On the Current and Future Capability of Parameter Estimation Methods of Compact Binary Coalescence events through measured Gravitational Waves via LIGO, VIRGO, and Cosmic Explorer

Sterling Scarlett*
Boston University Commonwealth Ave, Boston, MA 02215
LIGO Summer Undergraduate Research Fellowship program

Mentors: Alan Weinstein and Jacob Golomb,†
Division of Physics, California Institute of Technology,
1200 E. California Blvd, Pasadena, CA 91125, USA
(Dated: August 5, 2024)

We introduce, understand, and interpret parameter estimation (PE), examining its role in analyzing Compact Binary Coalescence events (CBC) via their emission of Gravitational Waves (GW) and the measured GW strains by instruments such as LIGO, VIRGO, and the future detector Cosmic Explorer (CE). We explain and list all fifteen parameters that characterize of CBC. Understanding and looking at current methods of parameter estimation *(Bayesian inference, Fisher Matrix)*, we investigate how each method is currently used *(Bilby)*, future possibilities of the methods via machine learning *(Dingo)*, and advancements in detectors *(Cosmic Explorer)*. Conducting a comparative quantitative analysis with a methodology considering the speed, accuracy, and precision of the different methods and detectors, we aim to assess their efficiency and capabilities, offering insight on the future of PE of CBC events. We evaluate simulated events developed from a ***IMRPhenomXPHM*** waveform approximation and similar phenomenological frequency-domain waveform approximations to develop our injections of a full spectrum of CBC events. Doing this allows us to measure greater accuracy, precision, and characterization of potential strengths and weaknesses of each waveform model and detector.

Keywords: Parameter Estimation, Compact Binary Coalescence, Bayesian Inference, Fisher Matrix, Bilby, Dingo, Cosmic Explorer

I. INTRODUCTION

There exist a current established hierarchy of the parameter estimation methods on CBC events. The Bayesian inference method of Bilby stands as the industry standard and is the most accurate and precise method. Below it, Dingo represents a new frontier of PE not yet fully explored nor understood. Lastly, a Fisher matrix approach to PE represents merely a fast, rough, approximation. While this is the current standing, science is always evolving and changing. Understanding this prompts our interest in the comparison of these methods in this paper, with goals of providing quantitative and qualitative understanding of how this hierarchy is and provide future predictions on possible changes that it may hold.

This paper is structured as follows: Section II gives a brief introduction to GW and CBC, section **A** lists and defines all the astronomical parameters of CBC, highlighting both the intrinsic and extrinsic parameters. Section **B** offers an explanation of the population of CBC events and of exceptional events that exist as outliers in the distribution. Section **C** offers a discussion on the current PE methods, addressing briefly how both Bayesian inference and Fisher matrix methods work, highlighting the current codes that use them and where their strengths and weaknesses lie. Section **III** of this paper defines the objectives of the research, indicating what we hope to accomplish and how we will do it. Section **IV** defines the approach and overall methodology we take in our comparative analysis study of PE methods, understanding the process behind our study and defining the steps we take in data simulation, PE, analysis, and explaining anything abstained in our study. Section **V** discusses our data analysis and initial interpretations; Section VI gives our conclusions and final interpretations of the data. Section VII discusses the future implications of our work as a whole. In Section VIII, we give our thanks and acknowledgements. Lastly, Appendix A-B shows tables of numerical data from all PE's conducted in our study.

II. INTRODUCTION TO *CBC* AND PARAMETER ESTIMATION

Gravitational Waves (GW), predicted by General Relativity and detected by instruments like The Laser Interferometer Gravitational-Wave Observatory (LIGO), offer a window into the universe and provide data for understanding natural phenomena such as compact binary coalescence (CBC). Compact binaries are systems consisting of two compact astronomical objects. Members of this class include binary black holes (BBH), binary neu-

* sterling.scarlett3@gmail.com
† ajw@caltech.edu jgolomb@caltech.edu

tron stars (BNS), or one of each (NSBH). Coalescence describes all three stages of the collision of compact binaries. First, the objects' decaying orbits draw each other closer until they fully "Inspiral", in which the pair is losing energy and spiraling towards each other. Then they merge into one, producing GW, and lastly, the object stabilizes, forming one body emitting GWs; this stage is called "Ringdown". We therefore define CBC as a system of two compact objects inspiring, merging and stabilizing into ringdown (IMR), producing GW. These waves carry details about the merging systems, usually BBH, allowing for parameter estimations of the source and resulting object [1]. Though compact binaries are the focus of this research, compact mergers are not limited to binaries. There exist different classes of mergers, an example are mergers from hierarchical triple-star systems or triples. These are systems organized as an inner, binary pair with a more distant, outer (tertiary) component [2–4].

Parameter	Symbol
Primary mass	m_1
Secondary mass	m_2
Time of coalescence	t_c
Spin magnitudes	χ_1, χ_2
Sky position	α, δ
Reference phase	ϕ_c
Luminosity distance	d_L
Inclination angle	θ_{JN}
Spin angles	$\theta_1, \theta_2, \phi_{12}, \phi_{JL}$
Polarization angle	ψ

TABLE I. Astrophysical parameters of CBC events.

A. Astronomical parameters of *CBCs*

Focusing on CBC, as shown in TABLE I, the gravitational waveform formed by these events is characterized by fifteen parameters, eight "intrinsic" $(m_1, m_2, \chi_1, \chi_2, \theta_1, \theta_2, \phi_{12}, \phi_{JL})$, and seven "extrinsic" $(t_c, \alpha, \delta, \phi_c, d_L, \theta_{JL}, \psi)$:

- **Detector-frame component masses** (m_1, m_2) - the mass of the larger object involved in the merger (**primary mass** m_1) and mass of the smaller object (**secondary mass** m_2);
- **Time of coalescence** (t_c) - the time when the binary system merges, as measured at the geocenter (center of the Earth);
- **Sky position** (α, δ) - the location of the event on the celestial sphere in right ascension α and declination δ;
- **Reference phase** (ϕ_c) - a reference point in the waveform used to describe the phase of the gravitational wave signal at the time of coalescence;
- **Luminosity distance** (d_L) - the distance from the observer to the binary system, calculated based on the observed amplitude of the gravitational wave signal and its intrinsic strength;
- **Inclination angle** (θ_{JN}) - the angle between the total angular momentum vector of the binary system and the line of sight from the observer to the source;
- **Polarization angle** (ψ) - the angle that describes the orientation of the polarization of the gravitational wave signal relative to the line of sight from the observer to the source.
- **Spin magnitudes** (χ_1, χ_2) - the magnitudes of the spins of the two compact objects in the binary system, measured in a 3-vector quantity indicating how fast they are rotating;
- **Spin angles** ($\theta_1, \theta_2, \phi_{12}, \phi_{JL}$) - the various angles that are referenced to the orbital angular momentum $\hat{L}$ and the total angular momentum $\vec{J} = \vec{L} + \vec{S_1} + \vec{S_2}$.

While these fifteen parameters describe CBC, they are not the only quantities that can be measured from GW strain data. Other parameters, derived from the 15 can be defined, such as:

- **Effective Inspiral spin magnitude** (χ_{eff}) - a measurement of the effective spin of the binary system during the "Inspiral" phase, accounting for both individual spins and their orbital orientation, used to find the quantities determining spin parameters;
- **Cosmological redshift** (z) - the stretching of the wavelength of light as it travels through space; determines distance, and can be inferred using d_L;
- **Signal-to-noise ratio** (snr) - measurements of the quantity of strength of the gravitational wave signal compared to the background noise in the detectors;
- **Radiated energy** (E_{rad}) - the energy radiated in GW during the merger;
- **Peak luminosity** (L_{peak}) - the maximum brightness signal emitted;
- **Orbital precession** (δ, ϕ) - the combination of Apsidal precession, where the major axis of an elliptical orbit cycles its orientation within its orbital plane, and Nodal precession, where non-spherical objects cause orbiting objects to change their orbits;
- **Chirp mass** ($\mathcal{M}_\odot$) - the combination of the two object masses of the system;
- **Final Mass** ($M_\odot$) - the mass of the resulting black hole after the merger.

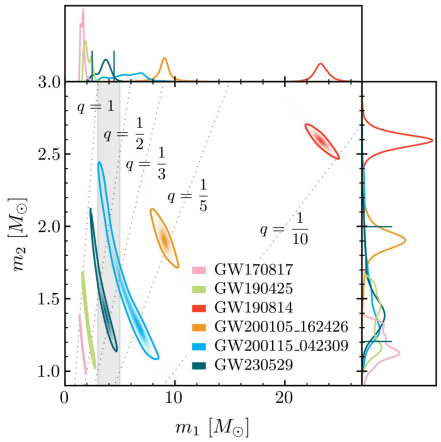

FIG. 1. Taken from [5]. The one- and two-dimensional posterior probability distributions for the component masses of the source binary of GW230529 (teal). The contours in the main panel denote the 90% credible regions with vertical and horizontal lines in the side panels denoting the 90% credible interval for the marginalized one-dimensional posterior distributions. Also shown are the two O3 NSBH events GW200105 162426 and GW200115 042309 (orange and blue respectively; Abbott et al. 2021a) with FAR < 0.25 yr ≈ 1 (Abbott et al. 2023a), the two confident BNS events GW170817 and GW190425 (pink and green respectively; Abbott et al. 2017a, 2019a, 2020a, 2024b), as well as GW190814 (red; Abbott et al. 2020c, 2024b) where the secondary component may be a black hole or a neutron star. Lines of constant mass ratio are indicated by dotted gray lines. The grey shaded region marks the 3–5 $M_\odot$ range of primary masses. The NSBH events and GW190814 use combined posterior samples assuming a high-spin prior analogous to those presented in this work. The BNS events use high-spin IMRPhenomPv2 NRTidal (Dietrich et al. 2019a) samples.

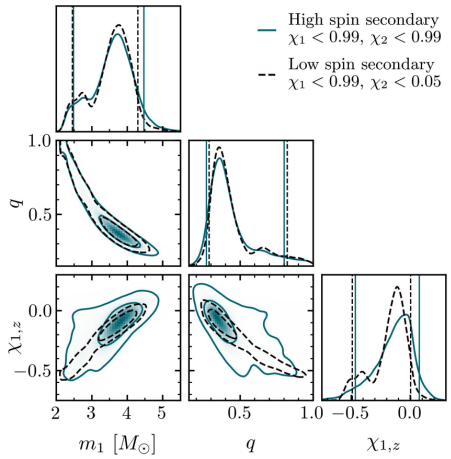

FIG. 2. Taken from [5]. Selected source properties of GW230529. The plot shows the one-dimensional (diagonal) and two-dimensional (off-diagonal) marginal posterior distributions for the primary mass m_1, the mass ratio q, and the spin component parallel to the orbital angular momentum $\chi_{1z} = \vec{\chi_1} \cdot \hat{L}$. The shaded regions denote the posterior probability with the solid (dashed) curves marking the 50% and 90% credible regions for the posteriors determined using a high spin (low spin) prior on the secondary of $\chi_2 < 0.99$ ($\chi_2 < 0.05$). The vertical lines in the one-dimensional marginal posteriors mark the 90% credible intervals.

Another set of parameters that extends the description (not derived from the 15) specifically for BNS is

- **Tidal distortion parameters** (Λ_1,Λ_2) - These refer to the effects of tidal forces on the shapes and orbits of the neutron stars in the binary system.

.

B. Population of events and exceptional events

By analyzing CBC events and conducting parameter estimations, scientists construct distributions of observed mergers, forming populations of events; these populations are used as the foundation for predictive models and simulations of potential future events. Populations also tell us about what nature is doing. We use of observations of CBC to infer, and gain understanding about the population of CBC systems in the universe, in nature. .3 shows a population of events developed from confirmed CBC events observed via LIGO by the component

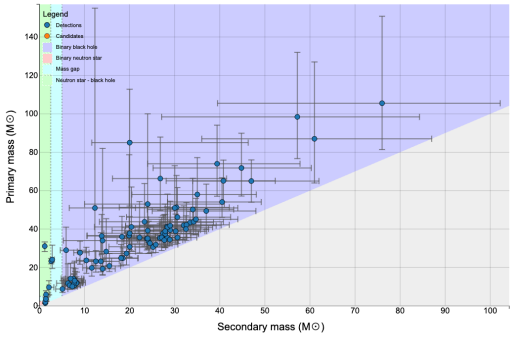

FIG. 3. Taken from [6] . Population distribution of confirmed CBC events via LIGO plotted by m_1 and m_2.

masses m_1 and m_2 However as shown in the figure, some observed events deviate from these populations, termed "exceptional events". These expected occurrences, such as black hole mergers resulting in unusually high final masses $M_\odot$, [7] or high component spins χ_1, χ_2 [8], challenge current theoretical frameworks. Presently, the origins and mechanisms behind these events are unknown. However, leading theories suggest scenarios involving hierarchical compact binary coalescence, in which coalescence binary black hole systems merge with another system [9]. A definitive understanding is lacking, leading

to uncertain predictability and parameter estimation of these events, prompting our data collection and further analysis of them. Addressing this knowledge gap is crucial for advancing our comprehension of accurate estimation of astrophysical processes to further understand the evolution of black hole systems. [1].

C. Parameter estimation

Given a specific GW source, such as BBH merger, PE aims to understand the 15 parameters that characterize the event by constructing a probability posterior distribution (PPD) of where each parameter may exist. However, there lies the difficulty in visualizing all fifteen parameters. As shown in FIG 1 taken from the LVK collaboration (LIGO, VIRGO and KAGRA) visualization of data measuring the one- and two-dimensional posterior probability distributions for the component masses of the source binary of GW230529, and the one-dimensional (diagonal) and two-dimensional (off-diagonal) marginal posterior distributions for the primary mass m_1, the mass ratio q, and the spin the component parallel to the orbital angular momentum $\chi_{1z} = \vec{\chi_1} \cdot \hat{L}$ [5], scientists tend to limit the visualization to small integer dimensional analysis.

As stated previously in section II, there exist different classes of mergers, triple-star system, hierarchical, dynamically captured system mergers, and many more. PE, not just focused on understanding the parameters that make up the merger, can tell us a lot about the forming mechanisms of the system in which the mergers existed in. An example of this is in the search for evidence of dynamically captured systems. These are systems in which the black holes of a BBH pair formed independently of each other, then captured each other's orbits, usually with the help of a third larger object, and become a BBH system. Evidence of such a merger being a dynamically captured system occurs when constructing PPD of the individual spins of the BBHs. Although these individual spins of the black holes (χ_1, χ_2) that make a BBH event are difficult to measure, if we develop methods to measure these two parameters accurate and measure an BBH merger event with much unaligned spins, this could serve as evidence to suggest that the BBH system was dynamically captured.

As stated previously, PE has a current hierarchy of methods, Bilby, Dingo, and Fisher matrix. In these methods there exist two fundamentally different methodologies that construct each of them, Bayesian inference, Fisher information matrix. .

1. *Bayesian inference*

Given a set of parameters $\vec{\theta}$ derived from a prior model and a set of data d, Bayesian inference aims to create a posterior distribution according to Bayes' theorem:

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

$p(\theta|d)$ can be interpreted as a conditional probability: the posterior probability distribution (PPD) for θ given the data d; $\mathcal{L}(d|\theta)$ is the probability of getting the data d given a set of model parameters θ. Some common assumptions are made of this function though; it is assumed that the noise is stationery and Gaussian, with zero mean. The Gaussian-noise likelihood function is:

$$\mathcal{L}(d|\theta) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{1}{2}\frac{|d-\mu(\theta)|^2}{\sigma^2}\right), \tag{2}$$

In which $\mu(\theta)$ represents an example gravitational strain waveform template derived from a given θ, σ is the detector noise, and $\pi(\theta)$ is the distribution of the prior on θ. This equation for likelihood $\mathcal{L}$ is a simplified version of the full equation, which can be derived as follows.

Given that data d is not constant but a function of time $d(t)$, we define the function of data as a combination of the signal $h(t)$ and the noise $n(t)$:

$$d(t) = h(t) + n(t). \tag{3}$$

The true gravitational wave signal in the detector is unknown, so we use signal models to describe it. These models are deemed to be a good description if the residuals $r = d - \mathcal{H}$, where $\mathcal{H}$ is our signal model, are consistent with the instrumental noise. Following this, simplifying assumptions, as stated previously, are made about the noise in the data $n(t)$. We can assume that the noise is Gaussian and that its probability density function (PDF) matches a uniform distribution. Using this assumption and the definitions of $d(t)$, we can now define the likelihood function as a noise model and rewrite it in the form:

$$p(d \mid h) = \frac{1}{\sqrt{\det(2\pi C)}} e^{-\frac{1}{2}\chi^2(d,h)}, \tag{4}$$

where C is the noise correlation matrix and:

$$\chi^2(d,h) = \mathbf{r}\cdot C^{-1}\cdot\mathbf{r} = (d_{I_k} - h_{I_k})C^{-1}(I_k)(J_m)(dJ_m - h_{J_m}). \tag{5}$$

Following the assumption of Gaussian noise and assuming that noise remains constant, or more specifically stationary, we can define the noise correlation matrix for each detector C to be diagonal in the Fourier domain:

$$\mathrm{S}^I_{km} = \delta_{km}\mathrm{S}^I_{(f_k)}. \tag{6}$$

We can now define $\chi^2(d,h) = (r \mid r)$, knowing that:

$$(a \mid b) = 2\int_0^\infty \frac{\tilde{a}(f)\tilde{b}^(f) + \tilde{a}^(f)\tilde{b}(f)}{S_n(f)}, df, \tag{7}$$

is a familiar inner product, in which it is understood that $S_n(f)$ represent the power spectral density also previously referred to as the detector noise σ . Given this

definition, we can expand the simplified version of our likelihood function $\mathcal{L}$ and write the full function for stationary Gaussian noise that is independent in detectors as:

$$\mathcal{L}(d \mid \theta) = \exp\left(-\frac{1}{2}\sum_I \left[\int \left(\frac{(d_I - h_I(\theta))^2}{S_n^I(f)} + \ln(S_n^I(f))\right) df\right]\right) \text{[10]}. \tag{8}$$

Reverting pack to our discussion of $p(\theta|d)$, $\mathcal{Z}$ is interpreted as the integral of the numerator or normalization factor called the "evidence" [11].

$$\mathcal{Z} = \int \mathcal{L}(d|\theta)\pi(\theta)d\theta. \tag{9}$$

Although not explored in this paper, the use of evidence $\mathcal{Z}$ can be used a source of different model comparisons; this topic is explored in [12]. Given that θ is the properties of the source, the efficacy of conducting the posterior distribution is determined by the number of parameters aiming to construct distributions for, with larger numbers leading to more imprecise approximations due to correlation between the parameters. It is noted that leaving out any known parameter when conducting Bayesian Inference leads to a more inaccurate, less precise measurement. This method suffers from "the curse of dimensionality" and is computationally difficult when the source has many parameters such as CBC events. Noting the less accurate nature of the measurement, scientists often construct a **Marginalized posterior distribution** to determine the distribution for a single parameter, in which you integrate over all measures of θ_k except the one we want

$$p(\theta_i|d) = \int \left(\prod_{k\neq i} d(\theta_k)\right) p(\theta|d) \tag{10}$$

$$p(\theta_i|d) = \frac{\mathcal{L}(d|\theta_i)\pi(\theta_i)}{\mathcal{Z}}. \tag{11}$$

This equation can be written in terms of the likelihood function of the specified parameter or **marginalized likelihood function** $\mathcal{L}(d|\theta_i)$

$$\mathcal{L}(d|\theta_i) = \int \left(\prod_{k\neq i} d(\theta_k)\right) \pi(\theta_k)\mathcal{L}(d|\theta). \tag{12}$$

Marginalizing over all but one θ_i results in a 1D posterior distribution on θ_i, which contains all of the uncertainty associated with the other parameters that are covariant with θ_i. An example of this is shown in FIG 4, which shows a marginalized posterior distribution for the well-known covariance between the luminosity distance of a merging compact binary from Earth D_L and the inclination angle θ_{JN}.

The previous most common tool used by LVK collaborations for Bayesian inference was LALInference [13]; this

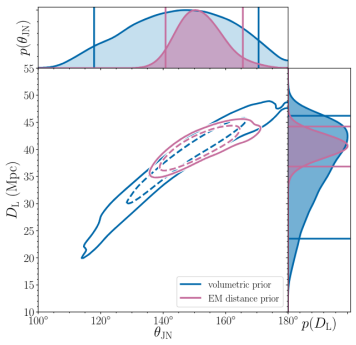

FIG. 4. Taken from [11]. The joint posterior for luminosity distance and inclination angle for GW170817 from (Abbott et al., 2019). The blue contours show the credible region obtained using gravitational wave data alone. The purple contours show the smaller credible region obtained by employing a relatively narrow prior on distance obtained with electromagnetic measurements. Publicly available posterior samples for this plot are available here: (LIGO/Virgo, LIGO/Virgo).

code is over a decade old, and an updated python-toolkit version of the code has been developed [14]. Improvements of these methods, Bilby [15] and Dingo [16] serve as the focus of our research.

a. Bilby is a likelihood-based Bayesian Inference Python toolkit used to conduct parameter estimation. It uses sampling of a prior distribution to produce a likelihood model. This code has more versatility than its predecessors *(LALInference)*; while mostly used in the estimation of compact binary events [17, 18], it is also used in other astronomical fields such as astrophysical inference in multimessenger astronomy, pulsar timing, and x-ray observations of accreting neutron stars [15]. It serves as not only a more versatile method of Bayesian interference compared to previous methods but also boasts improvements in efficiency, speed, accuracy, and simplicity over its predecessors [17].

Although true, Bilby's development of a likelihood causes it to suffer in terms of speed compared to the other methods of PE in our study *(Dingo, Fisher Matrix)*. This is due to various things, such as the integration over multiple parameters, potential intricate relationships and non-linear dependencies among variables, complex likelihood functions that are computationally expensive to evaluate, and sufficiently large datasets. All the deficiencies get expanded with the use of advanced sampling methods such as *dynasty* [19], the nested sampling method we use in this paper to construct our PEs. They do so because of a plethora of circumstances surrounding the method.

Nested sampling requires a number of live points to walk around the highest likelihood points, systematically weighting and sampling to build a posterior distribution of the parameter, and high dimensionality, complex likelihood surfaces, insufficient exploration settings, and large datasets can delay the construction of these posteriors.

All this, without sufficient computational power, makes the construction of marginalized posterior distributions for singular parameters with Bilby take anywhere from hours to days, to full fifteenth-dimensional PE, taking days to even months to complete on Bilby without advancements. Though this may be the case, Bilby is constantly in development and currently offers a multitude of settings to expedite the process of full PE, allowing posterior distributions of confirmed detected GW signals from LIGO to be constructed within minutes of observation. Some of these advancements include *Multibanding* [20], *Relative Binning* [21], and the use of *reduced order quadratures (ROQs)* [22]. [22] and[21] delve into the latter topics more and offer a full explanation of the methods; furthermore, we shall focus on *Multibanding* and *Relative Binning* in this paper and define them later in Section III as a result of their use in our construction of posteriors and subsequent final data analysis.

b. Dingo or Deep Inference for Gravitational-wave Observations, is a likelihood-free alternative approach to Bayesian Inference. It uses artificial intelligence (AI) machine learning instead of sampling to dramatically decrease the analysis time of gravitational wave inference and parameter estimation of CBC events, at the current somewhat cost of accuracy and precision. It uses the method of *neural posterior estimation* (NPE) [23, 24], in which it takes large simulated data sets with their associated parameters to train its type of neural network called a *normalizing flow* to produce a posterior distribution, generating distributions quickly after a detection is made, bypassing the need and cost of generating waveform at inference time [16]. A difference in how this method varies from the conventional Bayesian inference methods is how the likelihood is used. Conventional methods such as Bilby use the likelihood to determine density $p(d|\theta)$, while Dingo and NPE methods use the likelihood to simulate data $d \sim p(d|\theta)$ [16].

2. *Fisher Matrix*

A Fisher matrix or FIM (Fisher Information Matrix), approach is defined as

$$\Gamma_{\kappa\nu} = \left\langle \frac{\partial h}{(\partial\theta^k)} | \frac{\partial h}{(\partial\theta^\nu)} \right\rangle = -E\left[\frac{(\partial^2 lnL)}{(\partial\theta^k \partial\theta^\nu)}\right], \qquad (13)$$

assumes the waveform measured in CBC events can be seen as a linearizable approximation of the parameters, near enough to the true value, and the error distributions in the parameter errors are Gaussian, meaning the inverse of the FIM is said to give the variance-covariance matrix. Moreover, the square root of the diagonal components of the variance-covariance matrix gives an error estimation in the parameter estimation. As opposed to parameter codes that can produce multi-modal PPDs *(Bilby [15], Dingo [16])*, the FIM approach assumes a mono-modal posterior parameter distribution (PPD), assuming the 15 parameters of the observation describe a 15-dimensional space and the posterior distribution in that space is a 15-dimensional Gaussian with one peak. Rarely is this the case, as shown in both FIG 1 and 4, most events are multi-modal, consisting of multiple peaks in their distributions. Therefore, the FIM approach is only a rough, fast approximation to the most probable values of the parameters. This approach boasts much faster speeds than Bayesian interference in parameter estimation but suffers

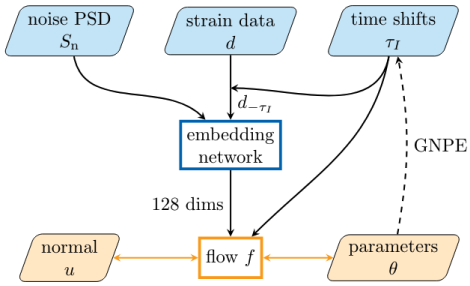

FIG. 5. Dingo flow chart taken from [16]. The posterior distribution is represented in terms of an invertible **normalizing flow** (orange), taking normally distributed random variables u into posterior samples θ. The flow itself depends on a (compressed) representation of the noise properties S_n and the data d, as well as an estimate τ_I of the coalescence time in each detector I. The data are time-shifted by τ_I to simplify the representation. For inference, the iterative *group equivariant neural posterior estimation* (GNPE) algorithm is used to provide an estimate of τ_I, as described in the main text.

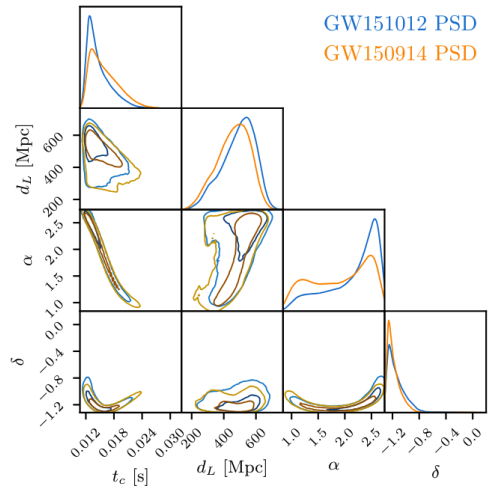

Figure 6. Comparison between DINGO evaluated on GW150914 using correct PSD as context, and using GW151012 PSD. These four parameters have a mean JSD of 0.020 nat.

FIG. 6. Taken from [16]. Comparison between Dingo evaluated on GW150914 using correct PSD as context, and using GW151012 PSD. These four parameters have a mean JSD of 0.020 nat.

greatly in accuracy and precision [25].

Currently their exist multiple of python libraries that use this method of PE, GWFish [26], GWFast [27], and Jimfisher to name a few. Though, the method of Fisher matrix we use in this paper is GWFish.

a. GWFish is a Fisher matrix simulation software used to evaluate the current and predict the future parameter estimation capabilities of gravitational-wave detector networks. Using the Fisher matrix method, GWFish estimates errors by employing a quadratic approximation of the likelihood $\mathcal{L}$, given by:

$$\mathcal{L} \propto \exp\left(-\frac{1}{2}\Delta\theta^i \mathcal{F}_{ij} \Delta\theta^j\right), \tag{14}$$

where $\Delta\theta = \theta - \vec{\theta}$ represents the vector of parameter estimation errors, with $\vec{\theta}$ being the true parameter values. The code computes the Fisher matrix $\mathcal{F}$ as:

$$\mathcal{F}_{ij} = \left(\frac{\partial h}{\partial \theta_i}\middle|\frac{\partial h}{\partial \theta_j}\right), \tag{15}$$

where h, following the same foundation in our equation for likelihood 8, refers to the frequency-domain strain at the detector, depending on the parameters θ. GWFish then calculates the covariance matrix by inverting $\mathcal{F}_{ij}$, which represents the variance in each parameter.

One of GWFish's main strengths is its general usability, especially when compared to GWFast, which has seen much of its code deprecated. Additionally, GWFish offers more versatility than JimFisher. While JimFisher currently uses JAX [28] for automatic differentiation of waveforms, not all waveforms—especially those that include higher-order modes, such as IMRPhenomXPHM—have been implemented in JAX. Until these waveforms are supported, numerical differentiation techniques are necessary, which GWFish implements but JimFisher does not.

Moreover, the use of waveform models that account for higher-order modes significantly affects the way the Fisher matrix produces posteriors. Accounting for higher-order modes reduces the degeneracy that can occur in the posteriors, making the distribution non-uniform. These degeneracies can manifest in various ways. An example of a well-known degeneracy was shown in Fig 4, where a high correlation was observed between the luminosity distance and orbital orientation posteriors. The inclusion of higher-order modes in GWFish not only decreases degeneracy in the posteriors but also constrains the variance in error estimates within the constructed posteriors. This effect is illustrated in Fig 7, which shows the difference in GWFish posterior distributions when accounting for higher-order modes in the waveform approximation versus when they are not accounted for. The usability and versatility of GWFish are the primary reasons for choosing it in our paper.

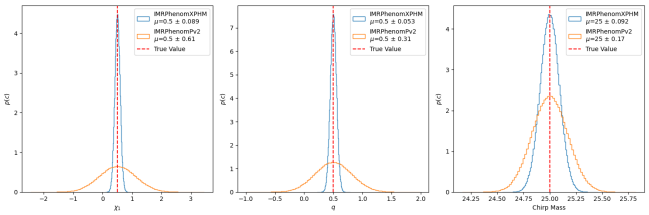

FIG. 7. GWFish posterior distributions using IMRPhenomXPHM *(blue)* and IMRPhenomPv2 *(orange)* waveform approximations for primary spin magntitude χ_1, mass ratio q, and chirp mass $\mathcal{M}$.

III. OBJECTIVES OF RESEARCH

Our primary goal is to conduct a comparative analysis of three parameter estimation methods—*Fisher Matrix*, *Bilby*, and *Dingo*—by employing simulated events as a benchmark, using a full spectrum of events as to assess the full capability of the models and detectors. As previously stated, we will also conduct this comparative analysis on the different methods depending on the detector being used to observe the event: *LIGO*, *Cosmic Explorer (CE)*. Furthermore, we will apply different Power Spectral Density's (PSD) to our detectors. Given the availability, we will simulate the data with a PSD of the previously observation LIGO run o3, and predictions of the PSD of observations o4 and o5.

Another goal of this paper is to ***predict the reliability of a Fisher matrix as an accurate method of PE in future detectors (CE)***. As a Fisher matrix approach to PE assumes the 15 distributions are Gaussian, as our detectors get more sensitive, via advancements of current detectors, or construction of higher sensitive detectors e.g CE, we predict that with much higher SNRs than we currently observe, the distributions will fit a Gaussian model. Another way this can occur is as stated previously with accounting for higher-order modes in the waveform approximate model. Higher order modes limit degeneracies that cause multi-modality, the use of a model such as **IMRPhenoXPHM** [29] can help construct Gaussian PPDs. All of this will allow Fisher Matrix to serve as an equivalent source of accuracy and precision to other PE models, while having much higher computational speed. If our predictions are correct, this will allow PE of simulated events to be conducted much computationally faster than current standards. Though, given the predicted improvements to sensitivity, the detectors will be able to observe much farther and fainter signals than previously in which the SNR will not be high enough for a Fisher Matrix to be appropriate, in this case Bilby or Dingo will be used for PE.

Overall, this comparative study will be done by a quantitative evaluation to discern the efficacy of each method in approximating parameters of interest. In our research, we aim to measure several key aspects:

1. Speed: We will quantify the computational effi-

ciency of each method, particularly focusing on the time taken to generate parameter estimation result for events, and the deviation of the speed depending on varying exceptional events and detectors being used.

2. Accuracy: We will evaluate the accuracy of parameter estimations provided by each model constructed by the method. Seeing how close the average value of the estimated distributions is to the true injected parameters. This will be done using a mean calculation of the parameter distributions, and then calculating the percentage error of the known simulated parameter versus the estimated.

3. Precision: We will determine how precise each model is, determine how uniform or wide the spread of the estimated distributions are. This will be done by evaluating the standard deviation of the each of the parameter distributions, then calculating the percentage error of the known simulated parameter versus the estimated.

4. Capability: We will assess the capability of each of method of maintaining solid speed, accuracy, and precision under various conditions. This includes different degrees of deviation from the observed distribution in simulated events, the number of parameters that each model can handle, the range of parameter space it can handle, and capability between the methods via the detector being used.

5. Gausianity: We will assess how Gaussian each method constructs the PPD for each parameter. This will be tested by various Gaussian test such as a K-L divergence test.

It is to be acknowledged that each simulated event deviates from the normal distribution to varying degrees, thereby challenging the efficacy of the methods in accurately capturing these deviations. Additionally, given the intrinsic characteristics of Fisher matrices, we anticipate that although this method offers speed advantages, it will compromise on accuracy compared to other approaches. Furthermore, Dingo claims to be much faster and more efficient than the alternative Bilby [16], therefore significant increases in productivity via Dingo are expected with a caveat of a decrease in overall accuracy. It is also known that Dingo is only trained on specific simulated events, such as mass ratios from 1:1 to 1:3; it does not have the current capability to construct meaningful distributions for mass relations 1:5 or higher, therefore decreases in the capability of the model in simulated events of this variety are expected. The next-generation detector, CE [30–32], is anticipated to significantly boost the signal-to-noise ratio of observed events. As a result, all parameter estimation methods are expected to show improvements in accuracy and precision.

IV. APPROACH TO COMPARATIVE ANALYSIS

A. Simulation of waveforms

We initially use the ***IMRPhenomXPHM*** [33] waveform approximation to simulate our preliminary waveforms and then continue using the same frequency-domain waveform approximation subsequently. The choice to construct our waveforms from an IMR-type model instead of an effective one-body-numerical-relativity (EOBNR) model***SEOBNRv3***, or a post-Newtonian gravitational wave approximation, like Taylor approximation models [34], stems from the varying times for waveform construction between the approximation models. IMRPhenomXPHM and its subsequent successors offer exponentially faster waveform generation than other models, although with a caveat of a decrease in both accuracy and precision. We feel the level of accuracy and precision increase other models have does not substantially change the analysis and conclusions of our data, nor does it warrant the exponentially larger waveform generation time, so we refrain from using such models.

B. Methodology of chosen injections

Given an understanding of the current distribution of CBC events, we mostly choose events that lie innately within this prior distribution, as shown in Table II.

.

TABLE II. Priors for Injection Parameters

Symbol	Distribution	Minimum	Maximum
m_1	Constraint	5	100
m_2	Constraint	5	100
q	Uniform	0.125	1
$\mathcal{M}$	Uniform	20	75
d_L	UniformSourceFrame	100	5000
δ	Cosine	-	-
α	Uniform	0	2π (periodic)
θ_{JN}	Sine	-	-
ψ	Uniform	0	π (periodic)
ϕ	Uniform	0	2π (periodic)
a_1	Uniform	0	0.99
a_2	Uniform	0	0.99
$\tilde{\theta}_1$	Sine	-	-
$\tilde{\theta}_2$	Sine	-	-
ϕ_{12}	Uniform	0	2π (periodic)
ϕ_{JL}	Uniform	0	2π (periodic)

We also choose "exceptional events" that lie in the outermost regions of these distributions, such as unusually high chirp masses and spins. We abstain from any PE of events below a chirp mass of $\mathcal{M} < 15$ due to Dingo's current inability to construct meaningful estimations below that threshold. Doing a combination of this allows

us to "stress-test" both the models and detector capabilities, giving us not only insightful visualization insights but also numerical insights to help build our final conclusions.

C. Use of Advancements in Models

As stated in I, Bilby offers a plethora of advancements that allow PE to speed up from taking days to mere minutes, such as relative binning, reduced order quadrature, and multibanding. The only advancement in Bilby we use in this paper for analysis is multibanding, which is defined as follows.

Multibanding Given the understanding that the inspirals of GW exist primarily at lower frequencies and that the emission of higher frequencies strongly correlates with the end of the inspiral, The technique of *MultiBanding* uses this morphology to develop waveforms at substantially lower frequencies while maintaining their accuracy. It accomplishes this by partitioning the frequency domain of the of the GW strain into bands

$$h(f) = \sum_b h_b(f). \tag{16}$$

At each band b, it evaluates the waveform at different resolutions. The bands and frequencies can be chosen based on leading post-Newtonian expressions (cite) for the time to merger as a function of frequency and chirp mass (cite) or general error bounds for the Taylor series (cite).

Relative Binning Heterodyning, more precisely referred to as relative binning, facilitates the integration of oscillatory functions within the waveform, particularly in likelihood construction at lower resolutions. This method operates on the assumption that waveforms supported by the posterior are similar to one another, given their alignment with the data. The technique maximizes the likelihood to select a reference waveform, $h_0(f)$, which is then used to describe any similar waveforms, $h(f : \theta)$. The reference waveform $h_0(f)$ is initially constructed at full resolution. Subsequently, smooth corrections are applied to adapt it for multiple values of θ at a lower frequency resolution. This approach allows for the frequency resolution and computational cost to remain independent of the signal's duration. As previously detailed,[21] provides a comprehensive discussion of this advanced method and the underlying process.

We refrain from using ROQs due to the fact that there is not currently an existing ROQ for CE use. The construction, or potential, of such a ROQ is currently unknown, making the use of ROQ for events detected in LIGO and the lack thereof in CE induce improper data.

D. Model enhancements/Training

Although Bilby can provide information and predictions about signals as if they were observed by CE, there is no trained Dingo network specifically for this purpose. To address this, we manually train a network tailored to ensure that parameter estimation (PE) behaves as if the gravitational wave (GW) signal were observed by CE. Given that we are predicting noise levels for CE, our Dingo model is trained on a single noise realization for CE, whereas most models are trained on datasets with varying amplitude spectral densities (ASDs). This choice is driven by the lack of implementations to generate artificial noise realizations, and the uncertainty involved in assuming varying noise levels based on predictions. The method Dingo uses to construct synthetic ASDs might introduce significant inaccuracies and biases, which would be challenging to analyze and could render the results non-meaningful. Consequently, we train our model on a dataset of waveform approximations constructed from **IMRPhenomXPHM**, ensuring consistency with the model used in other methods, In addition to this, the priors used in the training are those shown in Table II.

E. GWFish Prior Incorporation

GWFish currently lacks the innate implementation to incorporate priors in the sampling of its multivariate uniform distribution of parameters and construction of posteriors , which can lead to non-physical distributions such as mass ratios $q > 1$ or $q < 0$, and aligned spin magnitudes $a_i > 1$ or $a_i < 0$. This latter issue is illustrated in the Fig 7, which address the differences arising from using various waveform models, in the figure we see the distributions having a spin of less than 0 which is non physical. To incorporate priors into the distributions and fix these inaccuracies, we use two methods, but both follow the same initial process listed below.

1. Calculate the Probability Density Functions (PDFs) for Each Parameter:

For each parameter θ_i, compute the PDF $p_i(\theta_i)$ for each sample $\theta_i^{(j)}$:

$$\text{PDF}_{i,j} = p_i(\theta_i^{(j)}) \tag{17}$$

2. Compute the Product of the PDFs:

For a given set of samples $\{\theta_1^{(j)}, \theta_2^{(j)}, \ldots, \theta_N^{(j)}\}$, the product of all PDFs is:

$$P(\theta_1^{(j)}, \theta_2^{(j)}, \ldots, \theta_N^{(j)}) = \prod_{i=1}^{N} p_i(\theta_i^{(j)}) \tag{18}$$

3. Normalize the Product of PDFs:

Compute the normalization factor Z by summing the

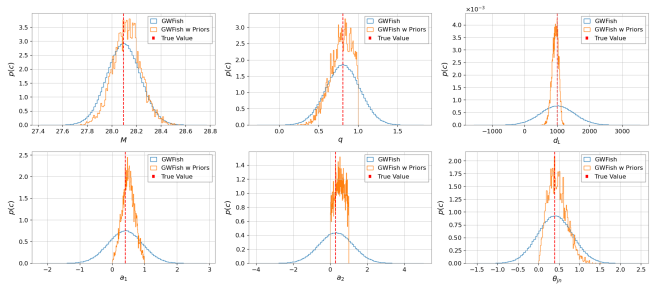

FIG. 8. Chirp mass M,mass ratio q, luminosity distance d_L,spin magnitudes a_1, a_2, and orbital orientation θ_{JN} posterior distributions constructed using GWFish. Blue posteriors indicated base GWFish not sampled with priors and orange shows GWFish with priors incorporated.

products of the PDFs over all samples:

$$Z = \sum_{j=1}^{M} \prod_{i=1}^{N} p_i(\theta_i^{(j)}) \tag{19}$$

Normalize the product for each sample:

$$w_j = \frac{\prod_{i=1}^{N} p_i(\theta_i^{(j)})}{Z} \tag{20}$$

where w_j is the weight associated with the j-th sample. After this, there are two approaches to incorporating priors into the distributions. First, we can use the already constructed parameter samples and simply add the weights to the graphing software used to visualize the distributions. Alternatively, we can adjust the process to incorporate priors directly into the sampling and distribution construction by

4. Resampling from the Multivariate Uniform Distribution Using the Weights:
Use the weights w_j to resample from the multivariate uniform distribution, obtaining new samples $\{\tilde{\theta}_1^{(j)}, \tilde{\theta}_2^{(j)}, \ldots, \tilde{\theta}_N^{(j)}\}$ that incorporate the prior distribution.

Following this process allows us to incorporate priors in our fisher matrix distributions, which is needed when making comparison between the various methods of PE. The effects of using priors and lack thereof are shown in FIG 8.

F. Preliminary Comparison of Models and Detectors

We conduct preliminary stress tests on both the PE models and detectors. This is done individually for each model (*Bilby,Dingo,Fisher*) and detector (*CE,LIGO*). This is implemented by manually choosing injections that lie within a known prior distribution of BBH events, simulating them with the ***IMRPhenomXPHM*** waveform approximation, then feeding the waveform to our detectors and conducting PE with our methods. Next, we construct corner plots similar to those shown in FIG. 2 of the desired parameters. Though an understanding of all 15 parameters will be numerically understood, for our test, we only visualize specific important parameters. These parameters are chirp mass $\mathcal{M}$, mass ratio q, effective spin χ_{eff}, sky localization (α, δ), and luminosity distance d_L. These are of importance because some of these parameters greatly affect the structure of the waveform, e.g., chirp mass $\mathcal{M}$ and effective spin χ_{eff}, which are more easily observed than parameters such as the individual spin of a BH χ_1, giving us better insight as a whole from the PE. After initial visualization, we build a table of all 15 estimated parameter distributions. The table has rows of the parameters and columns of time taken to generate PE, means of distributions, standard deviation, true value of parameters, accuracy's, calculated from the percent error of the mean, and precision's, calculated from the percent error of the standard deviation.

1. *Comparison via different detectors:*

LIGO currently has two detectors, one in Hanford, Washington, and the other in Livingston, Louisiana. Their respective, known performance is well documented, with their capability and spectral noise curve both respectively known [1, 5, 7–9, 13–18, 23, 24, 35]. CE is a next-generation observatory concept that will greatly deepen and clarify humanity's gravitational-wave view of the cosmos. It will have an L-shaped geometry and house a single interferometer. The CE facility will have two 40 km ultrahigh-vacuum beam tubes, roughly 1 m in diameter, and two 40 km arms that are 10 times longer than Advanced LIGO, offering the same amplitude of the observed signals but much lower noise, resulting in increases in the signal-to-noise ratio. Being built with all new technology based upon current LIGO technology, the observatory will offer increases in sensitivity and bandwidth of the instruments [30–32].

Estimating the improvement in observations via CE, we will simulate data as if it were detected by the observatory using both Bilby's innate CE configuration settings and our trained Dingo network, looking for events at much lower frequencies than LIGO has the means to detect. We will then simulate the same event for the current LIGO observatories and conduct parameter estimation on these events via all three of our methods, following the same format as explained previously. This will allow us to show the improvement of parameter estimation via the technological advancements of detectors and offer insight into the changes in efficiency and capability of each parameter estimation method due to different detectors.

Following the initial test, we inject the same parameters using the same waveform approximation model but using a different PSD for the detectors. Following the

same format as the initial test, we will visualize our important parameter and build a table using the same format as previously described. Next, we conduct comparisons between the different developed posteriors visually by constructing overlapping histograms of specific interesting parameters and constructing a table in the same format as before but with both estimation values and errors. We then numerically take both the numerical values and percent error values of accuracy and precision and compare the two PEs in totality. This process will be repeated for each possible combination of method and event *(Bilby-LIGO,DINGO-LIGO,Bilby-CE,etc.)*.

2. Exceptional Events

After preliminary data collection, model visualization, and numerical data comparison, we will simulate multiple high-mass, exceptional black hole merger events. After manually injecting a high mass and event into our waveform approximation, we will then use all three methods for PE. Using the same format as the initial preliminary comparison to avoid errors, we will conduct parameter estimation using each method, visualize specific parameters, construct tables of distribution values, repeat for varying PSD scenarios, visualize and compare the different PEs, and finally numerically compare the different PEs. Repeating for all possible combinations, ensuring accurate data collection.

G. Further comparisons between models and detectors

Using injected data, we conduct a qualitative and quantitative comparative analysis of all three methods, looking at the methods together and delineating where the strengths and weaknesses lie within each. Given our preliminary comparisons and stress tests of models and detectors, we then follow the same format as the preliminary comparison for further analysis, but instead of individually looking at specific combinations of models and detectors and measuring the degree of difference in the approximation to that of the simulated distribution, as stated previously, we will combine analysis for all three methods of PE. We will construct histograms comparing the distributions formed from specific parameters from each method and model, construct tables in the same format delineating accuracy, precision, and speed, numerically define the overall effectiveness of each method by using a comparison of the values in the table, and use said data to offer predictions on how this effectiveness may shift in the future.

V. DATA ANALYSIS

A. Preliminary Tests

All preliminary tests are developed using the injected parameters shown in Table III.

TABLE III. Injected Parameters for Preliminary Tests

Symbol	Test 1	Test 2	Test 5	Test 4	Test 3
m_1	36.0	40	40	85.0	85.0
m_2	29.0	25	25	70.0	70.0
q	0.805	0.625	0.625	0.823	0.823
$\mathcal{M}$	28	27.38	27.38	67.1	67.1
d_L	1000.0	4000.0	4000.0	4000.0	1000.0
δ	-1.2	0.6	0.6	0.6	0.6
α	1.375	4.307	4.307	4.307	4.307
θ_{jn}	0.4	2.0	2.0	0.6	2.0
ψ	2.659	3.0	3.0	3.0	3.0
ϕ	1.3	4.5	4.5	4.5	4.5
a_1	0.4	0.4	0.4	0.8	0.45
a_2	0.3	0.5	0.5	0.2	0.5
θ_1	0.5	2.7	2.7	2.7	2.5
θ_2	1.0	3.8	3.8	1.5	2.5
ϕ_{12}	1.7	5.5	5.5	5.5	5.5
ϕ_{jl}	0.3	5.2	5.2	3.892	5.3
t_c	1126259642.413	-	-	-	-

1. Bilby Tests

(currently in development, will be finished in future updates)

2. Dingo Tests

(currently in development, will be finished in future updates)

3. Fisher Matrix Test

Using the injections listed in Table III, we will initially address our impressions of the method before discussing the results. Although GWfish is the best option for Fisher matrix parameter estimation (PE) for our study, it still has some inadequacies. One such inadequacy, previously mentioned in Section IV, is GWfish's lack of innate implementation for incorporating priors in constructing distributions. This limitation significantly hindered the speed at which these tests could be conducted, as comparing and understanding results from incorporating priors and the base GWfish was necessary to gauge the overall efficiency of the method. Despite this issue, GWfish does offer a more streamlined process for Fisher matrix construction and distribution development compared to other software. This allowed for relatively

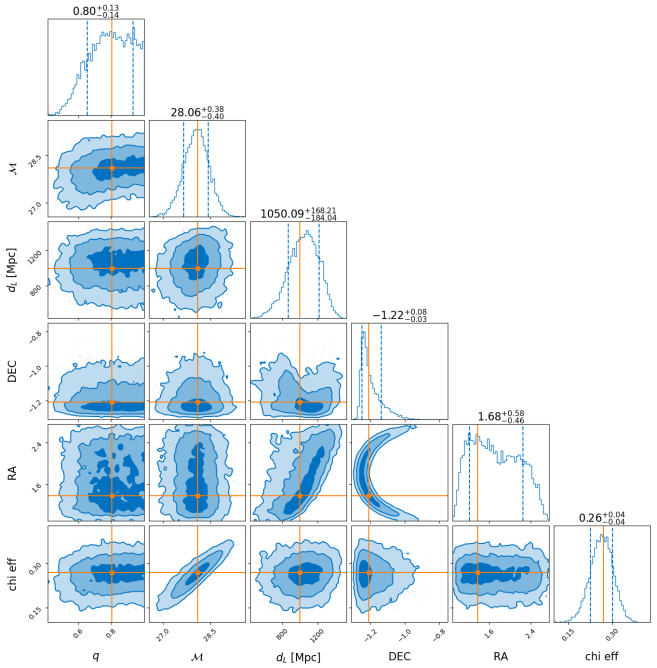

FIG. 9. Visualisation of $\mathcal{M}$, q, d_L, α, δ, $\chi_{\rm eff}$ Bilby PPDs using LIGO 04 PSD prediction and injection data from Test 1.

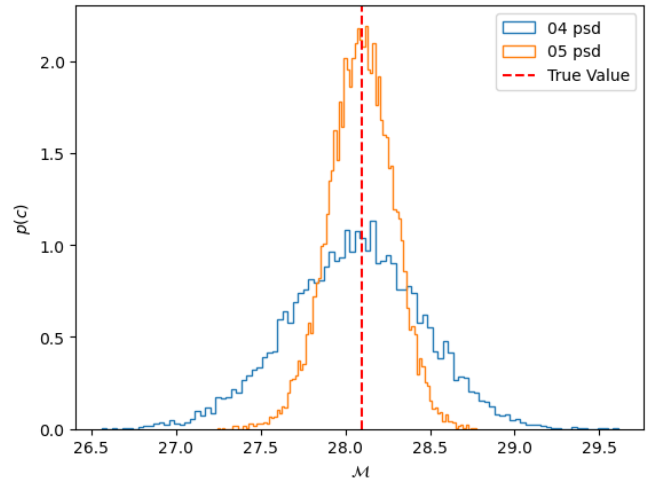

FIG. 10. Comparison of chirp mass $\mathcal{M}$ PPDs for o4 *(blue)* and o5 *(orange)* PSD predictions. Constructed using Bilby and injection parameters from Test 1.

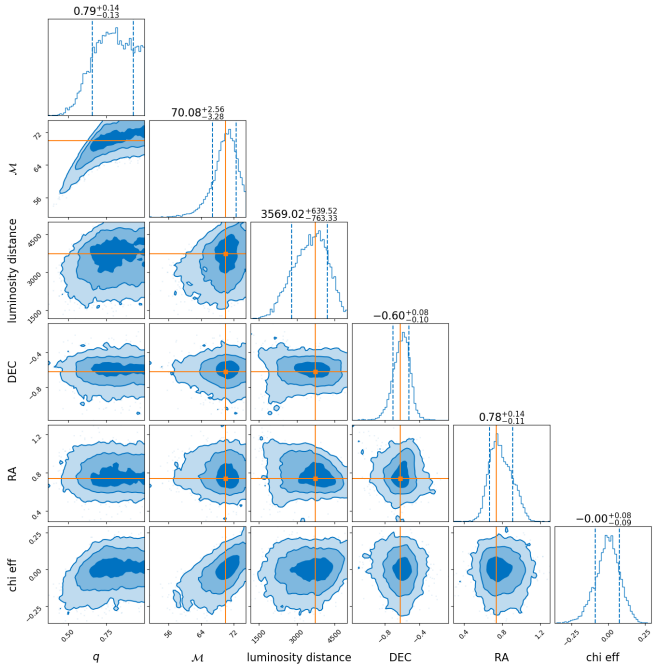

FIG. 11. Visualisation of $\mathcal{M}$, q, d_L, α, δ, $\chi_{\rm eff}$ Bilby PPDs using LIGO 05 PSD prediction and injection data from Test 2.

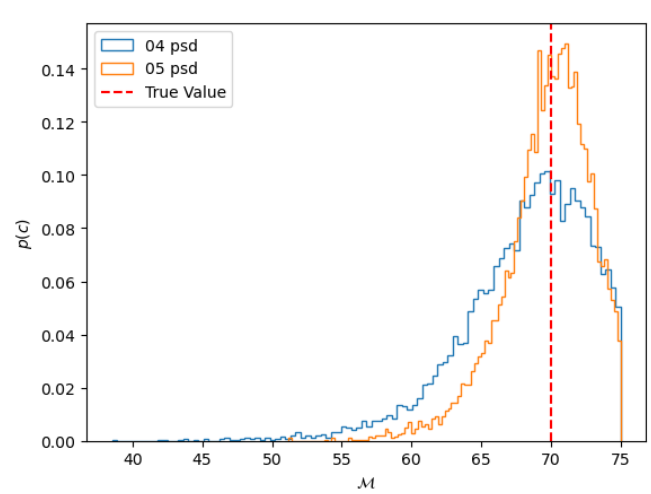

FIG. 12. Comparison of chirp mass $\mathcal{M}$ PPDs for o4 *(blue)* and o5 *(orange)* PSD predictions. Constructed using Bilby and injection parameters from Test 2.

easy and straightforward manual adaptation of the software in comparison with other Fisher matrix software.

Following these initial impressions, the preliminary tests showed the following results using current GW detectors (LIGO):

a. Current LIGO Detectors Tests: **Speed:** Without incorporating priors, using the five stated injections in Table III, GWfish took approximately 23 ± 4 seconds to construct distributions after each injection. With the incorporation of priors, due to the need to resample with new weights, the speed of GWfish posterior construction decreased to 32 ± 9 seconds. Although accuracy and precision statistics were not taken for GWfish without priors as the results are non-physical, it is important to provide an example of the average time that base GWfish takes to compute distributions. This serves as a foundational method for comparison with our implementation of priors.

Accuracy: Overall, averaging the results listed in Table IV across all parameters, it was shown that for

the preliminary test, GWfish had a percent error of 19.60% ± 44.09%, meaning that on average the parameters were off by approximately 20% with a wide range of varying accuracies. Among the parameters, the right ascension (α) had the highest accuracy (1.6%) and the lowest spread of accuracies (1.7%). Initially, the overall accuracy seems promising, but given the high variation in accuracies, indicated by a standard deviation of 18.26%, this highlights the overall lack of reliability in the Fisher matrix approach with modern detectors.

Precision: The overall precision of GWfish was a percent error of 20.00%±18.34%. Given that the Fisher matrix assumes Gaussian distributions, the deviation and spread from a Gaussian distribution using GWfish was expected. Additionally, it is noted that a Fisher matrix tends to underestimate error estimates, resulting in highly precise distributions. Given the relative newness of the software, it was expected that there would be some needed improvements before it is ready for full comparison. The results shown in the preliminary test, as presented in Table IV , suggest that adjustments in the use of priors and injection selection may be necessary to obtain meaningful results.

Parameter	Avg Accuracy ± Std Accuracy	Avg Precision ± Std Precision
m_1	0.83 (2.2%) ± 0.71 (2.4%)	3.2 (8.3%) ± 1.7 (6.2%)
m_2	1.4 (3.3%) ± 1.9 (5.0%)	4.4 (10.0%) ± 2.9 (8.3%)
d_L	560 (42%) ± 1100 (94%)	420 (14.0%) ± 570 (11.0%)
θ_{jn}	0.15 (19.0%) ± 0.16 (32.0%)	0.2 (18.0%) ± 0.09 (11.0%)
δ	0.039 (5.7%) ± 0.054 (9.4%)	0.093 (13.0%) ± 0.082 (20.0%)
α	0.056 (1.6%) ± 0.068 (1.7%)	0.11 (3.6%) ± 0.09 (3.4%)
ψ	0.15 (5.1%) ± 0.2 (7.2%)	0.19 (7.6%) ± 0.16 (8.5%)
ϕ	0.31 (11.0%) ± 0.38 (23.0%)	0.52 (15.0%) ± 0.36 (14.0%)
χ_1	0.13 (24.0%) ± 0.1 (21.0%)	0.18 (38.0%) ± 0.054 (17.0%)
χ_2	0.14 (52.0%) ± 0.11 (52.0%)	0.19 (47.0%) ± 0.071 (13.0%)
θ_1	0.49 (30.0%) ± 0.38 (39.0%)	0.4 (26.0%) ± 0.16 (15.0%)
θ_2	0.45 (21.0%) ± 0.39 (13.0%)	0.37 (22.0%) ± 0.17 (11.0%)
ϕ_{12}	1.0 (24.0%) ± 1.0 (25.0%)	1.1 (33.0%) ± 0.27 (19.0%)
ϕ_{jl}	0.24 (29.0%) ± 0.27 (77.0%)	0.6 (26.0%) ± 0.32 (23.0%)

TABLE IV. Table of accuracy and precision metrics of Fisher Matrix preliminary test for current LIGO detectors, showing the average values ± standard deviations, including percentage values, of all five test.

b. CE Tests: following the same methodology used for current detectors discussion for the prelimnary results, we now disucss the results and impressions for prelimnary fisher matrix CE tests.

Speed: The computational performance of GWfish when making detections with Cosmic Explorer (CE) is as follows: It took approximately 15 ± 3 seconds to construct distributions for injected signals without priors, and 25 ± 7 seconds when incorporating priors. The increase in computation time with CE could be attributed to various factors. One possibility is given that CE will produce higher-quality data with less noise, this could could simplify signal processing and analysis, thereby potentially reducing the computational time needed.

Accuracy: According to the results summarized in Table V, the average accuracy of GWfish using CE detections has a percent error of 10.36% ± 18.26%. This indicates that, on average, the distributions were off by 10%, with a narrower range of accuracies compared to current LIGO detectors. This improvement is evident in the accuracy metrics, showing a 9.24% enhancement and a 25.83% reduction in deviations when compared to current detectors. Notably, the parameter with the highest accuracy and lowest spread was the primary mass m_1, with 0.15(0.31%)±0.21(0.51%). These improvements are expected due to the advancements in detection technology with CE. However, if these trends remain consistent across further tests, the Fisher matrix approach appears to be a reliable source of parameter estimation when compared to Bilby and Dingo for third-generation detectors.

Parameter	Avg Accuracy ± Std Dev	Avg Precision ± Std Dev
m_1	0.15 (0.31%) ± 0.21 (0.51%)	0.68 (1.3%) ± 0.51 (1.2%)
m_2	0.10 (0.31%) ± 0.15 (0.58%)	0.46 (1.3%) ± 0.36 (1.3%)
d_L	870 (74%) ± 1500 (140%)	410 (13%) ± 550 (12%)
θ_{jn}	0.018 (2.6%) ± 0.032 (6.4%)	0.052 (5.9%) ± 0.049 (6.8%)
δ	0.20 (14%) ± 0.28 (41%)	0.11 (1.5%) ± 0.12 (26%)
α	0.24 (10%) ± 0.32 (20%)	0.14 (10%) ± 0.13 (22%)
ψ	0.35 (12%) ± 0.40 (15%)	0.24 (11%) ± 0.15 (8.5%)
ϕ	0.13 (3.4%) ± 0.16 (3.7%)	0.30 (11%) ± 0.23 (14%)
χ_1	0.019 (4.2%) ± 0.046 (12%)	0.034 (6.1%) ± 0.029 (6.0%)
χ_2	0.028 (10%) ± 0.041 (15%)	0.069 (22%) ± 0.051 (16%)
θ_1	0.025 (1.2%) ± 0.043 (1.7%)	0.070 (5.4%) ± 0.057 (7.6%)
θ_2	0.12 (5.4%) ± 0.25 (8.4%)	0.16 (9.5%) ± 0.18 (11%)
ϕ_{12}	0.037 (0.75%) ± 0.073 (1.3%)	0.25 (6.9%) ± 0.12 (6.0%)
ϕ_{jl}	0.056 (5.7%) ± 0.084 (19%)	0.13 (9.2%) ± 0.14 (14%)

TABLE V. Table of accuracy and precision metrics of Fisher Matrix preliminary test for future detectors (CE), showing the average values ± standard deviations, including percentage values, for all parameters.

A noteworthy issue arises with the accuracy of luminosity distance (d_L) and sky location parameters (δ, α), as shown in Table V. For these parameters, current detectors provided better estimates. The difficulty in accurately estimating sky locations with CE might be attributed to the fact that, despite its larger size, it remains a single-object detector, which could affect the triangulation of the signal. For d_L, the unexpected decrease in accuracy warrants further investigation to determine if this issue is consistent across different tests.

Precision: Overall, the average spread in the distribution was 8.14% ± 5.31%, with the most precise parameter distribution also being the primary mass m_1, with 0.68(1.3%) ± 0.51(1.2%). This represents a significant improvement compared to the precision of current detectors, which was 20.00% ± 18.34%. Given the higher signal-to-noise ratio of CE, it was anticipated that the distributions would be more confined and less spread out. These preliminary results are consistent with our expectations.

VI. CONCLUSIONS

(currently in development, will be finished in future updates)

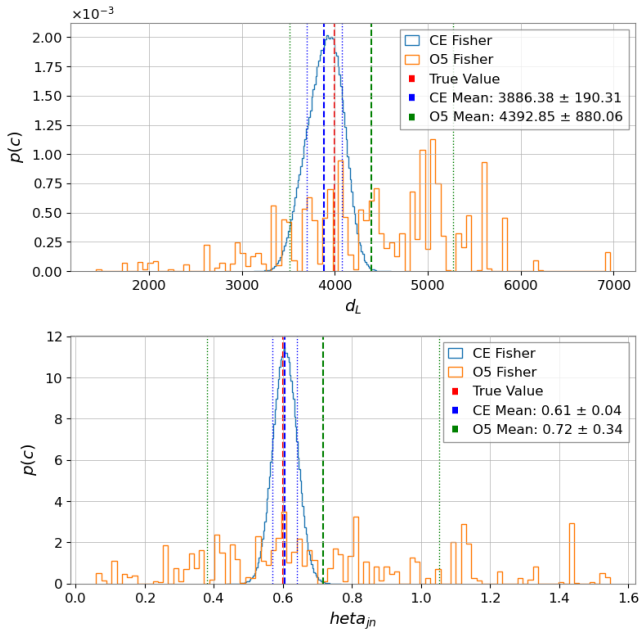

FIG. 13. GWfish comparison of detection networks, CE (blue) and Ligo (orange) for constructed posteriors for parameters for luminosity distance d_L and orbital orientation (θ_{JN}).

VII. FUTURE IMPLICATIONS

(currently in development, will be finished in future updates)

VIII. ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation Research Experience for Undergraduates (NSF REU) program, the LIGO Laboratory Summer Undergraduate Research Fellowship program (NSF LIGO), and the California Institute of Technology Student-Faculty Programs.

IX. APPENDIX A

[1] M. Spera, A. A. Trani, and M. Mencagli, *Compact Binary Coalescences: Astrophysical Processes and Lessons Learned*, Galaxies **10**, 10.3390/galaxies10040076 (2022).

[2] D. R. Czavalinga, T. Mitnyan, S. A. Rappaport, *et al.*, *New compact hierarchical triple system candidates identified using Gaia DR3*, Astronomy & amp; Astrophysics **670**, 10.1051/0004-6361/202245300 (2023).

[3] B. P. Abbott, R. Abbott, T. D. Abbott, *et al.*, *GW190425: Observation of a Compact Binary Coalescence with Total Mass ≈ 3.4 M⊙*, The Astrophysical Journal Letters **892**, L3 (2020).

[4] A. Fabio, T. Silvia, and A. S. Hamers, *Binary Black Hole Mergers from Field Triples: Properties, Rates, and the Impact of Stellar Evolution*, The Astrophysical Journal **841**, 77 (2017).

[5] The LIGO Scientific Collaboration, Virgo Collaboration, and the KAGRA Collaboration, *Observation of Gravitational Waves from the Coalescence of a 2.5 − 4.5 $M_\odot$ Compact Object and a Neutron Star* (2024), arXiv:2404.04248 [astro-ph.HE].

[6] https://catalog.cardiffgravity.org, Https://catalog.cardiffgravity.org/.

[7] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), *GW190521: A Binary Black Hole Merger with a Total Mass of 150 M⊙*, Physical Review Letters **125**, 10.1103/physrevlett.125.101102 (2020).

[8] Ng, Thomas C.K. Isi, Maximiliano, Wong, Kaze W.K. *et al.*, *Constraining gravitational wave amplitude birefringence with GWTC-3*, Physical Review D **108**, 10.1103/physrevd.108.084068 (2023).

[9] A. A. Trani, S. Rastello, *et al.*, *Compact object mergers in hierarchical triples from low-mass young star clusters*, Monthly Notices of the Royal Astronomical Society **511**, 1362 (2022), https://academic.oup.com/mnras/article-pdf/511/1/1362/42425280/stac122.pdf.

[10] B. P. Abbott, R. Abbott, T. D. Abbott, *et al.*, A guide to ligo–virgo detector noise and extraction of transient gravitational-wave signals, Classical and Quantum Gravity **37**, 055002 (2020).

[11] E. Thrane and C. Talbot, *An introduction to Bayesian inference in gravitational-wave astronomy: Parameter estimation, model selection, and hierarchical models*, Publications of the Astronomical Society of Australia **36**, 10.1017/pasa.2019.2 (2019).

[12] N. Jeffrey and B. D. Wandelt, Evidence networks: simple losses for fast, amortized, neural bayesian model comparison, Machine Learning: Science and Technology **5**, 015008 (2024).

[13] J. Veitch, V. Raymond, B. Farr, *et al.*, *Parameter estimation for compact binaries with ground-based gravitational-wave observations using the LALInference software library*, Phys. Rev. D **91**, 042003 (2015).

[14] C. M. Biwer, C. D. Capano, S. De, *et al.*, *PyCBC Inference: A Python-based Parameter Estimation Toolkit for Compact Binary Coalescence Signals*, Publications of the Astronomical Society of the Pacific **131**, 024503 (2019).

[15] G. Ashton, M. Hübner, P. D. Lasky, *et al.*, *Bilby: A User-friendly Bayesian Inference Library for Gravitational-wave Astronomy*, The Astrophysical Journal Supplement Series **241**, 27 (2019).

[16] M. Dax, S. R. Green, J. Gair, *et al.*, *Real-Time Gravitational Wave Science with Neural Posterior Estimation*, Physical Review Letters **127**, 10.1103/physrevlett.127.241103 (2021).

[17] C. Hoy and L. K. Nuttall, *BILBY in space: Bayesian inference for transient gravitational-wave signals observed with LISA* (2023), arXiv:2312.13039 [astro-ph.IM].

[18] K. Chandra, A. Pai, S. H. W. Leong, *et al.*, *Impact of Bayesian Priors on the Inferred Masses of Quasi-Circular Intermediate-Mass Black Hole Binaries* (2023), arXiv:2309.01683 [gr-qc].

[19] J. S. Speagle, dynesty: a dynamic nested sampling package for estimating bayesian posteriors and evidences, Monthly Notices of the Royal Astronomical Society **493**, 3132–3158 (2020).

[20] S. Morisaki, Accelerating parameter estimation of gravitational waves from compact binary coalescence using adaptive frequency resolutions, Physical Review D **104**, 10.1103/physrevd.104.044062 (2021).

[21] K. Krishna, A. Vijaykumar, A. Ganguly, C. Talbot, S. Biscoveanu, R. N. George, N. Williams, and A. Zimmerman, Accelerated parameter estimation in bilby with relative binning (2023), arXiv:2312.06009 [gr-qc].

[22] S. Morisaki, R. Smith, L. Tsukada, S. Sachdev, S. Stevenson, C. Talbot, and A. Zimmerman, Rapid localization and inference on compact binary coalescences with the advanced ligo-virgo-kagra gravitational-wave detector network (2023), arXiv:2307.13380 [gr-qc].

[23] D. Conor, B. Artur, M. Iain, *et al.*, *Neural Spline Flows* (2019), arXiv:1906.04032 [stat.ML].

[24] G. Papamakarios and I. Murray, *Fast ϵ-free Inference of Simulation Models with Bayesian Conditional Density Estimation* (2018), arXiv:1605.06376 [stat.ML].

[25] M. Vallisneri, *Use and abuse of the Fisher information matrix in the assessment of gravitational-wave parameter-estimation prospects*, Physical Review D **77**, 10.1103/physrevd.77.042001 (2008).

[26] U. Dupletsa, J. Harms, B. Banerjee, M. Branchesi, B. Goncharov, A. Maselli, A. Oliveira, S. Ronchini, and J. Tissino, gwfish: A simulation software to evaluate parameter-estimation capabilities of gravitational-wave detector networks, Astronomy and Computing **42**, 10.1016/j.ascom.2022.100671 (2023).

[27] F. Iacovelli, M. Mancarella, S. Foffa, and M. Maggiore, Gwfast: A fisher information matrix python code for third-generation gravitational-wave detectors, The Astrophysical Journal Supplement Series **263** (2022).

[28] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, JAX: composable transformations of Python+NumPy programs (2018).

[29] G. Pratten, C. García-Quirós, M. Colleoni, A. Ramos-Buades, H. Estellés, M. Mateu-Lucena, R. Jaume, M. Haney, D. Keitel, J. E. Thompson, and S. Husa, Computationally efficient models for the dominant and subdominant harmonic modes of precessing binary black holes (2021).

[30] R. Essick, S. Vitale, and M. Evans, *Frequency-dependent responses in third generation gravitational-wave detectors*, Phys. Rev. D **96**, 084004 (2017).

[31] S. Dwyer, D. Sigg, S. W. Ballmer, *et al.*, *Gravitational wave detector with cosmological reach*, Phys. Rev. D **91**, 082001 (2015).

[32] D. Reitze, R. Adhikari, S. Ballmer, *et al.*, *Cosmic Explorer: The U.S. Contribution to Gravitational-Wave Astronomy beyond LIGO* (2019), arXiv:1907.04833 [astro-ph.IM].

[33] J. Kwok, R. Lo, A. Weinstein, and T. Li, Investigation on the effects of non-gaussian noise transients and their mitigations on gravitational-wave tests of general relativity (2021).

[34] J. Roulet and T. Venumadhav, Inferring binary properties from gravitational-wave signals, Annual Review of Nuclear and Particle Science 10.1146/annurev-nucl-121423-100725 (2024).

[35] B. P. Abbott, A. Abel, C. Charles, *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), *GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo during the Second Part of the Third Observing Run*, Physical Review X **13**, 10.1103/physrevx.13.041039 (2023).