

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: July 17, 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 **IMR-PhenomPv2** 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 inspiraling, 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.

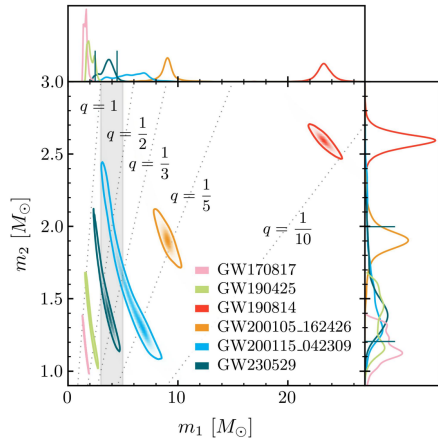


FIG. 1. Taken from [5]. The one- and two-dimensional posterior probability distributions for the component masses of the source binary of GW230529 (teal). The contours in the main panel denote the 90% credible regions with vertical and horizontal lines in the side panels denoting the 90% credible interval 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 $\text{FAR} < 0.25 \text{ yr} \approx 1$ (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.

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 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

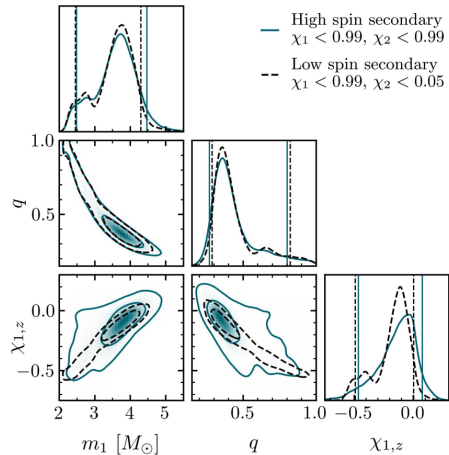


FIG. 2. Taken from [5]. Selected source properties of GW230529. The plot shows the one-dimensional (diagonal) and two-dimensional (off-diagonal) marginal posterior distributions for the primary mass m_1 , the mass ratio q , and the spin component parallel to the orbital angular momentum $\chi_{1,z} = \vec{\chi}_1 \cdot \hat{L}$. 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.

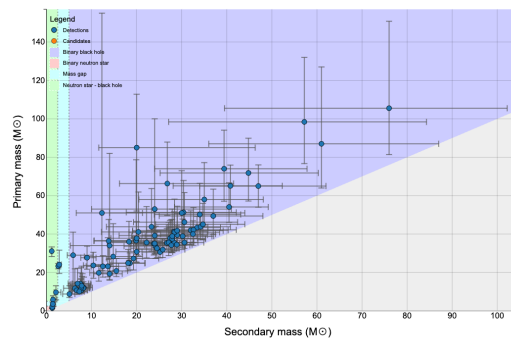


FIG. 3. Taken from [6]. Population distribution of confirmed CBC events via LIGO plotted by m_1 and m_2 .

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 accurately 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:

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

$p(\theta|d)$ 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 stationary and Gaussian, with zero mean. Since the data contains Gaussian noise: $d = h(t|\theta) + n$, the probability distribution is a Gaussian in data space. A Gaussian-noise likelihood function looks like is:

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

in which $\mu(\theta)$ is an example gravitational strain waveform template derived from a given θ and σ is the detector noise (power spectral density). $\pi(\theta)$ is distribution of the prior on θ , and $\mathcal{Z}$ is interpreted as the integral of the numerator or normalization factor called the ‘‘evidence’’ [10].

$$\mathcal{Z} = \int \mathcal{L}(d|\theta)\pi(\theta)d\theta. \quad (3)$$

Although not explored in this paper, the use of evidence $\mathcal{Z}$ can be used as a source of different model comparisons; this topic is explored in [11]. 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

$$p(\theta_i|d) = \int \left(\prod_{k \neq i} d(\theta_k) \right) p(\theta|d) \quad (4)$$

$$p(\theta_i|d) = \frac{\mathcal{L}(d|\theta_i)\pi(\theta_i)}{\mathcal{Z}}. \quad (5)$$

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

$$\mathcal{L}(d|\theta_i) = \int \left(\prod_{k \neq i} d(\theta_k) \right) \pi(\theta_k) \mathcal{L}(d|\theta). \quad (6)$$

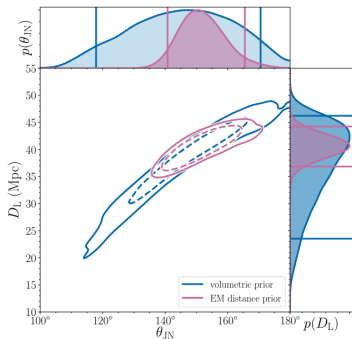


FIG. 4. Taken from [10]. The joint posterior for luminosity distance and inclination angle for GW170817 from (Abbott et al., 2019). The blue contours show the credible region obtained using gravitational wave data alone. The purple contours show the smaller credible region obtained by employing a relatively narrow prior on distance obtained with electromagnetic measurements. Publicly available posterior samples for this plot are available here: (LIGO/Virgo, LIGO/Virgo).

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 [12]; this code is over a decade old, and an updated python-toolkit version of the code has been developed [13]. Improvements of these methods, Bilby [14] and Dingo [15] 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 [16, 17], 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 [14]. It serves as not only a more versatile method of Bayesian inference compared to previous methods but also boasts improvements in efficiency, speed, accuracy, and simplicity over its predecessors [16].

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* [18], 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* [19], *Relative Binning* [20], and the use of *reduced order quadratures (ROQs)* [21]. [21] and [20] delve into the latter topics more and offer a full explanation of the methods; furthermore, we shall focus on *Multibanding* in this paper and define it later in Section III as a result of its 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)* [22, 23], 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 [15]. 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 $d \sim p(d|\theta)$ [15].

2. Fisher Matrix

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

$$\Gamma_{\kappa\nu} = \left\langle \frac{\partial h}{(\partial \theta^\kappa)} \middle| \frac{\partial h}{(\partial \theta^\nu)} \right\rangle = -E \left[\frac{(\partial^2 \ln L)}{(\partial \theta^\kappa \partial \theta^\nu)} \right], \quad (7)$$

assumes the waveform measured in CBC events can be seen as a linearizable approximation of the parameters,

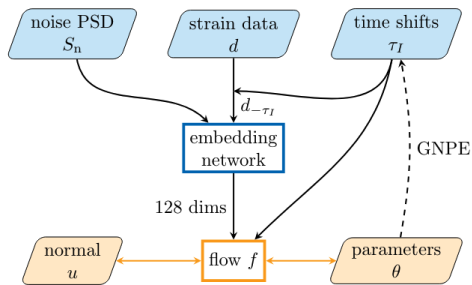


FIG. 5. Dingo flow chart taken from [15]. The posterior distribution is represented in terms of an invertible **normalizing flow** (orange), taking normally distributed random variables u into posterior samples θ . The flow itself depends on a (compressed) representation of the noise properties S_n and 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.

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* [14], *Dingo* [15]), 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 inference in parameter estimation but suffers greatly in accuracy and precision [24].

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

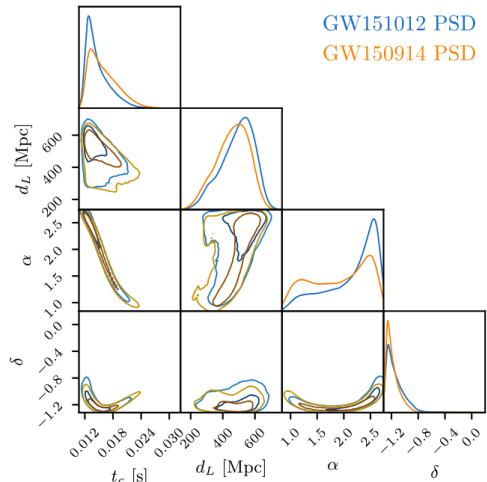


Figure 6. Comparison between DINGO evaluated on GW150914 using correct PSD as context, and using GW151012 PSD. These four parameters have a mean JSD of 0.020 nat.

FIG. 6. Taken from [15]. 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.

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 with accounting of 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** [25] 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 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. Gaussianity: 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 [15], 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 [26–28], 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 *IMRPhenomPv2* [29] waveform approximation to simulate our preliminary waveforms and then other *IMRPhenom* frequency-domain waveform approximations 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 [30], stems from the varying times for waveform construction between the approximation models. IMRPhenomPv2 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 *MultiBand-ing* 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

$$h(f) = \sum_b h_b(f). \quad (8)$$

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).

We refrain from using relative binning for its currently unclear improvement in generation speed and accuracy in the evaluation of the waveform. We also 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 has information and can make predictions about signals as if they were observed by CE, there is no trained Dingo network for such specifications. Understanding this, we may manually train a specific network to work and make PE behave as if the GW signal were observed by CE.

E. 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 *IMRPhenomPv2* 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, 12–17, 22, 23, 31]. 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 [26–28].

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 parame-

ters 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.

F. 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 3	Test 4
m_1	36.0	104.7	49.131	-
m_2	29.0	63	24.714	-
q	0.805	0.597	0.503	0.79
$\mathcal{M}$	28	70.0	29.985	63.443
d_L	1000.0	3736	4000.0	1939.171
δ	-1.2	-0.623	0.124	0.171
α	1.375	0.742	1.654	0.32
θ_{jn}	0.4	0.95	1.42	2.932
ψ	2.659	2.501	2.115	1.854
ϕ	1.3	1.401	5.986	3.058
a_1	0.4	0.815	0.195	0.99
a_2	0.3	0.126	0.812	0.99
τ_1	0.5	1.547	1.568	0.379
τ_2	1.0	0.396	1.203	2.31
ϕ_{12}	1.7	5.475	4.185	3.3
ϕ_{jl}	0.3	2.264	3.364	4.841
t_c	1126259642.413	0.0	-	-

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

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

VI. CONCLUSIONS

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

VII. FUTURE IMPLICATIONS

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

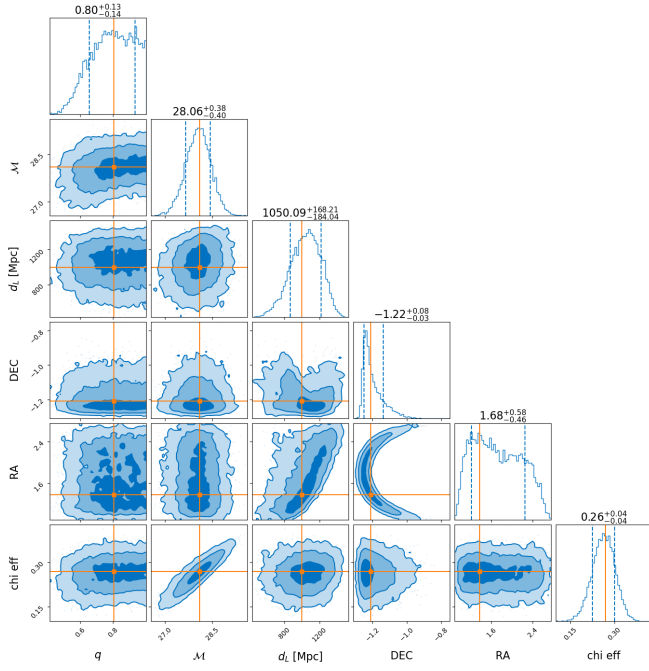


FIG. 7. Visualisation of $\mathcal{M}$, q , d_L , α , δ , χ_{eff} Bilby PPDs using LIGO 04 PSD prediction and injection data from Test 1.

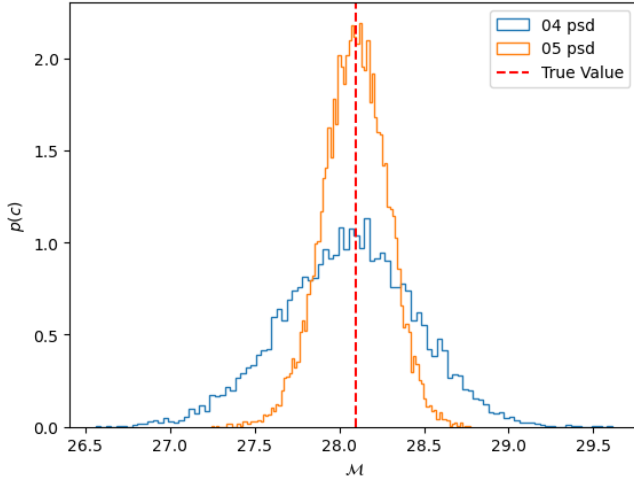


FIG. 8. Comparison of chirp mass $\mathcal{M}$ PPDs for o4 (blue) and o5 (orange) PSD predictions. Constructed using Bilby and injection parameters from Test 1.

VIII. ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation Research Experience for Undergraduates (NSF REU) program, the LIGO Laboratory Summer Un-

dergraduate Research Fellowship program (NSF LIGO), and the California Institute of Technology Student-Faculty Programs.

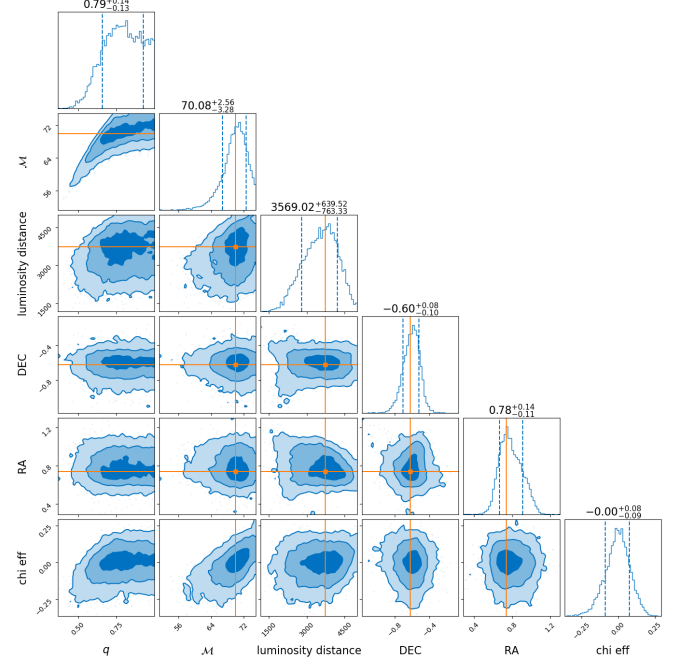


FIG. 9. Visualisation of $\mathcal{M}$, q , d_L , α , δ , χ_{eff} Bilby PPDs using LIGO 05 PSD prediction and injection data from Test 2.

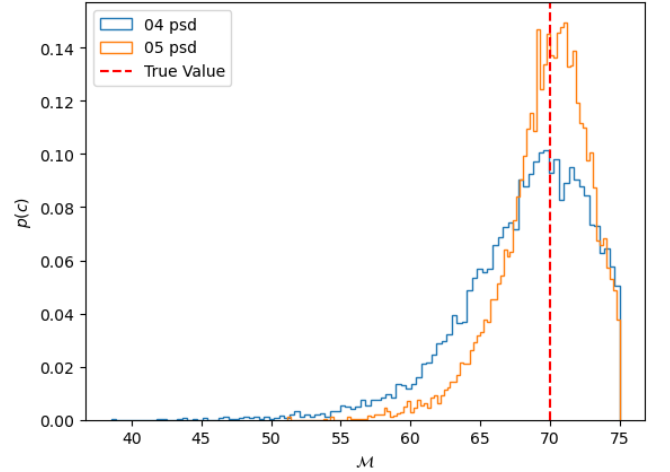


FIG. 10. Comparison of chirp mass $\mathcal{M}$ PPDs for o4 (blue) and o5 (orange) PSD predictions. Constructed using Bilby and injection parameters from Test 2.

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).

Parameter	Mean $\pm$ Std Dev	04Mean $\pm$ Std Dev	05	True Value	Accuracy 04	Precision 04	Accuracy 05	Precision 05
mass_1	36.623 $\pm$ 3.072	35.870 $\pm$ 2.374	36.0	36.0	0.623 (1.73%)	3.072 (8.39%)	0.130 (0.36%)	2.374 (6.62%)
mass_2	28.665 $\pm$ 2.378	29.235 $\pm$ 1.812	29.0	29.0	0.335 (1.16%)	2.378 (8.30%)	0.235 (0.81%)	1.812 (6.54%)
mass_ratio	0.793 $\pm$ 0.125	0.822 $\pm$ 0.104	0.805	0.805	0.021 (1.59%)	0.125 (15.71%)	0.017 (2.09%)	0.104 (12.67%)
chirp_mass	28.051 $\pm$ 0.397	28.089 $\pm$ 0.191	28.095557954604	28.095557954604	0.045 (0.16%)	0.397 (1.42%)	0.006 (0.02%)	0.191 (0.68%)
a_1	0.461 $\pm$ 0.241	0.434 $\pm$ 0.212	0.4	0.4	0.061 (15.34%)	0.241 (52.16%)	0.034 (8.58%)	0.212 (48.75%)
a_2	0.497 $\pm$ 0.267	0.439 $\pm$ 0.250	0.3	0.3	0.177 (58.84%)	0.267 (86.05%)	0.129 (42.84%)	0.250 (83.25%)
tilt_1	0.925 $\pm$ 0.554	0.822 $\pm$ 0.519	0.5	0.5	0.425 (85.05%)	0.554 (59.86%)	0.322 (64.44%)	0.519 (63.09%)
tilt_2	1.129 $\pm$ 0.651	1.102 $\pm$ 0.657	1.0	1.0	0.139 (12.88%)	0.651 (57.66%)	0.102 (10.17%)	0.657 (59.27%)
phi_12	3.142 $\pm$ 1.791	3.163 $\pm$ 1.823	1.7	1.7	1.442 (84.83%)	1.791 (57.01%)	1.463 (86.05%)	1.823 (57.64%)
phi_j	2.738 $\pm$ 2.066	2.374 $\pm$ 2.194	0.5	0.5	2.438 (812.69%)	2.066 (75.45%)	2.074 (691.25%)	2.194 (92.44%)
luminosity_distance	0.00347 $\pm$ 110.933	0.00333 $\pm$ 117.185	1000.0	1000.0	40.547 (4.05%)	170.933 (16.43%)	33.799 (3.38%)	117.185 (11.33%)
theta_jn	0.421 $\pm$ 0.249	0.340 $\pm$ 0.203	0.4	0.4	0.021 (5.31%)	0.249 (59.01%)	0.060 (14.94%)	0.203 (59.77%)
phase	3.140 $\pm$ 1.827	3.148 $\pm$ 1.818	1.3	1.3	1.868 (141.53%)	1.827 (58.02%)	1.848 (142.12%)	1.818 (57.66%)
ra	1.772 $\pm$ 0.454	1.519 $\pm$ 0.274	1.375	1.375	0.347 (55.26%)	0.454 (26.36%)	0.144 (10.46%)	0.274 (18.03%)
dec	-1.197 $\pm$ 0.062	-1.214 $\pm$ 0.040	-1.2108	-1.2108	0.014 (-1.15%)	0.062 (5.17%)	0.003 (-0.27%)	0.040 (-3.27%)
chi_off	0.283 $\pm$ 0.039	0.265 $\pm$ 0.019	0.355334615384615	0.355334615384615	0.082 (28.02%)	0.039 (14.12%)	0.060 (25.35%)	0.019 (7.32%)
psi	1.592 $\pm$ 0.906	1.591 $\pm$ 0.907	2.659	2.659	1.067 (40.12%)	0.906 (56.90%)	1.068 (40.17%)	0.907 (57.05%)

FIG. 11. Data table showing mean, standard deviation, accuracy, and precision of both 04 and 05 distributions of 1

- [2] D. R. Czaulanga, T. Mitnyan, S. A. Rappaport, *et al.*, *New compact hierarchical triple system candidates identified using Gaia DR3*, *Astronomy & 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 $\approx 3.4 M_{\odot}$* , *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 $M_{\odot}$* , *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] 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).
- [11] 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).
- [12] 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).
- [13] 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).
- [14] 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).
- [15] 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).
- [16] 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].
- [17] 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].
- [18] 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).
- [19] 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).
- [20] 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].
- [21] 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].
- [22] D. Conor, B. Artur, M. Iain, *et al.*, *Neural Spline Flows* (2019), arXiv:1906.04032 [stat.ML].
- [23] G. Papamakarios and I. Murray, *Fast ϵ -free Inference of Simulation Models with Bayesian Conditional Density Estimation* (2018), arXiv:1605.06376 [stat.ML].
- [24] 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).
- [25] 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).
- [26] R. Essick, S. Vitale, and M. Evans, *Frequency-dependent responses in third generation gravitational-wave detectors*, *Phys. Rev. D* **96**, 084004 (2017).
- [27] S. Dwyer, D. Sigg, S. W. Ballmer, *et al.*, *Gravitational wave detector with cosmological reach*, *Phys. Rev. D* **91**, 082001 (2015).
- [28] 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].
- [29] 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).
- [30] 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).
- [31] 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).