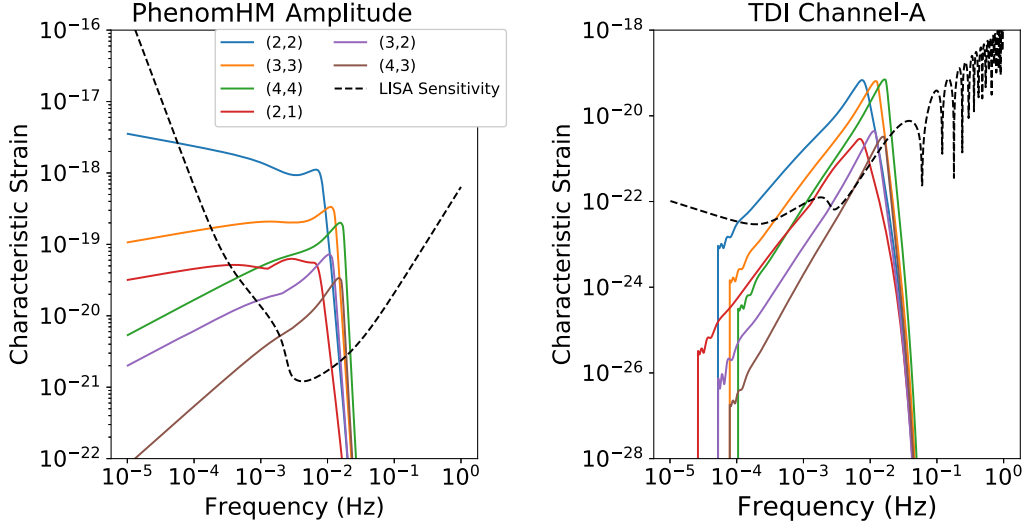


FIG. 2. The plots above display an example signal for Binary 1 from Sec. IV. Each harmonic mode is shown in a different color in the characteristic strain representation. The characteristic strain, h_c is given as $h_c^2 = f^2 |\tilde{h}(f)|^2$ [82,83]. The left plot displays the amplitude of each mode determined by PhenomHM [80] as well as the LISA sensitivity from [45] with a contribution from the Galactic background noise after one year of observation described in Babak *et al.* [84]. The left plot effectively displays this signal before the LISA response from Marsat and Baker [68] is applied. The right plot shows the same signal put through the LISA response. Here, we display TDI channel A. This is also shown in the characteristic strain representation. The modulations shown at low frequencies are due to the motion of the LISA constellation in its orbit about the Sun. The modes shown in the right plot are the same as those labeled in the left plot.

transformations of the original X, Y, and Z Michelson TDI observables [91] given by

$$A = \frac{1}{\sqrt{2}}(Z - X), \quad (8)$$

$$E = \frac{1}{\sqrt{6}}(X - 2Y + Z), \quad (9)$$

$$T = \frac{1}{\sqrt{3}}(X + Y + Z). \quad (10)$$

Employing channels A, E, and T, we assume the noise in each channel is uncorrelated, giving a diagonalized noise matrix [92,93]. $\mathcal{T}(f, t_{lm}(f))$ is determined using the extrinsic parameters, $\{\phi_{\text{ref}}, \iota, \lambda, \beta, \psi, t_{\text{ref}}\}$ (D_L is factored into h_{lm}); the time-frequency correspondence of the waveform; and the orbital and rotational properties of the LISA constellation. $\{\lambda, \beta, \psi, t_{\text{ref}}\}$ are used in the solar system barycenter (SSB) frame when determining $\mathcal{T}(f, t_{lm}(f))$; however, during MCMC runs, these parameters are sampled in the LISA frame and then converted to the SSB frame before waveform generation. For details on the construction and methodology of $\mathcal{T}(f, t_{lm}(f))$, please see Marsat and Baker [68] and Marsat *et al.* [40]. An example of the PhenomHM waveform amplitudes fed through the response function can be seen in the right plot in Fig. 2.

With the waveforms in each TDI channel, we need the PSD of the noise in each channel, $S_n^{A,E,T}$. We use the TDI package from the LISA Data Challenges Working Group

software collection to generate our PSD in each channel. Within the TDI package, we analyze the “SciRDv1” model [45] for the PSD with a contribution from the Galactic background noise predicted for 1 year into LISA observation. This Galactic background noise is described in Babak *et al.* [84]. Figure 2 shows the $\text{PSD}^{1/2}$ in the left plot, which includes the contribution from the Galactic background. In the right side plot of the same figure, the $\text{PSD}^{1/2}$ is shown in the TDI A channel. In Eq. (2), the inner product is really the sum of the inner products over all three channels:

$$\langle a|b \rangle = \sum_{j=A,E,T} \langle a^j|b^j \rangle. \quad (11)$$

We analyze both datasets with noise [$\tilde{n}(f) \neq 0$] and without noise [$\tilde{n}(f) = 0$]. However, the S_n weighting factor in Eq. (3) applies in both cases as it defines the SNR. When $\tilde{n}(f) = 0$, we are assuming standard LISA noise assumptions while analyzing the unlikely scenario the noise assumes its average value of zero. When we do generate noise as part of the data, we generate it in the frequency domain given by

$$|\tilde{n}(f)| = \mathcal{N}\left(0, \frac{1}{2\sqrt{\Delta f}}\right) \sqrt{S_n(f)}, \quad (12)$$

where $\mathcal{N}(\mu, \sigma)$ represents a normal distribution with mean μ and standard deviation σ , and Δf is the Fourier bin width equivalent to $1/T_{\text{obs}}$. It is then assigned a random phase from 0 to 2π based on a uniform distribution.

With the tools to calculate the likelihood, we will now discuss acceleration of the process using GPUs.

A. GPU accelerated waveforms and likelihoods

A GPU is a special piece of hardware designed for parallel computation. They were originally designed for graphics cards related to video games and other video-related areas; however, as of the mid-2000s, it was realized these units could be repurposed for general computational and academic use under the description: general purpose computing on graphics processing units (GPGPU). Here, we implement our PhenomHM waveforms, LISA response calculation, and likelihood computation in NVIDIA's proprietary GPU programming language CUDA [94]. GPUs have been used in gravitational wave analysis in Talbot *et al.* [95], where they have implemented PhenomPv2 [96], a waveform for the (2,2) mode with precessing spins, directly in CUDA for the LIGO Algorithmic Library [LAL; 97]. We take a different approach to our waveform implementation. Additionally, we have the first GPU implementation for the LISA fast response from Marsat and Baker [68]. We will introduce our implementation and then provide some timing results compared to a similar CPU-based program. For our timing tests, we used a Tesla V100 GPU and a Xeon Gold 6132 2.60 GHz CPU.

The main goal of our implementation is to perform as much of the computation as is feasible and beneficial on the GPU, while reducing the number and size of necessary memory transfers. Before beginning sampling where we will calculate the likelihood $>10^7$ times, we perform the aspects of this computation that only need to be completed once. These include inputting our data stream, d , PSD information, and transferring this information to the GPU. Following these initial steps, we begin the calculation process that occurs each time we call the likelihood function.

The first part of the PhenomHM creation process, which involves determining fitting constants from input parameters, is implemented in the C++ programming language. Structures containing these constants are then copied to the GPU. From here, the actual waveform creation from these fitting constants is implemented in parallel across the GPU because each frequency point is independent of one another. With the waveform constants, the amplitude and phase are generated and stored on the GPU. We then calculate $t(f)$ using Eq. (6). $t(f)$ is input into the response function with the extrinsic parameters. $\mathcal{T}(f, t_{lm}(f))$ is calculated on the GPU and stored on the GPU for each frequency point. At this stage, $A(f)$, $\phi(f)$, and $\mathcal{T}(f, t_{lm}(f))$ are represented as smooth functions with $\sim 2^{10}$ frequency points. In LISA analysis, the data streams can be represented by upwards of 10^6 – 10^7 points depending on the sampling frequency and duration of observation. Therefore, to maintain speed in data analysis, these smooth functions describing the waveform and detector response

are interpolated to the resolution desired. We designed a cubic spline interpolation algorithm, for each of these smooth functions, based on the tridiagonal nature of the cubic spline coefficient solution in the SciPy library [98] and the sparse matrix computational tools included in the cuSPARSE library provided by NVIDIA [99]. This allowed us to avoid memory transfers necessary to perform spline interpolation on the CPU. The final step of waveform creation is to interpolate all input smooth functions to create the signal in each TDI channel. In this step, we weight by the PSD. With the waveform calculated and on the GPU, we use the cuBLAS library [100] to compute complex dot products for the discretized inner product in Eqs. (2) and (11). The likelihood is then calculated and returned to the sampling program.

For ease of usage, our program is wrapped into PYTHON using CYTHON [101]. This allows us to run PYTHON sampling packages using our GPU-accelerated likelihood calculation. CYTHON generally does not interface with the CUDA-based compiler NVCC. However, we wrote a special wrapping for it based on McGibbon and Zhao [102].

Speed comparisons are shown in Fig. 3. Three configurations are shown: 1 CPU; 1 CPU and 1 GPU (host and device); and 28 CPUS and 2 GPUs. The last configuration vectorizes the computation over an array of sources. Therefore, these timings are run in vectorized form and then divided by the number of binaries in the array. As the points in the waveform increase, the difference between the CPU and GPU grows. The time of GPU evaluations barely increases until $\sim 10^6$ points. Above this number, the GPU is saturated, running a high percentage of its available threads constantly.

This increase in speed allows us to perform our analysis in a unique way. Traditionally, in a zero-noise representation, each spherical harmonic mode is treated separately in the likelihood computation (see Marsat *et al.* [40] for an example of this method). CPUs are not fast enough to efficiently compute the number of points necessary for combining all modes at a high enough resolution to preserve the fine structure created during the mixing of the higher modes. However, this is a balancing act because as more modes are added, the cross terms between modes in Eq. (2) become extensive and prohibitive. With GPUs, the time for the calculation allows us to forgo this question and produce templates at a high enough resolution to combine the modes into a one-dimensional data stream. This prevents the calculation of cross terms and allows us to calculate one inner product per TDI channel. Additionally, CPU speed prohibits the timely evaluation of likelihoods with noise infused in the injection. With $T_{\text{obs}} \sim \text{years}$, Δf in the Fourier transform is $\sim 10^{-8}$ Hz, leading to data streams with $\sim 10^7$ points. Similarly, for a typical source expected for LISA with $M_T = 10^6 M_{\odot}$, the waveform must be calculated at $\sim 5 \times 10^6$ points. For lower masses, this is even larger. Therefore, CPUs have large difficulty

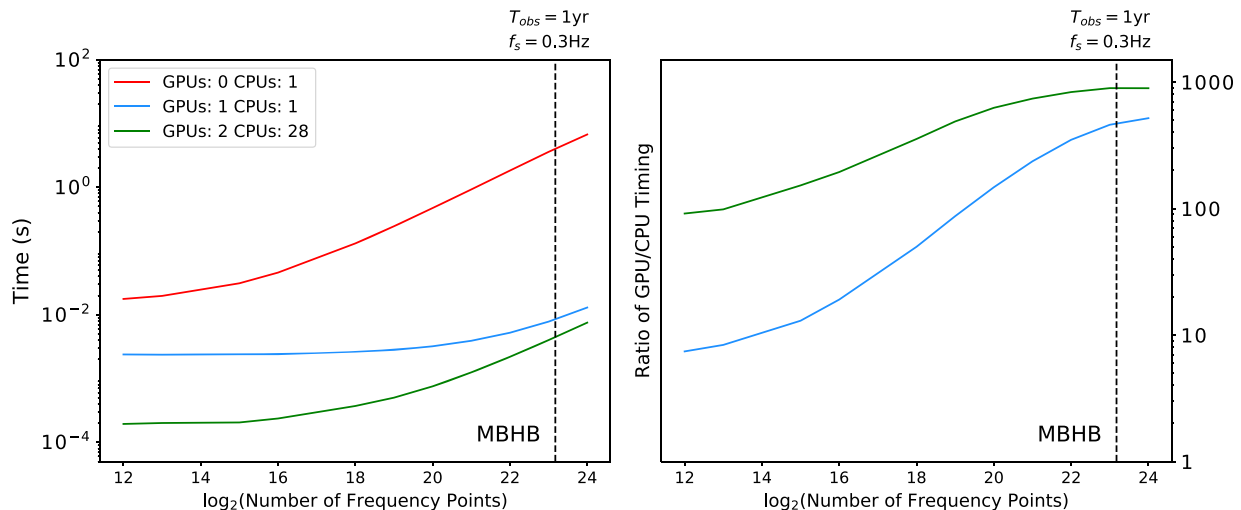


FIG. 3. Timing results for our GPU-accelerated likelihood computations are shown above. The timing shown represents the total time to generate the waveform, apply the response, and calculate the likelihood in all three channels. The left plot represents the actual time per likelihood calculation. The CPU-only results are shown in red. For these results, no likelihood approximations were included; the calculation is performed in a so-called “brute-force” way to compare directly with GPU timings. The blue line shows the usage of one CPU and one GPU. This timing is greatly affected by the time spent on the CPU performing preprocessing for the PhenomHM [80] waveform constants. The green line shows the speedup using 28 CPUs and 2 GPUs. This allows us to parallelize the preprocessing and then let the GPUs split the waveform production reducing the calculation time even further. The right plot translates the left plot into an “effective number of CPUs” by dividing the CPU-only time by the tests with at least one GPU. The vertical dashed line shows the number of data points in the Fourier transform of a data stream sampled at 0.3 Hz for one year of observation.

performing this calculation with noise involved without including approximations and other tricks to speed up the calculation. The GPUs, on the other hand, facilitate the use of the brute-force calculation, without approximations. In the future, we may test different approximations in our GPU implementation to gain even more performance.

III. SAMPLING METHODS

There has been a large variety of methods developed to illuminate the posterior distribution, $p(\vec{\Theta}|D)$, as well as to calculate the evidence, Z . Two commonly used methods include nested sampling [103,104] and MCMC methods, for which the most commonly used type is the Metropolis-Hastings (M-H) algorithm [105–107]. Nested sampling is generally used to estimate the evidence, Z , by effectively performing a numerical integration across the prior volume. MCMC, on the other hand, is used to sample directly from the posterior to produce marginalized posterior distributions across the parameters tested. In this work, we will employ only MCMC techniques to illuminate our posterior distributions for the injected signals we examine. However, we have tested our accelerated likelihood with nested sampling techniques and have confirmed its possible usage as a search method. This will be included in future work. For actual sampling, we use variants of the typical MCMC techniques. Below, we will give a brief description of the algorithms used, but refer the interested reader to the cited papers for further details on the specific algorithms and their implementations.

A. Markov chain Monte Carlo

The implementation we use for MCMC methods is based on the Python packages EMCEE [108] and PTMCEE [109]. Please see their release papers and documentation for details on specific constructions. The most popular form of MCMC is the M-H algorithm. Starting with a position $\vec{\Theta}_t$ at step t , a new position, $\vec{\Theta}_{t+1}$, is sampled from a transition distribution $Q(\vec{\Theta}_{t+1}|\vec{\Theta}_t)$. This new point is then accepted with a probability given by

$$\min\left(1, \frac{\mathcal{L}(\vec{\Theta}_{t+1}) Q(\vec{\Theta}_t|\vec{\Theta}_{t+1})}{\mathcal{L}(\vec{\Theta}_t) Q(\vec{\Theta}_{t+1}|\vec{\Theta}_t)}\right). \quad (13)$$

A usual choice for Q is a multivariate Gaussian distribution centered around $\vec{\Theta}_t$. This procedure has strong dependence on tuning the distribution for Q (the covariance matrix for a multivariate Gaussian) with order $\sim D^2$ tuning parameters. Additionally, the M-H algorithm can be slow to convergence if tuning is not optimal [108].

Goodman and Weare [110] proposed a different method for MCMC sampling that significantly outperforms M-H in many situations. Specifically, this means the autocorrelation time (see Sec. III C) of the MCMC chains is much less, indicating fewer likelihood evaluations per independent sample. This method is referred to as affine invariant MCMC (AIMCMC). In AIMCMC, the sampler evolves an ensemble of K walkers $S = \{\vec{\Theta}_k\}$. The algorithm proposes a new position for the k th walker based on the

other $K - 1$ walkers in the ensemble ($S_{[k]} = \{\vec{\Theta}_j, \forall j \neq k\}$). First, a walker is chosen at random from $S_{[k]}$, represented by $\vec{\Theta}_j$. Then, a new position is proposed given by

$$\vec{\Theta}_{k,t} \rightarrow \vec{\Theta}_{k,t+1} = \vec{\Theta}_j + Y[\vec{\Theta}_{k,t} - \vec{\Theta}_j], \quad (14)$$

where Y is a random variable distributed according to the probability density $g(Y = y)$, given by [110]

$$g(y) = \begin{cases} \frac{1}{\sqrt{y}} & \text{if } y \in [\frac{1}{a}, a], \\ 0 & \text{otherwise,} \end{cases} \quad (15)$$

where a is a tuning parameter that Goodman and Weare [110] set to 2. This proposal is referred to as the ‘‘stretch’’ proposal. The proposed point will be accepted according to the probability, a_{str} , given by

$$a_{\text{str}} = \min \left(1, Y^{(D-1)} \frac{\mathcal{L}(\vec{\Theta}_{k,t+1})}{\mathcal{L}(\vec{\Theta}_{k,t})} \right). \quad (16)$$

B. Parallel tempering

In order to efficiently explore the entire prior range, we employ parallel tempering [111,112]. Specifically, we use the implementation from Vousden *et al.* [109] and its associated PYTHON package, PTEMCEE. Here, we present the main aspects of parallel tempering, but we refer the interested reader to Vousden *et al.* [109], and sources within, for more detailed information.

In parallel tempering, the likelihood in Eq. (1) becomes $\mathcal{L}(\vec{\Theta})^{1/T_i}$, where T_i is the temperature of a given chain. Groups of MCMC walkers are assigned to a specific temperature arranged in a ladder with N rungs: $T_i = T_1 < T_2 < \dots < T_N$. T_1 is 1, indicating chains assigned to that temperature represent the target distribution. As temperature increases, the tempered distribution becomes more and more representative of the prior distribution rather than the target distribution.

Each chain explores its tempered distribution using the AIMCMC algorithm with stretch proposals. At specified intervals, chains can swap temperatures according to an M-H acceptance criterion given by [109]

$$a_{\text{swap}} = \min \left(1, \left[\frac{\mathcal{L}(\vec{\Theta}_i)}{\mathcal{L}(\vec{\Theta}_j)} \right]^{\beta_j - \beta_i} \right), \quad (17)$$

where $\beta_i = 1/T_i$ is the inverse temperature of chain i . The form of Eq. (17) indicates that temperature swaps usually occur between chains at adjacent rungs in the temperature ladder. We set the max temperature to infinity per the PTEMCEE documentation. Setting the max temperature to infinity means the highest temperature chain is exclusively exploring the prior distribution. We also use the adaptive tempering option. For more details on the adaptive method, please see the original paper.

The parallel tempered run for each binary was performed under the same settings. With the highest temperature, T_{10} , set to ∞ , the remaining temperatures are geometrically spaced set by PTEMCEE defaults using log spacing between 1 and $(2.04807)^8$ (2.04807 is chosen based on the dimensionality of the parameter space). We set the scale factor (a) in Eq. (15) to 1.15. We found the default setting of 2 was showing an undesirably small acceptance fraction for proposed jumps. We run a burn in of 10^4 steps for each run. In Marsat *et al.* [40], it is shown that we can generally expect eight sky location posterior modes in the LISA reference frame (as opposed to the SSB reference frame). The longitudinal modes are located at values separated by $\pi/2$ from the true longitude value. Each of these four values has an associated latitudinal value that is either above or below the LISA orbital plane at $\pm\beta$. We use this information to ensure faster convergence in our sampler by placing four walkers in each temperature group at each of the eight sky modes. It is clear from our analysis that our burn in phase allows the walkers to spread out accordingly away from modes that are not likely. We also verified that if we start all the walkers on the correct mode, the walkers spread out and locate the other modes if there is posterior weight there thanks to the tempering technique.

C. MCMC autocorrelation time

The posterior samples generated in an MCMC analysis are not independent. Therefore, we estimate the autocorrelation within a chain, referred to as the autocorrelation time, τ_f , following Sokal [113] and Foreman-Mackey [114]. With multiple chains sampled based on methods described above, we work to estimate the mean, $\hat{\mu}$, and variance, $\hat{\sigma}^2$, of the posterior distribution. The sampling variance, σ^2 , on these estimators is given by

$$\sigma^2 = \frac{\tau_f}{N} \hat{\sigma}^2. \quad (18)$$

Therefore, the effective sample size (ESS) is N/τ_f , representing the number of samples necessary to reduce the variance in the estimator to an acceptable value.

We estimate the autocorrelation time with the estimator for the normalized autocorrelation function, $\hat{\rho}_f(\tau)$. This is based on the chain generated for each walker, $\{f_n\}_{n=1}^N$, and is given by

$$\hat{\rho}_f(\tau) = \frac{\hat{c}_f(\tau)}{\hat{c}_f(0)}, \quad (19)$$

with

$$\hat{c}_f(\tau) = \frac{1}{N - \tau} \sum_{n=1}^{N-\tau} (f_n - \mu_f)(f_{n+\tau} - \mu_f). \quad (20)$$

TABLE II. The injection parameters of each binary tested. The first row shows the parameter while the second row shows the units for each parameter. The final column displays the overall SNR for each source. It must be noted this table is meant to be a summary. The actual injection parameters are known to a much higher precision. As a reminder, the actual coalescence time is $t_{\text{coal}} = t_{\text{ref}} + t_0$, where t_0 is 1 year.

Binary	M_T	q	a_1	a_2	D_L	ϕ_{ref}	ι	λ	β	ψ	t_{ref}	SNR
...	M_{\odot}	$(q \leq 1)$	Gpc	s	...
1	2×10^6	1/3	0.0	0.0	36.59	2.14	1.05	-0.024	0.62	2.03	50.25	588
2 ^a	2×10^6	1/3	0.0	0.0	36.59	2.14	1.05	-0.024	0.62	2.03	50.25	588
3 ^b	2×10^6	1/3	0.0	0.0	36.59	2.14	1.05	-0.024	0.62	2.03	50.25	588
4	2×10^6	1/3	0.85	0.88	36.59	2.14	1.05	-0.024	0.62	2.03	50.25	878
5	2×10^6	1/3	-0.83	-0.89	36.59	2.14	1.05	-0.024	0.62	2.03	50.25	473
6	4×10^7	1/5	0.0	0.0	15.93	3.10	1.34	4.21	-0.73	0.14	23.93	310
7 ^a	3×10^5	7/10	0.0	0.0	70.58	1.22	1.32	2.24	-0.22	2.73	94.33	40

^aThis binary was injected without a noise realization.

^b a_1 and a_2 are fixed during sampling.

The mean of the chain, μ_f , is given by

$$\frac{1}{N} \sum_{n=1}^N f_n. \quad (21)$$

The integrated autocorrelation estimator is given by

$$\hat{\tau}_f(M) = 1 + 2 \sum_{\tau=1}^M \hat{\rho}_f(\tau), \quad (22)$$

where $M \ll N$. Sokal [113] suggests using the smallest value of M that satisfies $M \geq 5\hat{\tau}_f(M)$. With the multiple chains generated with EMCEE, the variance in the estimator is reduced. Therefore, according to Foreman-Mackey [114], this procedure is satisfactory for chains longer than $50\hat{\tau}_f$. We thin the chains by this value before plotting posteriors to achieve a set of effectively independent samples.

We run our sampler without checking the ESS. We ensure the ESS is greater than 10^4 at the culmination of sampling. We report the autocorrelation time determined with Eq. (22), the ESS, and the acceptance fraction for each run in Table VII in the Appendix.

IV. RESULTS AND DISCUSSION

Our first aim of this paper is to build upon Marsat *et al.* [40] by adding aligned spins to our waveform and injecting noise within our fabricated data streams. To achieve this we run four different sets of binary parameters. We begin with the main source used in Marsat *et al.* [40] with $\{M_T = 2 \times 10^6 M_{\odot}, q = 1/3, z = 4\}$ (injection parameters for all binaries tested are shown in Table II). Binary 1 is examined with Schwarzschild MBHs and injected noise. However, even with the $a_1 = a_2 = 0$ injection, we allow the sampler to examine other spins maintaining a $D = 11$ dimensional parameter space.

Binary 2 is the same as binary 1, but with a pure waveform injection with no noise realization (also referred to as the zero-noise representation). This setup allows us to determine what the posterior would look like averaged over many runs with noise.

For Binary 3, which is a similar injection to Binary 1, we fix the spins in the sampler so that we are sampling in a $D = 9$ dimensional space, which allows for a closer comparison with Marsat *et al.* [40] in terms of the sampling methods (we do use a different waveform model). In addition to Schwarzschild injections, we want to examine how including spinning MBHs changes the posterior distributions while holding the other parameters constant.

Binary 4 represents the same set of MBHs as Binaries 1–3 with $a_1 \approx 0.85$ and $a_2 \approx 0.88$. A set of antialigned spins is tested in Binary 5 with $a_1 \approx -0.83$ and $a_2 \approx -0.89$. The nonzero spins were chosen from a uniform distribution between $+(-)0.75$ and $+(-)0.9$ for the aligned (antialigned) binary.

We also test a larger and smaller binary in terms of the total mass. The larger binary has $\{M_T = 4 \times 10^7 M_{\odot}, q = 1/5, z = 2\}$ and the smaller binary has $\{M_T = 3 \times 10^5 M_{\odot}, q = 7/10, z = 7\}$. Both large and small binaries were injected with Schwarzschild MBHs, but the spins were allowed to vary in sampling. The large binary was chosen to examine MBH binaries that merge at lower frequencies than the center of the LISA band. It was also injected with noise.

The small binary was chosen to represent large seed MBHs at a higher redshift. During repeated attempts to analyze the small binary injected in noise with a variety of trial sampler settings, we were unable to attain a converged posterior distribution because the low temperature chains were unable to explore the likelihood surface efficiently and expand throughout the parameter space to the level expected. This is likely due to the combination of a lower SNR, as well as the inability to gain extra information from the higher-order modes. When we

TABLE III. The SNR contribution from each mode is shown below in addition to the full SNR of the source. The mode SNRs are computed in isolation to avoid confusion from mode mixing. This is to highlight in a general sense the contribution to the total SNR from each mode. Since the various modes mix with each other, the total SNR is not equal to the quadrature sum of all of the modes.

Binary	All	(2,2)	(3,3)	(4,4)	(2,1)	(3,2)	(4,3)
1-3	588	552	190	97	36	21	5
4	878	820	279	146	49	31	8
5	474	443	150	74	50	17	4
6	311	230	217	197	25	57	38
7	40	40	3	2	< 1	< 1	< 1

lowered the temperatures in an effort to suppress the effect of the posterior tails on the ability for the walkers to maneuver the likelihood surface, we found the chains exhibited $\tau_f = 61$ with an ESS of 15744, which are usually positive indicators. We believe the posteriors would converge to the proper distribution if we were able to run our sampler for a longer time. Even with our accelerated likelihood computations, this became difficult. Therefore, we present results for the small binary in the zero-noise representation. In future work we will further examine this problem and work toward analyzing noise injections at lower masses.

All extrinsic quantities for the large and small binaries were sampled from uniform distributions from the entirety of the prior domain. See Table II for a full breakdown of the parameters used. Table III shows the SNR of each harmonic mode for each binary tested. Please note the modes do not directly add in quadrature due to mode mixing. Therefore, the total SNR shown in the “all” column will not be the

quadrature sum of the singular mode values. This is purely to indicate the type of effect each mode has in the characterization of each binary. It is clear from the values seen in Table III that more higher modes will be needed to fully describe MBH binary systems since the likelihood is related to the SNR^2 [40].

In the following analysis, we will present subsections of the overall corner plots representing the parameter posterior distributions; we focus on specific interesting and/or important posterior distributions. First, we will discuss intrinsic parameters, followed by a discussion on sky localization constraints and other extrinsic quantities. The full corner plot for each binary is shown in Sec. A of Appendix. Recovered parameter values for $\ln M_T$, q , a_1 , a_2 , $\ln D_L$, ϕ_{ref} , and $\ln t_{\text{ref}}$, including their means and 1σ errors, are shown in Tables IV–VI. Binaries 1, 2, and 3 are shown in Table IV. Table V contains parameters for Binaries 4 and 5. Binaries 6 and 7 are shown in Table VI. These tables do not include the sky location and orientation because these parameters are inherently multimodal rendering their one-dimensional means and 1σ errors not representative of the true recovery values.

A. Intrinsic parameters

The first marginalized posterior we address is the two-dimensional posterior for $\ln M_T$ and q . Figure 4 shows this parameter space for each binary. There are two aspects to point out. First, Binaries 1–5 with the same injection mass and mass ratio, as well as the small binary (Binary 7), show an anticorrelated behavior, while the posterior for the larger binary (Binary 6) shows a positively correlated behavior. The difference involves the position of the merger and ringdown over the sensitivity curve, or more specifically,

TABLE IV. The recovered parameters in comparison to the injected parameters for Binaries 1, 2, and 3. The one-dimensional 1σ errors for each parameter are shown in addition to the recovered mean. The sky position and orientation parameters are not shown because they are inherently multimodal due to, at minimum, reflection across the LISA orbital plane (see Fig. 6). Binary 1 represents similar parameters to those shown in Marsat *et al.* [40] with Schwarzschild MBHs injected; however, the sampler is free to examine spins other than zero. This binary also included a noise realization generated according to Eq. (12). Binary 2 is the same injection and sampler settings as Binary 1. This binary, however, does not include a noise realization allowing the likelihood to be calculated in a zero-noise representation. Binary 3 has the same injection as Binary 1 with a noise realization. However, the spin parameters are fixed during sampling, therefore, covering a parameter space of dimensionality $D = 9$ rather than $D = 11$. This is why “N/A” is shown for the recovery of these spin parameters.

Parameter	Injection	Binary 1	Binary 2	Binary 3
		Recovered	Recovered	Recovered
$\ln M_T$	14.50866	14.50913 ^{+0.00075} _{-0.00075}	14.50857 ^{+0.00080} _{-0.00081}	14.50787 ^{+0.00061} _{-0.00054}
q	0.33333	0.33296 ^{+0.00085} _{-0.00085}	0.33330 ^{+0.00095} _{-0.00080}	0.33439 ^{+0.00072} _{-0.00092}
a_1	0.0	-0.0047 ^{+0.0049} _{-0.0046}	-0.0012 ^{+0.0048} _{-0.0047}	N/A
a_2	0.0	0.028 ^{+0.018} _{-0.019}	0.003 ^{+0.018} _{-0.018}	N/A
$\ln D_L$	3.581	3.603687 ^{+0.020} _{-0.020}	3.604 ^{+0.021} _{-0.021}	3.617 ^{+0.018} _{-0.018}
ϕ_{ref}	2.140	2.36 ^{+0.14} _{-0.15}	2.16 ^{+0.14} _{-0.14}	2.116 ^{+0.022} _{-0.018}
t_{ref}	3.9169	3.69 ^{+0.17} _{-0.20}	3.89 ^{+0.13} _{-0.15}	3.9238 ^{+0.0074} _{-0.0081}

TABLE V. Similar to Table IV, the parameter recovery for Binaries 4 and 5 are shown. Binary 4 has the same injection parameters and sampler settings as Binary 1 with the exception of the injected spin parameters. This binary is injected in a spin configuration with high aligned spins for each MBH. Binary 5 is similar to Binary 4, but it is injected in a high antialigned spin configuration.

Parameter	Binary 4		Binary 5	
	Injection	Recovered	Injection	Recovered
$\ln M_T$	14.508658	14.50865 ^{+0.00037} _{-0.00042}	14.5087	14.5090 ^{+0.0013} _{-0.0015}
q	0.33333	0.33335 ^{+0.00052} _{-0.00047}	0.3333	0.3323 ^{+0.0016} _{-0.0014}
a_1	0.85697	0.85678 ^{+0.00059} _{-0.00059}	-0.8298	-0.8291 ^{+0.0080} _{-0.0079}
a_2	0.8830	0.8846 ^{+0.0033} _{-0.0033}	-0.887	-0.895 ^{+0.027} _{-0.030}
$\ln D_L$	3.600	3.615 ^{+0.015} _{-0.015}	3.600	3.591 ^{+0.027} _{-0.026}
ϕ_{ref}	2.140	2.150 ^{+0.029} _{-0.029}	2.14	2.11 ^{+0.16} _{-0.17}
t_{ref}	3.917	3.907 ^{+0.022} _{-0.023}	3.92	3.95 ^{+0.18} _{-0.20}

where the majority of the signal is accumulated. For the binaries with $M_T = 2 \times 10^6 M_\odot$, the peak of the merger and ringdown occurs at frequencies around or above 10^{-2} Hz. In this region, the sensitivity curve has a positive slope (see Fig. 2). At small deviations from the highest likelihood point, the signal, as it changes with varying parameters, must follow the shape of the sensitivity in order to maintain a similar likelihood value, thereby creating this observed correlation. The effect of increasing the total mass by a small perturbation will move the signal slightly to higher strains and lower frequencies. Therefore, in order for this signal with a slightly larger mass to follow the sensitivity curve, we will need to decrease the mass ratio, which causes the signal to decrease in strain. For this reason, we see an anticorrelation between these two parameters. Once again, in the limit of small deviations, the actual signal shape will not change significantly.

This description also applies to the small binary because it evolves at higher frequencies than Binaries 1–5, indicating the sensitivity curve is also sloping upwards where a majority of its signal is accumulated. However, since the mass ratio is not strongly constrained due to the lack of information from higher modes, the morphology of the signal does change slightly over the range of mass ratios observed in the posterior distribution. This effect adds curvature to the posterior ellipse.

The opposite is true for the large binary. At the frequencies at which the large binary merges and rings down ($\sim 10^{-3}$ Hz), the sensitivity curve has a negative slope. As the total mass is increased, the signal will move to higher strains and lower frequencies. However, due to the slope of the sensitivity curve, an increase in the mass ratio is also needed to ensure the signal actually follows this slope. Hence, we see a positive correlation between the two quantities. In addition to the correlation observed, the small binary exhibits a curved ellipsoidal shape and a mass ratio

TABLE VI. Recovered parameters for Binaries 6 and 7 are shown below similar to Table IV. Binary 6 represents a completely different binary from Binaries 1–5 in terms of injection parameters. Its total mass is chosen to represent a larger binary about an order of magnitude larger than Binaries 1–5. It is, however, also injected with Schwarzschild MBHs in a noise realization. Binary 7 represents a binary with total mass about an order of magnitude smaller than Binaries 1–5. It is also injected with Schwarzschild MBHs. However, this binary, like Binary 2, is analyzed in a zero-noise representation. This is due to issues achieving convergence with our sampler on this binary when it was injected with a noise realization. Achieving convergence in the presence of noise for this binary is a topic for future work. The spin parameters and ϕ_{ref} for Binary 7 show a multimodal structure; therefore, their associated mean values and errors do not resemble the actual recovery that is observed in the posterior distributions. For this reason, these quantities are labeled with “Multi.” Please see the full posterior in Fig. 13 for more information.

Parameter	Binary 6		Binary 7	
	Injection	Recovered	Injection	Recovered
$\ln M_T$	17.5044	17.5033 ^{+0.0017} _{-0.0021}	12.612	12.609 ^{+0.012} _{-0.009}
q	0.2000	0.1990 ^{+0.0009} _{-0.0010}	0.700	0.716 ^{+0.079} _{-0.072}
a_1	0.0	-0.0044 ^{+0.0042} _{-0.0053}	0.0	Multi
a_2	0.0	-0.0038 ^{+0.0039} _{-0.0051}	0.0	Multi
$\ln D_L$	2.77	2.81 ^{+0.19} _{-0.13}	4.27	4.58 ^{+0.35} _{-0.24}
ϕ_{ref}	3.098	3.103 ^{+0.013} _{-0.036}	2.4	Multi
t_{ref}	3.2	2.6 ^{+1.4} _{-1.6}	6.849	6.857 ^{+0.010} _{-0.014}

that is not strongly constrained. This is most likely due to the low or negligible signal contributed from the higher modes (see Table III).

When comparing Binaries 1–5 with one another, we see that most of the posterior distributions per parameter set are similar in character. There are a few interesting differences. First, most parameters, for all of these binaries, have similar errors that vary in comparison to one another in proportion with the SNR for each source (see Table III). However, this is not true for the reference values, ϕ_{ref} and t_{ref} , for Binary 3 (fixed spins during sampling with $D = 9$). Even though this binary has the exact same injection waveform as Binaries 1 and 2, the reference values are constrained to a higher degree, with error values similar to the higher SNR Binary 4 (spin up). The reference frequency at which these reference values are set is determined in part by the symmetric combination of the mass-weighted spins [70,71]. Therefore, when we fix the spins during sampling, we are ensuring that two of four quantities (the others are total mass and mass ratio) that determine the reference frequency are fixed. This allows for better constraints on these reference values. Similarly, in Figs. 7–11 we can see that Binary 3 (fixed spins during sampling) has unique posterior shapes when comparing the total mass and mass ratio to these reference values. It can be seen these posteriors show a higher degree of

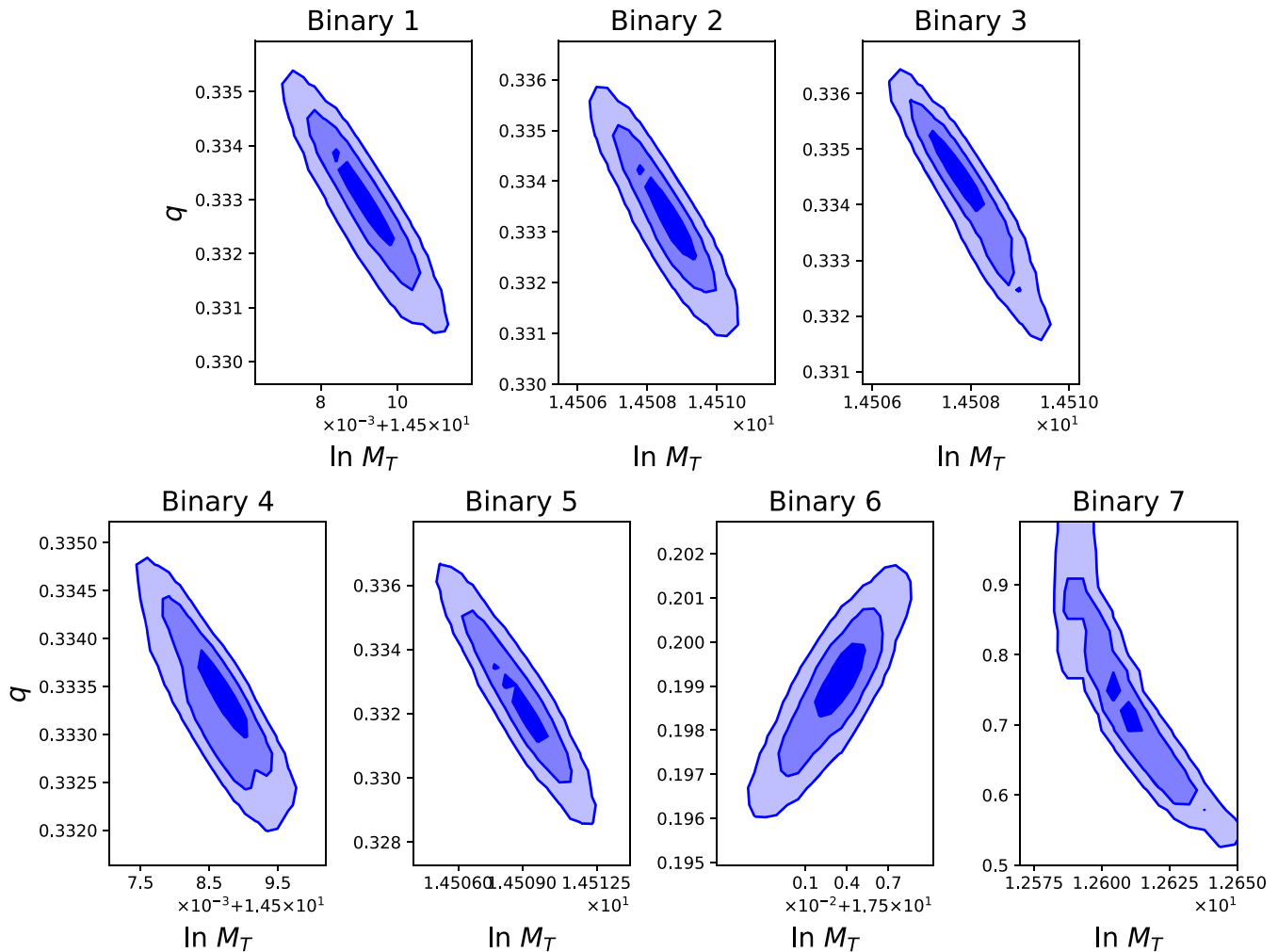


FIG. 4. The two-dimensional posterior distributions of $\ln M_T$ versus q for all binaries tested are shown here. The binary is labeled in the title of each plot. All binaries, with the exception of Binary 6, display an anticorrelation between these parameters. Binary 6 shows a positive correlation. As explained in Sec. IV, this is due to the slope of the sensitivity curve at frequencies around the merger of the binary where the signal is strongest.

correlation or anticorrelation. Once again, with the spins fixed, the mass ratio and total mass are the two quantities that determine the reference frequency, therefore, causing the magnitude of correlation with these quantities to increase.

When examining the large binary, it is clear its posterior distributions are much less ellipsoidal than those shown in Binaries 1–5. The base reason for this is the frequency range over which we observe this signal. We receive minimal information from the inspiral of this binary, receiving all information from the merger and ringdown. One peculiar intrinsic parameter posterior of note for the larger binary is the a_1 versus a_2 posterior as it is far from elliptical, and actually strongly asymmetric in relation to an antisymmetric combination of mass-weighted spins. To further investigate this observation, we analyzed two more injections with the same parameters as Binary 6, but with differing mass ratios. Figure 5 shows three binaries similar to Binary 6 in the plane of the mass-weighted spin values. The $q = 1/5$ binary is

Binary 6. The other two are injected with all the same parameters, but differing mass ratios. All three binaries exhibit similar behavior in this plane. Toward the upper left of the plot, the likelihood surface drops off steeply. On the contrary, toward the lower right, the likelihood surface displays a shallow fall. The dashed lines represent a line with the slope equal to the mass ratio. Above (below) this line, the mass-weighted spin of the primary is smaller (greater) than the secondary. The purpose of this line is to illustrate that this steep dropoff is generally parallel to a line with a slope equal to the mass ratio. For this large binary, where we receive minimal information from the inspiral, the swapping of the mass weighted spins causes the signal to change in comparison to the injection with varying rapidity that indicates the broken symmetry of the masses greatly affects the ability of the template to match the injection. At $q = 1/5$ and $q = 1/3$, the posteriors generally stay parallel in character to that line. As the mass ratio is increased to

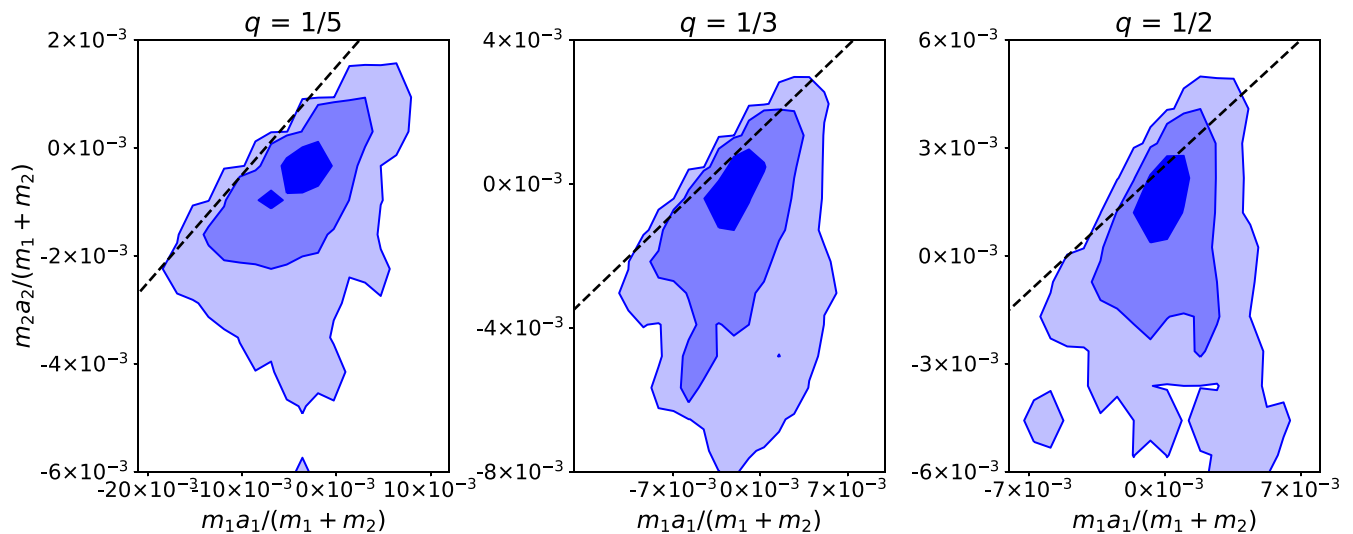


FIG. 5. The two-dimensional posterior distributions for the mass-weighted spin parameters for the large binary (Binary 6). The posteriors show a unique nonellipsoidal shape that is not observed for the other binaries. The dashed line in each plot has a slope equal to the mass ratio to show the steep likelihood dropoff toward the upper left is approximately parallel to this line. The large binary exhibits almost its entire signal in the merger and ringdown portion of coalescence. Therefore, it is more sensitive to quantities such as the reference frequency that sets the location of the merger ringdown. The reference frequency, f_{ref} , is set by the symmetric combination of the mass-weighted spins. Since this binary has similar values for a_1 and a_2 , the injection has a larger mass-weighted contribution from its primary. For this reason, we see the likelihood is very sensitive to changes in the spins when the mass-weighted spin contribution from the secondary is larger than the primary's contribution.

$q = 1/2$, the mass weighting tends toward a more even symmetry, causing this effect to decrease; however, it still shows this nonelliptical behavior.

The small binary (binary 7) has many unique posterior distributions due to its lower SNR and minimal information contributed from its higher harmonic modes (see Fig. 13). For example, the spin posteriors show a multimodal structure. a_1 shows a virtually equivalent mode at a slightly higher spin of ~ 0.2 . Similarly, a_2 shows a corresponding second mode at a lower spin of ~ -0.3 . The ratio of these secondary spin modes is approximately equal to the mass ratio due to the effect of the mass-weighted symmetric spin during waveform construction. Even with this multimodal structure, the a_1 versus a_2 posterior shows an ellipsoidal shape surrounding the two local maxima.

B. Sky location and other extrinsic parameters

In terms of the ecliptic longitude and latitude, we observe what is generally expected. Figure 6 shows the $1-3\sigma$ contours for the sky localization of each binary. For Binaries 1–5, with strong SNRs, signal at higher frequencies, and measurements of higher modes, we are able to precisely identify the ecliptic longitude, while strongly recovering the ecliptic latitude. Also, as expected, we are unable to differentiate between above and below the LISA orbital plane, even though more posterior weight is located at the correct latitudinal mode. The relative level to which we constrained the sky location for Binaries 1–5 is determined by their SNR, modulated slightly by the randomness of the noise.

The sky localization achieved for the large binary is effectively minimal. This lack of localization in this specific case is due to Binary 6 existing in a near-edge-on configuration. In an edge-on configuration, the information from higher modes in terms of constraining the sky location is minimized. With the same binary injection parameters but an inclination in a more face-on configuration ($i = \pi/8$), the sky localization is reduced to the eight degenerate sky modes expected for binaries at lower frequencies (see Marsat *et al.* [40]). This configuration is shown for reference in Fig. 6. In this specific case, seven of the eight modes are highlighted. The frequency spectrum for Binary 6 does not extend to high enough frequencies to constrain the longitude to one of the four longitudinal modes. However, if this binary was run in the “spin up” configuration performed for Binary 4, the frequency-domain spectrum would extend high enough allowing the localization to narrow to one longitudinal mode. We also observe uncertainty as to whether the source is truly above or below the plane.

Similar to the large binary, the sky location constraint for the small binary is weak. This is also due to the edge-on configuration. When analyzed in a more face-on configuration ($i = \pi/8$), the sky localization is much smaller, locating the correct longitudinal mode as is expected for a high-frequency binary. The sampler also located the correct latitude, but confusion as to whether the source is above or below the plane does exist. This face-on configuration is also shown in Fig. 6 for reference.

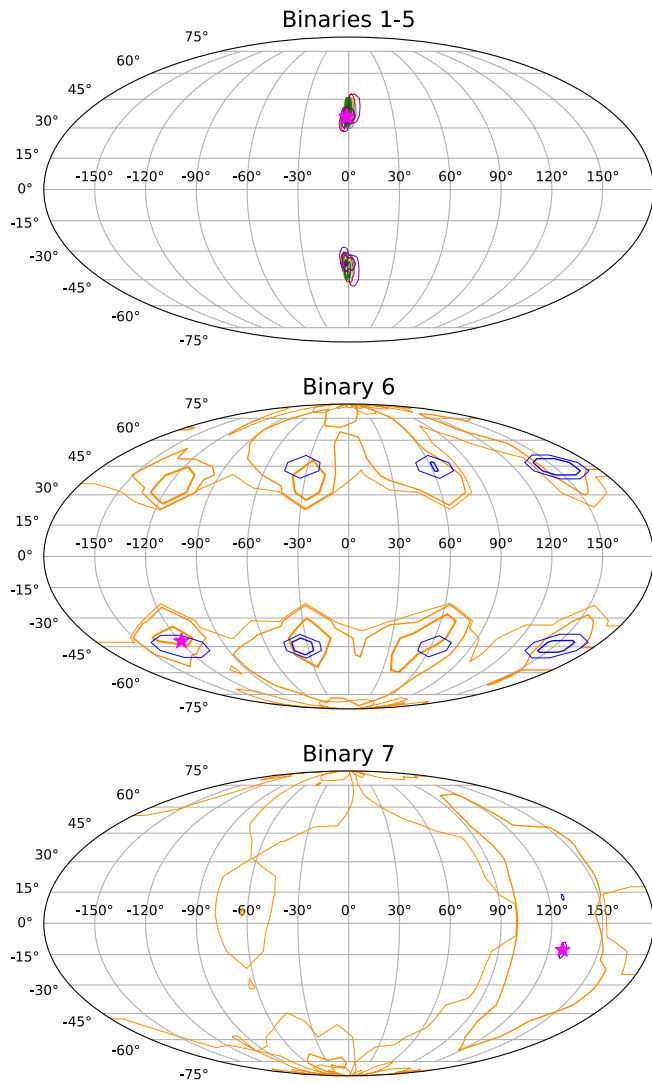


FIG. 6. The sky location from all binaries tested is shown above. Binaries 1–5 are shown in the top plot. Binaries 6 and 7 are shown in orange in the center and bottom plots, respectively. The 3σ , 2σ , and 1σ contours are shown with increasing line thickness and the true injection point is marked by a magenta star. The large binary (Binary 6) has effectively no localization. Similarly, Binary 7 also has a large localization area. The inclination for both Binaries 6 and 7 indicate they are close to an edge-on configuration. To illustrate the difference compared to a more face-on configuration, we add sky locations in blue for both large and small binaries with the exact same parameters, but with an injection inclination of $\pi/8$. Binaries 1–5 are all concentrated in the center of the map, precisely around the correct longitude value for those sources. The degeneracy above and below the plane also exists. It cannot be seen easily, but there is more posterior weight around the correct ecliptic latitude for all Binaries 1–5 compared to the position opposite the LISA orbital plane.

The other extrinsic orientation parameters follow a predictable structure related to the ecliptic longitude and latitude. The inclination is reflected at $-\cos i$ and ψ is reflected at $\pi - \psi$. Additionally, since more than one

longitudinal mode is observed for Binary 6, ψ is observed at deviations of $\pi/2$ from both the true and the reflected values. Binaries 1–5 show Gaussian distributions for $\ln D_L$, $\ln \phi_{\text{ref}}$, and $\ln t_{\text{ref}}$. However, it can be seen that t_{ref} correlates similarly to ϕ_{ref} as they both depend on the intrinsic parameters in terms of setting the reference frequency at which these values are set.

The large binary is the only injection where the reference time is not constrained well (see Fig. 12). Since the merger and ringdown are short relative to an inspiral evolution, the movement of the LISA constellation, as well as its rotation, is small during the observation of this signal. We are effectively observing a snapshot of the temporal aspect of the response function; therefore, moving the binary forward or backward in time will not change the response of the spacecraft by much, especially since we are operating on the order of seconds for the reference time parameter. The behavior of ϕ_{ref} is reflective of the unique spin behavior detailed previously: ϕ_{ref} directly correlates with the abnormal distribution for a_2 . Similar to the spin distributions, this is due to the setting of the reference frequency at which this phase is assigned. The luminosity distance for the large binary is constrained to within ~ 5 Gpc and shows the mean value and main peak near the injection value. However, the distribution of the distance measurements is not Gaussian, tending toward higher distances. Without much inspiral signal, we lose information that helps to constrain the distance to the binary.

We are able to constrain t_{ref} for the small binary quite well due to its long evolution in band (see Fig. 13). ϕ_{ref} seems to show a “split” posterior. However, it is really a symptom of wrapping the phase at 2π . If we were to unwrap those posteriors, the areas that seem to be disjointed and separated would fit as one multimodal posterior similar to those shown for the intrinsic parameters. The distance posterior for the small binary is unique. Similar to the large binary, this posterior tends toward larger distances than the injection. A difference with the small binary is that the mean value and main peak is not located at the true injection value. Since this binary is analyzed in a zero-noise representation, the random nature of noise did not cause this deviation. Without any noise, the maximum likelihood value would be found at the true distance value. However, since we do not see a main peak around this value, it means the parameter space volume centered around the true value does not contain a high percentage of posterior weight; our observations show there is more posterior weight around ~ 100 Gpc.

V. CONCLUSION

We present the first noise-infused analysis of MBH binaries with LISA employing full waveforms spanning inspiral, merger, and ringdown that include higher-order harmonic modes. Additionally, we expand on previous

work by adding the inclusion of aligned spins. This study was performed with `PhenomHM` [80] and the fast frequency domain response from Marsat and Baker [68]. Both of these processes, as well as the likelihood computation, were accelerated with GPUs, which is a first for the LISA response function. We showed the extreme improvement in speed these GPU devices achieve. In instances where full data streams are analyzed with noise-infused injections, GPUs can achieve speeds ~ 500 or more times faster than a single CPU.

With these accelerated likelihood computations, we used Bayesian inference techniques to extract full marginalized posterior distributions across the $D = 11$ dimensional parameter space. In particular, we employed a parallel-tempered implementation of the stretch proposal MCMC sampler implemented in Vousden *et al.* [109].

We obtained full posterior distributions for seven test binaries, which can be seen in Figs. 7–13 in Sec. A of Appendix. The parameters of the test binaries can be seen in Table II. For the intrinsic parameters, ϕ_{ref} , D_L , and t_{ref} , we show the recovered parameter values and their 1σ errors in Tables IV–VI. The errors generally follow proportionally to the SNR. However, when we fix the spins for Binary 3, we can further constrain the reference values because the spins help determine the frequency at which the reference values are set.

We highlighted the correlation between $\ln M_T$ and q as it varied for the different binaries. They exhibited a positive (negative) correlation if the slope of the sensitivity curve near the merger, where most of the signal is accumulated, is negative (positive).

The large binary (Binary 6) exhibits unique posteriors in a_1 versus a_2 . We show this posterior exhibits a non-ellipsoidal structure that follows a slope similar to the mass ratio (see Fig. 5). Since the large binary signal is observed during only the merger and ringdown, this binary is more susceptible to likelihood changes when the secondary mass-weighted spin begins to dominate over the primary.

The small binary (Binary 7) shows a multimodal distribution most likely due to its lower SNR and lack of strong observation of higher-order modes.

We also discuss our ability to localize each binary (see Fig. 6). The inclination of the binary holds strong influence over its localization. Face-on configurations provided much better sky localization for all binaries compared to edge-on configurations.

In future work, we plan to expand on the findings shown here by growing the binary parameter space that we have looked at, focusing on astrophysically motivated catalogs. Additionally, we will examine how the sky localization for a binary changes over time with the new LISA sensitivity curve, as well as with newer, advanced waveforms like `PhenomHM`, such as the coming waveform family for `PhenomX` [115–117].

With this work, we see the aligned spins add a new dimension to the sampling of binaries resulting in interesting configurations in spin posteriors, as well as varying the constraints on the other parameters. We see that injected noise, as expected, has varied our ability to recover parameters. The analysis of these noise injections was made possible by the implementation of GPU-accelerated likelihood calculations, a tool that will help to further the reach of these parameter estimation studies in the run up to the launch of the LISA mission.

ACKNOWLEDGMENTS

M. L. K. acknowledges support from the National Science Foundation under Grant No. DGE-0948017 and the Chateaubriand Fellowship from the Office for Science & Technology of the Embassy of France in the United States. M. L. K. thanks folks at the Laboratoire Astroparticule & Cosmologie in Paris, including Edward K. Porter, Marc Arene, Calum Murray, and the LISA group at APC for being gracious hosts. A. J. K. C. acknowledges support from the Jet Propulsion Laboratory Research and Technology Development program. This research was supported in part through the computational resources and staff contributions provided for the Quest/Grail high performance computing facility at Northwestern University. `Astropy`, a community-developed core `PYTHON` package for Astronomy, was used in this research [118]. Parts of this research were carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (80NM0018D0004). This paper also employed use of `SciPy` [98], `NUMPY` [119], and `MATPLOTLIB` [120].

APPENDIX: POSTERIOR DISTRIBUTIONS

Here we present the full corner plots for all binaries tested (see Table II) to accompany subsections of the corner plots shown in Sec. IV. Binaries 1–7 are shown in Figs. 7–13, respectively. The exact injection parameters for each binary can be seen in Table II, and are represented by a green point in the full posteriors. All sky position and orientation parameters are displayed over their entire domain (or prior range) due to their inherent multimodal structure. The other parameters are shown zooming in on the distributions that account for 99.9% of marginalized posterior points. The inclination, ι , and ecliptic latitude, β , are plotted as $\cos \iota$ and $\sin \beta$ because these are the parameters sampled during the MCMC runs. Similarly, the total mass, M_T ; luminosity distance, D_L ; and reference time, t_{ref} , are all sampled and plotted as the natural log of these values. Table VII shows the autocorrelation time (τ_f) and ESS according to Sec. III C for each MCMC run. It also gives the average acceptance fraction (AF) for each run.

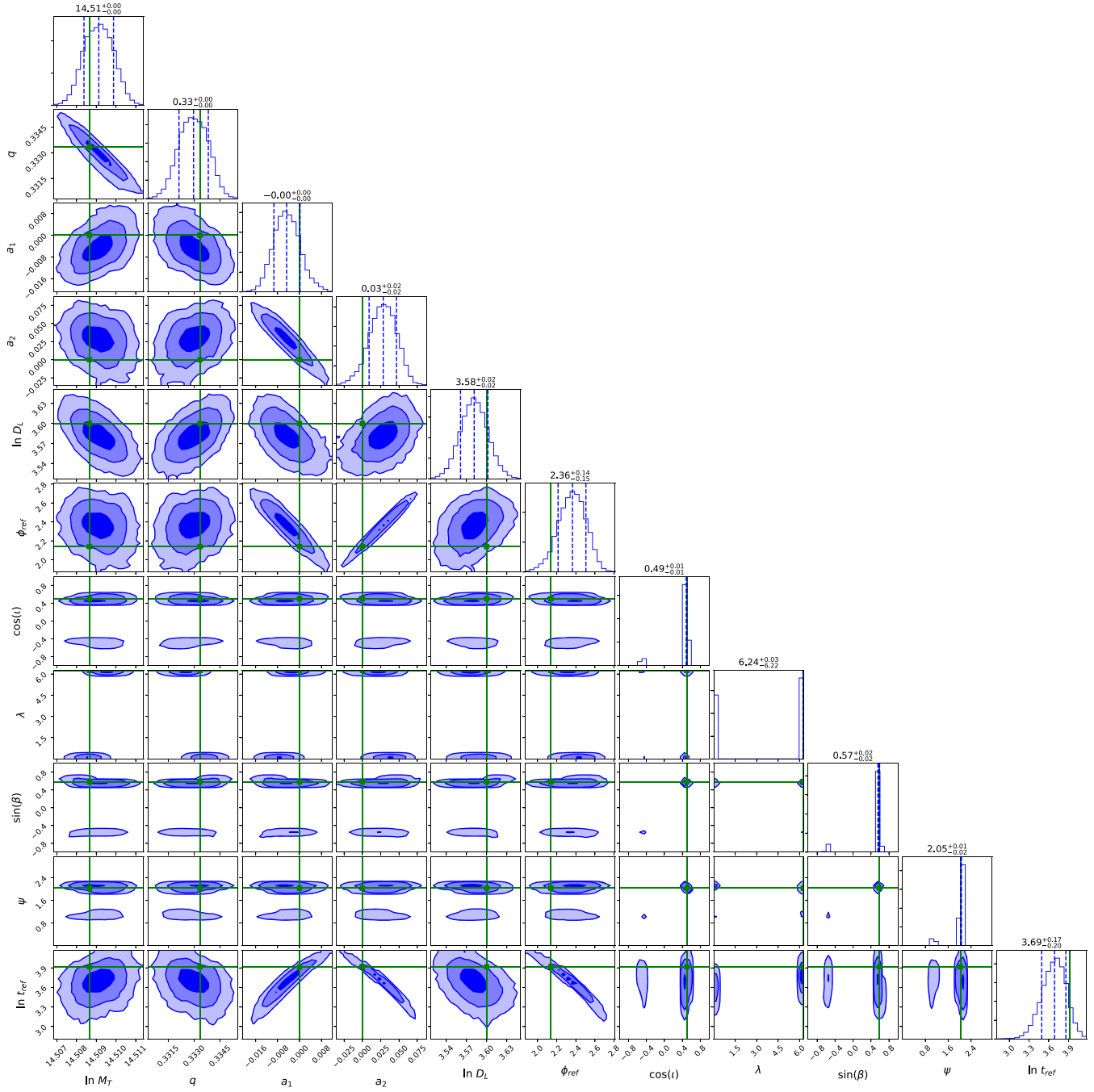


FIG. 7. The posterior distributions for all parameters for Binary 1 are shown above. Binary 1 was injected with Schwarzschild MBHs with a sampled noise realization. The sampler in this case was allowed to sample nonzero spins. Refer to the beginning of Sec. A of Appendix for information on the corner plots.

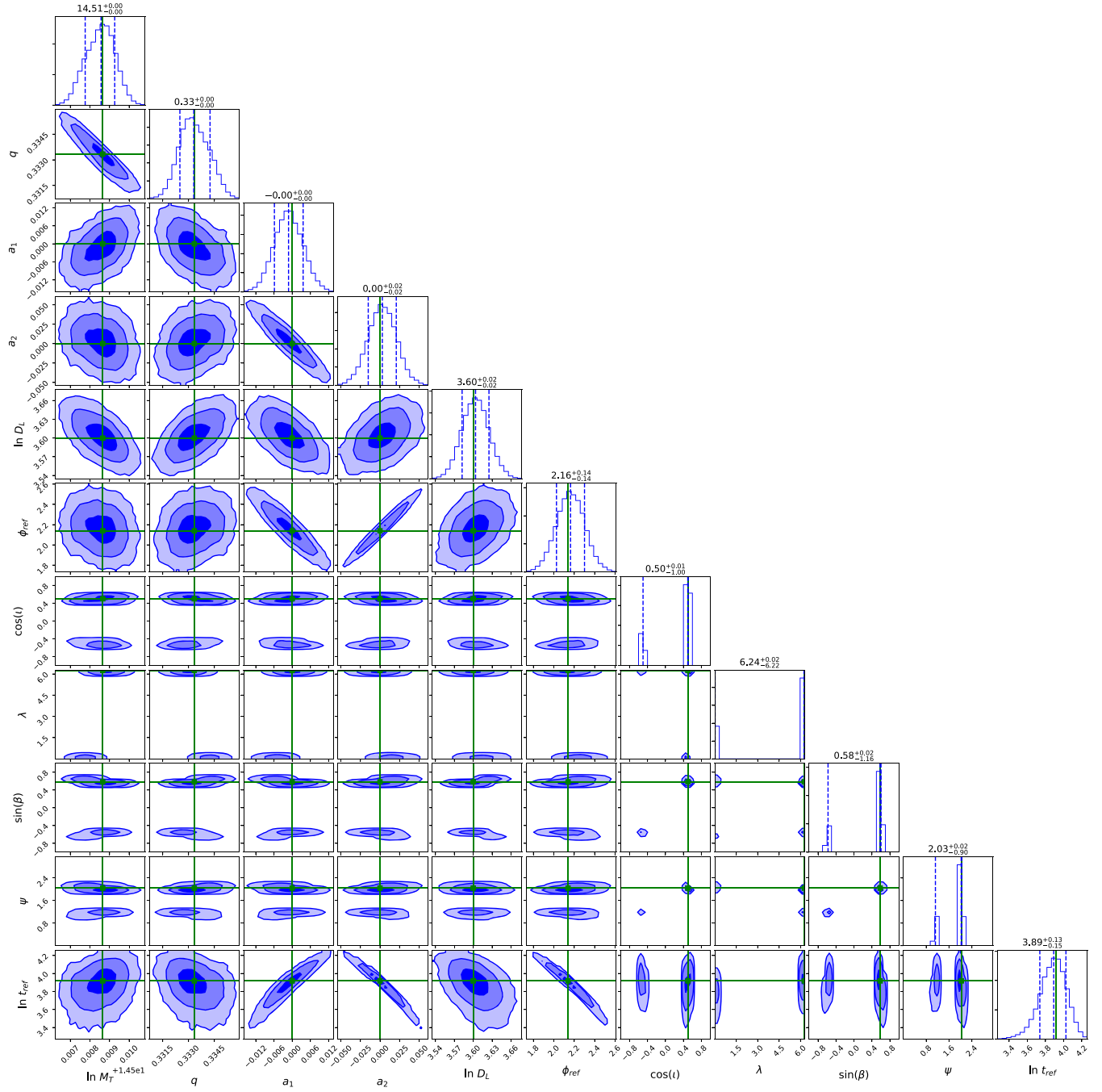


FIG. 8. Binary 2's posterior distributions are shown in this figure. Binary 2 was injected with Schwarzschild MBHs and analyzed in a zero-noise representation. Spins were allowed to vary during sampling. Please see Sec. A of Appendix for more information about the construction of the corner plots.

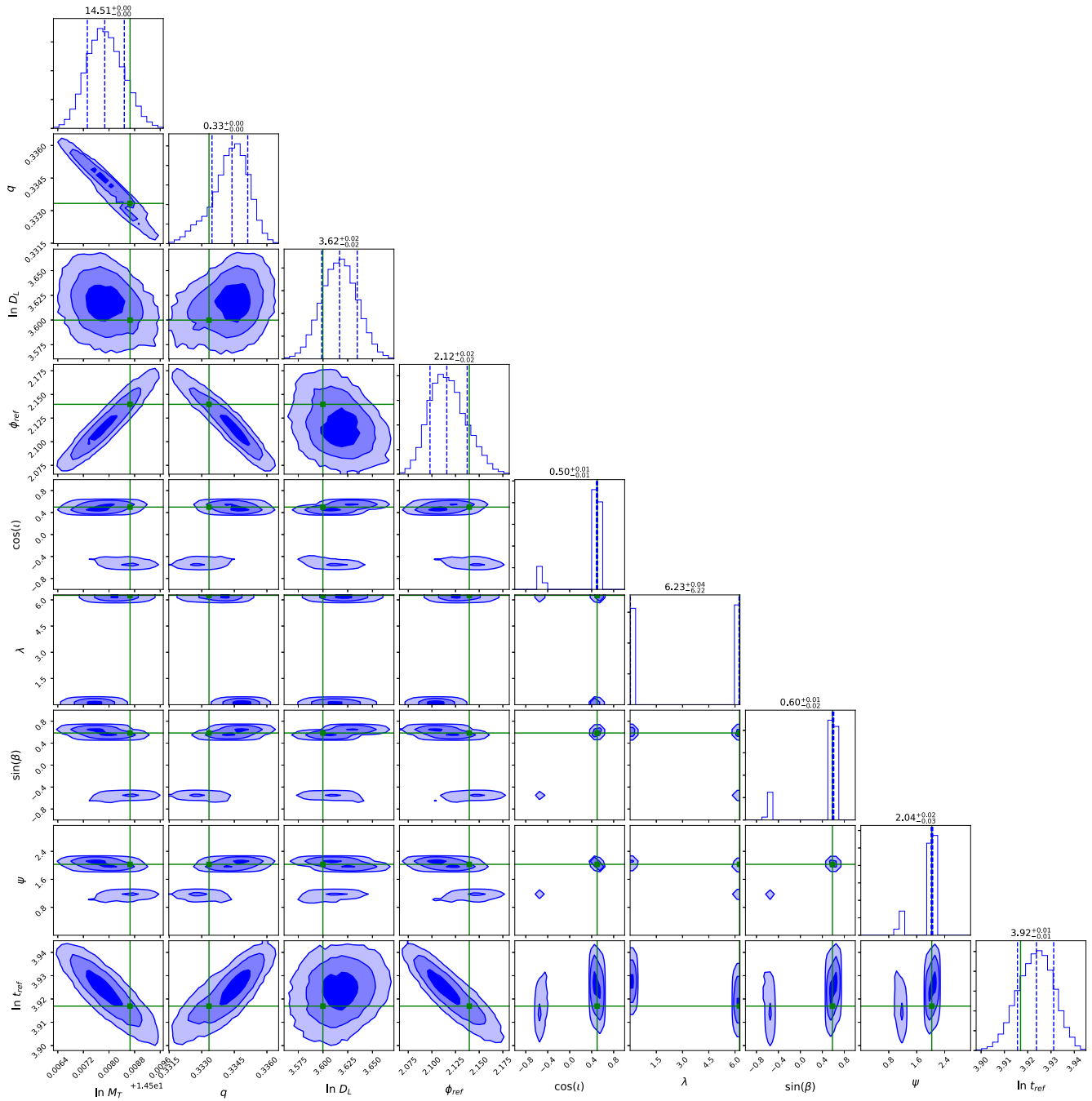


FIG. 9. The posterior distributions for Binary 3 are shown. Binary 3 was injected with Schwarzschild MBHs in a sampled noise realization. However, the sampler for this binary held the spins fixed, therefore, analyzing 9 parameters rather than 11. For this reason, spin distributions are not shown. Section A of Appendix discusses the constructions used in the corner plots.

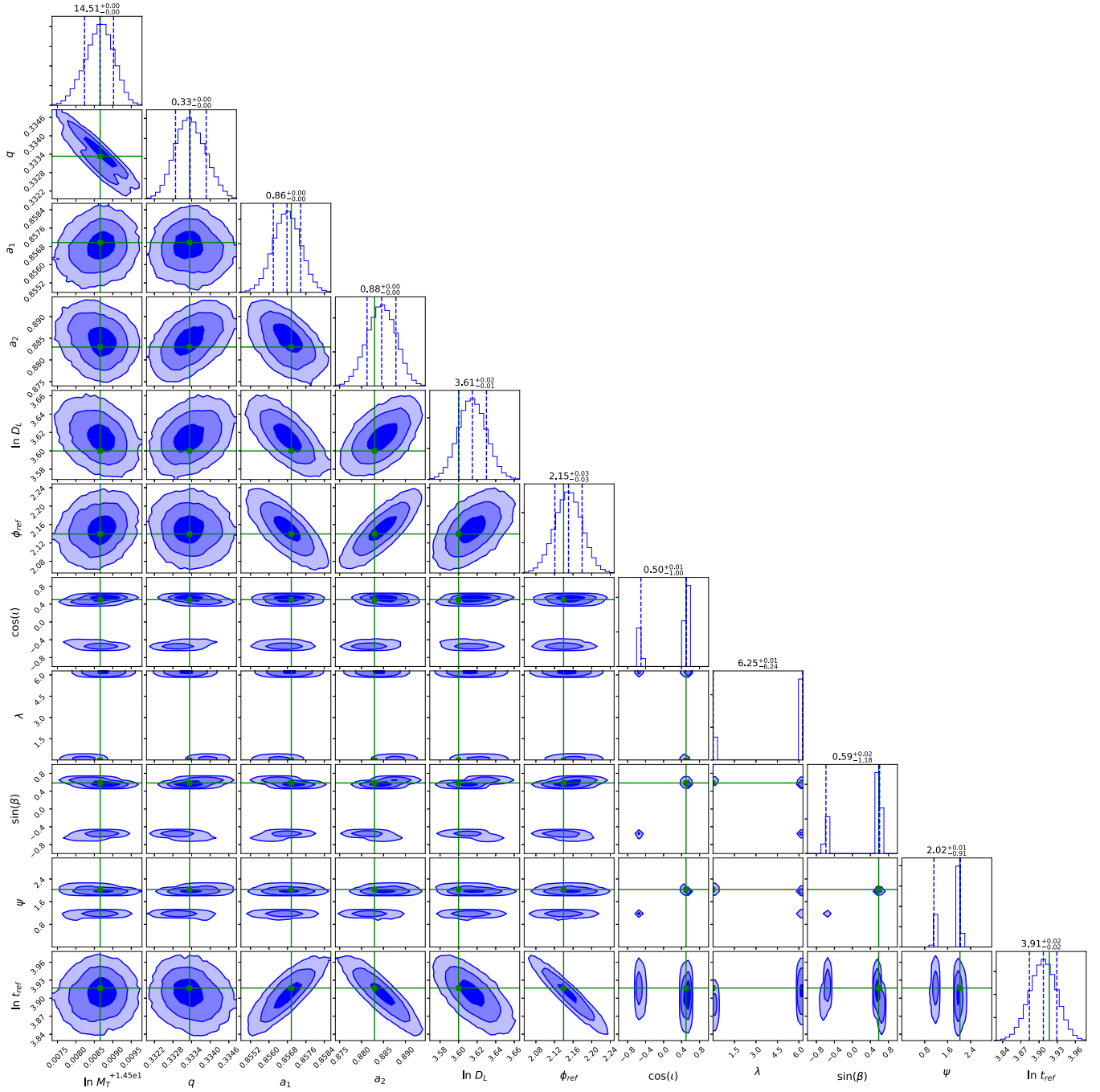


FIG. 10. The binary injected with strong aligned spins, Binary 4, is shown here. This binary was injected with a noise realization and the sampler was allowed to vary the spins during sampling. Please see Sec. A of Appendix for information on the constructions involved in this plot.

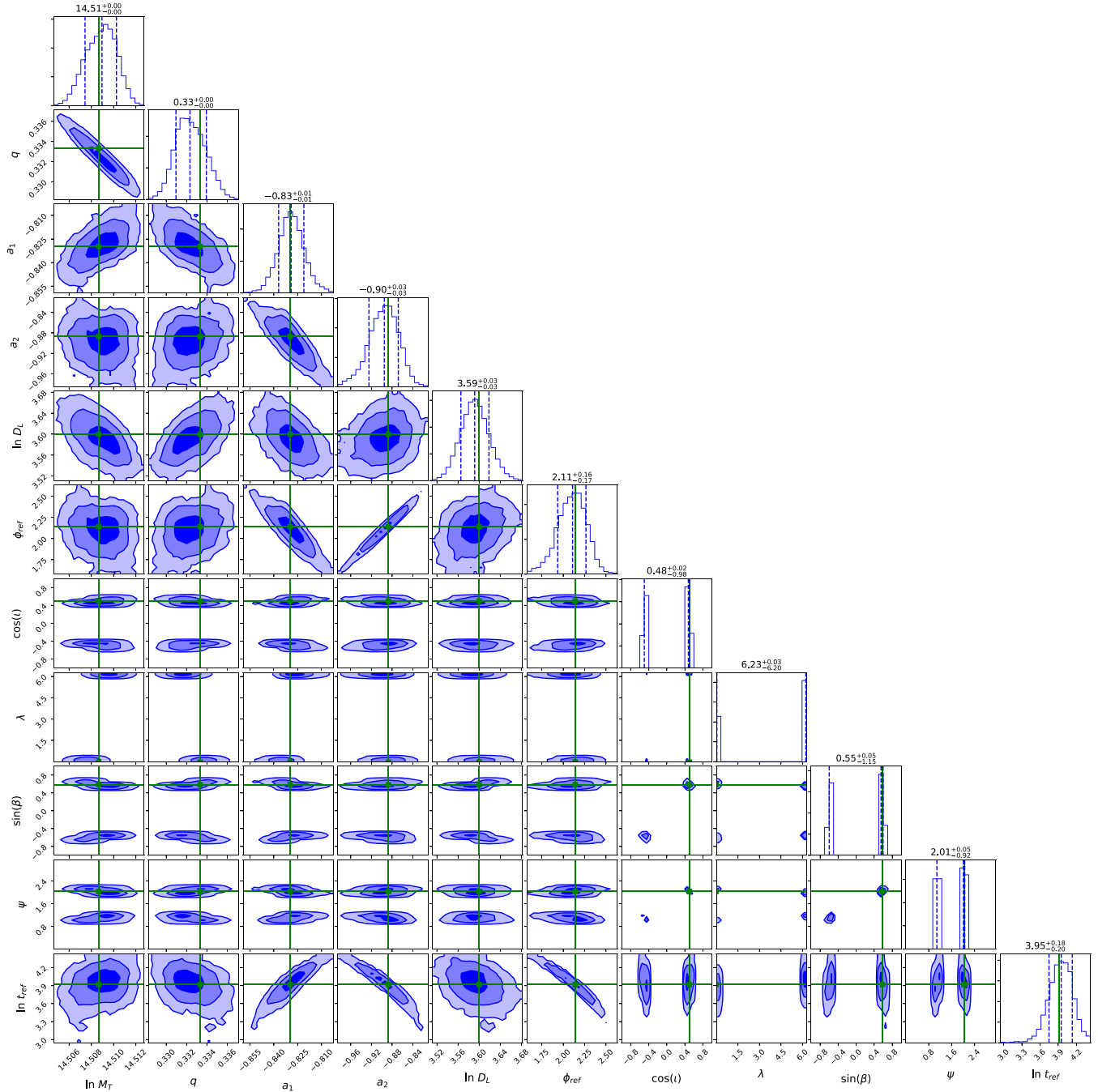


FIG. 11. The binary injected with strong antialigned spins, Binary 5, is shown in this plot. The sampler here was allowed to vary spins. The data stream contained a sampled noise realization. The constructions involved in this corner plot are explained in Sec. A of Appendix.

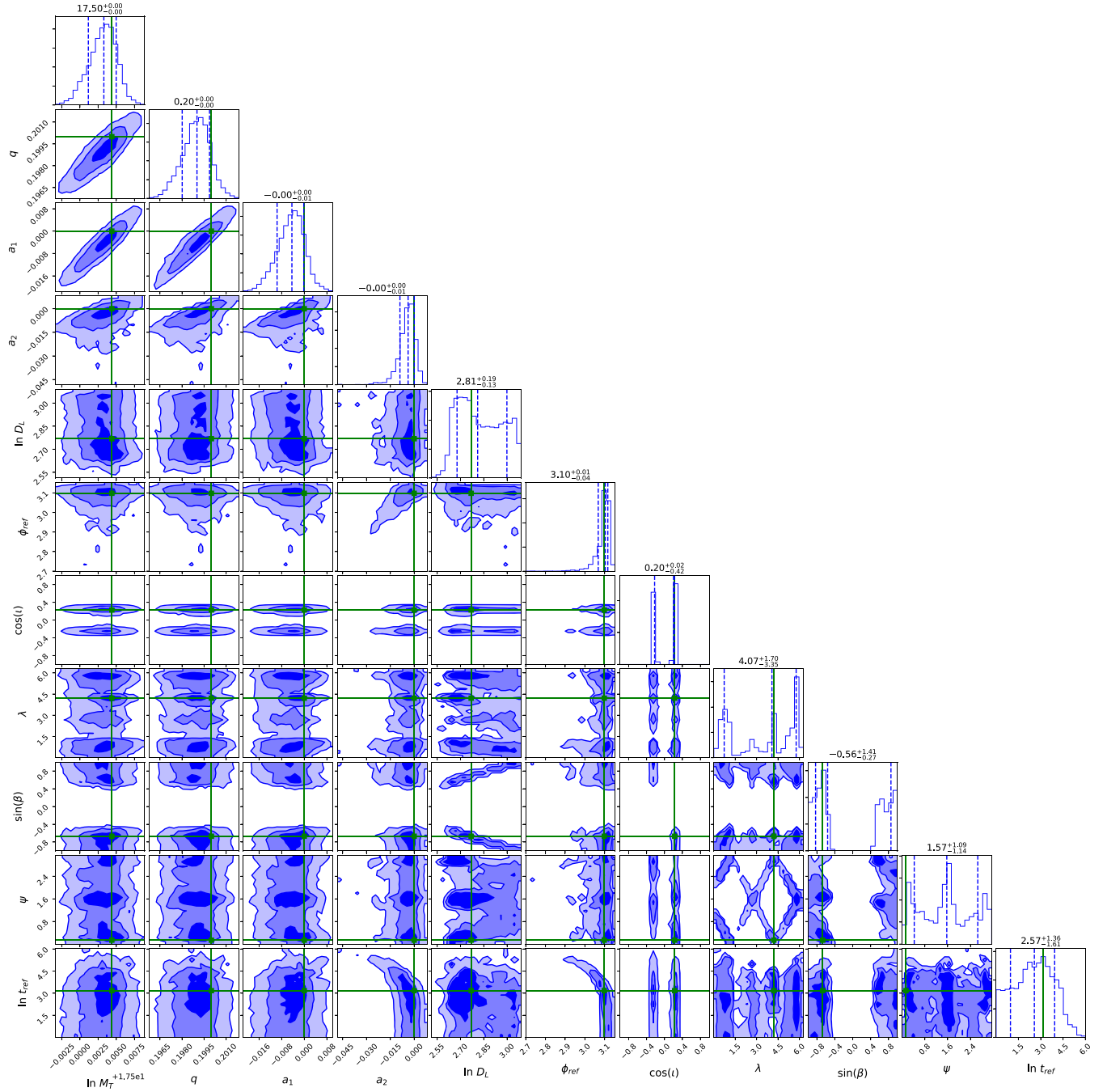


FIG. 12. The large mass binary, Binary 6, is shown above. The posterior distributions represent a signal with a noise realization and Schwarzschild MBHs. During sampling, the MCMC algorithm varied the spins, therefore, analyzing the full 11-dimensional parameter space. Please see Sec. A of Appendix for information on the corner plot constructions.

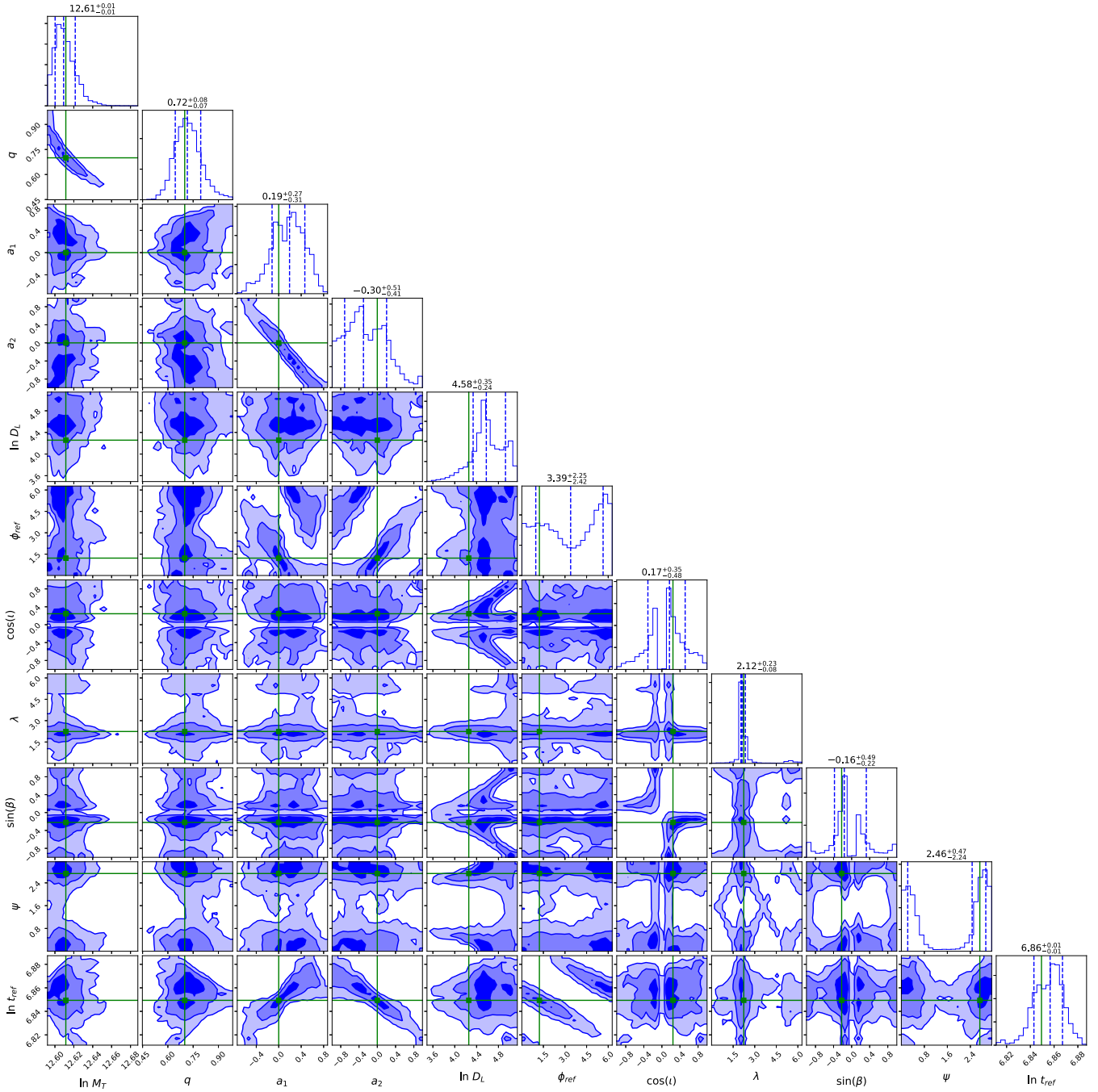


FIG. 13. The posterior distributions for the small mass binary, Binary 7, are shown in this plot. The small mass binary MCMC run was performed in a zero-noise representation due to issues with convergence of the sampler when adding noise. This is an issue that will be addressed in future work. This binary was injected with Schwarzschild MBHs and the spins were allowed to vary during sampling. Section A of Appendix describes the constructions used in the corner plots.

TABLE VII. The sampler output info for each run is shown below. τ_f is the autocorrelation time as discussed in Sec. III C. ESS is the effective sample size. This is the number of posterior density points used to create the plots shown below. The last column shows the average AF for each run.

Binary	τ_f	ESS	AF
1	97	79200	0.20
2	93	103232	0.19
3	63	50816	0.25
4	90	67584	0.24
5	80	52000	0.19
6	45	64000	0.18
7	8	600000	0.21

- [1] P. Amaro-Seoane *et al.*, Laser interferometer space antenna, [arXiv:1702.00786](#).
- [2] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs, *Phys. Rev. X* **9**, 031040 (2019).
- [3] LVK Collaboration, Prospects for observing and localizing gravitational-wave transients with advanced LIGO, advanced Virgo and KAGRA, *Living Rev. Relativity* **21**, 3 (2018).
- [4] A. Klein, E. Barausse, A. Sesana, A. Petiteau, E. Berti, S. Babak, J. Gair, S. Aoudia, I. Hinder, F. Ohme, and B. Wardell, Science with the space-based interferometer eLISA: Supermassive black hole binaries, *Phys. Rev. D* **93**, 024003 (2016).
- [5] E. Berti, A. Sesana, E. Barausse, V. Cardoso, and K. Belczynski, Spectroscopy of Kerr Black Holes with Earth- and Space-Based Interferometers, *Phys. Rev. Lett.* **117**, 101102 (2016).
- [6] J. Salcido, R. Bower, T. Theuns, S. McAlpine, M. Schaller, R. Crain, J. Schaye, and J. Regan, Music from the heavens—gravitational waves from supermassive black hole mergers in the EAGLE simulations, *Mon. Not. R. Astron. Soc.* **463**, 870 (2016).
- [7] M. L. Katz, L. Z. Kelley, F. Dosopoulou, S. Berry, L. Blecha, and S. L. Larson, Probing massive black hole binary populations with LISA, *Mon. Not. R. Astron. Soc.* **491**, 2301 (2020).
- [8] M. Bonetti, A. Sesana, F. Haardt, E. Barausse, and M. Colpi, Post-Newtonian evolution of massive black hole triplets in galactic nuclei—IV. Implications for LISA, *Mon. Not. R. Astron. Soc.* **486**, 4044 (2019).
- [9] eLISA Consortium, The gravitational Universe, [arXiv:1305.5720](#).
- [10] E. Barausse, J. Bellovary, E. Berti, B. Holley-Bockelmann, K. Farris, B. Sathyaprakash, and A. Sesana, Massive black hole science with eLISA, *J. Phys. Conf. Ser.* **610**, 012001 (2015).
- [11] M. Katz and S. Larson, Evaluating black hole detectability with LISA, *Mon. Not. R. Astron. Soc.* **483**, 3108 (2019).
- [12] A. Loeb and F. Rasio, Collapse of primordial gas clouds and the formation of quasar black holes, *Astrophys. J.* **432**, 52 (1994).
- [13] M. Begelman, M. Volonteri, and M. Rees, Formation of supermassive black holes by direct collapse in pre-galactic haloes, *Mon. Not. R. Astron. Soc.* **370**, 289 (2006).
- [14] M. Latif, D. Schleicher, W. Schmidt, and J. Niemeyer, The characteristic black hole mass resulting from direct collapse in the early Universe, *Mon. Not. R. Astron. Soc.* **436**, 2989 (2013).
- [15] M. Habouzit, M. Volonteri, M. Latif, Y. Dubois, and S. Peirani, On the number density of ‘direct collapse’ black hole seeds, *Mon. Not. R. Astron. Soc.* **463**, 529 (2016).
- [16] K. Ardaneh, Y. Luo, I. Shlosman, K. Nagamine, J. Wise, and M. Begelman, Direct collapse to supermassive black hole seeds with radiation transfer: Cosmological halos, *Mon. Not. R. Astron. Soc.* **479**, 2277 (2018).
- [17] G. Dunn, J. Bellovary, K. Holley-Bockelmann, C. Christensen, and T. Quinn, Sowing black hole seeds: Direct collapse black hole formation with realistic Lyman-Werner radiation in cosmological simulations, *Astrophys. J.* **861**, 39 (2018).
- [18] K. Omukai, R. Schneider, and Z. Haiman, Can supermassive black holes form in metal-enriched high-redshift protogalaxies?, *Astrophys. J.* **686**, 801 (2008).
- [19] B. Devecchi and M. Volonteri, Formation of the first nuclear clusters and massive black holes at high redshift, *Astrophys. J.* **694**, 302 (2009).
- [20] M. B. Davies, M. C. Miller, and J. M. Bellovary, Supermassive black hole formation via gas accretion in nuclear stellar clusters, *Astrophys. J. Lett.* **740**, L42 (2011).
- [21] H. Katz, D. Sijacki, and M. Haehnelt, Seeding high-redshift QSOs by collisional runaway in primordial star clusters, *Mon. Not. R. Astron. Soc.* **451**, 2352 (2015).
- [22] Z. Haiman, T. Abel, and M. Rees, The radiative feedback of the first cosmological objects, *Astrophys. J.* **534**, 11 (2000).

- [23] C. Fryer, S. Woosley, and A. Heger, Pair-instability Supernovae, gravity waves, and gamma-ray transients, *Astrophys. J.* **550**, 372 (2001).
- [24] A. Heger, C. Fryer, S. Woosley, N. Langer, and D. Hartmann, How massive single stars end their life, *Astrophys. J.* **591**, 288 (2003).
- [25] M. Volonteri, P. Madau, and F. Haardt, The formation of Galaxy stellar cores by the hierarchical merging of supermassive black holes, *Astrophys. J.* **593**, 661 (2003).
- [26] T. Tanaka and Z. Haiman, The assembly of supermassive black holes at high redshifts, *Astrophys. J.* **696**, 1798 (2009).
- [27] M. A. Alvarez, J. H. Wise, and T. Abel, Accretion onto the First Stellar-Mass Black Holes, *Astrophys. J.* **701**, L133 (2009).
- [28] A. Petiteau, S. Babak, and A. Sesana, Constraining the dark energy equation of state using LISA observations of spinning massive black hole binaries, *Astrophys. J.* **732**, 82 (2011).
- [29] J. Gair, M. Vallisneri, S. Larson, and J. Baker, Testing general relativity with low-frequency, space-based gravitational-wave detectors, *Living Rev. Relativity* **16**, 7 (2013).
- [30] B. Schutz, Determining the Hubble constant from gravitational wave observations, *Nature (London)* **323**, 310 (1986).
- [31] D. E. Holz and S. A. Hughes, Using gravitational-wave standard sirens, *Astrophys. J.* **629**, 15 (2005).
- [32] P. J. Armitage and P. Natarajan, Accretion during the merger of supermassive black holes, *Astrophys. J. Lett.* **567**, L9 (2002).
- [33] S. Burke-Spolaor, Multi-messenger approaches to binary supermassive black holes in the ‘continuous-wave’ regime, *Classical Quantum Gravity* **30**, 224013 (2013).
- [34] T. Bogdanović, Supermassive Black hole binaries: The search continues, in *Gravitational Wave Astrophysics* (2015), Vol. 40, p. 103.
- [35] T. Dal Canton, A. Mangiagli, S. C. Noble, J. Schnittman, A. Ptak, A. Klein, A. Sesana, and J. Camp, Detectability of modulated x-rays from LISA’s supermassive black hole mergers, *Astrophys. J.* **886**, 146 (2019).
- [36] N. J. Cornish and E. K. Porter, MCMC exploration of supermassive black hole binary inspirals, *Classical Quantum Gravity* **23**, S761 (2006).
- [37] N. J. Cornish and E. K. Porter, Catching supermassive black hole binaries without a net, *Phys. Rev. D* **75**, 021301 (2007).
- [38] E. K. Porter and J. Carré, A Hamiltonian Monte-Carlo method for Bayesian inference of supermassive black hole binaries, *Classical Quantum Gravity* **31**, 145004 (2014).
- [39] E. K. Porter and N. J. Cornish, Fisher versus Bayes: A comparison of parameter estimation techniques for massive black hole binaries to high redshifts with eLISA, *Phys. Rev. D* **91**, 104001 (2015).
- [40] S. Marsat, J. G. Baker, and T. Dal Canton, Exploring the Bayesian parameter estimation of binary black holes with LISA, *arXiv:2003.00357*.
- [41] N. Christensen and R. Meyer, Markov chain Monte Carlo methods for Bayesian gravitational radiation data analysis, *Phys. Rev. D* **58**, 082001 (1998).
- [42] E. K. Porter and N. J. Cornish, Effect of higher harmonic corrections on the detection of massive black hole binaries with LISA, *Phys. Rev. D* **78**, 064005 (2008).
- [43] M. Vallisneri, Use and abuse of the Fisher information matrix in the assessment of gravitational-wave parameter-estimation prospects, *Phys. Rev. D* **77**, 042001 (2008).
- [44] C. Cutler and M. Vallisneri, LISA detections of massive black hole inspirals: Parameter extraction errors due to inaccurate template waveforms, *Phys. Rev. D* **76**, 104018 (2007).
- [45] L. S. S. Team, LISA science requirements document (2018).
- [46] C. Cutler and É. Flanagan, Gravitational waves from merging compact binaries: How accurately can one extract the binary’s parameters from the inspiral waveform? *Phys. Rev. D* **49**, 2658 (1994).
- [47] A. Vecchio, LISA observations of rapidly spinning massive black hole binary systems, *Phys. Rev. D* **70**, 042001 (2004).
- [48] K. Arun, Parameter estimation of coalescing supermassive black hole binaries with LISA, *Phys. Rev. D* **74**, 024025 (2006).
- [49] E. Berti, A. Buonanno, and C. M. Will, Estimating spinning binary parameters and testing alternative theories of gravity with LISA, *Phys. Rev. D* **71**, 084025 (2005).
- [50] R. N. Lang and S. A. Hughes, Measuring coalescing massive binary black holes with gravitational waves: The impact of spin-induced precession, *Phys. Rev. D* **74**, 122001 (2006).
- [51] S. Babak *et al.* (Challenge 3 Participants) The mock LISA data challenges: From challenge 3 to challenge 4, *Classical Quantum Gravity* **27**, 084009 (2010).
- [52] D. A. Brown, J. Crowder, C. Cutler, I. Mandel, and M. Vallisneri, A three-stage search for supermassive black-hole binaries in LISA data, *Classical Quantum Gravity* **24**, S595 (2007).
- [53] J. Crowder, N. J. Cornish, and J. L. Reddinger, LISA data analysis using genetic algorithms, *Phys. Rev. D* **73**, 063011 (2006).
- [54] E. Wickham, A. Stroeer, and A. Vecchio, A Markov chain Monte Carlo approach to the study of massive black hole binary systems with LISA, *Classical Quantum Gravity* **23**, S819 (2006).
- [55] C. Röver, A. Stroeer, E. Bloomer, N. Christensen, J. Clark, M. Hendry, C. Messenger, R. Meyer, M. Pitkin, J. Toher, R. Umstätter, A. Vecchio, J. Veitch, and G. Woan, Inference on inspiral signals using LISA MLDC data, *Classical Quantum Gravity* **24**, S521 (2007).
- [56] F. Feroz, J. R. Gair, M. P. Hobson, and E. K. Porter, Use of the MULTINEST algorithm for gravitational wave data analysis, *Classical Quantum Gravity* **26**, 215003 (2009).
- [57] J. R. Gair and E. K. Porter, Cosmic swarms: A search for supermassive black holes in the LISA data stream with a hybrid evolutionary algorithm, *Classical Quantum Gravity* **26**, 225004 (2009).
- [58] A. Petiteau, Y. Shang, and S. Babak, The search for black hole binaries using a genetic algorithm, *Classical Quantum Gravity* **26**, 204011 (2009).

- [59] K. Arun, B. R. Iyer, B. Sathyaprakash, S. Sinha, and C. van den Broeck, Higher signal harmonics, LISA's angular resolution, and dark energy, *Phys. Rev. D* **76**, 104016 (2007).
- [60] M. Trias and A. M. Sintes, LISA observations of supermassive black holes: Parameter estimation using full post-Newtonian inspiral waveforms, *Phys. Rev. D* **77**, 024030 (2008).
- [61] S. T. McWilliams, J. I. Thorpe, J. G. Baker, and B. J. Kelly, Impact of mergers on LISA parameter estimation for nonspinning black hole binaries, *Phys. Rev. D* **81**, 064014 (2010).
- [62] J. Thorpe, S. McWilliams, B. Kelly, R. Fahey, K. Arnaud, and J. Baker, LISA parameter estimation using numerical merger waveforms, *Classical Quantum Gravity* **26**, 094026 (2009).
- [63] S. T. McWilliams, B. J. Kelly, and J. G. Baker, Observing mergers of nonspinning black-hole binaries, *Phys. Rev. D* **82**, 024014 (2010).
- [64] S. McWilliams, R. Lang, J. Baker, and J. Thorpe, Sky localization of complete inspiral-merger-ringdown signals for nonspinning massive black hole binaries, *Phys. Rev. D* **84**, 064003 (2011).
- [65] S. Babak, M. Hannam, S. Husa, and B. Schutz, Resolving super massive black holes with LISA, [arXiv:0806.1591](https://arxiv.org/abs/0806.1591).
- [66] S. Babak, J. Gair, and R. Cole, Extreme mass ratio inspirals: perspectives for their detection, *Fundam. Theor. Phys.* **179**, 783 (2015).
- [67] V. Baibhav, E. Berti, and V. Cardoso, LISA parameter estimation and source localization with higher harmonics of the ringdown, *Phys. Rev. D* **101**, 084053 (2020).
- [68] S. Marsat and J. G. Baker, Fourier-domain modulations and delays of gravitational-wave signals, [arXiv:1806.10734](https://arxiv.org/abs/1806.10734).
- [69] S. Larson, W. Hiscock, and R. Hellings, Sensitivity curves for spaceborne gravitational wave interferometers, *Phys. Rev. D* **62**, 062001 (2000).
- [70] S. Husa, S. Khan, M. Hannam, M. Pürrer, F. Ohme, X. Forteza, and A. Bohé, Frequency-domain gravitational waves from nonprecessing black-hole binaries. I. New numerical waveforms and anatomy of the signal, *Phys. Rev. D* **93**, 044006 (2016).
- [71] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. Forteza, and A. Bohé, Frequency-domain gravitational waves from nonprecessing black-hole binaries. II. A phenomenological model for the advanced detector era, *Phys. Rev. D* **93**, 044007 (2016).
- [72] N. Cornish and T. Robson, The construction and use of LISA sensitivity curves, *Classical Quantum Gravity* **36**, 105011 (2019).
- [73] P. Amaro-Seoane *et al.*, Laser interferometer space antenna, [arXiv:1702.00786](https://arxiv.org/abs/1702.00786).
- [74] M. C. Edwards, R. Meyer, and N. Christensen, Bayesian semiparametric power spectral density estimation with applications in gravitational wave data analysis, *Phys. Rev. D* **92**, 064011 (2015).
- [75] T. B. Littenberg and N. J. Cornish, Bayesian inference for spectral estimation of gravitational wave detector noise, *Phys. Rev. D* **91**, 084034 (2015).
- [76] S. Biscoveanu, C.-J. Haster, S. Vitale, and J. Davies, Quantifying the effect of power spectral density uncertainty on gravitational-wave parameter estimation for compact binary sources, *Phys. Rev. D* **102**, 023008 (2020).
- [77] P. M. Goggans and Y. Chi, Using thermodynamic integration to calculate the posterior probability in bayesian model selection problems, *AIP Conf. Proc.* **707**, 59 (2004).
- [78] N. Lartillot and H. Philippe, Computing Bayes factors using thermodynamic integration, *Syst. Biol.* **55**, 195 (2006).
- [79] P. Maturana-Russel, R. Meyer, J. Veitch, and N. Christensen, Stepping-stone sampling algorithm for calculating the evidence of gravitational wave models, *Phys. Rev. D* **99**, 084006 (2019).
- [80] L. London, S. Khan, E. Fauchon-Jones, C. García, M. Hannam, S. Husa, X. Jiménez-Forteza, C. Kalaghatgi, F. Ohme, and F. Pannarale, First Higher-Multipole Model of Gravitational Waves from Spinning and Coalescing Black-Hole Binaries, *Phys. Rev. Lett.* **120**, 161102 (2018).
- [81] C. Kalaghatgi, M. Hannam, and V. Raymond, Parameter Estimation with a spinning multi-mode waveform model: IMRPhenomHM, *Phys. Rev. D* **101**, 103004 (2020).
- [82] L. Finn and K. Thorne, Gravitational waves from a compact star in a circular, inspiral orbit, in the equatorial plane of a massive, spinning black hole, as observed by LISA, *Phys. Rev. D* **62**, 124021 (2000).
- [83] C. J. Moore, R. H. Cole, and C. P. L. Berry, Gravitational-wave sensitivity curves, *Classical Quantum Gravity* **32**, 015014 (2015).
- [84] S. Babak, J. Gair, A. Sesana, E. Barausse, C. Sopuerta, C. Berry, E. Berti, P. Amaro-Seoane, A. Petiteau, and A. Klein, Science with the space-based interferometer LISA. V. Extreme mass-ratio inspirals, *Phys. Rev. D* **95**, 103012 (2017).
- [85] N. J. Cornish and L. J. Rubbo, Publisher's Note: LISA response function [*Phys. Rev. D* **67**, 022001 (2003)]; *Phys. Rev. D* **67**, 029905 (2003).
- [86] M. Tinto and J. W. Armstrong, Cancellation of laser noise in an unequal-arm interferometer detector of gravitational radiation, *Phys. Rev. D* **59**, 102003 (1999).
- [87] J. Armstrong, F. Estabrook, and M. Tinto, Time-delay interferometry for space-based gravitational wave searches, *Astrophys. J.* **527**, 814 (1999).
- [88] F. B. Estabrook, M. Tinto, and J. W. Armstrong, Time-delay analysis of LISA gravitational wave data: Elimination of spacecraft motion effects, *Phys. Rev. D* **62**, 042002 (2000).
- [89] S. V. Dhurandhar, K. R. Nayak, and J.-Y. Vinet, Algebraic approach to time-delay data analysis for LISA, *Phys. Rev. D* **65**, 102002 (2002).
- [90] M. Tinto and S. V. Dhurandhar, Time-delay interferometry, *Living Rev. Relativity* **8**, 4 (2005).
- [91] M. Vallisneri, Synthetic LISA: Simulating time delay interferometry in a model LISA, *Phys. Rev. D* **71**, 022001 (2005).
- [92] M. Tinto and S. V. Dhurandhar, Time-delay interferometry, *Living Rev. Relativity* **17**, 6 (2014).
- [93] M. Muratore, D. Vetrugno, and S. Vitale, Revisitation of time delay interferometry combinations that suppress laser noise in LISA, [arXiv:2001.11221](https://arxiv.org/abs/2001.11221).

- [94] J. Nickolls, I. Buck, M. Garland, and K. Skadron, Scalable parallel programming with CUDA, *Queueing Syst. Theory Appl.* **6**, 40 (2008).
- [95] C. Talbot, R. Smith, E. Thrane, and G. B. Poole, Parallelized inference for gravitational-wave astronomy, *Phys. Rev. D* **100**, 043030 (2019).
- [96] M. Hannam, P. Schmidt, A. Bohé, L. Haegel, S. Husa, F. Ohme, G. Pratten, and M. Pürrer, Simple Model of Complete Precessing Black-Hole-Binary Gravitational Waveforms, *Phys. Rev. Lett.* **113**, 151101 (2014).
- [97] LIGO Scientific Collaboration, LIGO algorithm library—LALSuite, free software (GPL) (2018).
- [98] E. Jones, T. Oliphant, P. Peterson *et al.*, SciPy: Open source scientific tools for PYTHON.
- [99] cuSPARSE, <https://docs.nvidia.com/cuda/cusparse/index.html>.
- [100] cuBLAS, <https://docs.nvidia.com/cuda/cublas/index.html>.
- [101] S. Behnel, R. Bradshaw, C. Citro, L. Dalcin, D. S. Seljebotn, and K. Smith, Cython: The best of both worlds, *Comput. Sci. Eng.* **13**, 31 (2011).
- [102] R. McGibbon and Y. Zhao, nvidia-example.
- [103] J. Skilling, Nested sampling, *AIP Conf. Proc.* **735**, 395 (2004).
- [104] J. Skilling, Nested sampling for general Bayesian computation, *Bayesian Anal.* **1**, 833 (2006).
- [105] N. Metropolis and S. Ulam, The Monte Carlo method, *J. Am. Stat. Assoc.* **44**, 335 (1949).
- [106] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, Equation of state calculations by fast computing machines, *J. Chem. Phys.* **21**, 1087 (1953).
- [107] W. K. Hastings, Monte Carlo sampling methods using Markov chains and their applications, *Biometrika* **57**, 97 (1970).
- [108] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, EMCEE: The MCMC hammer, *Publ. Astron. Soc. Pac.* **125**, 306 (2013).
- [109] W. Vousden, W. Farr, and I. Mandel, Dynamic temperature selection for parallel tempering in Markov chain Monte Carlo simulations, *Mon. Not. R. Astron. Soc.* **455**, 1919 (2016).
- [110] J. Goodman and J. Weare, Ensemble samplers with affine invariance, *Commun. Appl. Math. Comput. Sci.* **5**, 65 (2010).
- [111] R. H. Swendsen and J.-S. Wang, Replica Monte Carlo Simulation of Spin Glasses, *Phys. Rev. Lett.* **57**, 2607 (1986).
- [112] D. J. Earl and M. W. Deem, Parallel tempering: Theory, applications, and new perspectives, *Phys. Chem. Chem. Phys.* **7**, 3910 (2005).
- [113] A. Sokal, Monte Carlo methods in statistical mechanics: Foundations and new algorithms, in *Functional Integration: Basics and Applications*, edited by C. DeWitt-Morette, P. Cartier, and A. Folacci (Springer, Boston, MA, 1997), pp. 131–192.
- [114] D. Foreman-Mackey, Autocorrelation time estimation, <https://dfm.io/posts/autocorr/> (2018).
- [115] G. Pratten, S. Husa, C. Garcia-Quiros, M. Colleoni, A. Ramos-Buades, H. Estelles, and R. Jaume, Setting the cornerstone for the IMRPhenomX family of models for gravitational waves from compact binaries: The dominant harmonic for non-precessing quasi-circular black holes, [arXiv:2001.11412](https://arxiv.org/abs/2001.11412).
- [116] C. García-Quirós, M. Colleoni, S. Husa, H. Estellés, G. Pratten, A. Ramos-Buades, M. Mateu-Lucena, and R. Jaume, IMRPhenomXHM: A multi-mode frequency-domain model for the gravitational wave signal from non-precessing black-hole binaries, [arXiv:2001.10914](https://arxiv.org/abs/2001.10914).
- [117] G. Pratten, C. García-Quirós, M. Colleoni, A. Ramos-Buades, H. Estellés, M. Mateu-Lucena, R. Jaume, M. Haney, D. Keitel, J. E. Thompson, and S. Husa, Let's twist again: Computationally efficient models for the dominant and sub-dominant harmonic modes of precessing binary black holes, [arXiv:2004.06503](https://arxiv.org/abs/2004.06503).
- [118] T. Robitaille *et al.* (Astropy Collaboration), Astropy: A community PYTHON package for astronomy, *Astron. Astrophys.* **558**, A33 (2013).
- [119] S. van der Walt, S. C. Colbert, and G. Varoquaux, The NumPy array: A structure for efficient numerical computation, *Comput. Sci. Eng.* **13**, 22 (2011).
- [120] J. D. Hunter, MATPLOTLIB: A 2D graphics environment, *Comput. Sci. Eng.* **9**, 90 (2007).