A Robust and Efficient $\mathcal{F}$-statistic-based Framework for Consistent Bayesian Inference of Compact Binary Coalescences

Hai-Tian Wang (王海天)[1,*]

[1] *School of Physics, Dalian University of Technology, Dalian 116024, People's Republic of China*

(Dated: May 15, 2026)

We present a comprehensive investigation of the $\mathcal{F}$-statistic method for parameter estimation of gravitational wave (GW) signals from compact binary coalescences. By analytically maximizing the likelihood over the luminosity distance and polarization angle, this approach reduces the dimensionality of the parameter space to enhance computational efficiency. We also introduce a novel formulation for calculating the Bayesian evidence for the $\mathcal{F}$-statistic, enabling a quantitative assessment of its performance against standard full frequency-domain (FFD) Bayesian inference. Applying these two methods to analyze several representative GW events (GW190412, GW190814, and GW170817), we find that the $\mathcal{F}$-statistic consistently yields results in good agreement with the FFD approach, while offering a significant reduction in computational cost. We demonstrate that including calibration uncertainty generally improves the agreement between the two methods. Furthermore, under the assumption of physical priors, the $\mathcal{F}$-statistic-based analyses consistently yield higher Bayesian evidence than the corresponding FFD analyses. While the $\mathcal{F}$-statistic produces slightly broader constraints on some parameters, we argue this represents a more honest uncertainty quantification, particularly in high-dimensional parameter spaces with complex posterior structures. These results highlight the significant advantages of the $\mathcal{F}$-statistic method for GW data analysis, positioning it as a powerful tool for the era of high-rate detections with future observatories.

arXiv:2509.14849v2 [gr-qc] 14 May 2026

I. INTRODUCTION

During the initial three observing runs conducted by the LIGO-Virgo-KAGRA Collaboration (LVKC), over 90 gravitational wave events were identified [1–3]. Notably, more than 90% of these detections are binary black hole (BBH) mergers, making them the dominant source type in the gravitational wave (GW) catalog. These BBH systems are invaluable for testing general relativity (GR), as their strong-field dynamics offer stringent constraints on deviations from Einstein's predictions [4–6].

GW signals are buried deep within detector noise, requiring sophisticated methods to extract their underlying physical parameters. Researchers typically employ Bayesian inference for this purpose. Although conventional Bayesian methods have been successfully applied in both real data analysis and simulated signal injection tests [7–10], they face significant challenges when analyzing the majority of detected GW events, which often have low signal-to-noise ratio (SNR). Furthermore, systematic uncertainties arising from detector calibration can bias the recovery of physical parameters if not properly accounted for [11–14]. For such events, the high-dimensional parameter space (up to 15 dimensions) leads to prolonged computational times for parameter estimation. Moreover, when the posterior distribution exhibits a multimodal structure, the accuracy of parameter estimation can diminish. These limitations highlight the need for more efficient and robust approaches in GW data analysis.

To address these challenges, researchers have developed various techniques to enhance the standard Bayesian framework, such as marginalization [15–17], reduced order quadrature algorithms [18–20], machine learning [21–24], and the $\mathcal{F}$-statistic [25–27]. The core innovation of the $\mathcal{F}$-statistic is its analytical maximization over certain time- and frequency-independent parameters, which reduces the dimensionality of the parameter space and enhances both the efficiency and accuracy of parameter estimation. The $\mathcal{F}$-statistic method has been widely applied in GW research, including searches for continuous waves from isolated neutron stars [28], binary white dwarf systems [29], and BBHs [30]; investigations of extreme mass-ratio inspirals [31]; and studies of ringdown signals [27, 32, 33] or the full inspiral-merger-ringdown (IMR) signal [34–36].

Previous applications of the $\mathcal{F}$-statistic in the context of IMR GW analyses have largely been developed with a focus on signal searches and detection studies. These early investigations typically employed waveform models appropriate to the computational and modeling capabilities available at the time, and were instrumental in establishing the feasibility and utility of the $\mathcal{F}$-statistic framework for GW data analysis [34]. Since then, GW modeling and inference techniques have undergone substantial advances. Contemporary IMR analyses routinely employ waveform families that incorporate higher-order modes and orbital precession, such as `IMRPhenomXPHM` [37] and `IMRPhenomXPHM-SpinTaylor` [38], which are now known to play an important role in accurately describing events like GW190412 [39] and GW190814 [40]. In parallel, Bayesian parameter estimation methods have matured significantly, supported by increasingly sophisticated sampling algorithms. Motivated by these developments, this work revisits the $\mathcal{F}$-statistic method in the context of modern, physically complete waveform models and focuses on its application to parameter estimation

* wanght9@dlut.edu.cn

for individual IMR events. We further introduce a consistent formulation for computing the Bayesian evidence associated with the $\mathcal{F}$-statistic, which enables direct and quantitative comparisons with standard Bayesian inference approaches. Our results demonstrate that, in this setting, the $\mathcal{F}$-statistic method provides a computationally efficient and robust alternative for GW parameter estimation.

This paper is organized as follows. In Sec. II, we detail the framework for full frequency-domain (FFD) Bayesian inference and the derivation of the $\mathcal{F}$-statistic method. We then present the results of our comparative analysis using data from different types of events in Sec. III. Finally, we provide a summary of our findings and discuss their implications in Sec. IV. Unless stated otherwise, we use geometric units where $G = c = 1$ throughout this paper.

II. METHODOLOGY

In this section, the statistical methods employed for parameter estimation are presented. We begin by outlining the conventional FFD framework, which serves as the benchmark for our analysis. Subsequently, we detail the derivation of the $\mathcal{F}$-statistic for a complete IMR signal. This derivation proceeds in two stages: first, the waveform model is reformulated to isolate parameters that appear linearly in the signal model; second, the profile likelihood is constructed by analytically maximizing the likelihood function with respect to these parameters. Following the derivation, we describe the procedure for reconstructing the posterior distributions of the analytically maximized parameters. A novel formulation for calculating the Bayesian evidence, and consequently the Bayes factor, for the $\mathcal{F}$-statistic method is also introduced. The section concludes with an outline of the prior distributions and the specific sampler configurations adopted for this work.

A. Full Bayesian Inference Framework

In Bayesian inference, the posterior probability density function (PDF) of a set of parameters Θ, given the data $\mathbf{d}$, is determined by Bayes' theorem:

$$P(\Theta|\mathbf{d}) = \frac{\mathcal{L}(\mathbf{d}|\Theta)P(\Theta)}{P(\mathbf{d})}, \tag{1}$$

where $\mathcal{L}(\mathbf{d}|\Theta)$ is the likelihood function, $P(\Theta)$ is the prior PDF representing our knowledge before analyzing the data, and $P(\mathbf{d})$ is the evidence. The evidence is a normalization constant for parameter estimation but is crucial for model comparison.

For a GW signal $\mathbf{h}(\Theta)$ embedded in stationary and Gaussian noise, the likelihood is constructed using the noise-weighted inner product, defined as:

$$\langle \mathbf{h}_1(f)|\mathbf{h}_2(f)\rangle = 4\int_{f_{\rm low}}^{f_{\rm high}} \frac{\mathbf{h}_1^*(f)\mathbf{h}_2(f)}{S_n(f)}\mathbf{d}f,$$

where $\mathbf{h}_{1,2}(f)$ are frequency-domain waveforms or data, $\mathbf{h}^*$ denotes complex conjugation, and $S_n(f)$ is the one-sided power spectral density (PSD) of the detector noise. The log-likelihood function, $\ln\mathcal{L}$, is then given by

$$\begin{aligned}\ln\mathcal{L}(\Theta) &= -\frac{1}{2}\langle \mathbf{d}-\mathbf{h}(\Theta)|\mathbf{d}-\mathbf{h}(\Theta)\rangle \\ &= \ln\Lambda(\Theta) - \frac{1}{2}\langle \mathbf{d}|\mathbf{d}\rangle,\end{aligned} \tag{2}$$

where the log-likelihood ratio, $\ln\Lambda(\Theta)$, is

$$\ln\Lambda(\Theta) = \langle \mathbf{d}|\mathbf{h}(\Theta)\rangle - \frac{1}{2}\langle \mathbf{h}(\Theta)|\mathbf{h}(\Theta)\rangle. \tag{3}$$

This approach requires numerical sampling over the *full* set of model parameters, Θ. Throughout this paper, we refer to this method as FFD.

B. The $\mathcal{F}$-statistic Method

The $\mathcal{F}$-statistic method offers an alternative to the FFD approach by reducing the dimensionality of the parameter space, thereby increasing computational efficiency. This is achieved by analytically maximizing the likelihood function over a subset of parameters that appear linearly in the waveform model. The derivation consists of two primary stages, which are detailed in the following subsections. First, the waveform template is reformulated to isolate these "linear" parameters from the remaining "non-linear" parameters, $\boldsymbol{\theta}$. Second, the profile likelihood, known as the $\mathcal{F}$-statistic, is constructed by analytically solving for the values of the linear parameters that maximize the likelihood for any given set of the non-linear parameters.

1. *Waveform Reformulation*

A frequency-domain IMR signal from a BBH merger is composed of two polarization modes, $\mathbf{h}_+(f)$ and $\mathbf{h}_\times(f)$. The signal observed in a detector, $\mathbf{h}(f)$, is a linear combination of these modes, weighted by the antenna pattern functions, F_+ and $F_\times$:

$$\mathbf{h}(f) = F_+\mathbf{h}_+(f) + F_\times\mathbf{h}_\times(f). \tag{4}$$

The antenna patterns are functions of the source's sky location (right ascension α, declination δ) and the polarization angle ψ. Specifically, they take the form:

$$\begin{aligned}F_+ &= F_1\cos 2\psi - F_2\sin 2\psi, \\ F_\times &= F_1\sin 2\psi + F_2\cos 2\psi,\end{aligned} \tag{5}$$

where $F_{1,2}$ are functions of α and δ. For the short-duration signals typical of LVKC detections, the time-variation of the antenna patterns is negligible, and they can be treated as independent of frequency.

The full waveform model depends on a total of 15 intrinsic and extrinsic parameters. The central idea of the $\mathcal{F}$-statistic is to reformulate the waveform to separate a subset of these parameters—the "linear" parameters—from the remaining "non-linear" parameters, denoted by the vector $\boldsymbol{\theta}$. The waveform is thus written as a linear combination of basis waveforms $\mathbf{h}_\mu$:

$$\mathbf{h}(f) = B^\mu \mathbf{h}_\mu(f), \tag{6}$$

where the Einstein summation convention is implied[1]. The coefficients B^μ are functions of the linear parameters, while the basis waveforms $\mathbf{h}_\mu(f)$ depend on $\boldsymbol{\theta}$.

Ideally, four parameters are candidates for this linear separation: luminosity distance (d_L), polarization angle (ψ), orbital inclination (ι), and a reference phase (ϕ_c). However, for realistic waveform models that include higher-order multipoles and orbital precession, the dependencies on ι and ϕ_c become highly complex and are entangled with the non-linear parameters. Therefore, in practice, only d_L and ψ can be analytically isolated.

By substituting the expressions for the antenna patterns into the detector response and separating the terms containing d_L and ψ, the waveform can be expressed as $h(f) = B^1\mathbf{h}_1(f) + B^2\mathbf{h}_2(f)$, where:

$$\begin{aligned} B^1 &= \frac{\cos 2\psi}{d_L}, \\ B^2 &= \frac{\sin 2\psi}{d_L}, \\ \mathbf{h}_1(f) &= F_1\mathbf{h}_+(f) + F_2\mathbf{h}_\times(f), \\ \mathbf{h}_2(f) &= F_1\mathbf{h}_\times(f) - F_2\mathbf{h}_+(f). \end{aligned} \tag{7}$$

This reformulation is the first essential step toward constructing the $\mathcal{F}$-statistic.

2. *Deriving the Profile Likelihood*

By substituting the reformulated waveform from Eq. (7) into the log-likelihood ratio of Eq. (3), the likelihood can be expressed in terms of the linear parameters B^μ and non-linear parameters $\boldsymbol{\theta}$:

$$\ln\Lambda(\Theta) = B^\mu \mathbf{s}_\mu(\boldsymbol{\theta}) - \frac{1}{2}B^\mu \mathbf{M}_{\mu\nu}(\boldsymbol{\theta})B^\nu, \tag{8}$$

where $\mathbf{s}_\mu \equiv \langle \mathbf{d}|\mathbf{h}_\mu\rangle$ is the projection of the data onto the basis waveforms, and $\mathbf{M}_{\mu\nu} \equiv \langle \mathbf{h}_\mu|\mathbf{h}_\nu\rangle$ is the covariance matrix of these basis waveforms.

To derive the profile likelihood, we analytically maximize this expression with respect to B^μ. This is achieved by setting the partial derivative of $\ln\Lambda(\Theta)$ with respect to B^ν to zero:

$$\frac{\partial \ln\Lambda(\Theta)}{\partial B^\nu} = \mathbf{s}_\nu - B^\mu \mathbf{M}_{\mu\nu} = 0. \tag{9}$$

Solving for B^μ yields the maximum likelihood estimator, $\hat{B}^\mu$:

$$\hat{B}^\mu = \mathbf{M}^{\mu\nu}\mathbf{s}_\nu, \tag{10}$$

where $\mathbf{M}^{\mu\nu}$ is the inverse of the matrix $\mathbf{M}_{\mu\nu}$. The matrix of second partial derivatives is negative-definite, ensuring that $\hat{B}^\mu$ corresponds to a unique maximum for any given $\boldsymbol{\theta}$. Note that the values of $\hat{B}^\mu$ do not depend on the priors of B^μ.

Following Wang *et al.* [32], substituting this result back into Eq. (8) allows the log-likelihood ratio to be rewritten as:

$$\ln\Lambda(\Theta) = \frac{1}{2}\mathbf{s}_\mu\mathbf{M}^{\mu\nu}\mathbf{s}_\nu - \frac{1}{2}(B^\mu - \hat{B}^\mu)\mathbf{M}_{\mu\nu}(B^\nu - \hat{B}^\nu). \tag{11}$$

This rearranged form is powerful because, according to the conjunction rule of probability, $P(\boldsymbol{\theta}, B^\mu|\mathbf{d}) = P(\boldsymbol{\theta}|\mathbf{d})P(B^\mu|\boldsymbol{\theta}, \mathbf{d})$, it separates the posterior into two distinct components:

$$P(\boldsymbol{\theta}|\mathbf{d}) \propto e^{\frac{1}{2}\mathbf{s}_\mu\mathbf{M}^{\mu\nu}\mathbf{s}_\nu} p(\boldsymbol{\theta}), \tag{12}$$

$$P(B^\mu|\boldsymbol{\theta}, \mathbf{d}) \propto e^{-\frac{1}{2}(B^\mu - \hat{B}^\mu)\mathbf{M}_{\mu\nu}(B^\nu - \hat{B}^\nu)} p(B^\mu|\boldsymbol{\theta}), \tag{13}$$

where $p(\boldsymbol{\theta})$ and $p(B^\mu|\boldsymbol{\theta})$ are the prior distributions.

From this separation, we can define the log-likelihood for the non-linear parameters $\boldsymbol{\theta}$ as:

$$\ln\Lambda(\boldsymbol{\theta}) = \frac{1}{2}\mathbf{s}_\mu\mathbf{M}^{\mu\nu}\mathbf{s}_\nu, \tag{14}$$

which is defined as the $\mathcal{F}$-statistic and is the log-profile likelihood used in this study for sampling $\boldsymbol{\theta}$. The log-likelihood for the conditional posterior of B^μ is:

$$\ln\Lambda(B^\mu|\boldsymbol{\theta})[2] = -\frac{1}{2}(B^\mu - \hat{B}^\mu)\mathbf{M}_{\mu\nu}(B^\nu - \hat{B}^\nu). \tag{15}$$

For an analysis involving multiple detectors, the network $\mathcal{F}$-statistic is obtained by summing the $\mathbf{s}_\mu$ and $\mathbf{M}_{\mu\nu}$ terms from each detector.

C. Posterior Reconstruction for Maximized Parameters

Once the likelihood is expressed in the separated form of Eq. (11), there are two distinct statistical approaches

[1] In this work, we adopt the Einstein summation convention.

[2] It is written in this form for simplicity but is not rigorously formal; the complete expression should be written as $P(B^\mu|\boldsymbol{\theta}, \mathbf{d})/p(B^\mu|\boldsymbol{\theta})$.

for parameter estimation. The first approach is to use Eq. (11) as the likelihood and marginalize it over B^μ for sampling. This method has been adopted in several studies [41, 42] and, importantly, has been shown to produce posterior distributions consistent with the standard FFD approach for high SNR cases [42]. The second approach, which we adopt in this work, leverages the probabilistic decomposition shown in Eqs. (12) and (13). By using only the profile likelihood (the $\mathcal{F}$-statistic from Eq. (14)) for the primary analysis, we sample only the space of the non-linear parameters $\boldsymbol{\theta}$. This reduction in the dimensionality of the sampled parameter space is the key to the enhanced computational efficiency of the method.

A consequence of this approach is that the posterior distributions for the linear parameters B^μ (and thus for luminosity distance d_L and polarization angle ψ) are not obtained directly from the sampler. Instead, they must be constructed in a post-processing step. The $\mathcal{F}$-statistic analysis yields N_s posterior samples, $\{\boldsymbol{\theta}_i\}$, drawn from $P(\boldsymbol{\theta}|\mathbf{d})$. For each sample $\boldsymbol{\theta}_i$, the conditional posterior for B^μ, given by Eq. (13) and assuming a uniform prior $p(B^\mu|\boldsymbol{\theta})$, is a multivariate Gaussian distribution:

$$P(B^\mu|\boldsymbol{\theta}_i, \mathbf{d}) \sim \mathcal{N}\left(\hat{B}^\mu(\boldsymbol{\theta}_i), \mathbf{M}^{\mu\nu}(\boldsymbol{\theta}_i)\right), \tag{16}$$

where the mean vector $\hat{B}^\mu(\boldsymbol{\theta}_i)$ and covariance matrix $\mathbf{M}^{\mu\nu}(\boldsymbol{\theta}_i)$ are calculated using the specific values from the sample $\boldsymbol{\theta}_i$. To obtain the full marginal posterior $P(B^\mu|\mathbf{d})$, we marginalize over the uncertainty in $\boldsymbol{\theta}$ via composition: for each posterior sample $\boldsymbol{\theta}_i$, we draw 500 corresponding samples of B^μ from its conditional Gaussian distribution.

This procedure initially yields posteriors corresponding to a uniform prior on B^μ. To apply a different, physically motivated target prior, we use importance resampling. The weight for each sample is the ratio of the target prior to the original sampling prior:

$$w_i = \frac{p(B^\mu_i|\boldsymbol{\theta}_i, \otimes)}{p(B^\mu_i|\boldsymbol{\theta}_i, \ominus)}, \tag{17}$$

where $\ominus$ and $\otimes$ denote the uniform and target priors on B^μ, respectively.

D. Priors

In our framework, the non-linear parameters $\boldsymbol{\theta}$ and the linear parameters B^μ are assumed to be independent in the prior distribution. The joint prior therefore factorizes as $P(\Theta) = p(\boldsymbol{\theta})p(B^\mu)$, and the conditional prior $p(B^\mu|\boldsymbol{\theta})$ simplifies to $p(B^\mu)$.

The reconstruction method described in Sec. II C most naturally operates with a uniform prior on the linear parameters B^μ. It is crucial, however, to understand the prior this implies for the physical parameters: the luminosity distance d_L and the polarization angle ψ. The relationship is described by the transformation:

$$p(B^\mu) = d_L^3 p(d_L, \psi). \tag{18}$$

A uniform prior on B^μ indicates a uniform prior on $\psi \in [0, \pi)$ and an unphysical prior on the luminosity distance, where $p(d_L) \propto 1/d_L^3$. The physically motivated prior, which we use as our target prior for reporting final results, assumes sources are distributed uniformly in comoving volume. For the non-linear parameters $\boldsymbol{\theta}$, we adopt the same priors as those used by the LVKC in their catalog analyses [1–3].

E. Evidence and Bayes Factor Calculation

A key metric for model comparison is the Bayesian evidence, $\mathcal{Z}$. For the FFD method, the evidence for the full parameter set Θ is a direct output of nested sampling algorithms:

$$\mathcal{Z}(\Theta) = \int p(\boldsymbol{\theta}, B^\mu)\mathcal{L}(\boldsymbol{\theta}, B^\mu)\mathbf{d}\boldsymbol{\theta}dB^\mu. \tag{19}$$

In contrast, the $\mathcal{F}$-statistic method, by sampling only the non-linear parameters $\boldsymbol{\theta}$, directly yields only the evidence for that reduced parameter space, $\mathcal{Z}(\boldsymbol{\theta}) = \int p(\boldsymbol{\theta})\mathcal{L}(\boldsymbol{\theta})\mathbf{d}\boldsymbol{\theta}$. Similar to Wang *et al.* [43], we propose a rigorous method for incorporating the prior of the maximized parameters, B^μ, into the evidence calculation for the $\mathcal{F}$-statistic.

Specifically, we introduce a novel formulation to compute the full evidence for the $\mathcal{F}$-statistic framework, thereby enabling a robust Bayes factor comparison. The full evidence under a target prior (denoted by $\otimes$) is

$$\begin{aligned}\mathcal{Z}(\Theta|\otimes) &= \int \mathcal{L}(\boldsymbol{\theta}, B^\mu)p(\boldsymbol{\theta}, B^\mu|\otimes)\mathbf{d}\boldsymbol{\theta}dB^\mu \\ &= \int \mathcal{L}(\boldsymbol{\theta})\left(\int \Lambda(B^\mu|\boldsymbol{\theta})p(B^\mu|\boldsymbol{\theta}, \otimes)dB^\mu\right)p(\boldsymbol{\theta})\mathbf{d}\boldsymbol{\theta} \\ &= \int f(\boldsymbol{\theta})\mathcal{L}(\boldsymbol{\theta})p(\boldsymbol{\theta})\mathbf{d}\boldsymbol{\theta}. \end{aligned} \tag{20}$$

This integral can be approximated as an expectation value over the posterior samples of $\boldsymbol{\theta}$:

$$\mathcal{Z}(\Theta|\otimes) \approx \frac{\mathcal{Z}(\boldsymbol{\theta})}{N(\boldsymbol{\theta})}\sum_{i=1}^{N(\boldsymbol{\theta})} f(\boldsymbol{\theta}_i), \tag{21}$$

where the sum is over the $N(\boldsymbol{\theta})$ posterior samples of $\boldsymbol{\theta}$. The function $f(\boldsymbol{\theta})$ is calculated for each sample $\boldsymbol{\theta}_i$ as follows:

$$\begin{aligned} f(\boldsymbol{\theta}) &= \int \Lambda(B^\mu|\boldsymbol{\theta})p(B^\mu|\boldsymbol{\theta}, \otimes)dB^\mu \\ &\approx \frac{\mathcal{Z}(B^\mu|\boldsymbol{\theta}, \ominus)}{N(B^\mu|\boldsymbol{\theta}, \ominus)}\sum_{j=1}^{N(B^\mu|\boldsymbol{\theta}, \ominus)}\frac{p(B^\mu_j|\boldsymbol{\theta}, \otimes)}{p(B^\mu_j|\boldsymbol{\theta}, \ominus)}. \end{aligned} \tag{22}$$

In this expression, the sum is over $N(B^\mu|\boldsymbol{\theta}, \ominus)$ samples (we use 500 in this study) drawn from the conditional posterior $P(B^\mu|\boldsymbol{\theta}, \mathbf{d}, \ominus)$, which corresponds to a simple uniform prior (denoted by $\ominus$). $\mathcal{Z}(B^\mu|\boldsymbol{\theta}, \ominus)$ is the evidence for the linear parameters under this uniform prior.

With the full evidence $\mathcal{Z}_{\mathrm{F}s}$ computed, the Bayes factor, $\mathcal{B}^{\mathrm{s1}}_{\mathrm{s2}} = \mathcal{Z}_{\mathrm{s1}}/\mathcal{Z}_{\mathrm{s2}}$, can be used to quantitatively assess the stability under different sampler configurations.

It is important to emphasize that this evidence comparison is strictly conditional on the assumption that the joint prior factorizes as $P(\Theta) = p(\boldsymbol{\theta})p(B^{\mu})$, which implies $p(B^{\mu} \mid \boldsymbol{\theta}) = p(B^{\mu})$. This assumption is satisfied in essentially all current LVK IMR analyses relevant to this work. Under this condition, the $\mathcal{F}$-statistic evidence computed via importance sampling with physically motivated priors can in principle be compared directly with the FFD evidence. However, if one adopts a prior where the luminosity distance is not independent of other parameters (for example, a prior jointly coupled to the inclination angle), the current $\mathcal{F}$-statistic evidence expression would require modification, and a direct comparison would no longer be automatic. Furthermore, while a systematically higher $\mathcal{F}$-statistic evidence is qualitatively consistent with higher likelihood values, we caution against overstating this result; Bayesian evidence is most safely used for Bayes-factor-based model comparison within the same inference framework. We note that consistency between Bayes factors computed in the $\mathcal{F}$-statistic framework and in full Bayesian inference has already been examined in previous ringdown studies [32, 43, 44].

F. Sampler Configuration

For the main analyses, Bayesian inference was performed using the Bilby package [v2.7.0; 9], with the nested sampling algorithm implemented in Dynesty [v3.0.0; 8]. We adopted a robust sampler configuration: 1000 live points were used, the primary sampling method was set to "rwalk", the number of autocorrelation lengths was fixed to 50, and new points were selected using the "live" method—settings consistent with those employed in the GWTC-2.1 [2] and GWTC-3 [3] catalog analyses. For completeness, results obtained with alternative configurations—including different versions of Bilby and Dynesty, as well as varied sampler settings—are presented in the Appendix; all yield consistent conclusions. All analyses were parallelized using 128 threads to reduce computational time.

III. ANALYSIS AND RESULTS

In this work, we present a detailed comparison of parameter estimation results using the $\mathcal{F}$-statistic and FFD methods across three selected GW events: GW170817 [45], GW190412 [39], and GW190814 [40]. Notably, GW190814 has a mass ratio close to 10, with the secondary compact object's mass around $2.6\,M_{\odot}$, leaving open the possibility that it is a neutron star-black hole (NSBH) system. Therefore, for GW190814 we perform independent analyses using two different waveform models: `IMRPhenomXPHM-SpinTaylor` (a general-purpose model for black-hole binary coalescences) and `IMRPhenomNSBH` (specifically designed for NSBH systems). For completeness and as an additional benchmark, the analysis of the classic BBH event GW150914 is included in the Appendix. A.

Taken together, the test set covers several representative classes of GW sources currently reported by LVKC:

- BBH systems: GW190412 & GW190814 (where both spin precession and higher harmonics contribute significantly), both analyzed with `IMRPhenomXPHM-SpinTaylor`;
- NSBH system: GW190814, analyzed with `IMRPhenomNSBH`;
- Binary neutron star (BNS) system: GW170817, analyzed with `IMRPhenomXP-NRTidalv3` (including tidal deformability effects).

Note that our $\mathcal{F}$-statistic workflow is fully compatible with calibration error: the additional parameters induced by calibration uncertainty follow the standard spline model used in LVKC analyses [11–14]. These parameters enter the waveform model in a non-linear manner and are therefore treated in the same way as other non-linear parameters. Therefore, we present the analysis results both with and without taking into account the effect of calibration uncertainties.

For each source type, analysis settings—including parameter priors, sampling rate, data segment length, and other configuration choices—have been chosen to closely match the original LVKC published analyses, ensuring fair and meaningful comparison. Detailed configuration parameters for each event are provided together with the posterior samples accompanying this work, allowing full reproducibility of the presented results [46].

Our analysis is structured as follows. We begin by evaluating the overall performance and efficiency of the two methods by comparing the distributions of the log-likelihood ratio, $\ln \Lambda$, shown in Fig. 1. In each comparison between the FFD method and the $\mathcal{F}$-statistic method across different cases, the distributions associated with the $\mathcal{F}$-statistic method are consistently shifted toward higher values. This indicates a better fit to the data, a result attributable to the optimization of two linear parameters. For a more definitive and quantitative comparison, we compute the evidence, $\mathcal{Z}$, using the procedure outlined in Sec. II.

Next, we present results of each set of analyses one by one. For each analysis, we examine the posterior distributions of the luminosity distance (d_L) and polarization angle (ψ), which are analytically maximized in the $\mathcal{F}$-statistic framework and subsequently reconstructed. Furthermore, we provide a comprehensive comparison of all remaining source parameters and assess the stability of each method across different sampler configurations using the Jensen-Shannon divergence (JSD) (see Ref. [47] for details).

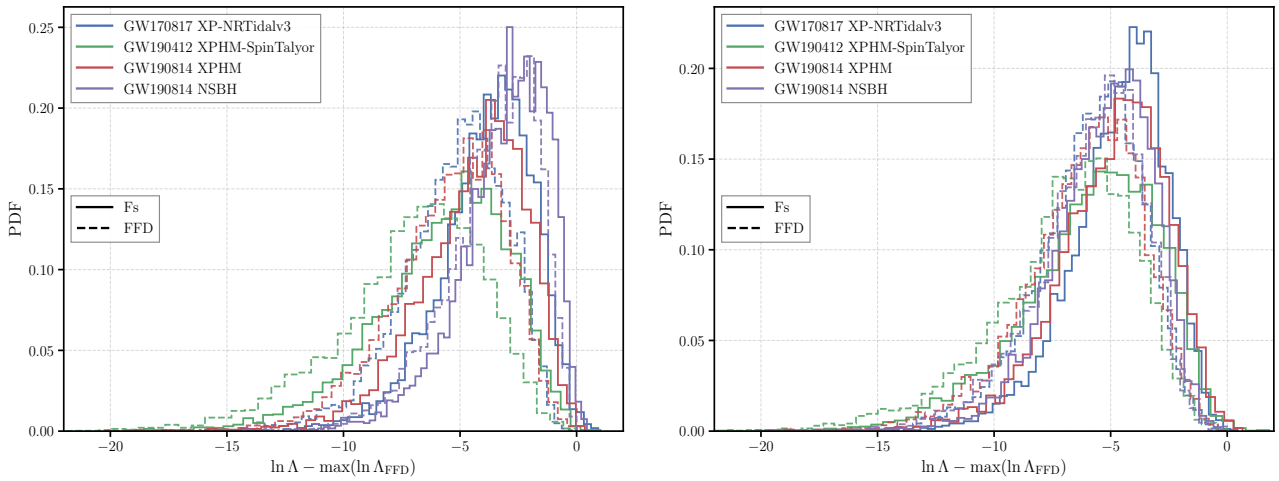


FIG. 1: Posterior distributions of the normalized log-likelihood ratio, $\ln\Lambda - \max(\ln\Lambda_{\rm FFD})$, for a set of representative gravitational-wave events, where the normalization is performed using the maximum log-likelihood obtained in the corresponding FFD analysis for each event. The dashed histograms show results from the FFD method, while the solid histograms correspond to the $\mathcal{F}$-statistic method (labeled as "Fs"). Different colors denote different events, as indicated in the legend. The left panel presents results obtained without accounting for calibration uncertainties, whereas the right panel shows results obtained when calibration uncertainty parameters are included and propagated through the inference via nested sampling. Interestingly, incorporating calibration uncertainties has a relatively minor impact on the differences between the two methods for a given event, while it significantly reduces the variation of the log-likelihood distributions across different events, leading to a greater degree of consistency among event-specific results.

A. Analysis of GW190412: An Asymmetric-Mass, Precessing Binary

We next apply our comparative framework to the BBH event GW190412 [39]. As a representative example of a complex merger alongside GW190814, this event features a highly asymmetric mass ratio, spin-induced orbital precession, and significant contributions from higher-order radiation multipoles. These physical properties make GW190412 an excellent test case for inference methods. In particular, the presence of higher-order modes helps break degeneracies involving extrinsic parameters, improving constraints on the luminosity distance and polarization angle. Because these are precisely the parameters that the $\mathcal{F}$-statistic analytically maximizes over, this event provides a stringent test of the method's ability to accurately reconstruct their posterior distributions while accelerating the analysis.

For GW190412, we conduct four parallel analyses, applying both the FFD and $\mathcal{F}$-statistic methods with and without marginalization over calibration uncertainties. All other settings are held identical to ensure a controlled comparison. This subsection first focuses on the inferred posterior distributions for the analytically maximized parameters—luminosity distance and polarization angle—before presenting a comprehensive comparison of all remaining source parameters.

1. *Reconstruction of Analytically Maximized Parameters*

The luminosity distance, d_L, and polarization angle, ψ, are treated fundamentally differently in the two frameworks. In the FFD approach, they are sampled directly along with all other parameters. In the $\mathcal{F}$-statistic method, they are analytically maximized, and their posteriors must be reconstructed in a post-processing step, as detailed in Sec. II C. Fig 2 presents a detailed comparison of the luminosity distance and polarization angle inferred using the $\mathcal{F}$-statistic and FFD approaches. These two parameters enter the detector response linearly and are therefore particularly sensitive to differences in likelihood construction and marginalization strategies. GW190412, with its asymmetric mass ratio and measurable higher-order mode content, provides sufficient information to partially break the distance–inclination–polarization degeneracies, making it an ideal system for such a comparison.

In the absence of calibration uncertainties, the two methods yield broadly consistent but not identical joint posteriors in the (d_L, ψ) plane. The residual discrepancies are most pronounced along the elongated degeneracy direction, reflecting the different treatments of the linear extrinsic parameters in the two inference schemes. When calibration errors are incorporated, the posterior distributions widen slightly, as expected, as these uncertainties are known to be partially degenerate with the luminosity distance and inclination [12–14]. However, the agreement between the $\mathcal{F}$-statistic and FFD results improves

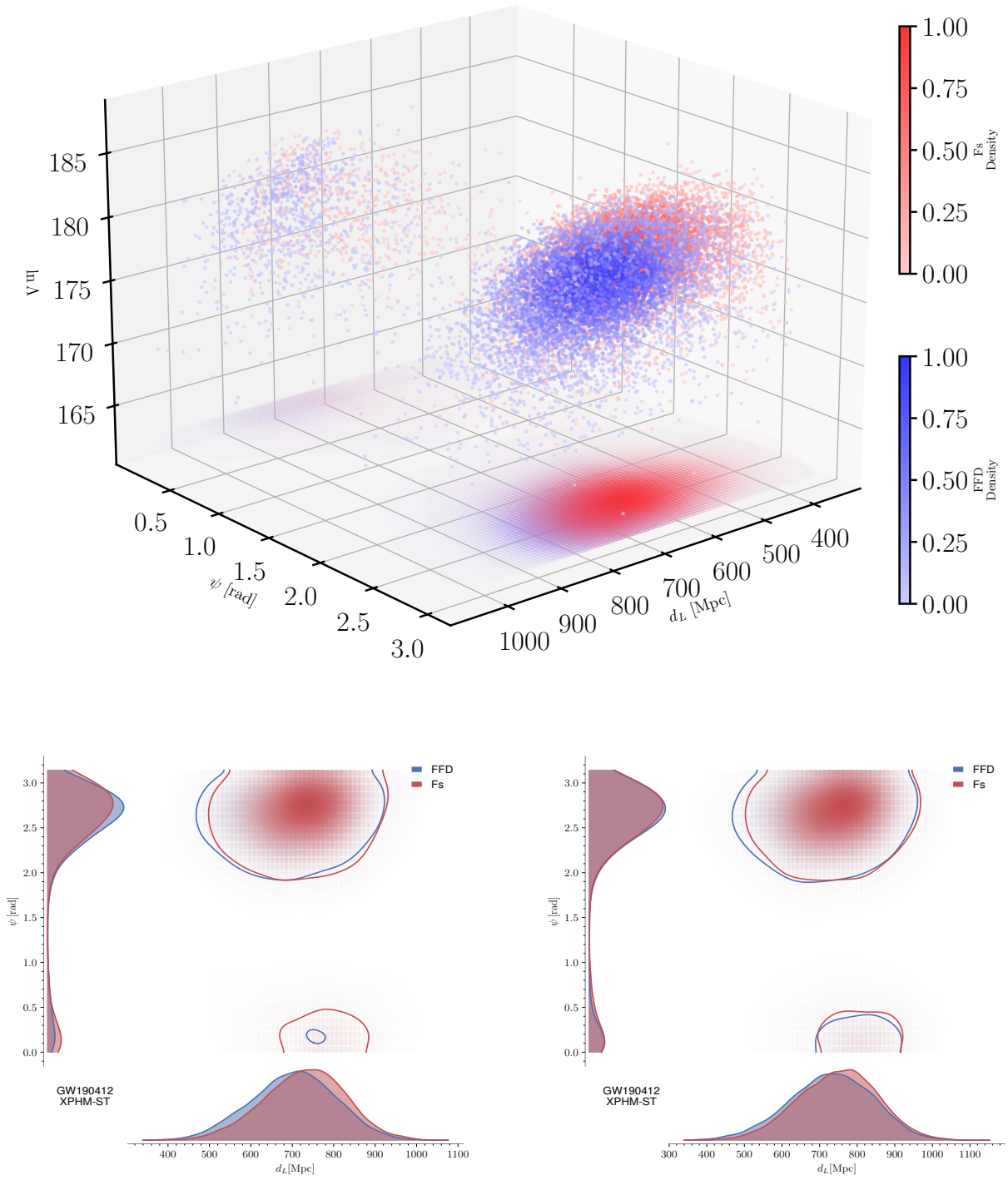


FIG. 2: Comparison of the luminosity distance d_L and polarization angle ψ inferred for GW190412 using the $\mathcal{F}$-statistic (Fs) and FFD likelihood approaches. The top panel shows a three-dimensional scatter plot of samples in the $(d_L, \psi, \ln \Lambda)$ space obtained with both methods when calibration uncertainties are included. Each point corresponds to a posterior sample, with the color indicating the relative sample density. This representation illustrates that the high-likelihood regions identified by the two approaches occupy consistent regions of parameter space. The lower left panel displays the joint posterior distributions of (d_L, ψ) obtained without accounting for calibration errors, while the lower right panel shows the corresponding posteriors when calibration uncertainties are included. Contours indicate the marginalized two-dimensional credible regions at 90% level, and the one-dimensional projections are shown along the axes. When calibration errors are neglected, small but noticeable differences between the $\mathcal{F}$-statistic and FFD posteriors are present, particularly along the degeneracy direction between d_L and ψ. Including calibration uncertainties broadens the posteriors and leads to substantially improved agreement between the two methods, indicating that calibration marginalization plays an important role in harmonizing likelihood-based inference schemes for these linear extrinsic parameters.

TABLE I: Posterior constraints for GW190412 estimated by $\mathcal{F}$-statistic (Fs) and FFD methods under two scenarios: without calibration error (noc) and with calibration error (c). For each parameter, we report the median and symmetric 90% credible intervals. The JSD quantifies the similarity between the posterior distributions inferred by the two methods under the same calibration assumptions. The lower rows list the number of posterior samples (N_s), total sampling time (T_s), and the log Bayesian evidence ($\ln \mathcal{Z}$) for each analysis.

	No Calibration Error (noc)			Calibration Error (c)		
Parameter	Fs-noc	FFD-noc	JSD	Fs-c	FFD-c	JSD
$t_c - t_0$[ms]	$0.96^{+0.93}_{-0.92}$	$1.01^{+0.98}_{-0.87}$	0.004	$0.96^{+0.91}_{-0.83}$	$0.95^{+0.89}_{-0.84}$	0.003
$\mathcal{M}[M_\odot]$	$15.25^{+0.30}_{-0.21}$	$15.26^{+0.31}_{-0.20}$	0.003	$15.25^{+0.33}_{-0.21}$	$15.27^{+0.36}_{-0.21}$	0.005
q	$0.28^{+0.09}_{-0.07}$	$0.28^{+0.08}_{-0.07}$	0.003	$0.28^{+0.10}_{-0.07}$	$0.27^{+0.09}_{-0.07}$	0.012
a_1	$0.37^{+0.16}_{-0.17}$	$0.38^{+0.16}_{-0.16}$	0.004	$0.36^{+0.16}_{-0.19}$	$0.39^{+0.16}_{-0.17}$	0.017
a_2	$0.54^{+0.41}_{-0.48}$	$0.54^{+0.39}_{-0.47}$	0.006	$0.53^{+0.41}_{-0.47}$	$0.51^{+0.44}_{-0.45}$	0.004
θ_1[rad]	$0.59^{+0.42}_{-0.35}$	$0.63^{+0.41}_{-0.34}$	0.009	$0.57^{+0.47}_{-0.37}$	$0.60^{+0.41}_{-0.36}$	0.007
θ_2[rad]	$1.33^{+1.21}_{-0.96}$	$1.30^{+1.15}_{-0.90}$	0.004	$1.32^{+1.21}_{-0.95}$	$1.34^{+1.24}_{-0.95}$	0.002
ϕ_{12}[rad]	$2.55^{+3.41}_{-2.28}$	$2.65^{+3.21}_{-2.28}$	0.005	$2.69^{+3.29}_{-2.39}$	$2.79^{+3.18}_{-2.48}$	0.003
ϕ_{JL}[rad]	$1.00^{+5.14}_{-0.89}$	$1.02^{+5.09}_{-0.88}$	0.005	$1.16^{+5.01}_{-1.04}$	$1.12^{+5.07}_{-1.01}$	0.010
θ_{JN}[rad]	$0.85^{+1.37}_{-0.29}$	$0.79^{+1.54}_{-0.27}$	0.027	$0.85^{+1.29}_{-0.31}$	$0.76^{+0.39}_{-0.28}$	0.044
ϕ[rad]	$2.00^{+1.15}_{-0.69}$	$2.06^{+1.22}_{-0.69}$	0.005	$2.04^{+1.30}_{-0.71}$	$2.04^{+1.35}_{-0.75}$	0.002
α[rad]	$3.81^{+0.04}_{-0.06}$	$3.81^{+0.04}_{-0.08}$	0.003	$3.81^{+0.04}_{-0.06}$	$3.81^{+0.03}_{-0.05}$	0.005
δ[rad]	$0.63^{+0.03}_{-0.04}$	$0.63^{+0.03}_{-0.05}$	0.004	$0.63^{+0.03}_{-0.04}$	$0.63^{+0.03}_{-0.04}$	0.003
d_L[Mpc]	$732.2^{+150.9}_{-172.5}$	$706.2^{+159.4}_{-175.7}$	0.014	$751.2^{+161.9}_{-179.1}$	$732.2^{+171.2}_{-192.8}$	0.008
ψ[rad]	$2.66^{+0.42}_{-2.54}$	$2.65^{+0.38}_{-2.24}$	0.017	$2.64^{+0.42}_{-2.53}$	$2.63^{+0.42}_{-2.53}$	0.007
N_s	7016	7559	-	7140	7141	-
$T_s[h]$	5.5	11.7	-	4.5	20.1	-
$\ln \mathcal{Z}$	146.7	145.9	-	146.7	146.6	-

1. Parameter definitions follow the conventions used in the PESummary package [48].
2. The geocentric time t_c is reported as an offset from the GW190412 trigger time, $t_0 = 1239082262.18$ s.

markedly.

These results demonstrate that calibration marginalization mitigates small method-dependent differences in the inference of linear extrinsic parameters, and leads to highly consistent posteriors between accelerated likelihood approaches. In particular, for d_L and ψ, the inclusion of calibration uncertainties effectively absorbs residual systematic differences between the $\mathcal{F}$-statistic and FFD formulations. This improved agreement is quantitatively reflected in Table I. Specifically, upon introducing calibration uncertainty, the JSD for the luminosity distance decreases from 0.014 to 0.008, while the JSD for the polarization angle drops from 0.017 to 0.007. These reductions confirm that including calibration uncertainty yields nearly indistinguishable posterior constraints for these parameters.

2. *Comparison of All Sampled Source Parameters*

Table I summarizes the full set of posterior constraints obtained for GW190412 using the $\mathcal{F}$-statistic and FFD approaches under identical analysis configurations. Across both calibration scenarios, we find excellent agreement between the two methods for all intrinsic and extrinsic parameters. This is quantitatively confirmed by the JSD values, which remain at the level of $\mathcal{O}(10^{-2})$ or below for the majority of parameters, indicating nearly indistinguishable posterior distributions.

Intrinsic parameters such as the chirp mass, mass ratio, and component spins show particularly strong consistency between the two methods, independent of whether calibration uncertainties are included. Extrinsic parameters, including sky location, coalescence phase, and polarization angle, also exhibit close agreement, with slightly larger but still small JSD values reflecting their broader and more degenerate posteriors. As already highlighted in the (d_L, ψ) comparison, the inclusion of calibration uncertainties systematically improves the consistency between the two approaches for parameters entering the detector response linearly.

In terms of computational performance, the $\mathcal{F}$-statistic method consistently yields a higher Bayesian evidence, as shown in Table I, demonstrating that the accelerated inference scheme does not compromise statistical fidelity. However, the sampling efficiency differs drastically between the two approaches. As detailed in Table I, in the absence of calibration errors, the $\mathcal{F}$-statistic method reduces the total sampling time (T_s) by approximately 53% (5.5 h vs 11.7 h). Notably, this computational advantage becomes significantly more pronounced when calibration uncertainties are included. While the computational cost

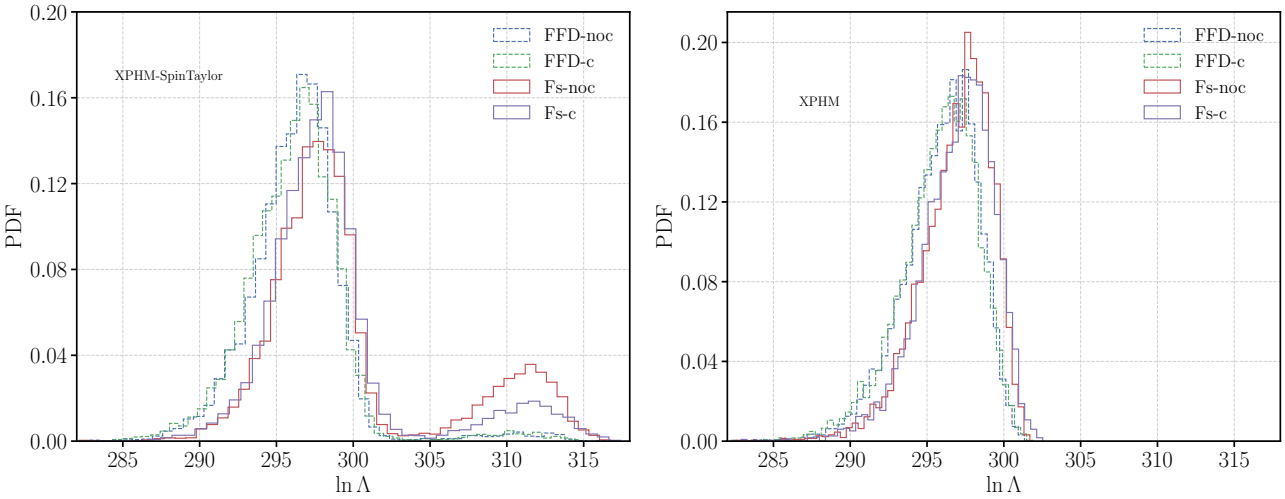


FIG. 3: Comparison of the log-likelihood distributions for GW190814 obtained using the $\mathcal{F}$-statistic (Fs) and FFD inference methods under different waveform and calibration assumptions. The left panel shows results obtained with the `IMRPhenomXPHM-SpinTaylor` waveform, while the right panel corresponds to the `IMRPhenomXPHM` waveform. For each waveform, we present four analyses: Fs and FFD, each performed with (c) and without (noc) including calibration uncertainties. For the SpinTaylor variant, the likelihood distribution exhibits a pronounced bimodal structure, whereas the `IMRPhenomXPHM` analysis yields a single dominant mode consistent with the LVK results. Including calibration uncertainty systematically suppresses the secondary likelihood peak in the Fs analysis, reducing the relative weight of the subdominant mode.

of the FFD analysis nearly doubles to 20.1 h in this scenario, the $\mathcal{F}$-statistic analysis completes in just 4.5 h—a speed-up factor of approximately 4.5. These results confirm that the $\mathcal{F}$-statistic method provides a robust and computationally efficient framework for complex GW parameter estimation, offering substantial time savings that become even more critical as the complexity of the inference problem increases.

Overall, the results presented here establish GW190412 as a stringent validation case for accelerated likelihood-based inference, and demonstrate that both the $\mathcal{F}$-statistic and FFD approaches can reliably recover high-dimensional posterior distributions even in the presence of higher-order modes, precession, and calibration uncertainties.

B. Analysis of GW190814: A High-Mass-Ratio System

Following the comparative study of GW190412, we extend our analysis to GW190814, a signal detected by the LVK detectors during the third observing run. GW190814 is a particularly distinct event characterized by the most unequal mass ratio observed to date ($q \approx 0.112$) and a secondary component mass of approximately $2.6\,\mathrm{M}_\odot$, which places it in the lower mass gap between known neutron stars and black holes. The highly asymmetric masses amplify the emission of higher-order multipoles, making their contribution to the signal significant. Furthermore, the unique parameter space of this event provides a stringent testing ground for waveform systematics and inference methods, particularly regarding the interplay between spin-precession effects and higher-order modes. To evaluate the performance of the FFD method and the $\mathcal{F}$-statistic method on this source, we perform four distinct analyses. These consist of applying both methods under two conditions: one neglecting calibration errors and one marginalizing over calibration uncertainties.

Regarding the choice of waveform approximants, our approach for GW190814 deviates slightly from the previous section. While the GW190412 analysis relied exclusively on the `IMRPhenomXPHM-SpinTaylor` model, our preliminary analysis of GW190814 using this waveform revealed a bimodal structure in the likelihood distribution, as shown in Fig. 3. In contrast, the posterior distributions reported by the LVK collaboration—based on `IMRPhenomXPHM`—exhibit a single peak. To investigate this discrepancy and ensure robust parameter estimation, we additionally employ the `IMRPhenomXPHM` approximant. We find that `IMRPhenomXPHM` yields a unimodal likelihood distribution consistent with LVK results, allowing for a direct comparison with the SpinTaylor variant. Furthermore, given the ambiguous nature of the secondary object—which may be either a heavy neutron star or a light black hole—we also conduct a separate analysis using the `IMRPhenomNSBH` waveform model to assess potential tidal signatures.

The analysis of this event is structured as follows: First, we compare the likelihood distributions obtained from `IMRPhenomXPHM` and `IMRPhenomXPHM-SpinTaylor` to address the observed bimodality. Second, we examine the joint posterior distributions of luminosity distance and polarization angle. Third, we present a comprehensive comparison of the constraints on intrinsic and extrin-

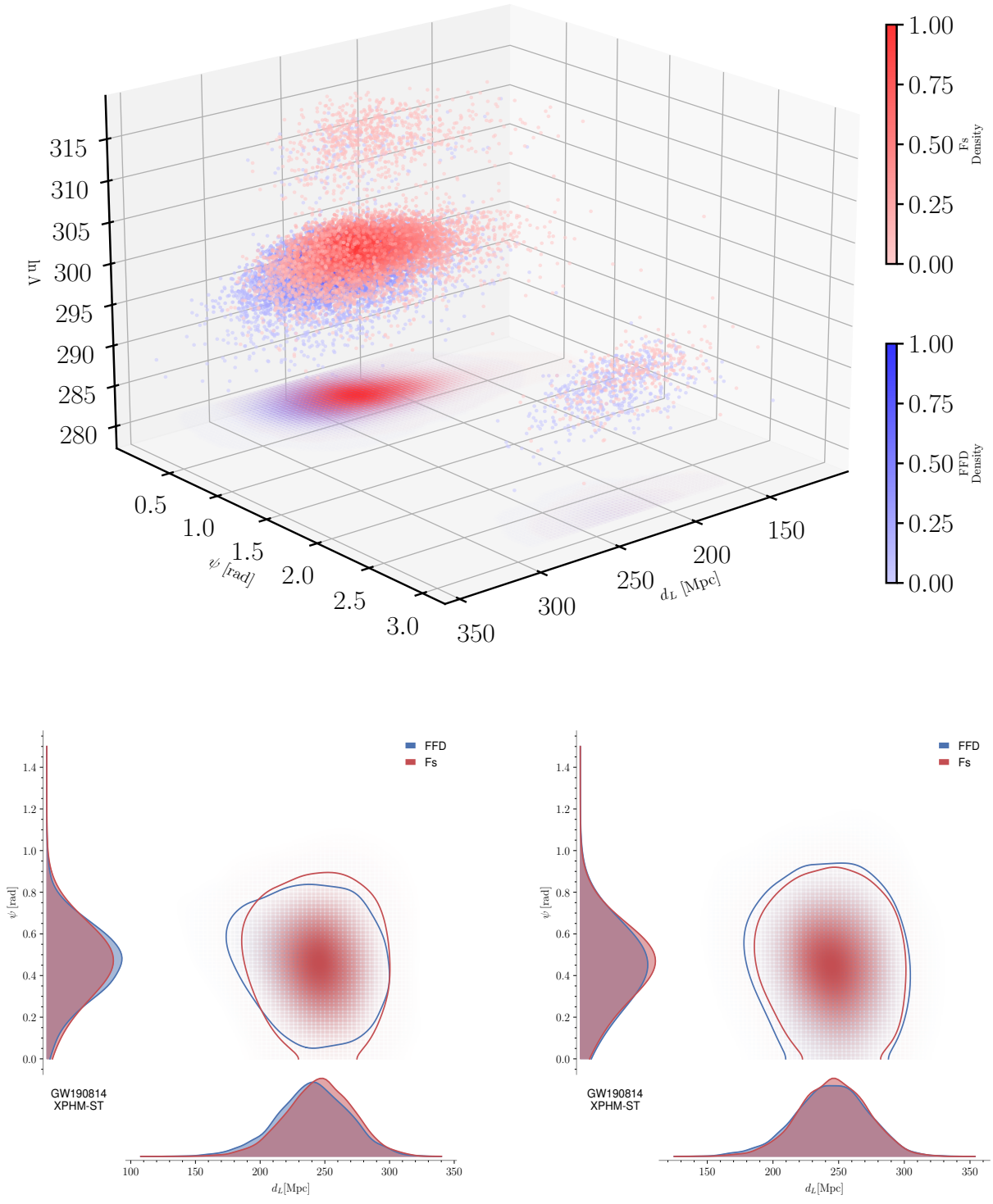


FIG. 4: Joint distributions of luminosity distance d_L and polarization angle ψ for GW190814 inferred using the `IMRPhenomXPHM-SpinTaylor` waveform. The arrangement of panels is the same as in Fig. 2. While both methods recover consistent primary modes, secondary structures associated with the bimodal likelihood are more visible in the $\mathcal{F}$-statistic cases. Marginalization over calibration uncertainties reduces the prominence of these secondary features and leads to improved agreement between the $\mathcal{F}$-statistic and FFD posteriors.

TABLE II: Posterior parameter estimates for GW190814 obtained using the `IMRPhenomXPHM-SpinTaylor` waveform with the $\mathcal{F}$-statistic (Fs) and FFD methods. Results are shown for analyses performed without (noc) and with (c) marginalization over calibration uncertainties. For each parameter, we report the median and symmetric 90% credible intervals, together with the JSD quantifying the agreement between the $\mathcal{F}$-statistic and FFD posterior distributions. Also listed are the number of posterior samples (N_s), total sampling time (T_s), and the log Bayesian evidence ($\ln \mathcal{Z}$). Including calibration uncertainty systematically reduces the JSD for most parameters, indicating improved consistency between the two inference methods.

	No Calibration Error (noc)			Calibration Error (c)		
Parameter	Fs-noc	FFD-noc	JSD	Fs-c	FFD-c	JSD
$t_c - t_0$[ms]	$-0.09^{+0.33}_{-0.33}$	$-0.11^{+0.33}_{-0.33}$	0.007	$-0.10^{+0.35}_{-0.34}$	$-0.09^{+0.38}_{-0.35}$	0.003
$\mathcal{M}[M_\odot]$	$6.41^{+0.02}_{-0.04}$	$6.41^{+0.02}_{-0.04}$	0.005	$6.41^{+0.02}_{-0.05}$	$6.41^{+0.02}_{-0.04}$	0.003
q	$0.11^{+0.01}_{-0.01}$	$0.11^{+0.01}_{-0.01}$	0.025	$0.11^{+0.01}_{-0.01}$	$0.11^{+0.01}_{-0.01}$	0.008
a_1	$0.03^{+0.08}_{-0.02}$	$0.04^{+0.07}_{-0.04}$	0.034	$0.03^{+0.09}_{-0.03}$	$0.04^{+0.08}_{-0.04}$	0.012
a_2	$0.51^{+0.42}_{-0.45}$	$0.49^{+0.43}_{-0.42}$	0.004	$0.50^{+0.44}_{-0.44}$	$0.48^{+0.45}_{-0.43}$	0.003
θ_1[rad]	$1.55^{+1.17}_{-1.15}$	$1.60^{+1.09}_{-1.12}$	0.005	$1.51^{+1.21}_{-1.15}$	$1.51^{+1.20}_{-1.14}$	0.004
θ_2[rad]	$1.61^{+0.94}_{-1.11}$	$1.51^{+1.01}_{-1.03}$	0.009	$1.59^{+1.03}_{-1.10}$	$1.56^{+1.06}_{-1.08}$	0.004
ϕ_{12}[rad]	$3.50^{+2.35}_{-3.05}$	$3.20^{+2.65}_{-2.77}$	0.045	$3.47^{+2.41}_{-3.06}$	$3.15^{+2.73}_{-2.80}$	0.016
ϕ_{JL}[rad]	$2.03^{+3.89}_{-1.76}$	$1.59^{+4.13}_{-1.30}$	0.018	$2.08^{+3.89}_{-1.77}$	$1.86^{+4.01}_{-1.58}$	0.006
θ_{JN}[rad]	$0.90^{+1.37}_{-0.24}$	$0.85^{+0.79}_{-0.22}$	0.024	$0.90^{+1.40}_{-0.24}$	$0.85^{+1.54}_{-0.24}$	0.035
ϕ[rad]	$2.75^{+0.51}_{-1.54}$	$2.74^{+0.51}_{-1.50}$	0.003	$2.76^{+0.51}_{-1.60}$	$2.75^{+0.53}_{-1.49}$	0.005
α[rad]	$0.22^{+0.16}_{-0.03}$	$0.22^{+0.08}_{-0.03}$	0.009	$0.23^{+0.17}_{-0.03}$	$0.23^{+0.17}_{-0.03}$	0.006
δ[rad]	$-0.44^{+0.03}_{-0.12}$	$-0.44^{+0.04}_{-0.07}$	0.006	$-0.44^{+0.04}_{-0.12}$	$-0.44^{+0.04}_{-0.13}$	0.007
d_L[Mpc]	$246.3^{+40.9}_{-44.3}$	$241.1^{+43.5}_{-48.5}$	0.009	$246.0^{+41.8}_{-43.6}$	$244.7^{+43.5}_{-48.1}$	0.005
ψ[rad]	$0.47^{+0.37}_{-0.32}$	$0.47^{+0.28}_{-0.30}$	0.011	$0.46^{+0.50}_{-0.34}$	$0.46^{+2.54}_{-0.35}$	0.004
N_s	7484	6348	-	6817	6780	-
$T_s[h]$	34.3	44.1	-	59.6	95.6	-
$\ln \mathcal{Z}$	261.3	261.0	-	261.4	260.1	-

1. The geocentric time t_c is reported as an offset from the GW190814 trigger time, $t_0 = 1249852256.99$ s.

sic parameters. Finally, we report the results obtained using the `IMRPhenomNSBH` template.

1. *Results of IMRPhenomXPHM-SpinTaylor*

When analyzed with the `IMRPhenomXPHM-SpinTaylor` waveform, the log-likelihood distribution exhibits a clear bimodal structure for both the $\mathcal{F}$-statistic and FFD runs (Fig. 3). Inspecting the high-dimensional likelihood landscape suggests that the subdominant mode, which occurs at higher likelihood values, is associated with a different region of extrinsic-parameter space, with luminosity distance and polarization angle being among the most affected directions (top panel of Fig. 4). Notably, the $\mathcal{F}$-statistic method appears to explore this subdominant, higher-likelihood mode more extensively. However, this bimodality observed in the distribution of $\ln \Lambda$ does not translate into a distinctly bimodal marginalized posterior in the (d_L, ψ) plane.

Such bimodality is absent in the LVK posterior results, which were obtained using the `IMRPhenomXPHM` waveform. To investigate this discrepancy, we repeat the analysis using `IMRPhenomXPHM`. In this case, the likelihood distribution becomes unimodal and closely matches the LVK results, indicating that the secondary likelihood peak observed with the SpinTaylor variant is waveform-dependent rather than a sampling artifact. This behavior highlights the sensitivity of GW190814 to subtle modeling differences in the treatment of spin precession and higher-order modes.

An important methodological observation is the role of calibration uncertainty in mitigating the effects of likelihood multimodality. As shown in Fig. 3, including calibration uncertainty systematically reduces the relative prominence of the secondary likelihood peak in the $\mathcal{F}$-statistic analysis. This improved robustness is also reflected quantitatively in Table. II. For the majority of parameters, the JSD between the $\mathcal{F}$-statistic and FFD posteriors decreases when calibration uncertainty is included, indicating better consistency between the two inference methods. For example, the JSD for the primary spin magnitude a_1 drops from 0.034 to 0.012, while for the spin phase angle ϕ_{12} it decreases from 0.045 to 0.016. Similarly, the agreement for luminosity distance d_L and polarization angle ψ improves, with JSDs falling from 0.009 to 0.005 and 0.011 to 0.004, respectively.

In terms of computational performance, the $\mathcal{F}$-statistic method retains its efficiency advantage. In the absence of calibration errors, the $\mathcal{F}$-statistic analysis completed in 34.3 hours, compared to 44.1 hours for the FFD method—a time saving of about 22%. Although this

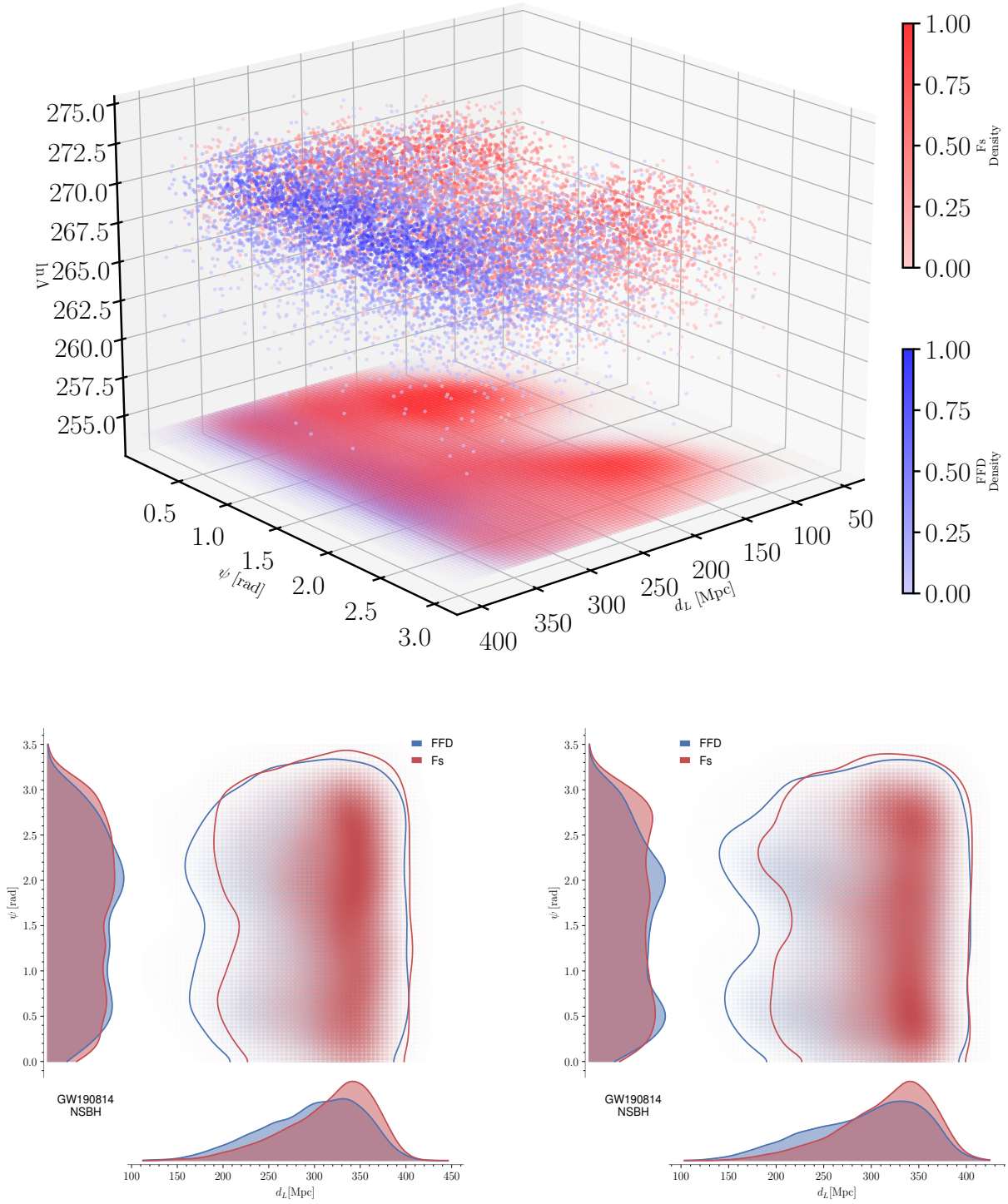


FIG. 5: Same as Fig. 2, but for the analysis of GW190814 under the NSBH hypothesis using the `IMRPhenomNSBH` waveform model.

TABLE III: Parameters of GW190814 inferred using the `IMRPhenomNSBH` waveform model under the NSBH hypothesis, estimated with the $\mathcal{F}$-statistic (Fs) and FFD methods. Results are shown for analyses without calibration uncertainty (noc) and with calibration uncertainty (c), including the tidal deformability parameter of the secondary object, λ_2. The JSD quantifies the agreement between the posterior distributions obtained with the two methods. The table also reports the number of posterior samples (N_s), sampling time (T_s), and log Bayesian evidence ($\ln \mathcal{Z}$) for each analysis.

	No Calibration Error (noc)			Calibration Error (c)		
Parameter	Fs-noc	FFD-noc	JSD	Fs-c	FFD-c	JSD
$t_c - t_0$[ms]	$1.40^{+2.19}_{-3.31}$	$1.58^{+1.99}_{-2.93}$	0.014	$1.11^{+2.50}_{-3.85}$	$1.36^{+2.27}_{-3.67}$	0.017
$\mathcal{M}[M_\odot]$	$6.51^{+0.44}_{-0.45}$	$6.52^{+0.41}_{-0.44}$	0.004	$6.51^{+0.44}_{-0.46}$	$6.52^{+0.41}_{-0.45}$	0.005
q	$0.19^{+0.63}_{-0.14}$	$0.20^{+0.58}_{-0.14}$	0.004	$0.19^{+0.64}_{-0.13}$	$0.20^{+0.60}_{-0.14}$	0.004
χ_1	$0.00^{+0.16}_{-0.13}$	$-0.01^{+0.09}_{-0.11}$	0.052	$0.01^{+0.22}_{-0.15}$	$-0.01^{+0.13}_{-0.13}$	0.047
χ_2	$0.00^{+0.42}_{-0.40}$	$-0.01^{+0.40}_{-0.40}$	0.008	$0.01^{+0.46}_{-0.44}$	$0.01^{+0.46}_{-0.42}$	0.011
θ_{JN}[rad]	$0.97^{+1.57}_{-0.66}$	$0.66^{+1.74}_{-0.45}$	0.105	$0.99^{+1.59}_{-0.68}$	$0.71^{+1.93}_{-0.50}$	0.075
ϕ[rad]	$3.13^{+2.78}_{-2.78}$	$2.98^{+2.75}_{-2.49}$	0.007	$3.16^{+2.80}_{-2.84}$	$2.65^{+3.05}_{-2.28}$	0.014
α[rad]	$0.23^{+0.18}_{-0.04}$	$0.22^{+0.16}_{-0.03}$	0.020	$0.23^{+0.18}_{-0.04}$	$0.23^{+0.18}_{-0.03}$	0.002
δ[rad]	$-0.44^{+0.04}_{-0.13}$	$-0.43^{+0.04}_{-0.12}$	0.019	$-0.44^{+0.04}_{-0.13}$	$-0.44^{+0.04}_{-0.13}$	0.004
d_L[Mpc]	$326.0^{+53.8}_{-104.0}$	$303.9^{+68.5}_{-104.0}$	0.027	$324.0^{+55.6}_{-101.9}$	$304.4^{+68.4}_{-115.4}$	0.031
ψ[rad]	$1.64^{+1.34}_{-1.48}$	$1.60^{+1.26}_{-1.36}$	0.011	$1.54^{+1.40}_{-1.38}$	$1.48^{+1.37}_{-1.29}$	0.012
λ_2	$2559.6^{+2157.9}_{-2263.2}$	$2463.1^{+2067.5}_{-2096.0}$	0.008	$2526.9^{+2191.9}_{-2250.0}$	$2564.7^{+2057.9}_{-2228.3}$	0.004
N_s	4921	5106	-	5660	5856	-
$T_s[h]$	9.7	12.8	-	17.9	40.8	-
$\ln \mathcal{Z}$	237.5	237.3	-	237.2	236.9	-

1. The geocentric time t_c is reported as an offset from the GW190814 trigger time, $t_0 = 1249852256.99$ s.

speed-up is less pronounced than in other analyses in this work, likely due to the additional sampling required to explore the bimodal likelihood landscape, the advantage of the $\mathcal{F}$-statistic method becomes more significant when calibration uncertainty is included. In this more complex scenario, the $\mathcal{F}$-statistic method took 59.6 hours versus 95.6 hours for the FFD, increasing the time saving to approximately 38%. Furthermore, the $\mathcal{F}$-statistic analysis yields a higher Bayesian evidence in both scenarios, suggesting a better exploration of the parameter space. This demonstrates that even for challenging signals with complex posterior structures, the $\mathcal{F}$-statistic method provides a notable and increasingly beneficial reduction in computational cost.

2. *Results of IMRPhenomNSBH*

While the preceding analysis treated GW190814 as a BBH coalescence, the true nature of the secondary component remains ambiguous. With a mass of approximately 2.6 $M_\odot$, this object sits within the lower mass gap, potentially representing either the lightest black hole or the heaviest neutron star ever observed in a compact binary system. Given this uncertainty, it is essential to investigate the signal under the hypothesis that the secondary is a neutron star, which introduces the possibility of tidal interactions. To this end, we repeat our analysis using the `IMRPhenomNSBH` waveform approximant, which includes the tidal deformability parameter (λ_2) for the secondary object. In this section, we present the results of this NSBH-specific analysis. We begin by illustrating the joint posterior distributions of the luminosity distance and polarization angle to facilitate comparison with the BBH models discussed previously. Following this, we report the detailed constraints on the full set of source parameters inferred from the `IMRPhenomNSBH` template.

Fig. 5 presents the joint posterior distributions of the luminosity distance and polarization angle for GW190814 under the NSBH hypothesis, inferred using the `IMRPhenomNSBH` waveform model. For both analyses with and without calibration uncertainty, the $\mathcal{F}$-statistic and FFD methods yield highly consistent posterior structures, with nearly identical support regions in the (d_L, ψ) plane. No additional multimodality is observed compared to the BBH-based analyses discussed previously.

The quantitative agreement between the two inference approaches is further confirmed in Table III. For most intrinsic and extrinsic parameters, including the tidal deformability λ_2, the JSD values remain small, indicating a high degree of consistency. However, we note a notable exception for the inclination angle θ_{JN}, where the JSD is 0.105 without calibration uncertainty. This discrepancy likely reflects the limitations of the `IMRPhenomNSBH` model, which does not account for the spin-precession and higher-order mode effects that are significant for GW190814, in contrast to the more consistent results obtained with the `IMRPhenomXPHM-SpinTaylor` waveform. This highlights the critical importance of using waveform models that fully capture the signal's physical content.

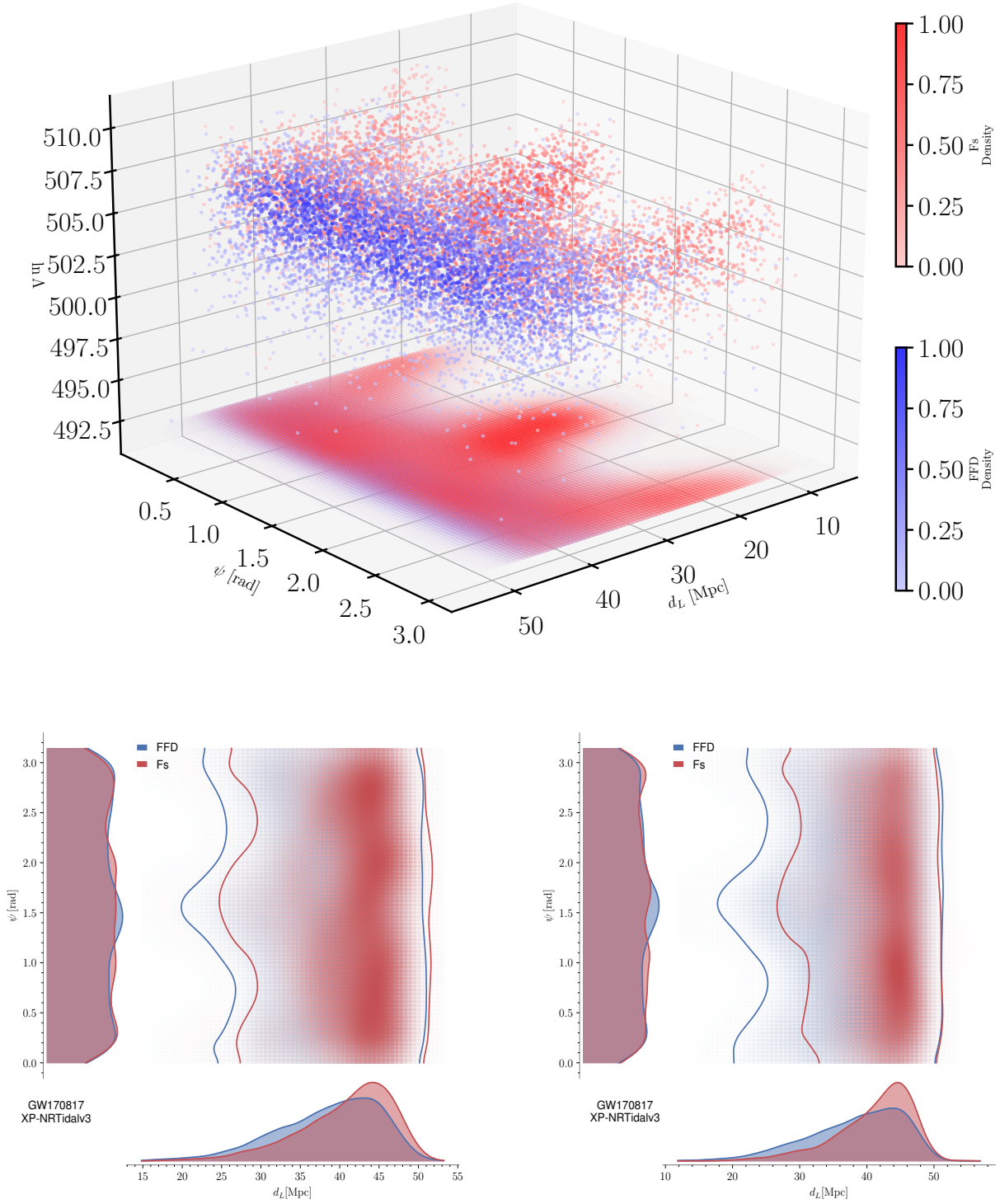


FIG. 6: Same as Fig. 2, but for the analysis of GW170817 under the binary neutron star hypothesis using the `IMRPhenomXP-NRTidalv3` waveform model.

TABLE IV: Parameters of GW170817 estimated using `IMRPhenomXP-NRTidalv3` waveform model by $\mathcal{F}$-statistic (Fs) and FFD methods. The table compares results without calibration error (noc) and with calibration error (c), including tidal deformability parameters λ_1 and λ_2. JSD values, sample size (N_s), sampling time (T_s), and log Bayes factor ($\ln \mathcal{Z}$) are included.

	No Calibration Error (noc)			Calibration Error (c)		
Parameter	Fs-noc	FFD-noc	JSD	Fs-c	FFD-c	JSD
$t_c - t_0$[ms]	$-1.97^{+1.60}_{-3.15}$	$-2.27^{+1.62}_{-3.25}$	0.015	$-2.12^{+1.88}_{-3.21}$	$-2.33^{+1.75}_{-3.34}$	0.012
$\mathcal{M} - \mathcal{M}_0[10^{-4} M_\odot]$	$5.9^{+1.5}_{-1.3}$	$6.0^{+1.5}_{-1.3}$	0.002	$6.1^{+1.6}_{-1.4}$	$6.0^{+1.5}_{-1.4}$	0.003
q	$0.87^{+0.12}_{-0.16}$	$0.87^{+0.11}_{-0.14}$	0.008	$0.87^{+0.11}_{-0.16}$	$0.87^{+0.11}_{-0.14}$	0.004
a_1	$0.02^{+0.02}_{-0.02}$	$0.02^{+0.02}_{-0.02}$	0.006	$0.02^{+0.02}_{-0.02}$	$0.02^{+0.02}_{-0.02}$	0.004
a_2	$0.02^{+0.02}_{-0.02}$	$0.02^{+0.02}_{-0.02}$	0.008	$0.02^{+0.02}_{-0.02}$	$0.02^{+0.02}_{-0.02}$	0.004
θ_1[rad]	$1.26^{+1.10}_{-0.88}$	$1.30^{+0.99}_{-0.80}$	0.010	$1.26^{+1.13}_{-0.89}$	$1.29^{+1.07}_{-0.87}$	0.004
θ_2[rad]	$1.37^{+1.12}_{-0.95}$	$1.33^{+1.01}_{-0.85}$	0.007	$1.32^{+1.13}_{-0.93}$	$1.33^{+1.06}_{-0.90}$	0.003
ϕ_{12}[rad]	$3.08^{+2.88}_{-2.77}$	$3.14^{+2.83}_{-2.84}$	0.002	$3.15^{+2.86}_{-2.85}$	$3.16^{+2.81}_{-2.87}$	0.003
ϕ_{JL}[rad]	$3.18^{+2.80}_{-2.88}$	$3.22^{+2.76}_{-2.90}$	0.003	$3.16^{+2.83}_{-2.83}$	$3.06^{+2.93}_{-2.76}$	0.003
θ_{JN}[rad]	$2.30^{+0.58}_{-0.37}$	$2.54^{+0.41}_{-0.43}$	0.103	$2.34^{+0.56}_{-0.41}$	$2.54^{+0.42}_{-0.48}$	0.060
ϕ[rad]	$3.13^{+2.72}_{-2.82}$	$3.15^{+2.78}_{-2.83}$	0.009	$3.18^{+2.71}_{-2.85}$	$3.19^{+2.73}_{-2.88}$	0.008
α[rad]	$3.42^{+0.04}_{-0.04}$	$3.43^{+0.04}_{-0.04}$	0.008	$3.43^{+0.04}_{-0.04}$	$3.43^{+0.04}_{-0.04}$	0.007
δ[rad]	$-0.37^{+0.06}_{-0.06}$	$-0.37^{+0.06}_{-0.07}$	0.004	$-0.37^{+0.07}_{-0.06}$	$-0.37^{+0.07}_{-0.06}$	0.004
d_L[Mpc]	$42.0^{+6.3}_{-12.8}$	$39.6^{+7.8}_{-13.1}$	0.025	$42.5^{+5.6}_{-12.8}$	$39.5^{+8.1}_{-14.4}$	0.038
ψ[rad]	$1.54^{+1.47}_{-1.40}$	$1.54^{+1.48}_{-1.42}$	0.004	$1.55^{+1.45}_{-1.44}$	$1.56^{+1.45}_{-1.42}$	0.005
λ_1	$319.0^{+831.0}_{-290.4}$	$375.3^{+757.0}_{-326.6}$	0.011	$351.5^{+880.8}_{-319.8}$	$388.4^{+783.9}_{-341.8}$	0.005
λ_2	$525.9^{+1108.3}_{-474.6}$	$544.8^{+919.5}_{-470.4}$	0.008	$589.5^{+1227.6}_{-525.7}$	$566.5^{+942.2}_{-495.1}$	0.010
N_s	5267	5785	-	5344	5863	-
$T_s[h]$	17.0	35.6	-	16.6	84.2	-
$\ln \mathcal{Z}$	474.7	474.1	-	473.6	473.3	-

1. The geocentric time t_c is reported as an offset from the GW170817 trigger time, $t_0 = 1187008882.43$ s.
2. The redshifted chirp mass $\mathcal{M}$ is reported as an offset from the reference mass, $\mathcal{M}_0 = 1.197\,\mathrm{M}_\odot$.

Nevertheless, including calibration uncertainty mitigates this disagreement, reducing the JSD for θ_{JN} to 0.075, underscoring the importance of accounting for instrumental effects.

In addition to statistical consistency, the $\mathcal{F}$-statistic method demonstrates a significant advantage in computational efficiency. As shown in Table III, without calibration uncertainty, the $\mathcal{F}$-statistic analysis required 9.7 hours compared to 12.8 hours for the FFD approach, a time saving of approximately 24%. This efficiency gain becomes substantially more pronounced when calibration is included: the $\mathcal{F}$-statistic analysis completed in 17.9 hours, while the FFD method required 40.8 hours, representing a speed-up factor of more than two (a time saving of $\sim 56\%$). As with the other analyses, the $\mathcal{F}$-statistic method yields a slightly higher Bayesian evidence, further supporting its robustness. This highlights the practical advantage of the $\mathcal{F}$-statistic method for computationally intensive IMR analyses involving complex waveform models and extended parameter spaces.

Overall, these results demonstrate that the $\mathcal{F}$-statistic method provides a computationally efficient yet statistically consistent alternative to FFD inference for NSBH analyses of GW190814, even in the presence of calibration uncertainty and tidal parameters.

C. Analysis of GW170817: A Binary Neutron Star Merger

Following the BBH and NSBH analyses presented above, we next consider a binary neutron star system to examine the broader applicability of the $\mathcal{F}$-statistic–based inference framework. We choose GW170817, the first observed BNS merger, as a representative and well-controlled test case. Owing to its high signal-to-noise ratio, long inspiral duration in the sensitive band, and well-established astrophysical interpretation, GW170817 provides an ideal benchmark for validating inference strategies beyond black-hole–dominated systems.

This analysis is designed to demonstrate that the $\mathcal{F}$-statistic approach is applicable across the full range of compact binary coalescences considered in this work, complementing the BBH and NSBH studies presented above. We perform parameter estimation using the `IMRPhenomXP-NRTidalv3` waveform model and adopt the same low-spin prior as used in the LVK analyses of

GW170817. As in the previous sections, we compare the $\mathcal{F}$-statistic–based method with the standard Bayesian inference both with and without calibration uncertainty, allowing for a consistent assessment of the impact of calibration errors in the BNS regime.

Fig. 6 and Table IV summarize the parameter-estimation results for GW170817 obtained with the $\mathcal{F}$-statistic–based and FFD likelihood methods, both without and with calibration uncertainty. Across the full set of intrinsic and extrinsic parameters, the two inference approaches yield statistically consistent posterior distributions. This consistency is quantitatively supported by the JSD values in Table IV. For instance, while there is a modest tension in the inclination angle θ_{JN} without calibration (JSD $= 0.103$), including calibration uncertainty improves the agreement significantly, reducing the JSD to 0.060. Similar improvements are seen for other key parameters, such as the mass ratio q (JSD drops from 0.008 to 0.004) and the primary tidal deformability λ_1 (JSD drops from 0.011 to 0.005), highlighting the importance of marginalizing over calibration errors for robust inference.

A significant practical difference between the two approaches emerges in their computational efficiency. As shown by the sampling times in Table IV, without calibration error, the $\mathcal{F}$-statistic-based analysis required 17.0 hours, while the FFD method took 35.6 hours, corresponding to a speed-up factor of over two. This computational advantage becomes dramatically more pronounced when calibration uncertainty is included. In this more complex scenario, the FFD analysis time more than doubled to 84.2 hours, whereas the $\mathcal{F}$-statistic analysis completed in just 16.6 hours. This represents a speed-up factor of more than five, demonstrating that the efficiency gains of the $\mathcal{F}$-statistic method are even more critical for realistic analyses of long-duration signals that include instrumental uncertainties and extended physical models. As seen in Table IV, the $\mathcal{F}$-statistic-based analysis consistently yields a higher Bayesian evidence than the FFD method.

Taken together, the GW170817 results demonstrate that the $\mathcal{F}$-statistic–based framework provides parameter estimates consistent with standard Bayesian inference, while offering a substantial reduction in computational cost. This confirms the robustness and efficiency of the method for binary neutron star systems, complementing the BBH and NSBH analyses presented above and establishing its applicability across a broad range of compact binary coalescences.

D. Comparison with Luminosity-Distance Marginalization in Bilby

In addition to the comparison between the FFD and $\mathcal{F}$-statistic methods presented above, it is also useful to benchmark the latter against an existing acceleration strategy that is already widely used in practical GW inference workflows, namely luminosity-distance marginalization in Bilby. In the standard implementation, this procedure marginalizes numerically over the luminosity distance through a Riemann sum evaluated on a precomputed grid in matched-filter SNR and optimal SNR, with a lookup table used to accelerate interpolation [8]. Since this treatment does not modify the likelihood function itself or the functional dependence among the remaining parameters, it is generally expected to preserve the parameter-estimation results of the full Bayesian analysis while often reducing the computational cost. At the same time, previous work has shown that the practical gain is not universal and may depend on the sampler configuration [49].

We further performed an additional set of analyses using Bilby with luminosity-distance marginalization enabled. To ensure a controlled comparison, we considered the same events, waveform models, and calibration-uncertainty treatment adopted in the corresponding analyses discussed above. Because the consistency of posterior estimation under luminosity-distance marginalization has already been examined in earlier studies and is not the main focus of the present paper, we restrict the discussion here to practical efficiency and evidence behavior.

The results are summarized in Table V. When compared with the corresponding full Bayesian analyses reported earlier, the behavior of luminosity-distance marginalization is found to be source dependent. For GW190814 analyzed with `IMRPhenomNSBH` and for GW170817 analyzed with `IMRPhenomXP-NRTidalv3`, the outcome is most consistent with the usual expectation: the differences in log Bayesian evidence are negligible, and the total sampling time is reduced relative to the corresponding FFD runs. These cases therefore confirm that luminosity-distance marginalization can provide a practical speed-up while preserving the overall statistical behavior of the inference.

By contrast, the cases based on `IMRPhenomXPHM-SpinTaylor` show a less uniform pattern. For both GW190412 and GW190814 under the BBH hypothesis, the differences in log evidence with respect to the corresponding FFD analyses are modest but visible, at the level of about 0.4. A plausible explanation is that these signals involve more complicated likelihood envelopes because of the combined effects of spin precession and higher-order modes, which may limit the extent to which standard distance marginalization alone improves the sampling performance.

This trend is particularly evident for GW190814 analyzed with `IMRPhenomXPHM-SpinTaylor`. In that case, enabling luminosity-distance marginalization does not reduce the total sampling time. We also note that this is the only case in which the number of posterior samples increases relative to the corresponding FFD analysis. Although this may suggest some connection between wall-clock time and sampling behavior in this particular case, we do not attempt to draw a strong conclusion, since the

TABLE V: Summary of the additional analyses performed with luminosity-distance marginalization enabled in Bilby, including the number of posterior samples N_s, the total sampling time T_s, and the log Bayesian evidence $\ln \mathcal{Z}$.

Event and source/model type	N_s	T_s [h]	$\ln \mathcal{Z}$
GW190412 (BBH, `IMRPhenomXPHM-SpinTaylor`)	6993	17.9	146.2
GW190814 (BBH, `IMRPhenomXPHM-SpinTaylor`)	6840	99.1	258.7
GW190814 (NSBH, `IMRPhenomNSBH`)	5799	37.1	236.9
GW170817 (BNS, `IMRPhenomXP-NRTidalv3`)	5624	60.2	473.3

final computational cost can also depend on waveform-evaluation expense and on the detailed behavior of the sampler. A more systematic investigation of these effects would be interesting, but it lies beyond the scope of the present work.

For comparison, the $\mathcal{F}$-statistic analysis for the same GW190814+`IMRPhenomXPHM-SpinTaylor` case also yields a larger number of posterior samples than the corresponding FFD analysis, yet it still remains substantially more efficient in total sampling time. This difference may reflect the fact that the $\mathcal{F}$-statistic method analytically maximizes over both the luminosity distance and the polarization angle in likelihood space, whereas the standard luminosity-distance-marginalization strategy removes only one of these two linear extrinsic parameters from the numerical sampling. In this sense, the present comparison further supports the practical advantage of the $\mathcal{F}$-statistic framework, especially for inference problems in which the likelihood structure is complicated and the cost of sampling extrinsic parameters remains non-negligible.

IV. SUMMARY AND CONCLUSION

In this study, we conducted a comprehensive investigation of the $\mathcal{F}$-statistic method for parameter estimation of complete IMR signals from a range of GW sources. Our findings demonstrate that the $\mathcal{F}$-statistic offers significant advantages in performance, stability, and computational efficiency over the traditional FFD approach. We presented a self-contained framework for the $\mathcal{F}$-statistic analysis that includes a method to reconstruct the posterior distributions for analytically maximized parameters under a physical prior, and a novel formulation to calculate the Bayesian evidence for robust model comparison.

Our comparative analysis across several landmark events—GW190412, GW190814, and GW170817—highlights several key findings. First, the $\mathcal{F}$-statistic method consistently produces posterior distributions that are in good agreement with those from standard FFD inference. Including calibration uncertainty tends to improve this agreement, suggesting that marginalization over instrumental effects helps harmonize the two approaches. Second, the $\mathcal{F}$-statistic method is computationally superior in the relative sense that, for each event and waveform configuration considered here, it requires less sampling time than the corresponding FFD analysis. This relative time saving is generally larger when calibration uncertainties are included, although the absolute wall-clock time of an individual run need not vary monotonically because it also depends on stochastic sampler behavior, convergence criteria, waveform-evaluation cost, and the detailed posterior geometry. Third, under the assumption of physical priors, the $\mathcal{F}$-statistic-based analyses consistently yield higher Bayesian evidence than the corresponding FFD analyses. While this systematic difference is intrinsically linked to the analytical maximization over two parameters, which allows for a more efficient exploration of high-likelihood regions, we caution that such evidence comparisons are most safely interpreted within the same inference framework. While the $\mathcal{F}$-statistic method produced slightly broader posteriors for some parameters, we interpret this as a more honest and conservative quantification of statistical uncertainty, especially in complex, high-dimensional problems.

Overall, our study demonstrates the efficiency and reliability of the $\mathcal{F}$-statistic method. As detector sensitivities improve and GW detection rates increase, the need for efficient and robust analysis methods will become paramount. The advantages demonstrated here establish the $\mathcal{F}$-statistic as a promising tool for the rapid and reliable analysis of large event catalogs [50]. Future work will focus on applying this framework to a wider range of GW events, including those with lower SNRs and more complex physical features, and on testing its performance.

ACKNOWLEDGMENTS

We thank the anonymous referee for their insightful comments. H.-T. Wang is supported by "the Natural Science Foundation of Liaoning Province" (Grant No. ZX20250217) and "the Fundamental Research Funds for the Central Universities" at Dalian University of Technology. This research has made use of data or software obtained from the Gravitational Wave Open Science Center [51], a service of LIGO Laboratory, the LIGO Scientific Collaboration, the Virgo Collaboration, and KAGRA [52]. LIGO Laboratory and Advanced

LIGO are funded by the United States National Science Foundation (NSF) as well as the Science and Technology Facilities Council (STFC) of the United Kingdom, the Max-Planck-Society (MPS), and the State of Niedersachsen/Germany for support of the construction of Advanced LIGO and construction and operation of the GEO600 detector. Additional support for Advanced LIGO was provided by the Australian Research Council. Virgo is funded, through the European Gravitational Observatory (EGO), by the French Centre National de Recherche Scientifique (CNRS), the Italian Istituto Nazionale di Fisica Nucleare (INFN) and the Dutch Nikhef, with contributions by institutions from Belgium, Germany, Greece, Hungary, Ireland, Japan, Monaco, Poland, Portugal, Spain. KAGRA is supported by Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan Society for the Promotion of Science (JSPS) in Japan; National Research Foundation (NRF) and Ministry of Science and ICT (MSIT) in Korea; Academia Sinica (AS) and National Science and Technology Council (NSTC) in Taiwan of China.

DATA AVAILABILITY

The posterior samples and analysis results supporting the findings of this study are publicly available [46].

Appendix A: Benchmark analysis: GW150914

As a benchmark for our methodology, we include a reanalysis of the public data for the GW150914 event. This event serves as a well-understood, high-SNR test case for validating inference methods. For this analysis, we used the `IMRPhenomXPHM` waveform model. The results presented here were obtained without including calibration uncertainty, allowing for a focused comparison of sampler performance and method stability, which complements the main text's discussion on calibration effects.

1. Sampler Configuration

The Bayesian inference was performed using the Bilby package [v2.4.0; 9], with the nested sampling algorithm implemented in Dynesty [v2.1.5; 8]. To assess the stability and robustness of both the FFD and $\mathcal{F}$-statistic methods, we employed two distinct sampler configurations, hereafter denoted 's1' and 's2'. Several key parameters were common to both configurations: the number of live points was set to 1000, and the primary sampling method was "rwalk". The configurations differed in the following parameters:

- **Configuration 's1':** The number of autocorrelation lengths was set to 50, and the method for selecting new points was "live". This configuration is similar to the default settings used in the GWTC-2.1 and GWTC-3 catalog analyses.
- **Configuration 's2':** The number of autocorrelation lengths was reduced to 20, proposals were drawn using a combination of "diff" and "volumetric" methods, and new points were selected using "live-multi". This alternative configuration was chosen for comparison, following standard practices outlined in the Bilby documentation.

All analyses were parallelized using 50 threads to reduce computational time.

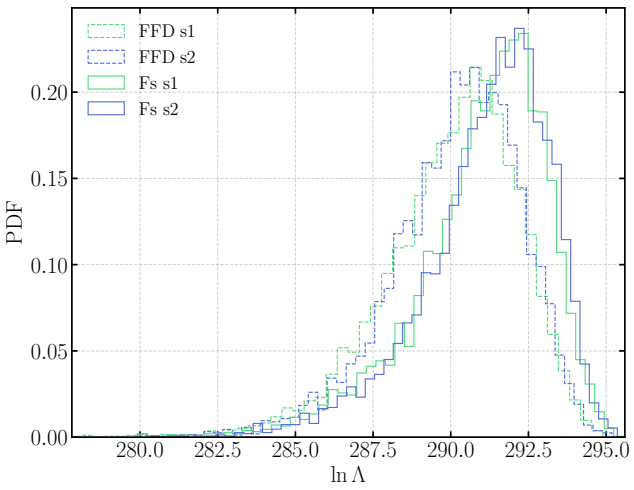


FIG. 7: Posterior distributions of the log-likelihood ratio, $\ln \Lambda$, comparing the performance of the FFD method (dashed histograms) and the $\mathcal{F}$-statistic method (solid histograms, denoted as 'Fs'). The two colors represent different sampler configurations: 's1' (green) and 's2' (blue).

2. Performance and Stability Analysis

In this section, we present a detailed comparison of the parameter estimation results obtained from the $\mathcal{F}$-statistic and the FFD methods for the GW150914 event. We evaluate the overall performance, efficiency, and stability of the two methods. We first compare their calculated Bayes factors and log-likelihood ratio distributions. Then, we examine posterior distributions for key parameter subsets: strongly degenerate extrinsic parameters (t_c, α), the analytically maximized parameters (d_L, ψ), and finally all remaining source parameters. We assess stability across different sampler configurations using the JSD.

3. Model Evidence and Performance Comparison

We begin the performance comparison by examining the posterior distributions of the log-likelihood ratio,

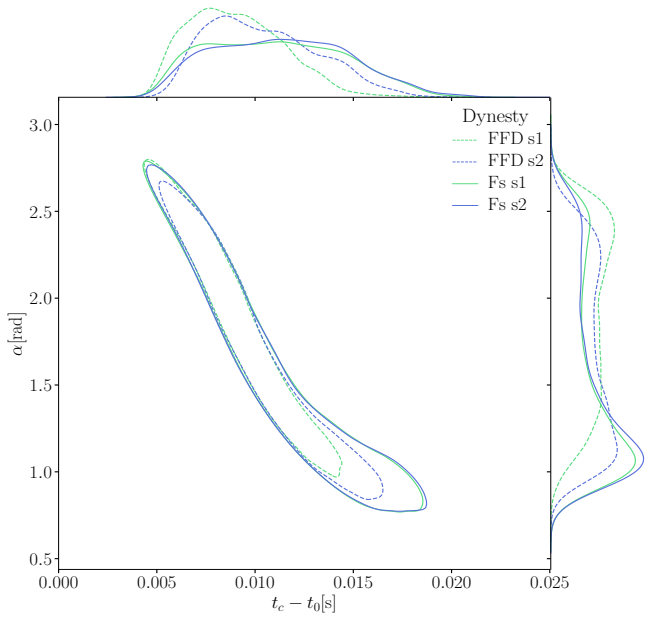


FIG. 8: Comparison of the joint posterior distribution for the geocenter time, t_c, and right ascension, α. The results from the FFD method (dashed lines) and the $\mathcal{F}$-statistic method (solid lines, denoted as 'Fs') are shown for two sampler configurations, 's1' (green) and 's2' (blue). The two-dimensional contours enclose the 90% credible regions. For clarity, the geocenter time is shown as an offset from the event trigger time, $t_0 = 1126259462.4$ s. The top and right panels display the corresponding one-dimensional marginalized posterior distributions for each parameter.

$\ln \Lambda$, shown in Fig. 7. For both the FFD and $\mathcal{F}$-statistic methods, the distributions are consistent across the two sampler settings ('s1' and 's2'). However, the distributions for the $\mathcal{F}$-statistic method are consistently shifted toward higher values, with median $\ln \Lambda$ values of 291.4 ('s1') and 291.5 ('s2'), compared to 290.3 ('s1') and 290.4 ('s2') for the FFD runs. This suggests a better fit to the data, enabled by maximizing two linear parameters.

For a more definitive and quantitative comparison, we compute the evidence, $\mathcal{Z}$, using the procedure outlined in Sec. II. Then, the log-Bayes factor between two configurations of each method is defined as $\ln \mathcal{B}^{\rm s1}_{\rm s2} \equiv \ln \mathcal{Z}_{\rm s1} - \ln \mathcal{Z}_{\rm s2}$. The difference in log-evidence between the two sampler settings for the $\mathcal{F}$-statistic is only 0.1, whereas for the FFD method, the difference is 0.4. This suggests that the $\mathcal{F}$-statistic method is more robust to changes in the sampler configuration for this event.

4. Constraints on Degenerate Extrinsic Parameters

The geocenter time (t_c) and right ascension (α) are crucial for localizing a gravitational wave source, but they often exhibit strong degeneracies that can challenge sampling algorithms. Fig. 8 compares the posterior distributions for these parameters obtained from both the FFD and $\mathcal{F}$-statistic methods.

We quantify the consistency between the 's1' and 's2' settings using the JSD. A key finding is the superior stability of the $\mathcal{F}$-statistic method against changes in sampler configuration. We quantify this using the JSD between the posteriors from the 's1' and 's2' settings. For the $\mathcal{F}$-statistic method, the JSD values are remarkably low: 0.004 for t_c and 0.005 for α. In contrast, the FFD method yields significantly higher JSD values of 0.067 for t_c and 0.058 for α. This represents an order-of-magnitude improvement in consistency for the $\mathcal{F}$-statistic method.

The median values and 90% credible intervals for each analysis further illustrate this point (where $t_0 = 1126259462.4$ s is the event trigger time):

- **$\mathcal{F}$-statistic ('s1'):** $t_c = t_0 + 0.011^{+0.006}_{-0.005}$ s; $\alpha = 1.46^{+1.09}_{-0.52}$ rad.
- **$\mathcal{F}$-statistic ('s2'):** $t_c = t_0 + 0.011^{+0.006}_{-0.005}$ s; $\alpha = 1.40^{+1.11}_{-0.46}$ rad.
- **FFD ('s1'):** $t_c = t_0 + 0.009^{+0.004}_{-0.003}$ s; $\alpha = 1.91^{+0.66}_{-0.77}$ rad.
- **FFD ('s2'):** $t_c = t_0 + 0.010^{+0.005}_{-0.004}$ s; $\alpha = 1.63^{+0.81}_{-0.63}$ rad.

The consistency of the credible intervals from the $\mathcal{F}$-statistic method, supported by the low JSD values, demonstrates its enhanced robustness when characterizing these degenerate parameters.

5. Reconstruction of Analytically Maximized Parameters

The luminosity distance, d_L, and polarization angle, ψ, are treated fundamentally differently in the two frameworks. In the FFD approach, they are sampled directly along with all other parameters. In the $\mathcal{F}$-statistic method, they are analytically maximized, and their posteriors must be reconstructed in a post-processing step, as detailed in Sec. II C. Fig. 9 shows the resulting posterior distributions from both methods using the physical target prior.

Despite the additional reconstruction step, the $\mathcal{F}$-statistic method demonstrates superior stability for the luminosity distance. For d_L, the JSD is only 0.003 for the $\mathcal{F}$-statistic method, an order of magnitude smaller than the JSD of 0.043 for the FFD method. For the polarization angle ψ, the JSD values are low and comparable (0.009 for $\mathcal{F}$-statistic vs. 0.003 for FFD). This is expected, as ψ is poorly constrained by the GW150914 event, leading to broad, featureless posteriors for both methods.

The median values and 90% credible intervals from the physical target prior are:

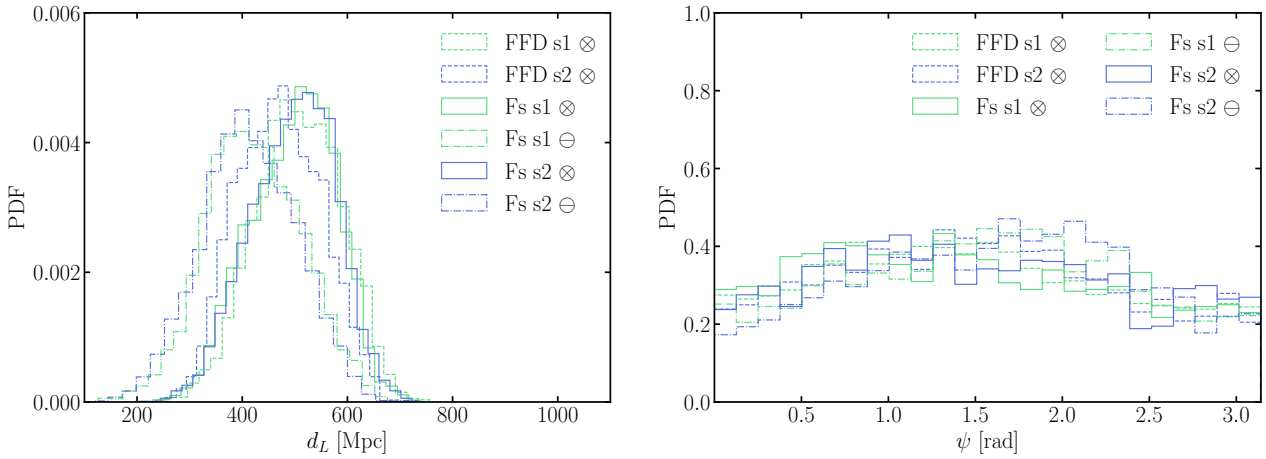


FIG. 9: Comparison of the marginalized posterior distributions for the luminosity distance, d_L (left panel), and the polarization angle, ψ (right panel). The posteriors from the FFD method are compared against the reconstructed posteriors from the $\mathcal{F}$-statistic (Fs) method for two sampler configurations ('s1' in green, 's2' in blue). For the $\mathcal{F}$-statistic results, posteriors are shown for two different prior assumptions: the initial reconstruction assuming a uniform prior on the linear parameters (marked ⊖), and the final result after re-weighting to the physical target prior (marked ⊗).

- **$\mathcal{F}$-statistic ('s1'):** $d_L = 504.8^{+115.0}_{-144.2}$ Mpc; $\psi = 1.41^{+1.52}_{-1.23}$ rad.
- **$\mathcal{F}$-statistic ('s2'):** $d_L = 502.0^{+117.0}_{-142.1}$ Mpc; $\psi = 1.48^{+1.47}_{-1.29}$ rad.
- **FFD ('s1'):** $d_L = 507.5^{+128.6}_{-134.1}$ Mpc; $\psi = 1.45^{+1.50}_{-1.27}$ rad.
- **FFD ('s2'):** $d_L = 469.7^{+130.3}_{-124.8}$ Mpc; $\psi = 1.50^{+1.43}_{-1.29}$ rad.

The greater consistency in the recovered constraints for d_L indicates that the $\mathcal{F}$-statistic approach performs more reliably, even for parameters that are not sampled directly.

6. Comparison of All Sampled Source Parameters

The posterior distributions for all remaining source parameters are shown in Fig. 10, with the quantitative constraints summarized in Table VI. A primary finding from this comprehensive comparison is the enhanced stability of the $\mathcal{F}$-statistic method. As shown in Table VI, the JSD values calculated between the 's1' and 's2' settings are systematically lower for the $\mathcal{F}$-statistic method across nearly all parameters, often by an order of magnitude. This demonstrates that the $\mathcal{F}$-statistic is significantly more robust to the specific configuration of the nested sampler.

We note that the constraints on some parameters, such as the chirp mass $\mathcal{M}$, are slightly broader for the $\mathcal{F}$-statistic method compared to the FFD method. This behavior is consistent with findings in other high-dimensional analyses, such as studies of black hole ringdown with multiple tones [27]. In such complex, often multimodal parameter spaces, standard nested sampling can sometimes struggle to explore the full posterior, leading to artificially narrow or biased constraints [27, 53]. The reduced dimensionality of the $\mathcal{F}$-statistic search space appears to mitigate these issues. Therefore, the slightly wider posteriors from the $\mathcal{F}$-statistic method represent a more conservative and honest quantification of the true statistical uncertainty.

Finally, the $\mathcal{F}$-statistic method provides a substantial improvement in computational efficiency. As detailed in the final rows of Table VI, for the 's1' ('s2') configuration, the FFD analysis required 63.7 (16.1) hours, while the $\mathcal{F}$-statistic analysis took only 18.5 (5.8) hours. This corresponds to a speed-up factor of approximately 3.4 (2.8), representing a reduction of about 70.6% in total sampling time while producing a similar number of posterior samples.

[1] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. X **9**, 031040 (2019).
[2] R. Abbott *et al.*, Phys. Rev. X **11**, 021053 (2021), arXiv:2010.14527 [gr-qc].
[3] R. Abbott *et al.* (KAGRA, VIRGO, LIGO Scientific), Phys. Rev. X **13**, 041039 (2023), arXiv:2111.03606 [gr-qc].
[4] B. P. Abbott *et al.* (LIGO Scientific, Virgo), Phys. Rev.

TABLE VI: Summary of parameter constraints for GW150914. The table compares the results from the FFD and $\mathcal{F}$-statistic (Fs) methods for the two sampler settings ('s1' and 's2'). For each parameter, we report the median and the 90% credible interval. The JSD quantifies the consistency between the posteriors obtained from the 's1' and 's2' configurations for each method. The final three rows report summary statistics for each run: the number of posterior samples (N_s), the total sampling time in hours (T_s), and the log-evidence ($\ln \mathcal{Z}$).

	FFD s1	FFD s2	JSD	Fs s1	Fs s2	JSD
$t_c - t_0$[s]	$0.009^{+0.004}_{-0.003}$	$0.010^{+0.005}_{-0.004}$	0.067	$0.011^{+0.006}_{-0.005}$	$0.011^{+0.006}_{-0.005}$	0.004
$\mathcal{M}[M_\odot]$	$30.59^{+1.51}_{-1.80}$	$30.66^{+1.63}_{-1.71}$	0.006	$30.39^{+1.67}_{-1.99}$	$30.50^{+1.56}_{-1.98}$	0.006
q	$0.83^{+0.15}_{-0.21}$	$0.88^{+0.11}_{-0.22}$	0.026	$0.84^{+0.15}_{-0.25}$	$0.85^{+0.14}_{-0.24}$	0.005
a_1	$0.37^{+0.53}_{-0.33}$	$0.42^{+0.46}_{-0.36}$	0.007	$0.40^{+0.52}_{-0.36}$	$0.43^{+0.49}_{-0.38}$	0.004
a_2	$0.41^{+0.51}_{-0.37}$	$0.45^{+0.47}_{-0.40}$	0.005	$0.43^{+0.50}_{-0.39}$	$0.44^{+0.49}_{-0.39}$	0.004
θ_1[rad]	$1.61^{+0.91}_{-1.11}$	$1.70^{+0.99}_{-0.99}$	0.017	$1.65^{+0.88}_{-1.12}$	$1.65^{+0.85}_{-1.02}$	0.003
θ_2[rad]	$1.76^{+0.98}_{-1.16}$	$1.71^{+0.99}_{-1.01}$	0.007	$1.79^{+0.96}_{-1.19}$	$1.73^{+1.00}_{-1.18}$	0.004
ϕ_{12}[rad]	$2.92^{+3.06}_{-2.64}$	$3.11^{+2.86}_{-2.80}$	0.004	$2.89^{+3.10}_{-2.59}$	$3.02^{+2.95}_{-2.71}$	0.004
ϕ_{JL}[rad]	$2.47^{+3.68}_{-2.35}$	$2.23^{+3.94}_{-2.09}$	0.003	$1.85^{+4.31}_{-1.71}$	$1.67^{+4.48}_{-1.54}$	0.004
θ_{JN}[rad]	$2.77^{+0.27}_{-0.43}$	$2.72^{+0.29}_{-0.40}$	0.013	$2.66^{+0.35}_{-0.62}$	$2.65^{+0.36}_{-0.57}$	0.006
ϕ[rad]	$3.40^{+2.38}_{-2.95}$	$3.68^{+2.04}_{-3.14}$	0.005	$3.47^{+2.22}_{-2.93}$	$3.37^{+2.31}_{-2.85}$	0.003
α[rad]	$1.91^{+0.66}_{-0.77}$	$1.63^{+0.81}_{-0.63}$	0.058	$1.46^{+1.07}_{-0.52}$	$1.40^{+1.11}_{-0.46}$	0.005
δ[rad]	$-1.24^{+0.19}_{-0.05}$	$-1.23^{+0.16}_{-0.05}$	0.008	$-1.20^{+0.21}_{-0.08}$	$-1.20^{+0.22}_{-0.08}$	0.004
d_L[Mpc]	$507.5^{+128.6}_{-134.1}$	$469.7^{+130.3}_{-124.8}$	0.043	$504.8^{+115.0}_{-144.2}$	$502.0^{+117.0}_{-142.1}$	0.003
ψ[rad]	$1.45^{+1.50}_{-1.27}$	$1.50^{+1.43}_{-1.29}$	0.003	$1.41^{+1.52}_{-1.23}$	$1.48^{+1.47}_{-1.29}$	0.009
N_s	5476	5162	-	5093	5027	-
$T_s[h]$	63.7	16.1	-	18.5	5.8	-
$\ln \mathcal{Z}$	261.2	261.6	-	261.4	261.3	-

1. Parameter definitions follow the conventions used in the PESummary package [48].
2. The geocenter time t_c is reported as an offset from the GW150914 trigger time, $t_0 = 1126259462.4$ s.

Lett. **116**, 221101 (2016), [Erratum: Phys.Rev.Lett. 121, 129902 (2018)], arXiv:1602.03841 [gr-qc].

[5] R. Abbott *et al.* (LIGO Scientific, Virgo), Phys. Rev. D **103**, 122002 (2021), arXiv:2010.14529 [gr-qc].

[6] R. Abbott *et al.* (LIGO Scientific, VIRGO, KAGRA), (2021), arXiv:2112.06861 [gr-qc].

[7] J. Veitch *et al.*, Phys. Rev. D **91**, 042003 (2015), arXiv:1409.7215 [gr-qc].

[8] I. M. Romero-Shaw *et al.*, Mon. Not. Roy. Astron. Soc. **499**, 3295 (2020), arXiv:2006.00714 [astro-ph.IM].

[9] G. Ashton *et al.*, Astrophys. J. Suppl. **241**, 27 (2019), arXiv:1811.02042 [astro-ph.IM].

[10] C. M. Biwer, C. D. Capano, S. De, M. Cabero, D. A. Brown, A. H. Nitz, and V. Raymond, Publ. Astron. Soc. Pac. **131**, 024503 (2019), arXiv:1807.10312 [astro-ph.IM].

[11] W. Farr, B. Farr, and T. Littenberg, *Modelling Calibration Errors In CBC Waveforms*, LIGO Technical Note LIGO-T1400682 (LIGO Scientific Collaboration, 2014).

[12] S. Vitale, W. Del Pozzo, T. G. F. Li, C. Van Den Broeck, I. Mandel, B. Aylott, and J. Veitch, Phys. Rev. D **85**, 064034 (2012), arXiv:1111.3044 [gr-qc].

[13] E. Payne, C. Talbot, P. D. Lasky, E. Thrane, and J. S. Kissel, Phys. Rev. D **102**, 122004 (2020), arXiv:2009.10193 [astro-ph.IM].

[14] L. Sun *et al.*, Class. Quant. Grav. **37**, 225008 (2020), arXiv:2005.02531 [astro-ph.IM].

[15] E. Thrane and C. Talbot, Publ. Astron. Soc. Austral. **36**, e010 (2019), [Erratum: Publ.Astron.Soc.Austral. 37, e036 (2020)], arXiv:1809.02293 [astro-ph.IM].

[16] C. Hoy, Phys. Rev. D **106**, 083003 (2022), arXiv:2208.00106 [gr-qc].

[17] J. Roulet, J. Mushkin, D. Wadekar, T. Venumadhav, B. Zackay, and M. Zaldarriaga, Phys. Rev. D **110**, 044010 (2024), arXiv:2404.02435 [gr-qc].

[18] H. Antil, S. E. Field, F. Herrmann, R. H. Nochetto, and M. Tiglio, J. Sci. Comput. **57**, 604 (2013), arXiv:1210.0577 [cs.NA].

[19] R. Smith, S. E. Field, K. Blackburn, C.-J. Haster, M. Pürrer, V. Raymond, and P. Schmidt, Phys. Rev. D **94**, 044031 (2016), arXiv:1604.08253 [gr-qc].

[20] R. Smith *et al.*, Phys. Rev. Lett. **127**, 081102 (2021), arXiv:2103.12274 [gr-qc].

[21] K. Cranmer, J. Brehmer, and G. Louppe, Proc. Nat. Acad. Sci. **117**, 30055 (2020), arXiv:1911.01429 [stat.ML].

[22] M. Dax, S. R. Green, J. Gair, M. Pürrer, J. Wildberger, J. H. Macke, A. Buonanno, and B. Schölkopf, Phys. Rev. Lett. **130**, 171403 (2023), arXiv:2210.05686 [gr-qc].

[23] K. W. K. Wong, M. Isi, and T. D. P. Edwards, Astrophys. J. **958**, 129 (2023), arXiv:2302.05333 [astro-ph.IM].

[24] E. Marx *et al.*, Phys. Rev. D **111**, 042010 (2025), arXiv:2403.18661 [gr-qc].

[25] K. D. Kokkotas, G. Tsegas, and A. Krolak, Class. Quant. Grav. **11**, 1901 (1994).

[26] P. Jaranowski, A. Krolak, and B. F. Schutz, Phys. Rev. D **58**, 063001 (1998), arXiv:gr-qc/9804014.

[27] H.-T. Wang, G. Yim, X. Chen, and L. Shao, Astrophys. J. **974**, 230 (2024), arXiv:2409.00970 [gr-qc].

[28] R. Abbott *et al.* (KAGRA, LIGO Scientific, VIRGO), Phys. Rev. D **106**, 102008 (2022), arXiv:2201.00697 [gr-qc].
[29] R. Prix and J. T. Whelan, Class. Quant. Grav. **24**, S565 (2007), arXiv:0707.0128 [gr-qc].
[30] Y. Wang, S. D. Mohanty, and F. A. Jenet, Astrophys. J. **795**, 96 (2014), arXiv:1406.5496 [astro-ph.IM].
[31] Y. Wang, Y. Shang, and S. Babak, Phys. Rev. D **86**, 104050 (2012), arXiv:1207.4956 [gr-qc].
[32] H.-T. Wang, Z. Wang, Y. Dong, G. Yim, and L. Shao, Phys. Rev. D **111**, 064037 (2025), arXiv:2411.13333 [gr-qc].
[33] X. Tao, Y. Wang, and S. D. Mohanty, (2024), arXiv:2412.07615 [astro-ph.HE].
[34] I. W. Harry and S. Fairhurst, Phys. Rev. D **83**, 084002 (2011), arXiv:1012.4939 [gr-qc].
[35] D. Keppel, Phys. Rev. D **86**, 123010 (2012), arXiv:1208.2340 [gr-qc].
[36] I. Harry, S. Privitera, A. Bohé, and A. Buonanno, Phys. Rev. D **94**, 024012 (2016), arXiv:1603.02444 [gr-qc].
[37] G. Pratten *et al.*, Phys. Rev. D **103**, 104056 (2021), arXiv:2004.06503 [gr-qc].
[38] M. Colleoni, F. A. R. Vidal, C. García-Quirós, S. Akçay, and S. Bera, Phys. Rev. D **111**, 104019 (2025), arXiv:2412.16721 [gr-qc].
[39] R. Abbott *et al.* (LIGO Scientific, Virgo), Phys. Rev. D **102**, 043015 (2020), arXiv:2004.08342 [astro-ph.HE].
[40] R. Abbott *et al.* (LIGO Scientific, Virgo), Astrophys. J. Lett. **896**, L44 (2020), arXiv:2006.12611 [astro-ph.HE].
[41] J. T. Whelan, R. Prix, C. J. Cutler, and J. L. Willis, Class. Quant. Grav. **31**, 065002 (2014), arXiv:1311.0065 [gr-qc].
[42] Y. Dong, Z. Wang, H.-T. Wang, J. Zhao, and L. Shao, (2025), arXiv:2502.01093 [gr-qc].
[43] H.-T. Wang, S.-P. Tang, P.-C. Li, and Y.-Z. Fan, (2025), arXiv:2509.02047 [gr-qc].
[44] H.-T. Wang, (2025), arXiv:2509.08657 [gr-qc].
[45] B. P. Abbott, R. Abbott, and T. D. e. a. Abbott (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. Lett. **119**, 161101 (2017).
[46] H.-T. Wang, Zenodo (2025), 10.5281/zenodo.16484750.
[47] H.-T. Wang and L. Shao, Phys. Rev. D **109**, 043027 (2024), arXiv:2401.13997 [gr-qc].
[48] C. Hoy and V. Raymond, SoftwareX **15**, 100765 (2021), arXiv:2006.06639 [astro-ph.IM].
[49] M. J. Williams, J. Veitch, and C. Messenger, Phys. Rev. D **103**, 103006 (2021), arXiv:2102.11056 [gr-qc].
[50] (2025), arXiv:2508.18082 [gr-qc].
[51] `https://gwosc.org`.
[52] R. Abbott *et al.* (KAGRA, VIRGO, LIGO Scientific), Astrophys. J. Suppl. **267**, 29 (2023), arXiv:2302.03676 [gr-qc].
[53] E. Higson, W. Handley, M. Hobson, and A. Lasenby, Bayesian Analysis **13**, 873 (2018).

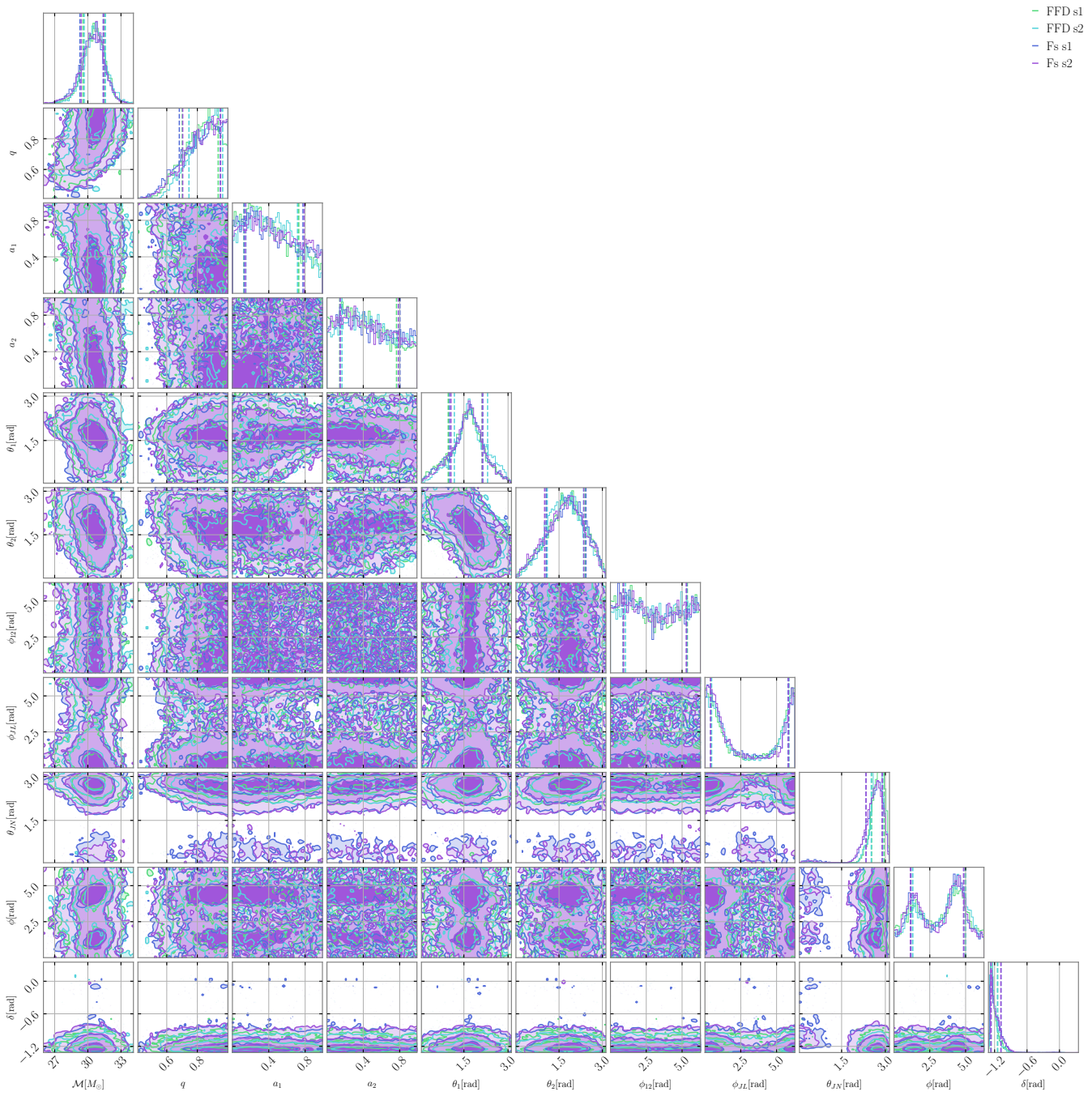


FIG. 10: Full corner plot showing the posterior distributions for the intrinsic source parameters, comparing the results from the FFD method and the $\mathcal{F}$-statistic (Fs) method under two different sampler settings ('s1' and 's2'). Two-dimensional contours indicate the 90% credible regions, while the diagonal panels show the marginalized one-dimensional posterior distribution for each parameter.