Gravitational-wave astronomy at the crossroads: from current to future detectors, from single events to populations

by

Viola De Renzis

A thesis submitted to the University of Milano-Bicocca for the degree of Doctor of Philosophy

Thesis advisor: Prof. Davide Gerosa

Department of Physics Giuseppe Occhialini
School Of Astronomy and Astrophysics
University of Milano-Bicocca

January 10, 2025

Thesis to be defended in January 2025 in front of a Board of Examiners composed by:

Prof. Stephen Green

Prof. Walter Del Pozzo

Prof. Alberto Sesana

Parameter Estimation for current and future gravitational-wave detectors

PhD thesis. University of Milano-Bicocca

This thesis has been typeset by LaTeX

Author's email: v.derenzis@campus.unimib.it

Dedicated to my two families,

one bound by blood,
the other chosen by heart.

Abstract

This PhD thesis explores the phenomenology of black-hole binary systems, with a particular focus on spin dynamics, instabilities as well as waveform systematic and population studies applied to the LIGO/Virgo catalog and to simulated gravitational-wave signals. It exploits a range of advanced statistical techniques, including Bayesian inference for parameter estimation, hierarchical analyses and model selection for both current and future gravitational-wave experiments.

The first part of this manuscript explores the impact of two subdominant spin effects in black-hole binaries and illustrates the prospects for their detection in upcoming observing runs. First, we investigate the ability of current interferometers, such as LIGO and Virgo, to detect the signature imprinted by two precessing spins on simulated data using a new formulation of the precession spin parameter χ_p. We show that the detection of two-spin effects is within reach of the current generation of gravitational-wave detectors. Second, we investigate the potential to detect a precessional instability that occurs in the so-called "up-down" black-hole binaries, where the primary (secondary) BH spin is aligned (anti-aligned) with the orbital angular momentum of the binary. Our simulations show that detecting this instability requires a high (but achievable) detector sensitivity, and that these sources must reside in specific regions of the parameter space where the instability can occur and become detectable.

The second part of the thesis moves beyond single-event analysis and focuses on population studies of merging black holes using third-generation detectors, such as the Einstein Telescope and Cosmic Explorer. Crucially, this is the first time that a population analysis with third-generation detectors has been conducted using the Fisher matrix approximation and allows for efficient forecasting of parameter uncertainties. This approach provides detailed prediction of how future detectors will improve our understanding of the mass, spin, and redshift distributions of black holes, providing critical insights into their astrophysical formation channels.

Contents

List of Figures

List of Tables

Chapter 1

Gravitational waves: the new frontier in astrophysics

Overview

Chapter 1 serves as a broad introduction of the field of gravitational wave astronomy, exploring the conceptual basis of General Relativity, the sources of gravitational waves, and the detection methods used by current observatories like LIGO, Virgo, and KAGRA. It also highlights the key results from the LIGO-Virgo-KAGRA analyses based on the third gravitational-wave transient catalog, and discusses the expectations for future observatories, such as the Einstein Telescope, Cosmic Explorer and the LISA space mission, in advancing our understanding of black holes and neutron stars.

1.1 A quick tour of General Relativity

The theory of General Relativity (GR), formulated by Albert Einstein in 1915, revolutionized our understanding of gravity [1]. Unlike the Newtonian framework, where gravity is a force acting at a distance, GR describes gravity as a manifestation

of the curvature of spacetime. Using a famous quote by the physicist John Archibald Wheeler, the essence of GR can be summarized as follows: *matter tells spacetime how to curve, and curved spacetime tells matter how to move* [2]. This relationship is mathematically encapsulated in Einstein's field equations, $G_{\mu\nu} \propto T_{\mu\nu}$, which describe how the distribution of matter and energy (through the stress-energy tensor $T_{\mu\nu}$) determines the curvature of spacetime (through the Einstein tensor $G_{\mu\nu}$) [3, 4, 5].

GR has been extensively tested over the years and is now regarded as the most comprehensive theory of gravity. One of the most compelling pieces of evidence supporting the validity of GR is the observational confirmation of the existence of black holes (BHs), regions of the spacetime where the gravitational pull is so intense that not even light can escape. Notably, the detection of gravitational waves (GWs) from a binary BH merger in 2015 (see Section 1.2.3 for further details) [6], along with the groundbreaking imaging of the supermassive BH shadows in the M87 galaxy in 2017 and Sagittarius A* in 2022 by the Event Horizon Telescope Collaboration [7, 8], are among the most direct confirmations of this revolutionary theory. On cosmological scales, GR represents the foundation of modern cosmology [9]. The Friedmann-Lemaître-Robertson-Walker metric is an exact solution of Einstein's field equations and describes an expanding, homogeneous, and isotropic universe [10, 11, 12, 13]. While cosmic expansion can also be described within the framework of Newtonian gravity under certain conditions [14] , the prediction of accelerated expansion in the late universe, confirmed by observations of distant supernovae in 1998 [15, 16], is one of the key implications of GR.

1.2 Gravitational Waves from compact binaries

1.2.1 Sources of gravitational waves

Just as electromagnetic waves carry information about the dynamics of charged systems through space, also changes in a mass-energy distribution should propagate as waves, conveying information about gravitational dynamics. Predicted by Einstein more than one century ago [17, 18, 19], GWs are perturbations in the fabric of the spacetime generated by the acceleration of massive objects and propagating at the speed of light. Unlike electromagnetic waves, which can radiate through monopole and dipole moments, GWs are emitted by systems with a time-varying quadrupole moment [20, 3]. The leading-order expression for the amplitude h of GWs, known as the quadrupole formula, is given by [2, 21]

$$h \sim \frac{2G}{c^4 r}\ddot{Q}_{ij}, \tag{1.1}$$

where $\ddot{Q}_{ij}$ is the second time derivative of the quadrupole moment tensor of the mass distribution, G is the gravitational constant, c is the speed of light, and r is the distance from the source to the observer. Eq. (1.1) describes the emission of GWs by a gravitating system evolving with time and shows that any form of energy and matter can be a source of GWs, as long as the second derivative of the quadrupole moment of the system is non-zero. The factor $2G/c^4 \sim 10^{-50}\mathrm{s}^2\ \mathrm{g}^{-1}\ \mathrm{cm}^{-1}$ affects the intensity of the emitted wave, explaining the typical weakness of GWs. In order to have detectable signals, GW sources need to be compact and with orbital velocity high enough to produce strong gravitational fields. Consequently, only the most violent astrophysical events in the Universe, such as the merger of extremely compact objects, can produce detectable signals [22]. The most compact objects known so far are BHs and neutron stars (NSs). When two compact objects orbit around each other, GWs are emitted carrying away energy and angular momentum from the system, and therefore causing

them to spiral towards each other down to the merger. Coalescing binary systems evolve through three stages: the inspiral, during which the gravitational fields and velocities are still relatively weak, the merger, which occurs when the two objects are close enough that they start to fuse, and finally the post-merger phase, when the newly formed object settles down to a stable state. In such systems, the orbital velocities of the objects can approach a significant fraction of the speed of light, especially in the final stages of the coalescence.

1.2.2 Gravitational-wave interferometers

The amplitude of GWs from merging compact binaries is typically of the order of $h \sim 10^{-21}$ or lower, which has posed a significant challenge for their discovery. Detecting GWs requires highly sensitive instrument known as laser interferometers, instruments with two perpendicular arms with proper length L and two suspended mirrors placed at the end of each arm [23, 22, 24, 25, 26, 27]. A laser beam is splitted and the two parts are sent down the perpendicular arms, then reflected off the mirrors and recombined. When a GW propagate, it stretches and compresses spacetime, causing the proper distance of these arms to change by an amount $\Delta L \sim hL$ and altering the interference pattern of the recombined laser beams that can be measured with high accuracy. If we assume an interferometer with 4-kilometer-long arms, the relative length change that those instrument need to measure is $\Delta L \sim 10^{-18}$ m (which is about one thousandth the size of a proton). Detecting such a tiny variation would require interferometer arms about 15-20 km long. However, with advanced techniques like power recycling mirrors and Fabry-Pérot cavities, the effective optical path length is significantly increased, allowing interferometers to achieve the required sensitivity to detect GWs with arms that are "only" a few kilometers long [28, 29, 30, 31].

The development of ground-based interferometric detectors dates back to the 1980s [32, 33, 34, 35], and since then, their sensitivity has been continuously improved,

addressing the main sources of noise, which are seismic noise, quantum noise, and thermal noise [36, 37, 38]. The currently operating detectors are (i) two LIGO (*Laser Interferometer Gravitational-wave Observatory*) interferometers built in the U.S. [35, 39], (ii) Virgo in Italy [40, 41] and (iii) KAGRA (*Kamioka Gravitational Wave Detector*) in Japan [42]. These detectors, with arm lengths of 3-4 km, operate in the frequency range of 10 Hz to a few kHz, making them sensitive to the mergers of stellar-mass BHs (around 10–100 $M_\odot$). Simultaneous detections of signals by multiple interferometers significantly enhance both the confidence of a detection and the precision of the sky localization and polarization measurements [43, 44].

In the near future, new interferometers will significantly expand our observational capabilities. The Einstein Telescope (ET) in Europe [45] and the Cosmic Explorer (CE) in the U.S. [46], that will become operational in $\sim$ 20 years, will operate in a wider frequency range, down to 2-5 Hz, and will be able to observe heavier BHs (up to several thousand $M_\odot$) and earlier stages of binary inspirals. These detectors will improve sensitivity by an order of magnitude compared to current ones, reducing quantum and seismic noise with improved cryogenic systems and underground facilities.

LISA (Laser Interferometer Space Antenna), a space-based detector set to launch in the 2030s, will observe GWs in the millihertz range (0.1 mHz to 1 Hz) [47, 48, 49], making it capable of detecting supermassive BH mergers (in the range of $10^4 - 10^7$ $M_\odot$).

Pulsar Timing Arrays (PTA) operate at nanohertz frequencies, allowing them to detect signals from supermassive BH binaries and background GWs. Recently, PTAs (including the European PTA [50] and NANOGrav [51]) announced potential evidence of a stochastic GW background, a significant breakthrough that may help study the evolution of massive BHs across cosmic time [52, 53, 54, 55].

1.2.3 Where we stand now and where we are going

In September 2015, the Advanced LIGO [39] began its first scientific observing run (O1). Shortly after, on September 14, 2015, the two LIGO interferometers simultaneously observed the first gravitational event, GW150914, starting a new era of GW astronomy [6, 56, 57, 58, 59, 60, 61]. This event matched the waveform predicted by GR for the inspiral and merger of a binary BH with component masses of about 36 and 29 $M_{\odot}$. Following the success of O1, LIGO and Virgo conducted their second observing run (O2) from November 2016 to August 2017 [62, 63, 64]. Several additional detections were made, including the landmark detection of the first binary NS merger, GW170817, which was accompanied by an electromagnetic counterpart and confirmed the connection between NS mergers and short gamma-ray bursts [65, 66, 67, 68, 69].

In the third observing run (O3), which spanned from April 2019 to March 2020, a variety of remarkable events were observed [70]. Among them, GW190521 stood out as the most massive binary BH merger detected, with a remnant mass of approximately $142M_{\odot}$, providing direct evidence for the existence of intermediate-mass BHs [71, 72, 73]. Another significant event, GW190814, exhibited the most unequal mass ratio observed to date. The primary BH had a mass of $23.3M_{\odot}$, while the secondary object, with a mass of $2.6M_{\odot}$, could be either the lightest BH or the heaviest NS ever observed, falling within the lower mass gap [74, 75, 76].

Currently, the fourth observing run (O4), which started in May 2023, is ongoing and is expected to conclude in June 2025 [77, 78]. The third gravitational-wave transient catalog (GWTC-3) released in November 2023 now contains $\sim$ 90 events produced by the coalescence of BHs and NSs. Over the coming years, many more detections of compact binary coalescences are expected for the fourth (O4) and fifth (O5) runs. New observations will be made possible by an increased sensitivity of currently operating GW detectors as well as the advent of new ground- and space- based facilities, such as the ET and the LISA space mission. These detectors will reach a level of sensitivity so

impressive that it will be possible not only to test GR in a genuine strong field regime (allowing to probe regions closer and closer to the event horizon of a BH) but also to measure with an extremely high accuracy the waveform parameters, thus placing strong constraints on the underlying astrophysical formation processes.

1.2.4 Gravitational-wave signal extraction and data analysis

The output of a GW detector, d, can be written as a linear combination of the GW signal, $h(\theta)$, and the noise, n, namely $d = h(\theta) + n$. The noise is usually assumed to be stationary and Gaussian and since it is not known, it is estimated by the power spectral density. The astrophysical contribution is encapsulated inside the parameters of the signal, θ, that characterize compact binary mergers (such as their masses, spins, sky localization or distance from the observer). Although great efforts are carried out in order to reduce the sources of noise affecting detectors, the GW signal is still buried in noisy data and needs to be extracted with efficient methods [79, 80]. This is typically done using a technique called matched filtering, where a discrete bank of theoretical waveforms (templates) covering the relevant parameter space θ is compared against the data to find signals that may originate from binaries with a wide range of masses and spins [81]. Triggers are generated when a signal matches a template above a certain signal-to-noise ratio (SNR) threshold. The highest SNR trigger is then selected for further analysis. Once a potential signal is found, the false alarm rate (FAR) is calculated to estimate how likely it is that the signal could be caused by noise fluctuations.

Accurate waveform modeling is crucial for GW detection. One key challenge is that the shape of the signal, $h(\theta)$, cannot be exactly predicted due to the nonlinearity of Einstein's field equations, requiring waveform approximations. Different approaches are used depending on the evolution stage of the binary. During the inspiral, when the binary components are far apart and the orbital velocities are small, Post-Newtonian

(PN) methods provide an analytical approximation of the emitted signal [82, 83]. For the late inspiral and merger, where strong gravity dominates, numerical relativity (NR) simulations are necessary, though computationally expensive [84, 85]. Complete waveforms can be constructed by combining PN results for the early inspiral with NR simulations for the later stages, including merger and post-merger. The two main approaches used to fill the gap between PN approximations and NR simulations are Effective One Body (EOB) methods [83, 86] and phenomenological waveform models [87, 88, 89]. The first combine PN expansions with non-perturbative techniques to model the full evolution of a compact binary system in a semi-analytical framework. EOB models are physically motivated but can be computationally intensive. On the other hand, phenomenological (*Phenom*) waveforms offer a purely phenomenological fit to NR data with faster generation times, making them more suitable for extensive parameter estimation. Additionally, NR surrogate models [90, 91] interpolate between existing NR simulations to produce accurate waveform approximations with much lower computational costs, effectively combining the precision of NR with the speed of analytical models. A key drawback of NR surrogate models is that their accuracy is limited by the density and frequency range of the available NR simulations they are built from, as well as the extent of the parameter space these simulations encompass.

After the identification of a signal through GW searches, the next step is to extract the parameters θ that characterize coalescing binary systems from the data. Parameter estimation is typically performed using Bayesian inference that aims at reconstructing the posterior distribution of the parameters θ, given the data and the a-prior knowledge of the expected astrophysical sources [92, 93, 94]. As the number of GW detections grows, it becomes increasingly valuable to analyze all events collectively to study the population properties of binary BHs and NSs (e.g. the BH mass spectrum or the NS equation of state). Hierarchical Bayesian inference has become the standard method for such population studies, targeting the posterior distribution of the "hyperparameters" (usually denoted ad λ), which characterize the underlying population from which individ-

ual events are drawn [95, 94, 96]. Many codes based on Bayesian inference are currently available to perform both single-event and population-level parameter estimation (e.g. LALSimulation [97], Bilby [98, 99], Dingo [100], PyCBC [101], RIFT [102]). Despite their high reliability, Bayesian inference methods are computationally very expensive and they will become prohibitive as we move toward the era of third-generation (3G) detectors that will observe thousands of sources every year, many of with high SNR. At the moment, the only feasible way to make forecasts for future observatories relies on the Fisher matrix approximation, which provides a faster, though less precise, method for estimating parameters in this high SNR regime [103]. As we will discuss in Chapter 4, various single-event parameter-estimation codes are now available (e.g. GWFast [104], GWBench [105], GWFish [106]) but there is a lack of tool to make forecasts about the population properties that will be observed with 3G detectors. There is currently an effort in performing single-event parameter estimation for 3G detectors with Bayesian inference methods, using techniques like relative binning –used also for current detectors– to enhance computational efficiency while preserving accuracy [107].

1.3 Astrophysical black holes

BHs of astrophysical origin span a wide range of masses that is determined by their formation mechanism and their evolutionary path [108, 109]. Current studies suggest that astrophysical BHs exists at least in two main categories:

- *stellar mass BHs* (from few up to hundreds of solar masses) are typically formed through the gravitational collapse of massive stars or in the coalescence or accretion-driven processes in binary systems [110, 111, 112, 6]. They have been indirectly observed through the X-ray emission from accretion disks and through GW emission, radial velocity measurements, and microlensing effects [113]. A binary system composed of two stellar mass BHs can form through two main processes:

(i) dynamical encounters in dense stellar clusters, or (ii) the evolution of isolated massive binary stars [109, 114, 115, 109]. In the isolated scenario, the formation is driven by stellar-evolution processes such as mass transfer, tidal interactions, and stellar winds, without external gravitational influences. In contrast, the dynamical pathway requires dense environments like globular or nuclear star clusters, where repeated gravitational interactions perturb BH orbits. The orientation of BH spins can help distinguish these formation channels, among other properties [116]. Dynamically formed binary BHs tend to have isotropic spin orientations, often leading to misaligned spins [117, 118, 119, 120]. In isolated binaries, instead, the common birthplace is expected to imprint a preferential alignment of the BH spins of the orbital angular momentum.

- *supermassive BHs* with masses $10^4 M_\odot < m < 10^{11} M_\odot$ can be found at the center of most galaxies and are expected to be produced as the result of galaxy mergers and accretion [121, 122]. So far, supermassive BHs have been observed indirectly by studying the motion of stars in close orbit around them (as in the case of the supermassive BH Sgr A^* in the center of the Milky Way) or by measuring the X-ray emission from accreting supermassive compact objects [123, 124, 125]. The mechanisms that lead to the formation of supermassive binary BHs are still not well understood. A complication is due to the so-called *final parsec problem*, which occurs when the binary inspiral slows as the BHs approach within a parsec, requiring additional mechanisms such as stellar scattering or gas interactions to drive the final merger [126]. Detecting GWs emitted during the merger of supermassive BHs is one of the main science goals of the LISA space mission.

There are both observational and theoretical indications that intermediate-mass BHs, which fill the gap between stellar-mass and supermassive BHs, do exist. The most compelling evidence so far comes from the detection of GW190521 (see Section 1.2.3). The advent of 3G detectors will enable observations at lower frequencies and detect BHs

with higher masses, which will greatly improve our understanding of intermediate-mass BHs. Additionally, it has been suggested that primordial BHs, with masses below that of the Sun, could have formed in the early universe. However, despite ongoing efforts to detect them, their existence remains purely speculative for now.

1.3.1 Population-level results from LVK analyses

In Section 1.2.4, we anticipated the importance of population-level analysis in refining theories of compact object formation and evolution. This section summarizes the key findings from the LIGO-Virgo-KAGRA (LVK) population analysis based on the GWTC-3 catalog, focusing on the mass, spin, and redshift distributions of BHs [64, 127, 128].

The BH mass spectrum spans from a few solar masses up to over $100M_{\odot}$, with a power-law trend favoring lower masses. The mass distribution shows peaks at $\sim 10M_{\odot}$ and $\sim 35M_{\odot}$. The first is likely due to isolated binary evolution, as models often predict a concentration near $10M_{\odot}$ [129, 130, 131]. Dynamical formation in young clusters is less favored to explain the peak at $10M_{\odot}$ because lighter BHs are typically –but not necessarily– ejected by supernova kicks [132, 133]. The second peak, at $\sim 35M_{\odot}$, is consistent with the formation of stellar-mass BHs via supernova explosions.Notably, LVK observations provide evidence supporting the presence of a mass gap between $50-120M_{\odot}$, which is consistent with predictions from stellar evolution models involving pair-instability supernovae as a possible mechanism. The absence of a sharp cutoff at higher masses suggests a possible dynamical origin, such as hierarchical mergers of BHs in dense stellar clusters.

The majority of the systems display low or moderate spin magnitudes, with the spins often misaligned relative to the orbital angular momentum of the binary. This suggests that a significant fraction of these mergers could be the result of dynamical formation in dense stellar environments. On the other hand, some events, such as GW151226 and GW190517, exhibit higher spins aligned with the orbital plane, indicating that

isolated binary evolution may also contribute to the observed population of BH mergers. Observations at the population level further suggest a preference for aligned spins.

In terms of redshift distribution, most GW detections occur at low redshifts ($z < 1$), corresponding to relatively recent cosmic history when massive stars would eventually collapse into BHs. However, binary BHs could have formed at earlier times, with their mergers being delayed due to evolutionary processes such as interactions in dense stellar environments or the time required for their orbits to decay. With the advent of 3G detectors, it is expected that more BH mergers at higher redshifts will be detected, allowing a better understanding of BH formation in the metal-poor, high-redshift universe.

1.4 Summary of my PhD activites

The main body of this thesis can be conceptually divided in two parts. The first one, contained in Chapter 2 and Chapter 3, is related to the exploration of two subdominant spin effects in both LVK data and simulated signals. In particular, I will illustrate the prospects for detecting (i) the signature imprinted by two precessing spins on GW signals coming from the coalescence of BH binaries using a carefully designed estimator and (ii) a precessional instability that occurs in unstable BH binaries which formed with aligned spins but enter the detector sensitivity window in precessing configurations. The material of these two chapters is based on Refs [134] and [135]. The main focus of the second part of this thesis is shifted from single-event parameter estimation to population-level analysis and from current detector to 3G interferometers. In Chapter 4, I present the first forecasts of the measurability of the population properties of binary BHs that will be observed in the future by ET and CE. The reference for this Chapter is [136].

During the course of my PhD, I also contributed to four additional projects. The first three that I will describe are related to precession and eccentricity in BH binaries.

In Ref. [137], we investigated the detectability of precession and nutation in BH binary mergers using GW data and injected signals. We found no evidence for those weak spin effects in current data but future observations (maybe already in O4) could impose stronger constraints on spin dynamics. In Ref. [138], we implemented the version 2 of the PRECESSION code, a python module useful to study the spin dynamics of precessing BH binaries using a multi-timescale approach. Ref. [139] highlight the relevance of residual eccentricity (below the current detection threshold of LVK) in reconstructing the formation history of the binary BHs that we observe. Finally, I collaborated to the work in Ref [140], in which we look for the optimal strategy to identify the two objects that compose a binary system. While the standard approach labels the objects according to their masses, we propose a new data-driven strategy for the identification of object "1" and object "2". This strategy is based on a semi-supervised machine learning algorithm known as spectral clustering [141]. The results that we obtain are relevant since with this new method, spin measurements are way more accurate if compared to the standard mass-labeling approach, the posterior distributions of masses and spins become closer to a Gaussian distribution and multimodalities tend to disappear.

During the three year of my PhD, I had the opportunity to supervise four Bachelor and one Master students. With two of them I started working on the optimal labeling problem which culminated in the publication of Ref [141]. Another research project focused on the up-down instability, a topic I will discuss in detail in Chapter 3. This student was included as a co-author of Ref. [135] for his contribution. In a separate project I supervised, we applied Bayesian model selection to the binary BHs contained in GWTC-3 to assess whether the events favored the zero-spin or isotropic spin orientation hypothesis. Lastly, in the final work I supervised, we investigated whether statistical or systematic errors are the primary source of inaccuracies in recovering the spins of highly rotating BHs.

Chapter 2

Characterization of merging black holes with two precessing spins

Overview

Spin precession in merging BH binaries is a treasure trove for both astrophysics and fundamental physics. There are now well-established strategies to infer from GW data whether at least one of the two BHs is precessing. In this thesis we tackle the next-in-line target, namely the statistical assessment that the observed system has two precessing spins. We find that the recently developed generalization of the effective precession spin parameter $\chi_{\rm p}$ is a well-suited estimator to this task. With this estimator, the occurrence of two precessing spins is a necessary (though not sufficient) condition to obtain values $1 < \chi_{\rm p} \leq 2$. Confident measurements of GW sources with $\chi_{\rm p}$ values in this range can be taken as a conservative assessment that the binary presents two precessing spins. We investigate this argument using a large set of > 100 software injections assuming anticipated LIGO/Virgo sensitivities for the ongoing fourth observing run, O4. Our results are very encouraging, suggesting that, if such binaries exist in nature

and merge at a sufficient rate, current interferometers are likely to deliver the first confident detection of merging BHs with two precessing spins. We investigate prior effects and waveform systematics and, though these need to be better investigated, do not find any confident false-positive case among all the configurations we tested. Our assessment should thus be taken as conservative. The material of this Chapter is based on Ref. [134].

2.1 Introduction

BH binary spin precession is a key feature of the relativistic two-body problem [142, 143]. Spin-spin and spin-orbit couplings in GR cause the orbital angular momentum $\mathbf{L}$ and the BH spins $\mathbf{S}_{1,2}$ to jointly precess about the direction of the total angular momentum $\boldsymbol{J} = \boldsymbol{L} + \mathbf{S}_1 + \mathbf{S}_2$. This motion induces modulations to both the amplitude and the phase of the emitted gravitational waves (GWs).

Measurements of spin precession have important repercussions in both astrophysics and fundamental physics. For the stellar-mass BH binaries observed by LIGO and Virgo [35, 41], spin precession provides unique leverage to discriminate between BH binaries formed in isolation and those assembled dynamically in stellar clusters [117, 118, 119, 120]. For the supermassive BH binaries targeted by LISA [47], spin measurements will provide information on, e.g., the occurrence of prolonged phases of disk accretion [144, 145]. GW observations of precessing binary BHs also allow us to constrain modified theories of gravity, especially those with parity-violating interactions caused by additional fields [146].

While the masses of LIGO/Virgo events are usually well measured, spin effects provide a subdominant contribution to the emitted radiation and are thus considerably more challenging to characterize. At present, an unambiguous measurement of BH-binary spin precession is one of the holy grails of observational GW astronomy.

Data from the first three observing runs of the LIGO/Virgo network have provided

some evidence for individual BH mergers with highly precessing spins [62, 63, 147, 148]. The most suggestive indication is that from GW200129_065458, where Refs. [149, 150] found strong evidence for BH-binary spin precession, while Ref. [151] raised potential issues in the glitch mitigation analysis. For the case of GW190521, a strong precession signature was also reported [63], though potential degeneracies with the eccentricity still need to be fully understood [152]. Collective evidence for spin precession was reported in the context of BH binary populations, with all current fits requiring some misaligned spins at high confidence [153, 154, 155].

Upcoming instrumental upgrades to the LIGO/Virgo (and hopefully KAGRA) network [44] are posed to provide increasingly accurate spin measurements. It is therefore not unreasonable to predict that the next observing run will deliver a confident, unambiguous identification of BH-binary spin precession. Crucially, measuring orbital-plane precession corresponds to inferring that at least one of the two BHs has a misaligned spin [142]. Inferring the presence of two misaligned spins requires extracting even feebler signatures from the signal, which are related to spin-spin (as opposed to spin-orbit) terms in the BH binary equations of motion.

This chapter of the thesis tackles such a next-in-line target. We perform > 100 software injections with realistic LIGO/Virgo sensitivity and demonstrate that signals with large-but-not-extreme signal-to-noise ratio (SNR) $\gtrsim 20$ allow us to detect two-spin effects *already in the next LIGO/Virgo observing run* (O4). Of course, this statement relies on the assumption that merging binaries with two large precessing spins exist and can merge efficiently. But if such GW sources are out there in the Universe, the next LIGO/Virgo run might provide the first observational constraints of their properties.

Compared to previous analyses which include two precessing spins (e.g., [156, 157, 158]) our investigation relies on a state-of-the-art reformulation of the precession estimator $\chi_{\rm p}$ [159]. This generalizes the commonly used expression [160] by employing a rigorous post-Newtonian(PN) average over the joint evolution of *both* spins. Measurements of such an augmented $\chi_{\rm p}$ for current GW events have been presented in

Refs. [159, 161]. Crucially for this work, the precession-averaged estimator presents an exclusion region $1 < \chi_{\rm p} \leq 2$ that can *only* be populated by binaries with two precessing spins. Measuring a binary with $\chi_{\rm p} > 1$ at some large confidence (GW astronomers often use the 90% Bayesian credible interval), would allow us to claim the first detection of binary BH physics sourced by two precessing spins.

Chapter 2 is organized as follows. Section 2.1 introduces the main topic of this Chapter, namely the effect of relativistic spin precession in BH binary mergers, together with current evidence for spin precessing from GWTC-3 both at the single-event and population level. In Section. 2.2 we present our methodology, including details on the adopted precession estimator as well as the implemented parameter-estimation pipeline. In Section 2.3 we present the results of our analysis. In particular, we characterize (i) the SNR dependence on the resulting $\chi_{\rm p}$ posterior distributions, (ii) the statistical behavior of large ensembles of sources, (iii) the impact of the prior, and (iv) the relevance of waveform systematics. Our conclusions are reported in Sec. 2.4. In the following we employ geometric units $G = c = 1$.

2.2 Methods

2.2.1 Spin precession estimators

While the full BH-binary spin properties are in principle described by six degrees of freedom (three components for two spin vectors), a considerable amount of effort has been devoted to identifying a reduced number of parameters that encapsulate most of the information. These are often derived in a PN framework, with the most widely used quantities being the effective aligned spin $\chi_{\rm eff}$ [162, 163, 164] and the effective precessing spin $\chi_{\rm p}$ [159, 160]. Alternative approaches include extending the precession estimator to a two-dimensional vector [165], exploiting the precession/nutation amplitudes and frequencies [166, 137], and computing the fraction of the SNR contained in the spin

modulations [167].

Let us consider a BH binary where $q = m_2/m_1 \leq 1$ is the mass ratio, $\chi_i \in [0, 1]$ are the dimensionless spin magnitudes, θ_i are the angles between the spins and the orbital angular momentum, and $\Delta\Phi$ is the angle between the projections of the two spins onto the orbital plane.

The effective aligned spin is defined as [162]

$$\chi_{\rm eff} = \frac{\chi_1 \cos\theta_1 + q\chi_2 \cos\theta_2}{1+q} \,. \tag{2.1}$$

This is the spin quantity that affects the GW phase at the leading spin-dependent order and is a constant of motion at 2PN [163, 164]. The effective spin $\chi_{\rm eff}$ was recognized as the best measured spin parameter since the very first GW detections, the key reason being that it directly impacts the length of the signal.

The spin-precession parameter $\chi_{\rm p} \propto |d\hat{\boldsymbol{L}}/dt|$ tracks the change of the direction of the orbital angular momentum $\boldsymbol{L}$ over time t [160, 159]. It was originally introduced by Schmidt *et al.* [160] as a building block toward the construction of precessing waveforms. Their definition reads

$$\chi_{\rm p}^{\rm (heu)} = \max\left(\chi_1 \sin\theta_1, q\frac{4q+3}{4+3q}\chi_2 \sin\theta_2\right) \,, \tag{2.2}$$

which in this work we refer to as "heuristic $\chi_{\rm p}$." This precessing spin parameter is defined in the domain $\chi_{\rm p}^{\rm (heu)} \in [0, 1]$. Unlike $\chi_{\rm eff}$, the parameter $\chi_{\rm p}$ depends on the projections of the spins onto the orbital plane, $\chi_i \sin\theta_i$, implying that a confident measurement of $\chi_{\rm p} > 0$ requires that at least one of the two BH spins was misaligned before merger, and hence that the system was precessing.

Gerosa *et al.* [159] recently pointed out that Eq. (2.2) was derived by preferentially selecting some terms when averaging over the spin motion. Mathematically, this is reflected in the maximization operation reported in Eq. (2.2), which selects one of the

two BHs as dominant to the precession dynamics, thus obfuscating two-spin effects. Relaxing this approximation yields a generalized parameter [159]

$$\chi_{\rm p}^{\rm (gen)} = \left[\left(\chi_1 \sin\theta_1 \right)^2 + \left(q \frac{4q+3}{4+3q} \chi_2 \sin\theta_2 \right)^2 + 2q \frac{4q+3}{4+3q} \chi_1 \chi_2 \sin\theta_1 \sin\theta_2 \cos\Delta\Phi \right]^{\frac{1}{2}}, \tag{2.3}$$

where the angles $\theta_1(t)$, $\theta_2(t)$, and $\Delta\Phi(t)$ all vary jointly with time. This can be averaged over a single precession cycle to obtain

$$\chi_{\rm p}^{\rm (av)} = \frac{1}{\tau} \int_0^\tau \chi_{\rm p}^{\rm (gen)}(t) dt, \tag{2.4}$$

where τ is the precession period. We argue this should be regarded as a more solid estimator because, although it is not a constant of motion like $\chi_{\rm eff}$, it at most varies only over the longer radiation-reaction timescale. In the following, we refer to Eq. (2.4) as the "averaged $\chi_{\rm p}$" parameter. In practice, we perform the integral in Eq. (2.4) using a 2PN quasiadiabatic approach where the precession cycle is parametrized by $S(t) = |\mathbf{S}_1(t) + \mathbf{S}_2(t)|$ [168, 164]. We refer the reader to Ref. [164] for details on the derivation of Eqs. (2.3) and (2.4), but stress that the starting point is simply the derivative $d\hat{\boldsymbol{L}}/dt$.

The reformulation of the precession parameter defines an extended range $\chi_{\rm p}^{\rm (av)} \in [0, 2]$. As shown in Ref. [159], the heuristic and averaged definitions of $\chi_{\rm p}$ have the same single-spin limit, which implies that the range of the latter cannot be freely absorbed with a normalization factor. From Eq. (2.3), it is immediate to see that $\chi_{\rm p} > 1$ requires both $\chi_1 \sin\theta_1 \neq 0$ and $\chi_2 \sin\theta_2 \neq 0$, i.e., the binary must have two precessing spins. Such sources *can* lie in $0 \leq \chi_{\rm p} \leq 1$, but both spins being misaligned is requisite in the two-spin domain. From Eq. (2.3), there is a larger volume of parameter space where $\chi_{\rm p} > 1$ for comparable mass binaries $q \lesssim 1$ compared to asymmetric sources with

$q \ll 1$. This is expected, as two-spin effects are highly suppressed in the in the low-mass ratio limit where $S_2/S_1 \propto q^2 \ll 1$ (cf. Ref. [169] for more work on spin precession in asymmetric binaries).

2.2.2 Parameter estimation pipeline

As is common practice in GW parameter estimation, we employ the following fifteen parameters to describe compact-binary coalescences: detector-frame total mass $M = m_1 + m_2$, mass ratio $q = m_2/m_1$, dimensionless spin magnitudes $\chi_{1,2}$, tilt angles $\theta_{1,2}$, azimuthal spin angle $\Delta\Phi$, azimuthal angle ϕ_{JL} between the total and orbital angular momenta, luminosity distance D_L, right ascension α, declination δ, polar angle θ_{JN} between total angular momentum and the line of sight, polarization ψ, time t_c and phase ϕ_c of coalescence [97].

We explore the joint Bayesian posterior distribution of these parameters under a Gaussian noise likelihood (e.g., [170]) using the parallelized Bilby pipeline [98, 99] and its underlying Dynesty implementation of nested sampling [171]. Our runs make use of 2048 live points, a number of autocorrelation equal to 50, a random walk sampling method, and a likelihood marginalized over time and distance. Runs are halted when the log-evidence gain falls below 0.1.

We consider a three-detector network consisting of LIGO Livingston, LIGO Hanford, and Virgo with their projected sensitivities for the ongoing fourth observing run O4 [44]. We consider data segments of 4 s, set a lower frequency cutoff of 20 Hz, assume a sampling frequency of 2048 Hz, and zero noise. Time-varying quantities are quoted when the detector-frame emission frequency of the dominant mode is 20 Hz. This choice has a negligible impact on the averaged $\chi_{\rm p}$ estimator because it only varies on the long inspiral timescale of the binary evolution; see Ref. [159]. Unless stated otherwise, we quote our results using medians and 90% equal-tailed credible intervals.

We adopt uninformative priors as commonly used in current LIGO/Virgo analyses [62,

63, 147, 148]. Specifically, priors on the detector-frame masses are chosen to be uniform in $m_{1,2} \in [5, 100]M_\odot$, with further constraints imposed on the mass ratio $q \in [1/8, 1]$ and detector-frame chirp mass $M_c \in [10, 60]M_\odot$. For most of our runs, priors on the spins are taken to be unform in magnitude $\chi_{1,2} \in [0, 0.99]$ and isotropic in directions. In the following, we will refer to this as our "standard" spin prior. To better explore prior effects, some of our runs are performed with a "volumetric" spin prior $p(\chi_i) \propto \chi_i^2$, corresponding to spin vectors that are uniformly drawn in volume (e.g., [172, 173, 62]). The luminosity distance prior is taken to be uniform in comoving volume with $D_L \in [100, 5000]$ Mpc.

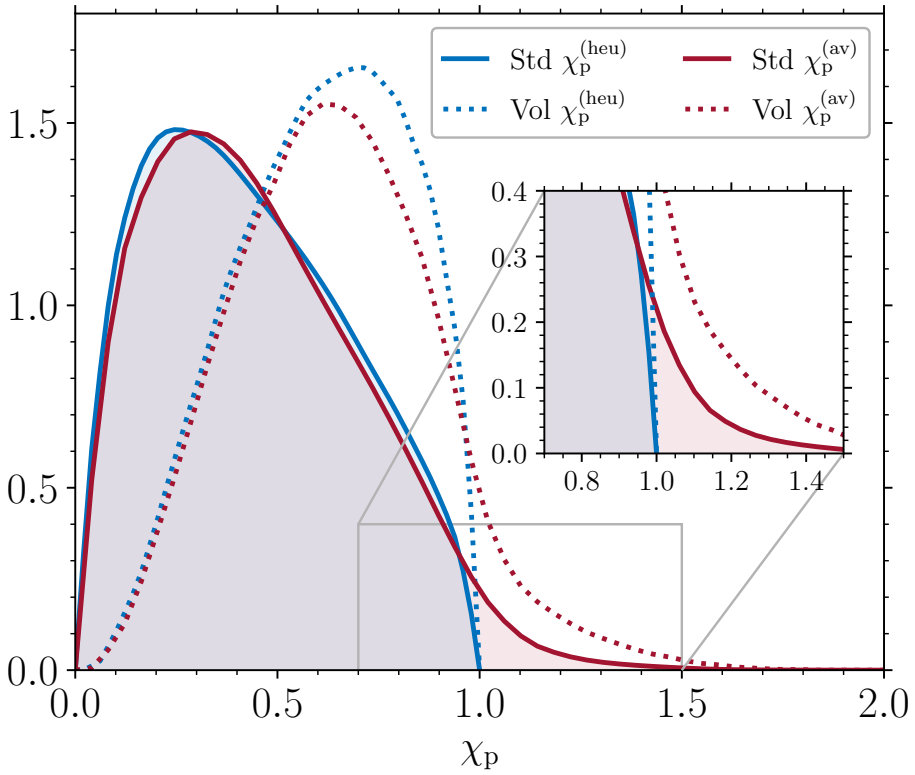

Figure 2.1: Prior distributions for the heuristic (blue) and averaged (red) χ_p estimators. Dotted curves are computed assuming the standard spin prior $p(\chi) = \text{const.}$; dashed curves instead assume a volumetric prior $p(\chi) \propto \chi^2$.

Figure 2.1 shows the resulting prior probability density for the heuristic and the averaged χ_p definitions. At low values of χ_p, the prior distributions of the two estimators

are qualitatively very similar. This behavior was explicitly imposed in Ref. [159] when generalizing the $\chi_{\rm p}$ definition. By construction, the region of $1 < \chi_{\rm p} \leq 2$ is not allowed for the heuristic formulation and, consequently, the prior distribution of $\chi_{\rm p}^{\rm (heu)}$ is steeply truncated at $\chi_{\rm p} = 1$. On the other hand, the prior distribution of $\chi_{\rm p}^{\rm (av)}$ extends into the two-spin region $\chi_{\rm p} > 1$. However, under these commonly used assumptions, the tail at large $\chi_{\rm p}$ values is very sparsely populated. From Eq. (2.3), reaching $\chi_{\rm p} \approx 2$ requires systems with $q \approx 1$, $\chi_{1,2} \approx 1$, $\theta_{1,2} \approx \pi/2$, and $\Delta\Phi \approx 0$. Such a strong prior suppression is a key element of our analysis and suggests that current GW data are being analyzed with a prior that strongly disfavors the region of parameter space that is exclusive to two-spin physics. Although still present, this effects is less prominent for the volumetric spin prior. More quantitatively, we find $p(\chi_{\rm p} > 1) = 0.02$ and 0.07 for the averaged estimator under the standard and volumetric prior, respectively.

For the majority of our runs, we employ the IMRPhenomXPHM [174, 175, 89] waveform model for both injection and recovery. This is a state-of-the-art frequency-domain approximant that captures spin precession without relying on a single-spin approximation. Selecting the same model for both injection and recovery allows us to first isolate statistical effects without systematics. Waveform systematics are then explored with a dedicated analysis where we select different models for injection and recovery. In particular, we use the time-domain model IMRPhenomTPHM [176] as well as the numerical-relativity surrogate NRSur7dq4 [177]. Unlike its frequency-domain counterpart, IMRPhenomTPHM does not rely on the stationary phase approximation, making it better suited for short signals or those dominated by the merger and ringdown phases. When recovering with NRSur7dq4, we restrict our prior to $q > 1/6$, which corresponds to the extended range of validity of the model. Because this approximant only covers ~ 20 orbits before merger, we also restrict our priors to $m_{1,2} \in [35, 150] M_\odot$ and $M_{\rm c} \in [40, 60] M_\odot$ to ensure the signal is fully above the low-frequency cutoff of 20 Hz.

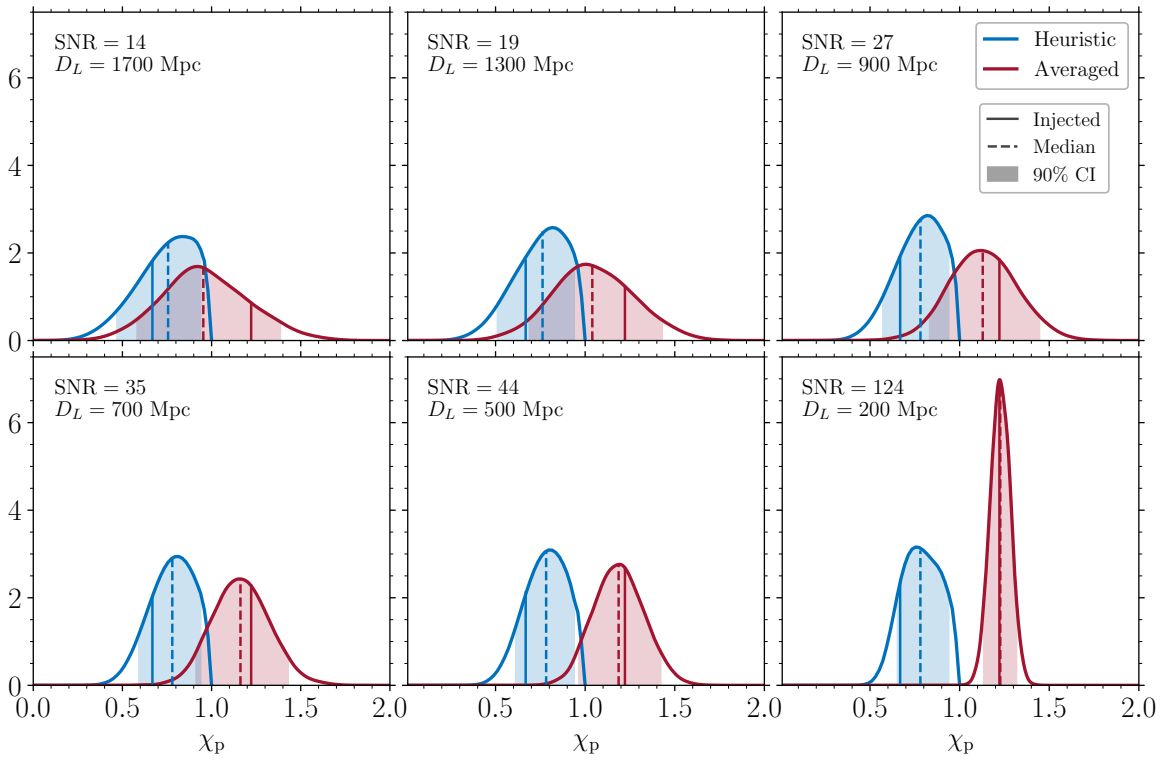

Figure 2.2: Posterior distribution of the heuristic (blue) and averaged (red) χ_p precession estimator for the single-system series described in Sec. 2.3.1. Panels from left to right and top to bottom shows results for the same source injected at increasing SNRs and decreasing luminosity distance D_L. Solid and dashed lines indicate the true value and the median of the recovered posterior. The shaded areas indicate the 90% (CI) credible intervals.

2.3 Results

2.3.1 Single-system series

As a first step, we highlight the main implications of the χ_p reformulation on GW parameter estimation. To this end, we present a series of six software injections where the same binary is observed at different SNRs. We select a source with $\chi_p^{(heu)} = 0.67$ and $\chi_p^{(av)} = 1.22 > 1$, which thus contains two prominently precessing spins. In particular, the injected system has $M = 54.1\,M_\odot$, $q = 0.96$, $\chi_1 = 0.56$, $\chi_2 = 0.7$, $\theta_{1,2} = \pi/2$,

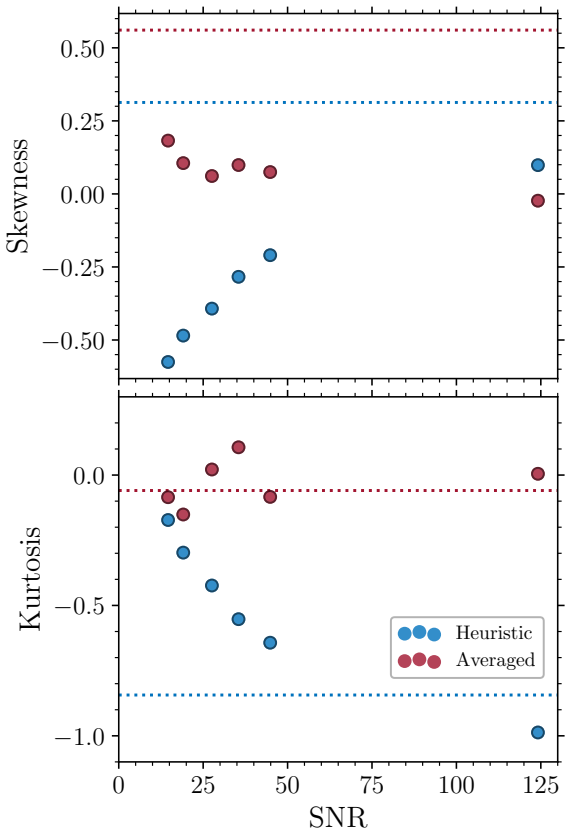

Figure 2.3: Skewness (upper panel) and kurtosis (lower panel) of the $\chi_{\rm p}$ posterior distribution as a function of the SNR for the six injections described in Sec. 2.3.1. Blue and red scatter points refer to the heuristic and averaged $\chi_{\rm p}$ definition, respectively. The dotted lines represent the values of the skewness and kurtosis calculated from the priors.

$\Delta\Phi = 0.1$, $\theta_{JN} = 1.0$, $\phi_{JL} = 1.0$, $\alpha = 0.75$, $\delta = 0.5$, $\psi = 1.0$, $\phi_c = \pi/4$, $t_c = 0.0$. We select increasing values of the luminosity distance $D_L = 200$, 500, 700, 900, 1300, 1700 Mpc while keeping the detector-frame mass M fixed. The corresponding three-detector network SNRs are $\rho = 124$, 44, 35, 27, 19, and 14. We use the IMRPhenomXPHM waveform model for both injection and recovery and employ standard uninformative priors.

Our results are illustrated in Fig. 2.2. As one moves from the lowest to the highest value of the SNR, the recovered posteriors of both the averaged and the heuristic $\chi_{\rm p}$

converge to the injected values. This is expected because we have used the same signal model for injection and recovery and we are not considering a specific noise realization.

For the system with the lowest SNR $\rho = 14$, the posteriors of our two $\chi_{\rm p}$ definitions largely overlap. Quoting median and 90% credible interval, we find $\chi_{\rm p}^{(\rm av)} = 0.95^{+0.43}_{-0.37}$, which implies that we cannot confidently tell that the source has two misaligned spins. As the SNR increases, so does our ability to infer that the binary has two precessing spins. For the system with the largest SNR $\rho = 124$, the two marginalized $\chi_{\rm p}$ distributions are almost completely detached. The posterior of the heuristic $\chi_{\rm p}$ is by definition truncated at $\chi_{\rm p} = 1$ because Eq. (2.2) allows only for the contribution from a single, dominant spin. On the contrary, considering our averaged definition yields $\chi_{\rm p} = 1.22^{+0.09}_{-0.09}$ for $\rho = 124$, implying one infers the presence of two precessing spins with a credibility of $p(\chi_{\rm p} > 1) = 99.9\%$.

Figure 2.2 also shows that the posterior of the averaged $\chi_{\rm p}$ is closer to a Gaussian compared to that of the heuristic estimator. This indicates that, if a significant non-Gaussianity in the heuristic $\chi_{\rm p}$ posterior were to appear in GW data, it could be taken as a potential indication that some additional two-spin physics is present but is being missed because of the suboptimality of the employed estimator.

This argument is further explored in Fig. 2.3, where we show the skewness and kurtosis for the same six injections presented in Fig. 2.2. These quantities are related to the third and fourth moments of the distribution and describe the departure from Gaussianity; both are zero for normally distributed data, with the skewness quantifying the left-right asymmetry and the kurtosis quantifying the weight of the tails [178]. Figure 2.3 shows that both skewness and kurtosis of the averaged $\chi_{\rm p}$ are approximately distributed around 0. On the other hand, the skewness (kurtosis) of the heuristic $\chi_{\rm p}$ strongly increases (decreases) with the SNR. This indicate that (i) the heuristic $\chi_{\rm p}$ posteriors have thinner tails compared to a normal distribution and that (ii) their left tail is more pronounced compared to the right tail. These features can be taken as a quantification of the artificial cutoff at $\chi_{\rm p} = 1$, an assumption that is naturally relaxed

when considering the averaged $\chi_{\rm p}$ estimator.

2.3.2 Parameter-space exploration

Using the same settings, we now target the statistical properties emerging from a large number of injected signals. Ideally, one would want to inject signals drawn from the prior (this is necessary, for instance, to present a probability-probability plot [179]). In our case, such a procedure would be highly suboptimal and ultimately computationally intractable because, as shown in Fig. 2.1, the two-spin region with $\chi_{\rm p} > 1$ corresponds to a very low prior volume (where from now on in this chapter we only refer to the average formulation of $\chi_{\rm p}$). Most of the injections would thus be placed in the region where only one of the two spins dominates. We thus opt for an injection distribution with $\chi_{\rm p}$ uniform in $[0, 2]$ which, although of dubious astrophysical relevance, is well suited to assess the statistical property of the proposed estimator. More precisely, we draw values of $\chi_{\rm p}$ and then reweight samples of the intrinsic binary properties drawn from the uninformative prior (Sec. 2.2.2) to the injection distribution using an acceptance/rejection scheme with an absolute numerical tolerance of 0.04 between the original and resampled values of $\chi_{\rm p}$. We have verified that this choice does not significantly impact our results. Since precession effects are subdominant in the waveform, when selecting the extrinsic properties for injections we only consider sources with $\rho > 20$, i.e., ≈ 2 times larger than the current detection threshold [148].

Figure 2.4 shows the recovered posteriors, obtained using the standard priors, of the averaged $\chi_{\rm p}$ parameter as a function of the true values for 100 such injections. One can divide the parameter space into four distinct regions, acting much like a confusion matrix in statistics.

(i) *True negatives* (bottom-left quadrant in Fig. 2.4): injected $\chi_{\rm p} < 1$ and recovered $\chi_{\rm p} < 1$. The injected configurations are not unique to sources with two precessing spins and are recovered as such.

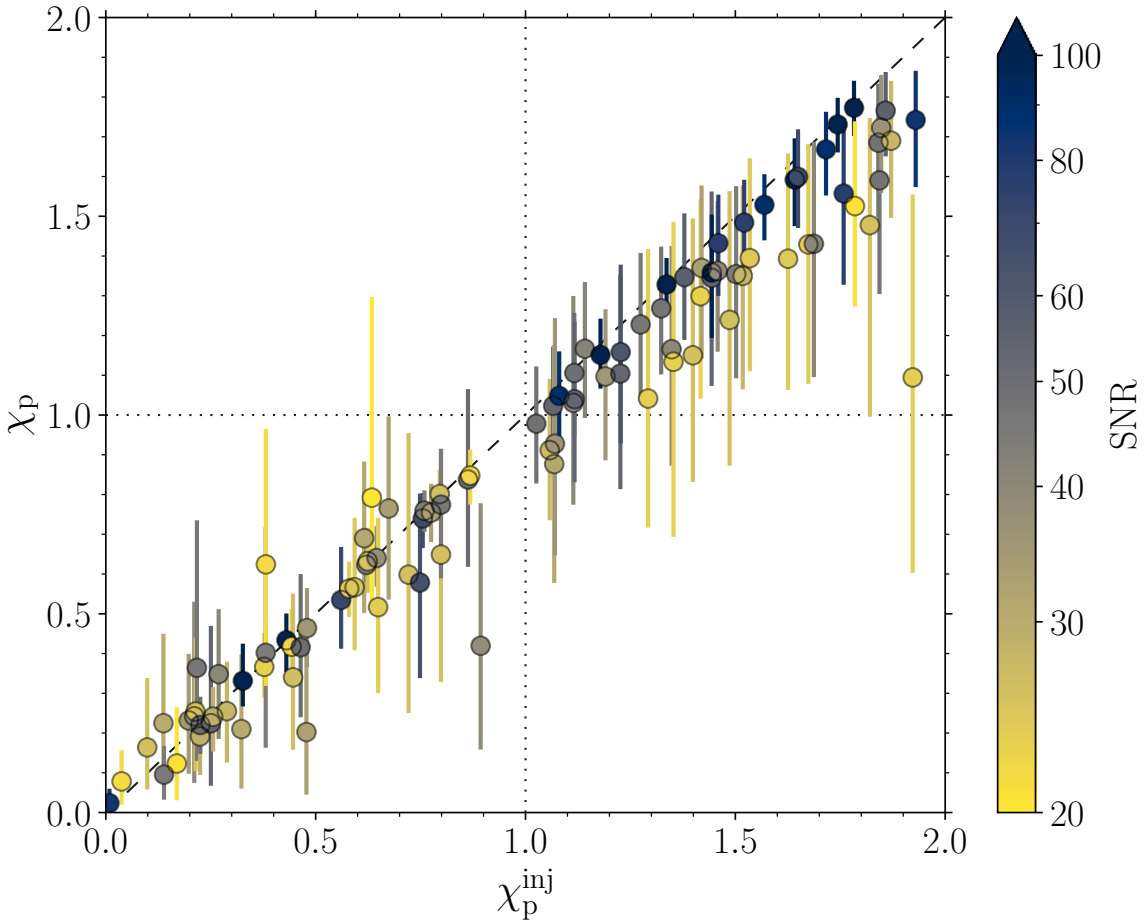

Figure 2.4: Set of 100 injections obtained by reweighting the averaged $\chi_{\rm p}$ prior toward a uniform distribution in $[0, 2]$. The medians (scatter points) and symmetric 90% credible intervals (error bars) of the recovered posteriors are plotted against the true values $\chi_{\rm p}^{\rm inj}$. Vertical and horizontal dotted lines indicate $\chi_{\rm p} = 1$ while the dashed diagonal line corresponds to $\chi_{\rm p} = \chi_{\rm p}^{\rm inj}$, i.e., successful recovery. The three-detector SNRs of the injected sources are reported on the color scale.

(ii) *False positives* (top-left quadrant in Fig. 2.4): injected $\chi_{\rm p} < 1$ and recovered $\chi_{\rm p} > 1$. For these sources, one infers the presence two precessing spins even if they might not be present.

(iii) *False negatives* (bottom-right quadrant in Fig. 2.4): injected $\chi_{\rm p} > 1$ and recovered $\chi_{\rm p} < 1$. In this region sources have two precessing spins but one is not able infer

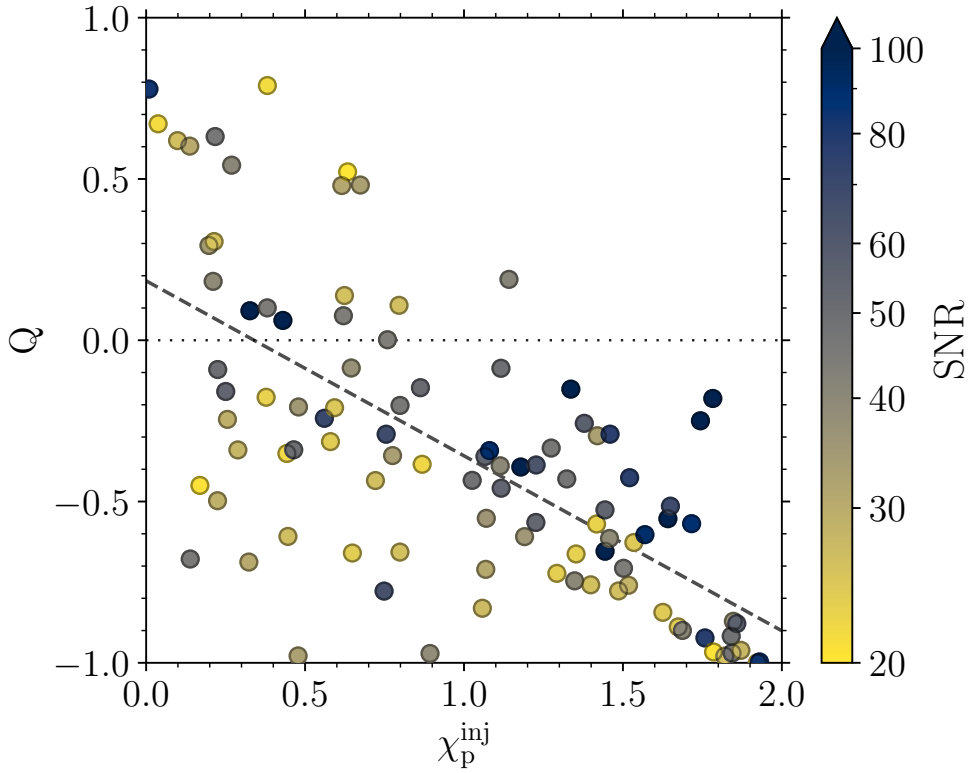

Figure 2.5: Adjusted posterior quantile Q for the posterior distribution of the averaged $\chi_{\rm p}$ parameter. Sources above (below) the horizontal dotted line indicate cases cases where $\chi_{\rm p}$ is overestimated (underestimated). To guide the eye, the diagonal dashed line shows a simple linear fit $Q = -0.54\chi_{\rm p}^{\rm inj} + 0.18$. The color scale indicates the SNRs of the sources.

their occurrence from the signal.

(iv) *True positives* (top-right quadrant in Fig. 2.4): injected $\chi_{\rm p} > 1$ and recovered $\chi_{\rm p} > 1$. These sources are characterized by two precessing spins and one can successfully infers that this is the case.

For each posterior distribution, we compute the fraction of the samples in each of these four regions and then compute the arithmetic mean over the injected sample (this is equivalent to assuming a flat population prior on $\chi_{\rm p}$ because our injections are distributed uniformly). We report 47.45% of true negatives, 0.55% of false positives,

7.01% of false negatives, and 44.98% of true positives.

From Fig. 2.4, the signals with higher SNR lie closer to the injected values and present thinner posterior distributions, as expected. In the true negative region, the recovered posteriors are distributed around the true value without evident systematic trends. On the other hand, when $\chi_{\rm p} > 1$, the recovered posteriors systematically underestimate the true value. While this is, in general, true for most systems, in a few cases this is sufficient to cause false negatives.

One can further quantify this behavior using the adjusted posterior quantile

$$Q = 2\int_0^{\chi_{\rm p}^{\rm inj}} p(\chi_{\rm p})\,{\rm d}\chi_{\rm p} - 1 \in [-1, 1], \tag{2.5}$$

where $p(\chi_{\rm p})$ is the posterior distribution and $\chi_{\rm p}^{\rm inj}$ is the true value. The ideal case where the median of $p(\chi_{\rm p})$ coincides with the true value corresponds to $Q = 0$. Obtaining $Q > 0$ ($Q < 0$) instead implies that the amount of precession in the system is being overestimated (underestimated), and $100|Q| < X$ implies that the injected value is inside the X% symmetric confidence interval of the recovered posterior. The values of Q for our 100 injections are shown in Fig. 2.5. We find a strong decreasing trend of Q for increasing values of $\chi_{\rm p}$, which becomes particularly evident in the $\chi_{\rm p} > 1$ region. Our analysis indicates that, in general, statistical errors cause an underestimate of $\chi_{\rm p}$ whenever $\chi_{\rm p} > 1$. In other words, given a waveform model, sources with two-spin precession require larger SNR for accurate measurement (cf. [180, 181]).

With a completeness of 86.5% and a contamination of 1.2%, our results are, overall, are extremely encouraging.[1] The broader conclusion is that sources with $\chi_{\rm p} > 1$ and sufficiently high SNR $\rho \gtrsim 20$ in O4 can, in principle, be correctly identified as affected by two precessing spins.

[1] As common in binary classification [178], we define completeness = true positives / (true positives + false negatives) and contamination = false positives / (true positives + false positives).

2.3.3 Impact of the prior

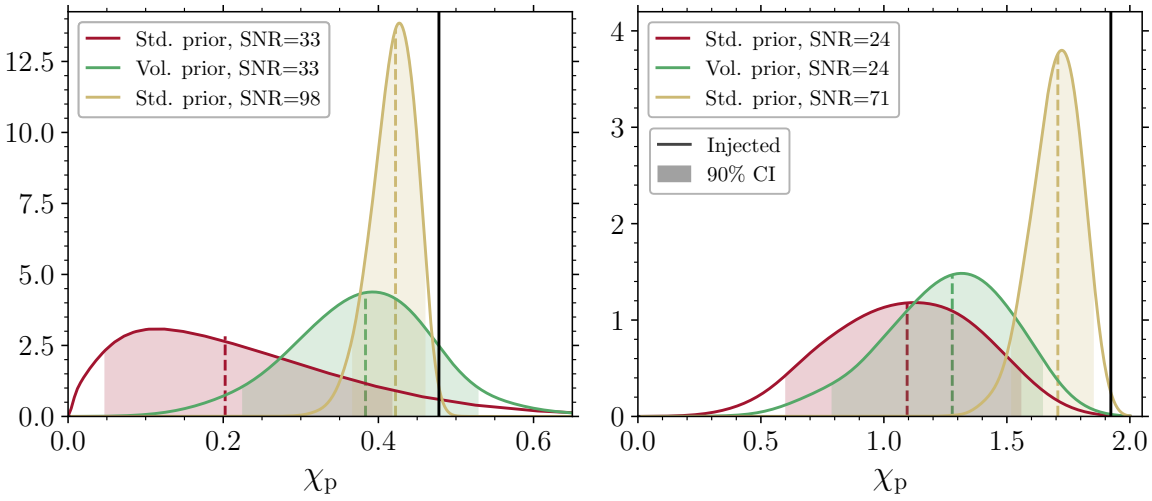

Figure 2.6: Recovery of the averaged χ_p estimator with different priors and SNRs. The left and right panel shows results for sources injected with $\chi_p = 0.48$ and $\chi_p = 1.92$, respectively (black vertical lines). The red and green posterior distributions are obtained under the standard and volumetric priors, respectively, and the same SNRs used in Sec. 2.3.2. For the yellow distributions, the SNR was boosted by a factor ≈ 3. Dashed lines indicate the medians of the posteriors while the shaded area indicates the 90% credible interval.

Our parameter-space exploration highlights a generic tendency to underestimate precession effects whenever $\chi_p > 1$. The steep feature at $\chi_p \approx 1$ shown in Fig. 2.1 strongly suggests that this statistical bias is driven by the employed prior. To verify this, we select two injections among the 100 we have just presented with posterior quantile $Q \approx -1$, i.e. where the displacement between the injected and recovered values of χ_p is maximized. More specifically, the two systems we consider have $\chi_p \in \{0.48, 1.92\}$, $Q \in \{-0.98, -1\}$, and $\rho \in \{32.8, 23.6\}$, respectively.

In Fig. 2.6 we compare the posterior distributions obtained under the standard uniformative prior as in Sec. 2.3.2 against additional inference runs where we instead take a volumetric prior on the spins. The latter choice enhances the prior weight assigned

to configurations with large spins (cf. Fig. 2.1). For the injection with $\chi_{\rm p} = 0.48$ (left panel), we recover $\chi_{\rm p} = 0.20^{+0.22}_{-0.16}$ with the standard prior and $\chi_{\rm p} = 0.38^{+0.14}_{-0.16}$ with the volumetric prior. This example shows that simply changing the prior to an alternative that is equally well motivated —why should vectors like the spins be distributed uniformly in magnitude instead of volume?— can significantly mitigate the inferred bias. In this case, the posterior quantile increases from $Q = -0.98$ to $Q = -0.75$. The improvement is less evident, but still present, for the injection with $\chi_{\rm p} = 1.92$, where a volumetric prior yields $\chi_{\rm p} = 1.28^{+0.37}_{-0.49}$ compared to $\chi_{\rm p} = 1.10^{+0.46}_{-0.49}$ for the standard prior.

Figure 2.6 also shows additional runs where the same sources are considered at higher SNR, larger by a factor ≈ 3, using the standard priors. As already shown in Sec. 2.3.1, the posterior tends toward the true values for louder sources. Prior effects are still evident, with the true value remaining outside the 90% credible interval. More specifically, for these high-SNR runs we find $\chi_{\rm p} = 0.42^{+0.04}_{-0.05}$ and $\chi_{\rm p} = 1.71^{+0.14}_{-0.19}$ for the $\chi_{\rm p} = 0.48$ and $\chi_{\rm p} = 1.92$ case, respectively.

2.3.4 Waveform systematics

All the analyses illustrated so far were performed using the same waveform model for both injection and recovery and, therefore, do not capture systematic errors due to any mismodeling of the signal. Binaries with prominent spin effects are harder to model, implying that the $\chi_{\rm p} > 1$ region we are interested in is also where discrepancies between different approximants are more likely to appear.

Figure 2.7 and Table 2.1 illustrate the posterior distribution of the averaged $\chi_{\rm p}$ parameter when different models are used in injection and recovery. We test various combinations of the IMRPhenomXPHM [89], IMRPhenomTPHM [176], and NRSur7dq4 [177] waveform models. We concentrate on two systems selected from the 100 injections presented in Sec. 2.3.2. In particular, we consider one source with

$\chi_{\rm p} = 1.57$ characterized by two prominently precessing spins as well as a control case with $\chi_{\rm p} = 0.43$. The SNRs computed using IMRPhenomXPHM are $\rho = 90.3$ and $\rho = 107.3$, respectively. Both sources have sufficiently large detector-frame total masses $M \gtrsim 125 M_\odot$ such that the signal is short enough to be simulated with NRSur7dq4. The cases where the signal is injected with NRSur7dq4 are arguably more realistic as this model is proven to be more accurate, i.e., closer to numerical-relativity simulations [177], though the model does require extrapolation at the low-q and high-$\chi_{1,2}$ edges of the parameter space we consider. For both the analyzed cases, the posteriors are relatively well centered on the true values whenever the injection and recovery are performed with the same waveform model.

The most evident feature from Fig. 2.7 is that systematic biases increase dramatically for higher values of $\chi_{\rm p}$. This statement holds even though our analyzed low-$\chi_{\rm p}$ (high-$\chi_{\rm p}$) case has a higher (smaller) SNR and should thus be more (less) susceptible to waveform systematics.

For the $\chi_{\rm p} = 0.43$ source (left panels in Fig. 2.7), the injected value lies inside the 90% credible interval of the posterior for most the waveform combinations we tested. The only exception is the case where we inject with the NRSur7dq4 and recover with IMRPhenomXPHM. This run shows the largest quantile $Q = -0.93$ which tentatively suggests a lower accuracy of that model to spin precession, at least for this specific set of parameters. This conclusion is reinforced by our results obtained when the source is generated with IMRPhenomTPHM but recovered with IMRPhenomXPHM: the injected value is barely inside the 90% credible interval with posterior quantile is $Q = -0.88$.

For the second case studied here with $\chi_{\rm p} = 1.57$ (right panels in Fig. 2.7), waveform systematics are severe. All waveform combinations where we inject and recover with different models return posterior distributions that are inconsistent with the true value at extremely high confidence (so high that we cannot meaningfully quantify it with the samples at our disposal). The worst cases appears to be those when we recover with

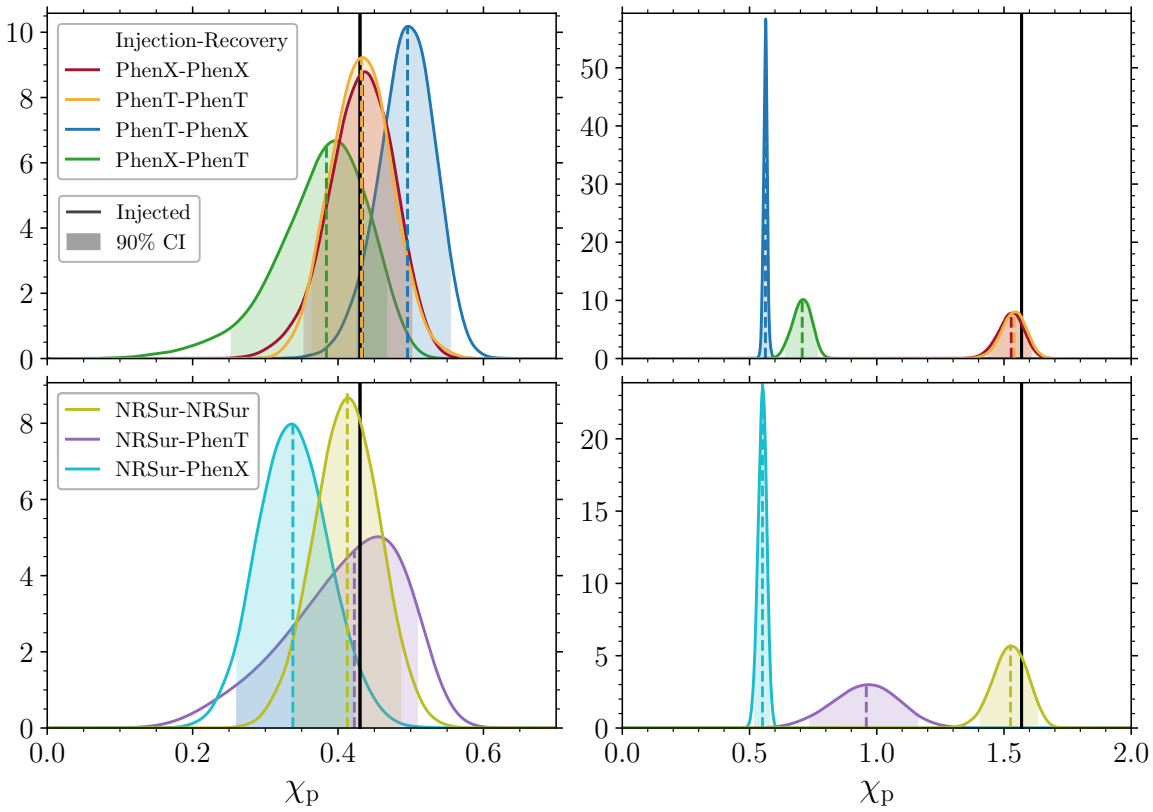

Figure 2.7: Posterior distributions of the averaged χ_p parameter obtained with seven different combinations of waveform models. The left (right) panels show results for a source injected with $\chi_p = 0.43$ and $M = 131.1M_\odot$ ($\chi_p = 1.57$ and $M = 130.8M_\odot$) The top panels show results obtained with the two phenomenological models IMRPhenomXPHM ("PhenX") for $\chi_p = 0.43$ ($\rho = 107.3$) and for $\chi_p = 1.57$ ($\rho = 90.3$) and IMRPhenomTPHM ("PhenT") for $\chi_p = 0.43$ ($\rho = 93.6$) and $\chi_p = 1.57$ ($\rho = 81.7$). The bottom panels show results obtained in combination with the numerical-relativity surrogate model NRSur7dq4 ("NRSur") for $\chi_p = 0.43$ ($\rho = 100.2$) and for $\chi_p = 1.57$ ($\rho = 75.6$). For each case, the label reported before (after) the hyphen in the legend refers to the waveform model used at the injection (recovery) stage. The injected values are indicated with black vertical lines. Medians and 90% credible interval of the posterior distribution are indicated with dashed lines and shaded areas, respectively.

IMRPhenomXPHM, which produce a χ_p posterior that is entirely below unity.

While a more complete investigation on waveform systematics is beyond the scope of

$\chi_{\rm p}^{\rm inj} = 0.43$	PhenX		PhenT		NRSur	
	Q	$\chi_{\rm p}$	Q	$\chi_{\rm p}$	Q	$\chi_{\rm p}$
PhenX	0.06	$0.43^{+0.07}_{-0.08}$	-0.6	$0.38^{+0.08}_{-0.13}$	–	–
PhenT	0.88	$0.5^{+0.06}_{-0.07}$	0.04	$0.43^{+0.07}_{-0.07}$	–	–
NRSur	-0.93	$0.34^{+0.08}_{-0.08}$	-0.08	$0.42^{+0.09}_{-0.16}$	-0.3	$0.41^{+0.07}_{-0.08}$

$\chi_{\rm p}^{\rm inj} = 1.57$	PhenX		PhenT		NRSur	
	Q	$\chi_{\rm p}$	Q	$\chi_{\rm p}$	Q	$\chi_{\rm p}$
PhenX	-0.6	$1.53^{+0.08}_{-0.09}$	-1	$0.71^{+0.06}_{-0.07}$	–	–
PhenT	-1	$0.56^{+0.01}_{-0.01}$	-0.47	$1.54^{+0.08}_{-0.08}$	–	–
NRSur	-1	$0.55^{+0.03}_{-0.03}$	-1	$0.96^{+0.2}_{-0.22}$	-0.48	$1.53^{+0.1}_{-0.12}$

Table 2.1: Posterior quantiles Q, medians, and 90% credible intervals of the averaged $\chi_{\rm p}$ estimator from analyses performed with three different waveform models: IMRPhenomXPHM ("PhenX"), IMRPhenomTPHM ("PhenT"), and NRSur7dq4 ("NRSur"). The top (bottom) table shows results for injections with $\chi_{\rm p} = 0.43$ ($\chi_{\rm p} = 1.57$). In each table, the rows (columns) indicate the waveform used for signal injection (recovery).

this work, the selected cases studied here tentatively indicate that current state-of-the-art approximants struggle at providing a consistent modeling of the signal in the $\chi_{\rm p} > 1$ region, to a level that will be significant for the heavy, loud sources expected in O4. The discrepancies between the waveform models in the high-precession limit are potentially expected due to the differences in the prescriptions for the spin dynamics between the models. In IMRPhenomXPHM, the precession angles are calculated by applying the stationary phase approximation to the multi-timescale scale analysis of the precession equations [164, 182]. The prescription is then artificially extended through the merger and ringdown beyond its regime of validity. In IMRPhenomTPHM, the precession angles are calculated by direct integration of the equations of motion, coupled to a semianalytical approximation for the merger-ringdown that relies on an angular velocity

determined by the quasinormal mode frequencies of the remnant BH [183].

Waveform developers are actively working toward calibrating PN-based waveform models using numerical-relativity simulations with precessing spins (e.g., [184]), which will hopefully alleviate the systematic deviations reported here. It is also important to stress that, for this exercise, we had to select sources with high-enough mass such that the signal is fully covered by NRSur7dq4, but these are also the systems where the precession signature is expected to be weaker. This is because precession cycles are contained in the low-frequency part of the signal that gradually falls out of band as the total mass increases.

On a more positive note, the conclusion that emerges here is that waveform systematic do not produce false positive: if a future observation will deliver $\chi_{\rm p} > 1$ at high confidence, it appears safe to claim that the BH binary had two precessing spin.

2.4 Final remarks

If an incoming LIGO/Virgo source is composed of merging BHs with two precessing spins, will we able to tell? In this chapter we provide a statistical assessment of this question using a large set of software injections.

For dimensionality reduction and interpretation purposes, it is useful to have a single parameter that can capture the effect of precession in GW data. We employ a recent generalization [159] of the effective precessing spin $\chi_{\rm p}$ [160]. Unlike its predecessor, the augmented formulation does not assume that one of the two spin dominates the dynamics. In particular, the region of the parameter space $1 < \chi_{\rm p} \leq 2$ is exclusive to binaries with two precessing spins. Because spin-spin couplings in GW data provide a weak contribution to the waveform, measuring a source with $\chi_{\rm p} > 1$ also requires sensitive detectors. While such a detection has not occurred in current data [159, 161], our software injections at O4 sensitivity demonstrate that this goal is well within our reach.

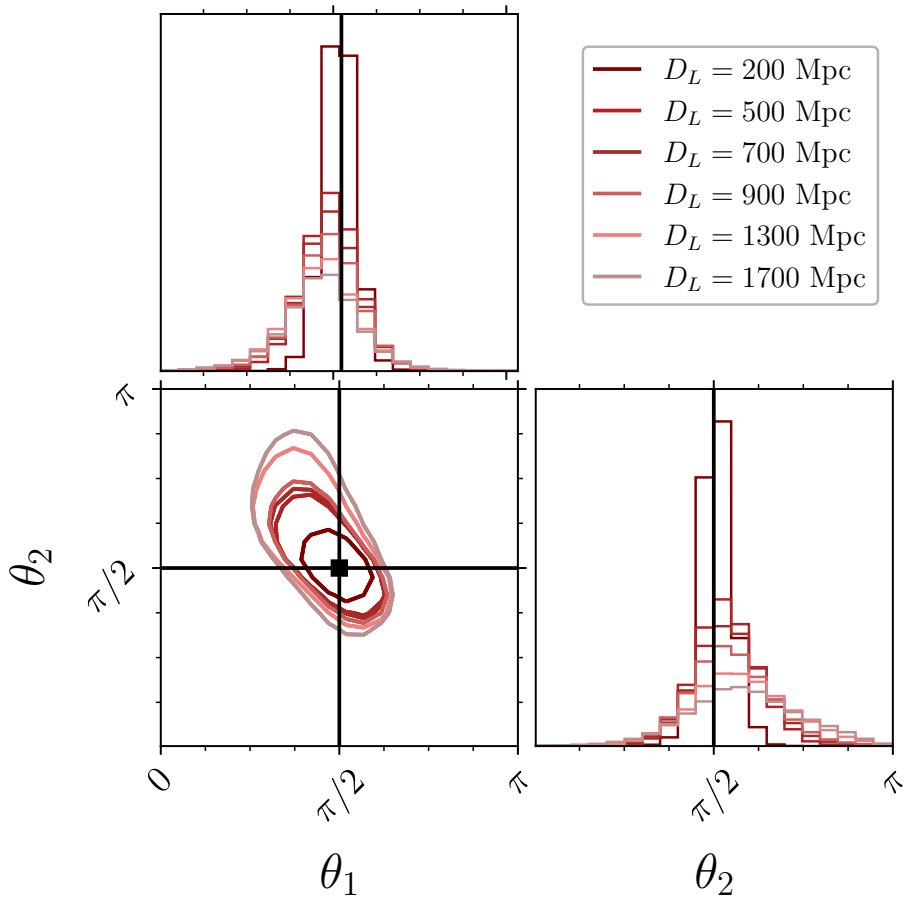

Figure 2.8: Joint posterior distribution of the tilt angles θ_1 and θ_2 for the single-system series described in Sec. 2.3.1. Darker (lighter) curves and histograms refer to sources at smaller (larger) distance. Contour level correspond to 90% credible intervals.

In this chapter, we have concentrated solely on inference of precession with the $\chi_{\rm p}$ estimator. With the BH-binary parameter space spanning 15 dimensions, our large set of injections naturally contains much more information that could potentially be extracted (including, but not limited to, different spin precession estimators, correlation between the effective spins and other binary parameters, and the vector spin components themselves). In order to facilitate further exploitation, our posterior chains are made publicly available in their entirety at github.com/ViolaDeRenzis/twoprecessingspins [185]. The

total computational budget to collect these data amounts to about half a million CPU hours.

It is important to stress that the occurrence of two precessing spins is a necessary, but not sufficient, condition to obtain values $\chi_{\rm p} > 1$, (i.e., being in this region implies the source has two precessing spins, but not vice versa). Such large values of $\chi_{\rm p}$ require a considerable fine-tuning of the binary's intrinsic parameters (large spin magnitudes, mass ratios close to unity, spins coplanar with the orbit and aligned with each other). This makes our assessment very conservative and lets us identify sources with smoking-gun evidence of two-spin precession.

As the detectors' sensitivities increase and one moves beyond effective-spin parametrizations, inference on the higher-dimensional spin parameter space will hopefully allow us to relax such conservative assumptions. Looking ahead in this direction, constraining a source away from the edges of the θ_1–θ_2 plane can also be taken as a telltale sign of two-spin precession (recall that $0 \leq \theta_i \leq \pi$). Figure 2.8 shows the joint posterior distribution of the spin tilts for the series of injections described in Sec. 2.3.1 where, indeed, nonprecessing configurations can be largely excluded. One point worth stressing is the short-timescale dependence of the tilt angles, which results in deeper issues when performing population studies [186]. In contrast, the formulation of $\chi_{\rm p}$ explored here only varies on the longer radiation-reaction timescale —which is the best one can hope for in the absence of additional constant of motions besides $\chi_{\rm eff}$ [163]. While future third-generation detectors will improve spin measurements, constraining the individual tilt angles is expected to remain challenging (see the discussion in Chapter 4.4.4). Thus, for the current generation of detectors, the parameter $\chi_{\rm p}$ provides a pragmatic and robust approach to capture double-spin precession, even if more detailed constraints may eventually become achievable.

Our large-scale injection study shows that BH binaries with two prominently precessing spins at sufficiently high SNR can be generically identified as such. We also pointed out how the departure from Gaussianity of the $\chi_{\rm p}$ distribution can be a precious

indicator of the suboptimality of the adopted indicator. At the same time, prior effects and waveform systematics introduce some interpretation issues that need to be further explored.

Of all the software injections we performed with different source parameters and waveform models, we did not detect a single confident false positive (i.e., a source with $\chi_{\rm p} < 1$ which is erroneously recovered as having $\chi_{\rm p} > 1$). Our study strongly indicates that, should a confident detection with $\chi_{\rm p} > 1$ be made in O4, this would provide a conservative and safe claim of the first observation of a merging BH binary with two precessing spins –except for possible noise issues such as glitches.

Chapter 3

Parameter estimation of binary black holes in the endpoint of the up–down instability

Overview

Black-hole binary spin precession admits equilibrium solutions corresponding to systems with (anti-) aligned spins. Among these, binaries in the up-down configuration, where the spin of the heavier (lighter) black hole is co- (counter-) aligned with the orbital angular momentum, might be unstable to small perturbations of the spin directions. The occurrence of the up-down instability leads to gravitational-wave sources that formed with aligned spins but are detected with precessing spins. We present a Bayesian procedure based on the Savage-Dickey density ratio to test the up-down origin of gravitational-wave events. This is applied to both simulated signals, which indicate that achieving strong evidence is within the reach of current experiments, and the LIGO/Virgo events released to date, which indicate that current data are not informative enough. The material of this Chapter is based on Ref [135].

3.1 Introduction

Gravitational-wave (GW) detections provide measurements of the intrinsic properties of astrophysical black holes (BHs), notably their masses and spins. At the time of writing, ground-based interferometers LIGO and Virgo have observed about mergers of stellar-mass BHs with false alarm rates $< 1\ \mathrm{yr}^{-1}$ [62, 63, 147, 148] and substantially more detections are expected from the upcoming observing runs [187, 44].

GWs emitted during the inspiral of BH binaries are mostly beamed along the direction of the orbital angular momentum $\boldsymbol{L}$. If the spins of the two BHs $\boldsymbol{S}_{1,2}$ are misaligned with $\boldsymbol{L}$, couplings between these three momenta cause them to precess [142, 143]. The resulting motion imparts characteristic modulations to the amplitude and phase of emitted GWs. From an astrophysical perspective, measuring spin precession is important to elucidate the possible astrophysical formation pathways of BH binaries, with large spin misalignments thought to be indicative of sources formed via dynamical interactions [115, 188].

Configurations with spins that are either aligned or anti-aligned with the orbital angular momentum are equilibrium solutions of the relativistic spin-precession equations. This means that binaries that are *exactly* aligned will remain so. There are four such cases, which we refer to as up–up, down–down, down–up, and up–down, where "up" ("down") indicates spins that are parallel (anti-parallel) to the orbital angular momentum and the direction before (after) the hyphen refers to the more (less) massive BH. Crucially, equilibrium does not imply stability. Reference [189] showed that, while up–up, down–down, and down–up binaries are always stable, up–down binaries can be unstable to spin precession. For these sources, infinitesimal perturbations to the spin directions cause large precession cycles. In particular, up–down binaries are stable at

early times and turn unstable at the critical orbital separation [189]

$$r_{\rm UD+} = \frac{\left(\sqrt{\chi_1} + \sqrt{q\chi_2}\right)^4}{(1-q)^2} M \,, \tag{3.1}$$

where $\chi_i = S_i/m_i^2$ are the Kerr parameters of the BHs, $q = m_2/m_1 \leq 1$ is the mass ratio, and $M = m_1 + m_2$ is the total mass of the system.[1] The up–down instability was first derived using a Post-Newtonian (PN) approach [189] and then confirmed using both independent PN codes [190, 191] and numerical-relativity simulations [192].

Measuring the up–down instability in GW data would provide a direct observation of an exquisite feature of the two-body problem in general relativity. At the same time, the up–down instability might also dilute the effectiveness of the spin orientations in discriminating BH-binary formation channels: GW sources that are observed with precessing spins in the LIGO/Virgo band did not necessarily form with misaligned spins. Rather, the spins used to be (anti-) aligned and became misaligned before merger. The flip side of the same coin is that observing unstable binaries will point toward a formation channel that can conceivably explain binaries with up–down spins. Notably, this might include active galactic nuclei (AGN) disks surrounding supermassive BHs [193, 194], where the spins of embedded stellar-mass BH binaries are expected to either align or anti-align with the angular momentum of the disk [195].

The up–down instability provides a testable prediction for GW observations. Reference [196] showed that unstable up–down BHs do not disperse in the available parameter space but converge to a well-defined endpoint late in the inspiral. This is a precessing configuration where all three angular momenta $\boldsymbol{S}_1, \boldsymbol{S}_2$, and $\boldsymbol{L}$ are coplanar, and furthermore, the two BH spins are collinear, namely [196],

$$\cos\theta_1 = \frac{\chi_1 - q\chi_2}{\chi_1 + q\chi_2} \,, \tag{3.2}$$

[1] Throughout the chapter we use natural units where $c = G = 1$.

$$\cos\theta_2 = \frac{\chi_1 - q\chi_2}{\chi_1 + q\chi_2}\,, \tag{3.3}$$

$$\phi_{12} = 0\,, \tag{3.4}$$

where θ_i indicate the tilts angles between $\boldsymbol{S}_i$ and $\boldsymbol{L}$, and ϕ_{12} indicates the azimuthal angle between the two BH spins measured in the orbital plane. After the instability is triggered, binaries reach this analytical endpoint after the orbital separation has decreased by only $\lesssim 100M$ [196]. Therefore, binaries that form as up–down and become unstable will appear in our detectors with spin orientations that are well approximated by Eqs. (3.2–3.4).

In this chapter of the thesis, we perform Bayesian parameter estimation of precessing BH binaries in the endpoint of the up–down instability. Should an unstable up–down binary enter the LIGO band, can we tell that this source was originally stable and aligned? In statistical terms, this is a model-selection problem between a broader hypothesis where binaries are generically precessing and a narrower hypothesis with constraints given by Eqs. (3.2–3.4). We apply this line of reasoning to both simulated signals and the current catalog of GW events. By employing the Savage-Dickey density ratio, we compute the odds in favor of the up–down hypothesis over that of generically precessing BH binaries. Crucially, this only requires an inference run with the uninformative prior, with the odds computed by post-processing the recovered posterior samples.

Chapter 3 is organized as follows. In Sec. 3.2 we derive the statistical framework and describe how it can be used to assess whether observed binaries are in the endpoint of the up–down instability. In Sec. 3.3 we present our results for an injection campaign and real sources, and also demonstrate that evolving binary BH spin posteriors backwards in time is a useful diagnostic when investigating the up–down instability. We finish with our conclusions in Sec. 3.4.

3.2 Methods

3.2.1 Gravitational-wave signals

We first consider synthetic GW signals from individual binary BH coalescences on quasi-circular orbits and target the statistical inference of all 15 parameters of the problem. These are two detector frame masses $m_{1,2}$, six spin degrees of freedom (magnitudes $\chi_{1,2}$, tilts $\theta_{1,2}$, azimuthal angles ϕ_{12} and ϕ_{JL}), and seven extrinsic parameters (luminosity distance D_L, sky location α, δ, polar angle θ_{JN}, polarization ψ, coalescence time t_c, and phase ϕ_c).

Signals are analyzed using the parallel version of the BILBY inference code [98, 99]. We use the IMRPhenomXPHM approximant [89] for both injection and recovery. We consider a three-detector network made of LIGO Livingston, LIGO Hanford, and Virgo at the sensitivity expected for the ongoing O4 run. We use data segments of 4 s, a sampling frequency of 2048 Hz, a low-frequency cutoff of 20 Hz, and zero noise. Spin orientations are quoted at a reference frequency of 20 Hz. We use the DYNESTY sampler [171] with 2048 live points, a random walk sampling method, a number of autocorrelation equal to 50, and a likelihood that is marginalized over time and distance.

Our priors are those commonly used in the standard LIGO/Virgo analyses [62, 63, 147, 148]. In particular, detector-frame component masses are distributed uniformly in $m_{1,2} \in [5, 100] M_\odot$ with bounds in mass ratio $q \in [1/8, 1]$ and detector-frame chirp mass $\mathcal{M} \in [10, 60] M_\odot$ while spins are distributed uniformly in magnitude $\chi_{1,2} \in [0, 0.99]$ and isotropically in directions.

In the following, we also postprocess GW data using publicly available posterior samples for the GWTC-2.1 [147] and the GWTC-3 [148] data releases. Among the available datasets, we use results from the IMRPhenomXPHM waveform model where the merger rate is uniform in comoving volume and source-frame time. We consider binary BH mergers with false alarm rates < 1 yr^{-1} in at least one of the detection

pipelines. From these, we exclude all the events that potentially contain a neutron star. The resulting list of 69 events is reported in Table 3.1.

When needed, we covert between PN orbital separation r and GW frequency $f_{\rm ref}$ using the 2PN expressions from Ref. [143].

3.2.2 Savage-Dickey density ratio

Given the data d associated with a measurement, and model hypothesis $\mathcal{H}$ characterized by parameters θ, the Bayesian evidence is defined as

$$\mathcal{Z}(d|\mathcal{H}) = \int \mathcal{L}(d|\theta,\mathcal{H})\pi(\theta|\mathcal{H})\,\mathrm{d}\theta\,, \tag{3.5}$$

where $\mathcal{L}$ is the likelihood and π is the prior distribution. Model selection in favor of, say, a "narrow" model $\mathcal{H}_{\rm N}$ over a "broad" model $\mathcal{H}_{\rm B}$ requires computing the posterior odds

$$\mathcal{O} = \frac{\mathcal{Z}(d|\mathcal{H}_{\rm N})}{\mathcal{Z}(d|\mathcal{H}_{\rm B})}\frac{\pi(\mathcal{H}_{\rm N})}{\pi(\mathcal{H}_{\rm B})}\,, \tag{3.6}$$

where the first term (ratio of the evidences) is the Bayes factor $\mathcal{B}$. Values of the posterior odds are often associated to descriptive terms using the so-called Jeffrey scale [197], where $|\ln\mathcal{O}| < 1$ is classified as "inconclusive," $1 < |\ln\mathcal{O}| < 2.5$ is classified as as "weak" evidence, $2.5 < |\ln\mathcal{O}| < 5$ is classified as "moderate" evidence, and $|\ln\mathcal{O}| > 5$ is classified as "strong" evidence. The sign of the log Bayes factor indicates which of the two models is statistically favored, with $\ln\mathcal{O} > 0$ signaling a preference for $\mathcal{H}_{\rm N}$ over $\mathcal{H}_{\rm B}$. In the following, we consider equal model priors such that $\mathcal{O} = \mathcal{B}$.

Let us now assume that model $\mathcal{H}_{\rm N}$ is nested within $\mathcal{H}_{\rm B}$. That is, among the parameters $\theta = \{\varphi, \gamma\}$, a subset of parameters φ is common to both models, while the other parameters γ are constrained to $\gamma_{\rm N}(\varphi)$ in the narrow model. Let us also assume

Event	$\ln \mathcal{B}$	Event	$\ln \mathcal{B}$
GW150914	0.14	GW190731_140936	0.11
GW151012	0.54	GW190803_022701	0.11
GW151226	0.50	GW190805_211137	0.61
GW170104	−0.02	GW190828_063405	0.3
GW170608	0.18	GW190828_065509	0.15
GW170729	0.47	GW190910_112807	−0.06
GW170809	0.26	GW190915_235702	0.29
GW170814	−0.06	GW190924_021846	0.31
GW170818	0.58	GW190925_232845	0.24
GW170823	0.26	GW190929_012149	−0.15
GW190408_181802	0.02	GW190930_133541	0.59
GW190412	0.6	GW191103_012549	0.58
GW190413_052954	−0.01	GW191105_143521	0.06
GW190413_134308	0.07	GW191109_010717	−0.83
GW190421_213856	0.09	GW191127_050227	0.31
GW190503_185404	−0.04	GW191129_134029	0.33
GW190512_180714	0.33	GW191204_171526	0.79
GW190513_205428	0.48	GW191215_223052	0.11
GW190514_065416	−0.01	GW191216_213338	0.27
GW190517_055101	0.53	GW191222_033537	−0.16
GW190519_153544	0.35	GW191230_180458	0.25
GW190521	−0.26	GW200112_155838	0.07
GW190521_074359	−0.42	GW200128_022011	0.46
GW190527_092055	0.23	GW200129_065458	0.63
GW190602_175927	0.44	GW200202_154313	0.1
GW190620_030421	0.52	GW200208_130117	−0.04
GW190630_185205	−0.15	GW200209_085452	0.21
GW190701_203306	0.05	GW200216_220804	0.26
GW190706_222641	0.8	GW200219_094415	0.05
GW190707_093326	0.04	GW200224_222234	0.2
GW190708_232457	0.15	GW200225_060421	−0.11
GW190720_000836	0.58	GW200302_015811	0.05
GW190725_174728	0.39	GW200311_115853	0.32
GW190727_060333	0.44	GW200316_215756	0.57
GW190728_064510	0.32		

Table 3.1: Current GW events and their Bayes factors in favor of the up–down hypothesis over generic spin precession. We select events with false alarm rates $< 1\ \mathrm{yr}^{-1}$ in at least one of the LIGO/Virgo searches, excluding those that can potentially include a neutron star.

that the prior on φ is the same for the two models. In symbols, this is

$$\pi(\varphi|\mathcal{H}_{\rm N}) = \pi(\varphi|\gamma = \gamma_{\rm N}(\varphi), \mathcal{H}_{\rm B})\,. \tag{3.7}$$

Within these assumptions, the Bayes factor in favor of the narrow model reduces to

$$\mathcal{B} = \int \frac{p(\varphi, \gamma = \gamma_{\rm N}(\varphi)|d, \mathcal{H}_{\rm B})}{\int \pi(\varphi', \gamma = \gamma_{\rm N}(\varphi)|\mathcal{H}_{\rm B}){\rm d}\varphi'}\,{\rm d}\varphi\,. \tag{3.8}$$

A formal proof of Eq. (3.8) is presented in Appendix 3.4. For the specific case where γ_N does not depend on φ, one has

$$\mathcal{B} = \frac{p(\gamma = \gamma_{\rm N}|d, \mathcal{H}_{\rm B})}{\pi(\gamma = \gamma_{\rm N}|\mathcal{H}_{\rm B})}\,, \tag{3.9}$$

where the numerator (denominator) corresponds to the posterior (prior) marginalized over the common parameters φ. Equation (3.9) is the so-called Savage-Dickey density ratio [198]. The key, practical advantage of both these expressions is that they only depend on the broad model $\mathcal{H}_{\rm B}$. One does *not* need to perform inference in the narrow model $\mathcal{H}_{\rm N}$, which can be challenging for non-trivial submanifolds $\gamma_{\rm N}(\varphi)$. It is sufficient to sample the broad model $\mathcal{H}_{\rm B}$ and then evaluate the resulting posterior and prior probability densities at the location prescribed by the narrow model.

3.2.3 Application to up–down binaries

For the specific case we are addressing here, the broad model $\mathcal{H}_{\rm B}$ is that of generically precessing BH binaries described in Sec. 3.2.1. The narrow model $\mathcal{H}_{\rm N}$ consists of binaries in the endpoint of the up–down instability, which are subject to the three constraints

of Eqs. (3.2–3.4). From these, we define the parameters $\gamma = \{\gamma_1, \gamma_2, \gamma_3\}$, where

$$\gamma_1 = \frac{\cos\theta_1 - \cos\theta_{\rm UD}(q, \chi_1, \chi_2)}{2}\,, \tag{3.10}$$

$$\gamma_2 = \frac{\cos\theta_2 - \cos\theta_{\rm UD}(q, \chi_1, \chi_2)}{2}\,, \tag{3.11}$$

$$\gamma_3 = \frac{1}{\pi}\arctan\left(\frac{\sin\phi_{12}}{\cos\phi_{12}}\right), \tag{3.12}$$

and

$$\cos\theta_{\rm UD}(q, \chi_1, \chi_2) = \frac{\chi_1 - q\chi_2}{\chi_1 + q\chi_2}\,. \tag{3.13}$$

While not unique, we find this parametrization convenient because all the γ_i are defined[2] in $[-1, 1]$ and the up–down endpoint is mapped to $\gamma = (0, 0, 0)$. We apply the transformations of Eqs. (3.10–3.12) to both prior and posterior samples, estimate the corresponding probability density functions using three-dimensional Kernel Density Estimation (KDE), and evaluate the Bayes factor from Eq. (3.9). We use Gaussian kernels and a bandwidth of 0.2 [199].

An example of this procedure is shown in Fig. 3.1. We consider a synthetic source in the endpoint of the up–down instability with tilt angles $\cos\theta_1 = \cos\theta_2 = \cos\theta_{\rm UD} = 0.103$ and $\phi_{12} = 0$. The injected system has $m_1 = 49.5M_\odot$ $m_2 = 39.4M_\odot$, $\chi_1 = 0.92$, $\chi_2 = 0.94$, $D_L = 845$ Mpc, $\theta_{JN} = 0.37$, $\phi_{JL} = 5.71$, $\alpha = 6.11$, $\delta = 0.24$, $\psi = 2.28$, $t_c = -0.069$ s (in GPS time), and $\phi_c = 5.12$. The prior and posterior KDEs are evaluated at the origin of the $\{\gamma_1, \gamma_2, \gamma_3\}$ cube (black lines in Fig. 3.1). The Savage-Dickey estimate of the Bayes factor is $\ln\mathcal{B} = 5.11$. For equal priors, this corresponds to strong evidence that the source is indeed in the up–down endpoint. Figure 3.1 also shows that the posteriors of the rescaled parameters γ_i are somewhat close to a multivarate Gaussian distribution; this not a generic feature but rather a consequence of the relatively high signal-to-noise ratio (SNR), which for this specific injection is 60.

[2]The trigonometric manipulation in Eq. (3.12) is necessary because $\phi_{12} \in [0, 2\pi]$.

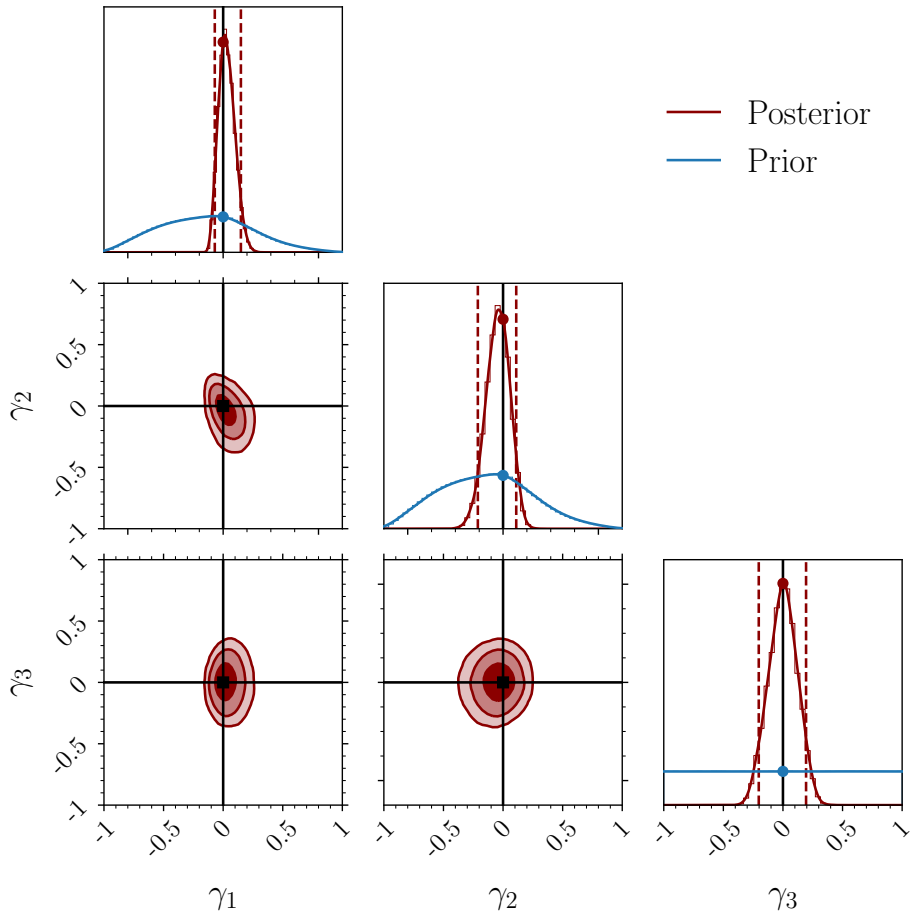

Figure 3.1: Joint posterior distributions of the rescaled parameters $\gamma = \{\gamma_1, \gamma_2, \gamma_3\}$ defined in Eqs. (3.10–3.12). Contour levels correspond to 50%, 90%, and 99% credible regions. Red dashed lines in the 1D marginals indicate the 90% credible intervals. Solid black lines mark the location of the narrow model $\gamma = 0$, i.e., the endpoint of the up–down instability. Black scatter points indicate the value of the posterior (red) and prior (blue) distributions at the endpoint, which are the key ingredients entering the Savage-Dickey evaluation of the Bayesian odds.

3.3 Results

3.3.1 Comparing posteriors

Before reporting Bayes factors, it is informative to compare posterior distributions against the predictions of Eqs. (3.2–3.4). This a preliminary step which is often used to

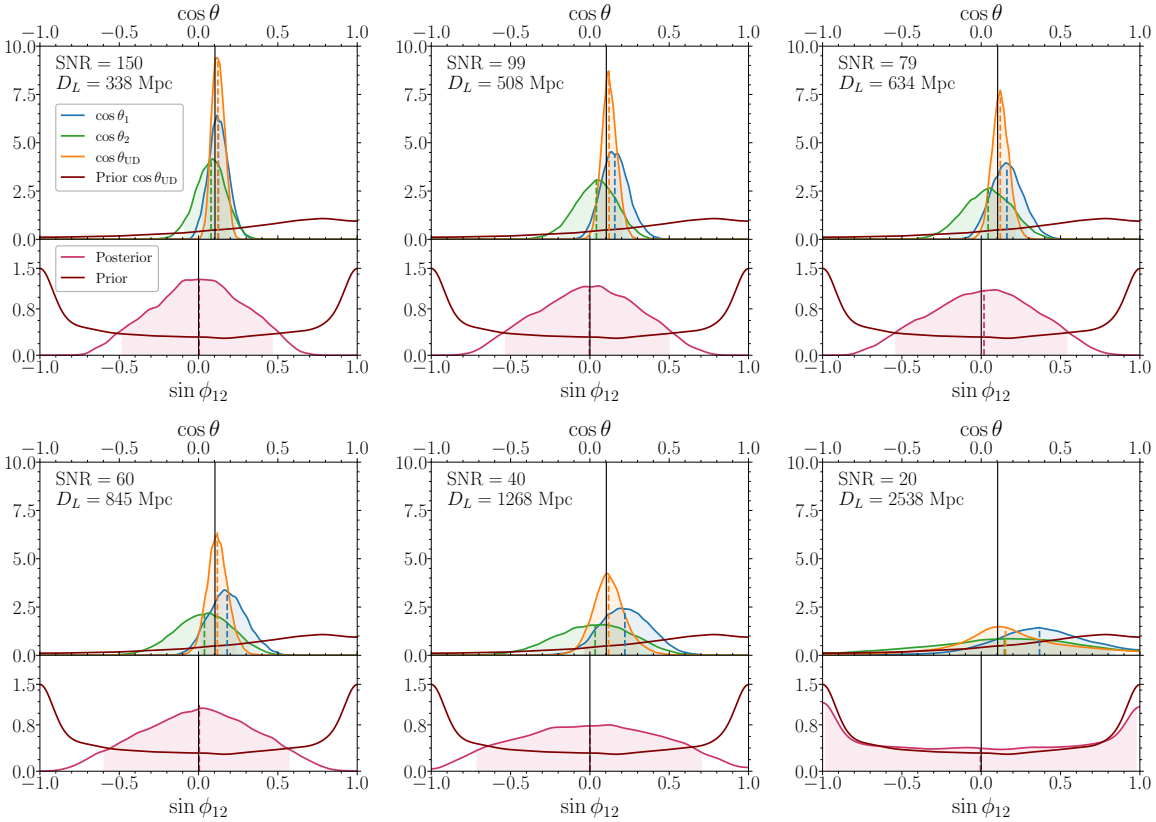

Figure 3.2: Panels from left to right and from top to bottom show parameter-estimation results for the same GW source injected at decreasing (increasing) values of the SNR (luminosity distance D_L). The upper subpanels show posterior distributions of $\cos\theta_1$ (blue), $\cos\theta_2$ (green), $\cos\theta_{\rm UD}$ (orange) and the prior distribution of $\cos\theta_{\rm UD}$ (dark red); the prior distributions of $\cos\theta_{1,2}$ are flat. The lower subpanels show posterior (pink) and prior (dark red) distributions of $\sin\phi_{12}$. Black vertical lines indicate the injected values. Dashed vertical lines mark the medians of each distribution while shaded areas indicate the 90% credible intervals.

identify promising candidates for a model-selection analysis.

We consider six synthetic signals describing binary BHs that are in the endpoint of the up–down instability when entering the LIGO band at the reference frequency of 20 Hz. We use the same set of source parameters as in Fig. 3.1. In particular, we fix the detector-frame masses and inject source waveforms with SNR $= 150, 100, 80, 60, 40, 20$, corresponding to luminosity distances $D_L = 338, 508, 634, 845, 1268, 2538$ Mpc. The

PN orbital separation of the binary at $f_{\rm ref} = 20$ Hz is $r_{\rm 20Hz} \simeq 10M$, while the critical separation for the instability is $r_{\rm UD+} = 266M$. The condition $r_{\rm UD+} - r_{\rm 20Hz} > 100M$ ensures that the predicted endpoint well describes these unstable up–down sources [196].

Our results are shown in Fig. 3.2, where each panel correspond to a different source. The upper subpanels compare the posterior distributions of $\cos\theta_{1,2}$ (as obtained from our parameter-estimation analysis) against that of $\cos\theta_{\rm UD}$ [as obtained from substituting the posterior samples of q, χ_1, χ_2 into Eq. (3.13)]. Note how the prior distribution of $\cos\theta_{\rm ud}$ peaks toward positive values, while those of $\cos\theta_{1,2}$ are flat. Close agreement between the posteriors of $\cos\theta_1$, $\cos\theta_2$, and $\cos\theta_{\rm UD}$ provide a qualitative (but not quantitative) indication that the theoretical prediction of the up–down instability is a reasonable description of the data. The lower subpanels report the posterior distribution of $\sin\phi_{12}$, where values close to zero 0 indicate a preference for the up–down hypothesis.

As expected, posteriors for the lowest SNRs tend to cover a large portion of prior range. As the SNR increases, the recovered posteriors approach the injected values that define the endpoint of the up–down instability. In particular, for the case of the highest SNR $= 150$, we find $\cos\theta_{\rm UD} = 0.122^{+0.068}_{-0.061}$ and $\phi_{12} = 0.004^{+0.460}_{-0.485}$ (where we quote the median and 90% credible interval), compared to the injected values $\cos\theta_{\rm UD} = 0.103$ and $\phi_{12} = 0$.

Note that systematic effects are not captured in both these results and the rest of the chapter because we perform zero-noise runs and use the same waveform model for both injection and recovery. Waveform systematics in the specific region of parameter space where the up–down instability take place still need to be investigated. Based on the discussion in Section 2.3.4, the IMRPhenomXPHM waveform model may encounter limitations in accurately modeling systems with high precession. Such scenarios typically arise for mass ratios close to unity, high spin magnitudes, and tilt angles near to 90 degrees.

We further note a common feature that characterize all cases shown in Fig. 3.2,

including those at low SNR. While the recovered values of $\cos\theta_{1,2}$ depart from the injected values as the SNR decreases, the medians of $\cos\theta_{\rm UD}$ tend to remain closer to that of the injected endpoint. This seems to indicate that, if the source is truly in the endpoint of the up–down instability, the estimator $\cos\theta_{\rm UD}$ might be more accurate than $\cos\theta_{1,2}$. We interpret this as a consequence of more accurate measurements of q and $\chi_{1,2}$ compared those of the spin tilts. This implies we can measure what the endpoint of a binary *would* be from the q–χ posteriors. However, inferring that the given source is in fact in its endpoint requires computation of posterior odds.

3.3.2 Model selection

While comparing posteriors as in Fig. 3.2 provides a useful indication of a potential up–down signature, this statement needs to be quantified with a full Bayesian model selection. For the same series of six injections, Fig. 3.3 shows the Bayes factor in favor of the up–down hypothesis over that of generic BH binaries computed using the Savage-Dickey density ratio (orange points). The Bayes factor increases from $\ln\mathcal{B}\sim 1.96$ for SNR $= 20$ (weak evidence) to $\ln\mathcal{B}\sim 6.89$ for SNR $= 150$ (strong evidence). While this is a controlled experiment where the true source parameters are injected in the up–down configuration, the successful recovery of a large value of $\mathcal{B}$ indicates that data are informative about this property in a concrete measurement setting.

We repeat the same study for six additional series of BH binaries in the up–down endpoint with different parameters θ (gray points) which are part of the broader set of injections described in Sec. 3.3.4. As expected, the Bayes factor increases with the SNR in all cases, though the overall normalization depends on the other source parameters. For the case discussed above and shown with orange scatter points, a strong evidence in favor of the nested model is achieved at SNR $\gtrsim 60$ —values within reach of next LIGO-Virgo observing run [44]. However, this is not generic. We find that the distinguishability power critically depends on the source parameters. Even among

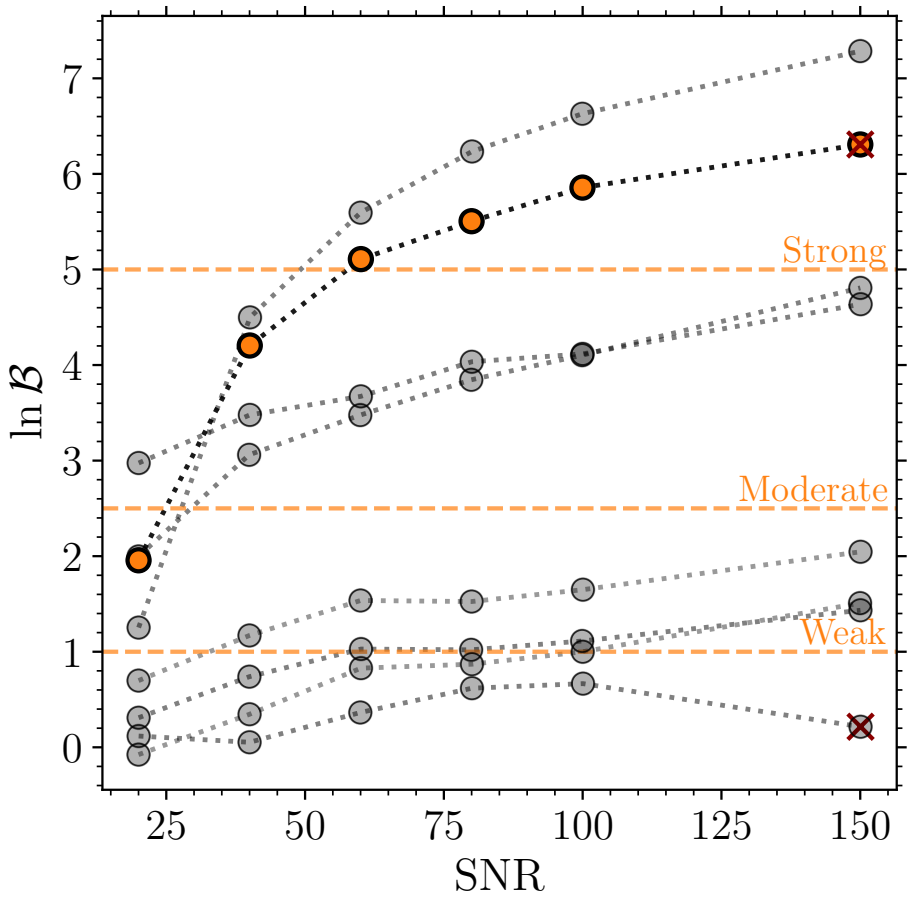

Figure 3.3: Natural logarithm of the Bayes factor in favor of the up–down hypothesis as a function of the SNR. We consider the same sources as in Fig. 3.2 (orange scatter points) as well as six other series of BH binaries in the up–down endpoint (gray scatter points). Horizontal dashed lines indicate the threshold values of the Bayes factor in the Jeffrey scale. Crosses indicate the sources shown in Fig. 3.4.

this limited set, there are cases that provide only weak or even inconclusive evidence even at SNR = 150.

3.3.3 Backpropagation

We can further visualize the up–down signature of BH binaries by back-propagating posteriors samples [186, 191]. If a detected source is truly an unstable up–down binary,

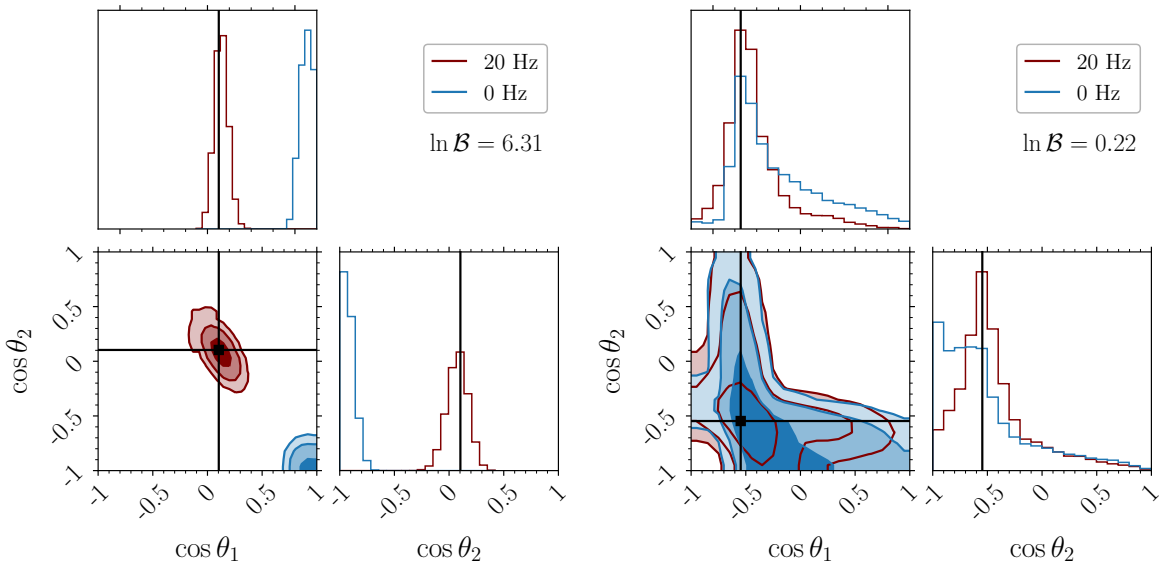

Figure 3.4: Joint posterior distribution of the tilt angles θ_1 and θ_2 for the sources described in Sec. 3.3.3 and marked with crosses in Fig. 3.3. The left (right) panel shows a case that presents strong (inconclusive) evidence in favor of the up–down hypothesis. Posterior samples are evolved numerically from $f_{\rm ref} = 20$ Hz (red) to 0 Hz (blue). Solid black lines indicate the injected values. Contour levels mark the 50%, 90%, and 99% credible regions.

evolving it backward in time should allow us to see it in the up–down spin configuration instead of the particular precessing configuration as observed. For a given injection, we numerically evolve each posterior sample backward from detection at $f_{\rm ref} = 20$ Hz to past-time infinity at $f_{\rm ref} = 0$ Hz using precession-averaged PN equations as implemented in Refs. [200, 201]. This procedure requires q, $\chi_{1,2}$, $\theta_{1,2}$, ϕ_{12}, and r at $f_{\rm ref} = 20$ Hz as inputs and returns the values of the tilt angles θ_{12} at 0 Hz (ϕ_{12} does not enter the dynamics at infinitely large orbital separations [164, 201]).

Figure 3.4 shows two examples which were selected from those of Fig. 3.3. Both sources have SNR = 150; one provides strong evidence in favor of the up–down endpoint (left panel, $\ln \mathcal{B} = 6.31$) while the other returns an inconclusive result (right panel, $\ln \mathcal{B} = 0.22$). The parameters of the former are listed in Sec. 3.2.3 while those of the latter are $m_1 = 26 M_\odot$, $m_2 = 26 M_\odot$, $\chi_1 = 0.17$, $\chi_2 = 0.57$, $\theta_{12} = 2.15$, $\phi_{12} = 0$,

$D_L = 190.06$ Mpc, $\psi = 2.89$, $\phi = 3.33$, $\alpha = 3.78$, $\delta = -0.081$, $\theta_{JN} = 0.41$, $\phi_{JL} = 3.71$, and $t_c = -0.01$ s.

For the binary with large $\mathcal{B}$ (left panel in Fig. 3.4), the posterior distribution at 0 Hz is constrained to be close to an aligned binary with up–down spins. In particular, we find $\cos\theta_1 > 0.80$ and $\cos\theta_2 < -0.99$ at 90% confidence. This result is an additional, visual indication that data taken at ~ 20 Hz are well described by a BH binary that *used to be* aligned but is being observed precessing.

On the other hand, for the inconclusive case (right panel in Fig. 3.4), the joint distribution of $\cos\theta_1$ and $\cos\theta_2$ at $f_{\rm ref} = 0$ Hz occupies a much broader region of the prior volume ($\cos\theta_1 > -0.65$ and $\cos\theta_2 < 0.41$ at 90% confidence). As indicated by the Bayes factor, this is a source where data are compatible with a variety of precessing configurations, some that did and some that did not form with up–down spin directions.

3.3.4 Injection campaign

We now investigate the distinguishability of up–down sources in a wider region of the parameter space. We construct a set of injections by drawing binaries from the standard uninformative priors; we sample q and $\chi_{1,2}$ and enforce $\cos\theta_{1,2}$ and ϕ_{12} from Eqs. (3.2–3.4). We then impose the following constraints:

(i) We only consider binaries with $r_{\rm UD+} - r_{\rm 20Hz} > 200M$, which is a conservative condition to ensure that the analytical instability endpoint well describes binaries that formed in the up–down configuration.

(ii) We further require sources to have SNR > 20, thus adopting a threshold that is about twice the current detection limit [62, 63, 147, 148]. Spin effects are known to be challenging to measure [180, 181, 134] and the model-selection problem tackled here inevitably requires loud signals.

Our results are shown in Fig. 3.5, where we report the Bayes factor as a function of

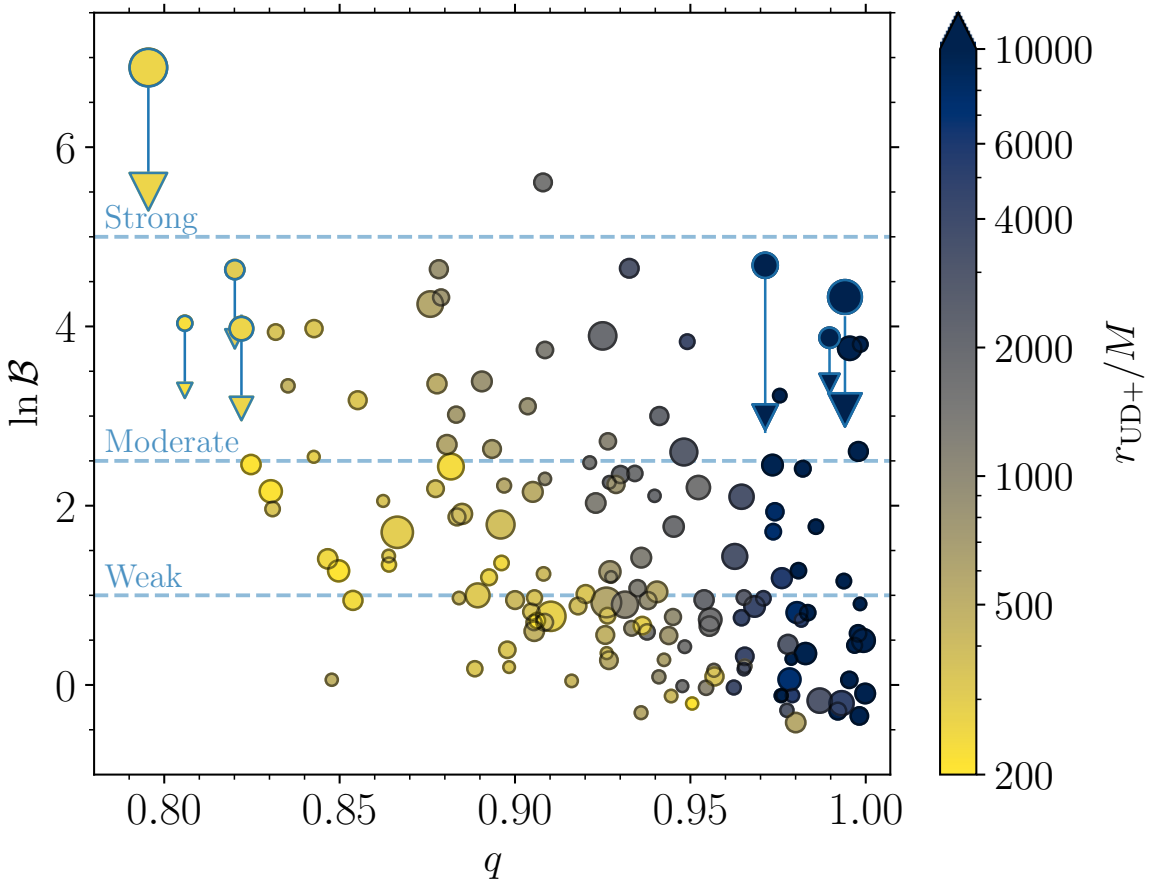

Figure 3.5: Natural logarithm of the Bayes factor $\mathcal{B}$ as a function of the mass ratio q for a set of 151 GW signals injected in the endpoint of the up–down instability. The critical orbital separation $r_{\mathrm{UD+}}$ is reported on the color bar and the size of each scatter point is directly proportional to the three-detector SNR. Horizontal dashed blue lines correspond to the threshold values of the Jeffrey scale for weak, moderate, and strong evidence. The scatter points connected by vertical lines are sources that were injected and recovered both with (upper markers, circles) and without (lower markers, triangles) higher-order modes.

the mass ratio q, the critical separation $r_{\mathrm{UD+}}$, and the SNR. It is immediate to note that all injections have mass ratios $q \gtrsim 0.8$; this is a direct consequence of selecting binaries with a large value of $r_{\mathrm{UD+}} \propto (1-q)^{-2}$ [cf. Eq. (3.1)] and is largely independent of the total mass M which only enters the source-frame/detector-frame conversion of

the frequency.

Among the 151 sources we select, we find that 31 present inconclusive evidence in favor of the up–down origin, 45 sources present weak evidence, 73 present moderate evidence, and 2 present strong evidence (recall that we are assuming equal model priors such that the posterior odds and the Bayes factor coincide).

We find a broad trend indicating that binaries with more unequal masses tend to have larger Bayes factors while binaries with close-to-equal masses cover a larger range of Bayes factors. The value of q is closely correlated with $r_{\rm UD+}$ from Eq. (3.1), which implies that pinpointing the up–down origin of binaries with lower values of the critical separation $r_{\rm UD+}$ is going to be somewhat easier (as long as $r_{\rm UD+}$ is still sufficiently large that the analytical endpoint provides a reasonable prediction, see above).

Figure 3.6 shows Bayes factors and SNRs for the same set of injections (blue triangles). As expected the two are positively correlated (cf. Fig. 3.3), though with a large dispersion, including several loud sources that still return an inconclusive model selection. Even SNRs as large as ~ 200 do not guarantee a decisive model selection result since the value of $\mathcal{B}$ strongly depends on the specific parameters of the source.

A key ingredient to this analysis is the inclusion of higher-order emission modes in the adopted waveform model. Higher harmonics can break degeneracies between the mass and spin parameters [202, 203, 204, 89], thus aiding our model selection problem. We further investigate this point by considering seven sources among those with the smaller and larger values of q from our set and repeat their analysis without higher-order modes. As expected, we find that the the Bayes factor decreases, with differences (in logarithmic scale) that are up to ~ 1.5.

3.3.5 Current gravitational-wave data

Finally, we apply our model-selection analysis as described in Sec. 3.2.2 to current GW events reported up to GWTC-3. We analyze the 69 binary BH coalescences listed

in Table 3.1 (see Sec. 3.2.1).

Figure 3.6 (orange circles) compares the Bayes factor and the source SNR (estimated using the median of the optimal network SNR posterior samples).

The Bayes factor in favor of the up–down hypothesis for current GW signals lies within the range $\ln \mathcal{B} \in [-0.8, 0.8]$, which is inconclusive. None of the current events support the up–down endpoint model, but they do not allows us to exclude it either. This is somewhat expected given that SNRs of current event are $\lesssim 30$, which is unlikely to provide meaningful constraints (cf. Fig. 3.3 and 3.6). Our finding agrees with previous analyses [62, 63, 147, 148] indicating that current data provide loose constraints on the orientations of individual BH spins, which in turn are key ingredients in the up–down model selection problem. We conclude that the current catalog of GW events does not contain promising up–down candidates.

At the same time, we note that the Bayes factor for the entire observed catalog $\sum_i \ln B_i \simeq 15$ shows a preference for the narrow hypothesis $\mathcal{H}_{\rm N}$. Properly quantifying the astrophysical relevance of this finding requires a deeper investigation on the systematics of the single-event B_i's as well as additional population modeling to include selection effects.

3.4 Final remarks

In this work, we performed parameter estimation of BH binaries that have encountered the up–down instability [189]. Binaries that are formed with the spin of the heavier (lighter) BH aligned (anti-aligned) with the orbital angular momentum might enter the LIGO/Virgo band with significant spin precession. Their final configuration (i.e., the endpoint of the up–down instability) can be computed in closed form [196] and allows us to test the up–down origin of precessing binary BHs. More ambitiously, one could also target up–down binaries as they become unstable (i.e. $r = r_{\rm UD+}$) and start precessing. While worthy of further investigation, the rate of these events is presumably

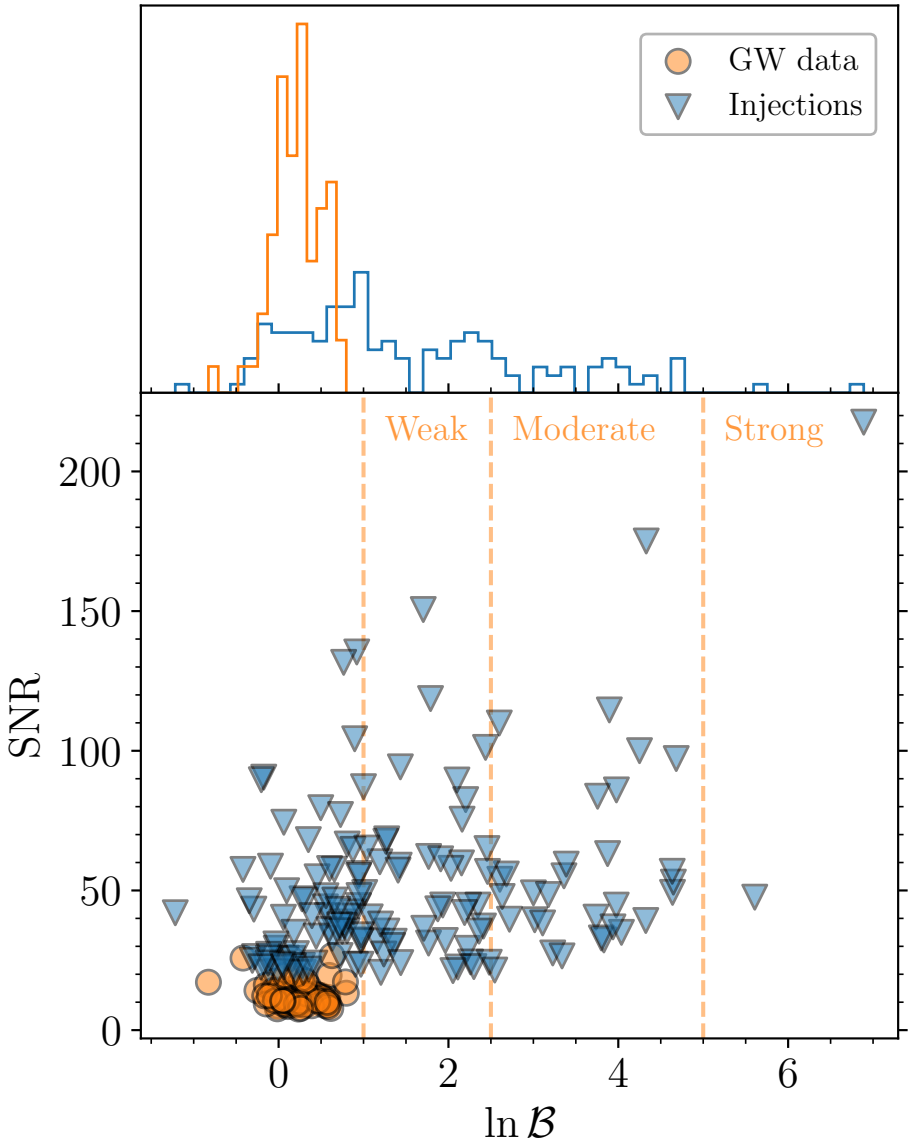

Figure 3.6: Natural logarithm of the Bayes factor as a function of the SNR for 151 simulated sources (blue triangles) and 69 GW events from GWTC-3 (orange circles). Vertical dashed orange lines indicate the threshold values of the Jeffrey scale for weak, moderate, and strong evidence. The upper panel shows an histogram of the Bayes factors.

very low.

We presented a statistical approach based on the Savage-Dickey density ratio for the

calculation of the Bayes factor and applied it to both simulated signals (which act as a control set) and current GW events. The identification of unstable up–down binaries depends on the source SNR, with higher-order emission modes providing an important contribution. At least within the limited set of injections performed here, we find that SNRs greater than ~ 100 are required. However, this is a necessary but not sufficient condition for the up–down origin to be distinguishable, as the resulting posterior odds strongly depends on the source parameters. Our model selection analysis is slightly more discriminative for sources with unequal masses and, consequently, with smaller values of $r_{\mathrm{UD}+}$. Posterior samples for all the injections presented in this chapter are publicly available at github.com/ViolaDeRenzis/updowninjections [205].

Among the current LIGO/Virgo events, we do not find promising candidates that could be interpreted as binary systems that were originally aligned in the up–down configuration. This result is not surprising, given the present SNRs which are $\lesssim 30$.

Future LIGO/Virgo upgrades as well as new facilities will largely increase the available statistical sample [187, 44]. The methodology developed in this work provides a straightforward, post-processing operation that can be performed on posterior samples from future GW catalogs. Looking ahead, testing the up–down hypothesis is particularly relevant in the context of supermassive BH binaries observed by LISA. Some of those sources are expected to have SNRs as large as ~ 3000 [47] and their spins might be brought to the up–down configuration by interactions with galactic-scale accretion disks [195, 206, 207].

A future detection of the up–down instability presents the opportunity to confirm this prediction of the general-relativistic two-body problem.

Appendix

Savage-Dickey density ratio

Following the notation introduced in Sec. 3.2.1, let us assume that we have some observed data d and two hypotheses such that

$$\mathcal{H}_{\rm N} : \mathcal{H}_{\rm B} \wedge \gamma = \gamma_{\rm N}(\varphi)\,. \tag{3.14}$$

With this definition, the evidence of the narrow model is

$$\begin{aligned}\mathcal{Z}(d|\mathcal{H}_N) &= \int \mathcal{L}(d|\varphi, \mathcal{H}_{\rm N})\, \pi(\varphi|\mathcal{H}_{\rm N}) \mathrm{d}\varphi \\ &= \int \mathcal{L}(d|\varphi, \gamma{=}\gamma_{\rm N}(\varphi), \mathcal{H}_{\rm B})\, \pi(\varphi|\gamma{=}\gamma_{\rm N}(\varphi), \mathcal{H}_{\rm B})\, \mathrm{d}\varphi\,.\end{aligned} \tag{3.15}$$

One can manipulate the first term in the integrand using Bayes' theorem,

$$\mathcal{L}(d|\varphi, \gamma{=}\gamma_{\rm N}(\varphi), \mathcal{H}_{\rm B}) = \frac{p(\varphi, \gamma{=}\gamma_{\rm N}(\varphi)|d, \mathcal{H}_{\rm B})\mathcal{Z}(d|\mathcal{H}_{\rm B})}{\pi(\varphi, \gamma{=}\gamma_{\rm N}(\varphi)|\mathcal{H}_{\rm B})}\,, \tag{3.16}$$

and write the Bayes factor in favor of the narrow model as

$$\begin{aligned}\mathcal{B} &= \frac{\mathcal{Z}(d|\mathcal{H}_{\rm N})}{\mathcal{Z}(d|\mathcal{H}_{\rm B})} \\ &= \int \mathrm{d}\varphi\, p(\varphi, \gamma{=}\gamma_{\rm N}(\varphi)|d, \mathcal{H}_{\rm B}) \frac{\pi(\varphi|\gamma{=}\gamma_{\rm N}(\varphi), \mathcal{H}_{\rm B})}{\pi(\varphi, \gamma{=}\gamma_{\rm N}(\varphi)|\mathcal{H}_{\rm B})}.\end{aligned} \tag{3.17}$$

The rule of conditional probability implies

$$\begin{aligned}\frac{\pi(\varphi, \gamma{=}\gamma_{\rm N}(\varphi)|\mathcal{H}_{\rm B})}{\pi(\varphi|\gamma{=}\gamma_{\rm N}(\varphi), \mathcal{H}_{\rm B})} &= \pi(\gamma{=}\gamma_{\rm N}(\varphi)|\mathcal{H}_{\rm B}) \\ &= \int \pi(\varphi', \gamma{=}\gamma_{\rm N}(\varphi)|\mathcal{H}_{\rm B}) \mathrm{d}\varphi'\,,\end{aligned} \tag{3.18}$$

where in the second equality we have explicitly indicated the marginalization over the common parameters. This yields

$$\mathcal{B} = \int \frac{p(\varphi, \gamma{=}\gamma_{\rm N}(\varphi)|d, \mathcal{H}_{\rm B})}{\int \pi(\varphi', \gamma{=}\gamma_{\rm N}(\varphi)|\mathcal{H}_{\rm B}){\rm d}\varphi'}\, {\rm d}\varphi\,, \tag{3.19}$$

which is equal to Eq. (3.8).

The Savage-Dickey density ratio is recovered by a suitable change of variables:

$$\{\varphi, \gamma\} \longrightarrow \{\bar{\varphi} = \varphi, \bar{\gamma}{=}\gamma - \gamma_{\rm N}(\varphi)\}\,. \tag{3.20}$$

The determinant of the resulting Jacobian is

$$\det \begin{pmatrix} \partial\bar{\varphi}/\partial\varphi & \partial\bar{\varphi}/\partial\gamma \\ \partial\bar{\gamma}/\partial\varphi & \partial\bar{\gamma}/\partial\gamma \end{pmatrix} = \det \begin{pmatrix} 1 & 0 \\ -{\rm d}\gamma_{\rm N}/{\rm d}\varphi & 1 \end{pmatrix} = 1 \tag{3.21}$$

such that, for any probability distribution P, one can simply write

$$P(\varphi, \gamma{=}\gamma_{\rm N}(\varphi)) = P(\varphi, \bar{\gamma} = 0)\,. \tag{3.22}$$

With this transformation, Eq. (3.19) reduces to

$$\mathcal{B} = \frac{\int p(\varphi, \bar{\gamma} = 0|d, \mathcal{H}_{\rm B}){\rm d}\varphi}{\int \pi(\varphi', \bar{\gamma} = 0|\mathcal{H}_{\rm B}){\rm d}\varphi'} = \frac{p(\bar{\gamma} = 0|d, \mathcal{H}_{\rm B})}{\pi(\bar{\gamma} = 0|\mathcal{H}_{\rm B})}\,, \tag{3.23}$$

as reported in Eq. (3.9), see also Ref. [208].

Chapter 4

Forecasting the population properties of merging black holes

Overview

Third-generation gravitational-wave detectors will observe up to millions of merging binary black holes. With such a vast dataset, stacking events into population analyses will arguably be more important than analyzing single sources. We present the first application of population-level Fisher-matrix forecasts tailored to third-generation gravitational-wave interferometers. We implement the formalism first derived by Gair *et al.* [MNRAS 519, 2736 (2023)] and explore how future experiments such as Einstein Telescope and Cosmic Explorer will constrain the distributions of black-hole masses, spins, and redshift. Third-generation detectors will be transformative, improving constraints on the population hyperparameters by several orders of magnitude compared to current data. At the same time, we highlight that a single third-generation detector of triangular shape and a network of separated detectors will deliver qualitatively similar performances. Obtaining precise measurements of some population features (e.g. peaks

in the mass spectrum) will require only a few months of observations while others (e.g. the fraction of binaries with aligned spins) will instead require years if not decades. We argue population forecasts of this kind should be featured in white papers and feasibility studies aimed at developing the science case of future gravitational-wave interferometers.

The material contained in this Chapter is based on Ref. [136].

4.1 Introduction

Gravitational-wave (GW) astronomy is quickly turning into a big-data discipline. We now have hundreds of detections, which will become thousands in a few years and millions in a few decades. The focus of GW inference is thus shifting from inferring the parameters of single events to those of the entire population of detected sources.

Current population analyses of LIGO/Virgo data [209, 128, 127] are providing precious information on the physical processes that drive black holes (BHs) and neutron stars to merger [115, 188, 194]. In particular, there is strong indication that the merger rate of stellar-mass BHs falls to $\sim$zero for sources of masses $\gtrsim 60\ M_\odot$ but piles up at $\sim 35\ M_\odot$. The spins of merging stellar-mass BHs show some preference for dimensionless magnitudes $\lesssim 0.3$ and some degree of co-alignment with the orbital angular momentum of the respective binaries. Besides marginal distributions, the growing GW dataset is now allowing us to explore correlations between parameters [210, 211, 212, 213, 214, 215, 216].

State-of-the-art population analyses are tackled using hierarchical Bayesian statistics [217, 96]. In short, one assumes an underlying inhomogeneous Poisson process and marginalizes over the parameters of the single events (e.g. masses, spins, etc.) while retaining information on the "hyperparameters" that describe the entire population of sources (e.g. mass cutoffs, spectral indexes, etc.). Current state-of-the-art implementations rely on recycling single-event posterior distributions for the computation of marginal integrals, as well as on injection campaigns to estimate selection biases via

reweighted Monte Carlo integration [218, 219]. This requires careful handling of both finite-sampling and selection effects [220]. While under control for the time being, the feasibility of these techniques with much larger catalogs remains an open question.

The future of ground-based GW astronomy is under active planning, with flagship projects such as Einstein Telescope (ET) [221] and Cosmic Explorer (CE) [46] that promise to observe virtually all the BH binaries merging in the visible Universe, with detection rates of $\mathcal{O}(10^5/\mathrm{yr})$ [222, 223]. At present, running full Bayesian schemes on such large catalogs is both too costly and premature, if feasible at all. Science forecasts largely rely on the Fisher information matrix formalism [103], which approximates the single-event likelihood in the high signal-to-noise ratio (SNR) limit. Fisher codes tailored to third-generation (3G) detectors include GWFast [104], GWBench [105], GWFish [106], TiDoFM [224], and the code used in Ref. [225]. All such codes provide forecasts for the parameter estimation of single events rather than of population parameters.

On the other hand, in the third-generation (3G) era the focus of GW science will definitely be on populations. With the exception of prominent outliers, it is easy to image how the analysis of a single event will not be relevant when there are millions of similar sources sitting next to it. Despite this quite obvious statement, current forecasts presented in "colored books" to funding agencies for the identification of the ET and CE science cases still rely on single-event analyses [226, 227, 228, 223]. We believe this is a severe limitation.

In this chapter of the thesis, we present the first population-level Fisher matrix implementation tailored to 3G detectors. The relevant equations have been developed by Gair *et al.* [229], who first wrote down the Fisher-like expansion of the population likelihood including selection effects. Their expressions have so far only been applied to toy models. Here we significantly broaden the scope of the analysis to state-of-the-art parameterized population models, presenting detailed predictions for 3G detectors. We argue forecasts of this kind should be featured in future design documents to

appropriately assess the outstanding science we will be able to deliver with facilities of the caliber of ET and CE.

Chapter 4 is organized as follows. Section 4.2 introduces the Fisher formalism for both individual-event and population inference. Section 4.3 provides details on the implementation of our code, alongside a description of the population models used throughout this study. Section 4.4 presents our population-inference results, with a particular focus on the detector network as well as the astrophysical distributions of masses, spins, and redshift. Finally, Sec. 4.5 summarizes our main findings and discusses their implications.

4.2 Formalism

In the following, d indicates the data of a single GW signal, $\{d\}$ indicates data of a set of signals, θ indicates the parameters of individual events, and λ indicates the hyperparameters describing the populations.

4.2.1 Single-event inference

Under the assumption that the noise is stationary and Gaussian, the likelihood of the event parameters is given by [230]

$$p(d|\theta) \propto \exp\left[-\frac{1}{2}\big(d-h(\theta)|d-h(\theta)\big)\right], \tag{4.1}$$

where $h(\theta)$ is the gravitational signal and $(\cdot|\cdot)$ indicates the noise-weighted inner product. For a given BH binary with true parameters $\bar{\theta}$, the elements of the Fisher information matrix are given by

$$\Gamma_{\theta,ij} \equiv -\left\langle \frac{\partial^2 \ln p(d|\theta)}{\partial\theta_i \partial\theta_j} \right\rangle \Bigg|_{\theta=\bar{\theta}} = \left(\frac{\partial h}{\partial\theta_i} \middle| \frac{\partial h}{\partial\theta_j} \right) \Bigg|_{\theta=\bar{\theta}}, \tag{4.2}$$

where the expectation value $\langle\cdot\rangle$ is taken over noise realizations. The covariance matrix is given by Γ_θ^{-1} such that the errors on each parameter θ_i are given by the diagonal elements $(\Gamma_\theta^{-1})_{ii}^{1/2}$. The Fisher matrix provides a Gaussian approximation to the likelihood in the high-SNR limit. This is equivalent to approximating the posterior distribution $p(\theta|d)$ if one assumes that the prior over θ is flat across the support of the likelihood itself.

4.2.2 Population inference

The chosen population model is encoded in the probability distribution $p_{\rm pop}(\theta|\lambda)$ that predicts the likelihood of compact objects merging with parameters θ for a given choice of the population λ. The population likelihood marginalized over the event rate is given by [217, 96]

$$p\left(\{d\}|\lambda\right) \propto p_{\rm det}(\lambda)^{-N_{\rm det}} \prod_{k=1}^{N_{\rm det}} \int p(d_k|\theta) p_{\rm pop}(\theta|\lambda)\, {\rm d}\theta\,, \tag{4.3}$$

where $N_{\rm det}$ is the number of detected events in the catalog. While state-of-the-art, this expression assumes that GW events are statistically independent. This is an excellent approximation for LIGO/Virgo but a poorer one for ET and CE where multiple signals might overlap in the same stretch of data [231, 232]. Performing GW population analyses with overlapping signals is an unsolved problem.

The term $p_{\rm det}(\lambda)$ in Eq. (4.3) is the fraction of detected events in the population described by λ, i.e.

$$p_{\rm det}(\lambda) = \frac{N_{\rm det}}{N} = \int p_{\rm det}(\theta) p_{\rm pop}(\theta|\lambda)\, {\rm d}\theta \leq 1\,, \tag{4.4}$$

where N is the total number of mergers and $p_{\rm det}(\theta)$ is the probability of observing a merging binary with parameters θ, which in turn encodes a model for the detector. For simplicity and as common practice in the field, we use the SNR as a ranking statistics

and set

$$p_{\rm det}(\theta) = \frac{1}{2}\left\{1 + {\rm erf}\left[\frac{\rho(\theta) - \rho_{\rm th}}{\sqrt{2}}\right]\right\}, \tag{4.5}$$

where $\rho_{\rm th}$ is a fixed threshold and $\rho(\theta) = \sqrt{(h(\theta)|h(\theta))}$ is the optimal SNR. Equation (4.5) includes the variance of the SNR due to noise realizations but not the effect of multiple detectors [233]. We neglect this subtlety in favor of an efficient numerical implementation and use Eq. (4.5) also for detector networks, adding the SNRs of each instrument in quadrature. In particular, we set $\rho_{\rm th} = 12$.

The Fisher-matrix approximation of the likelihood in Eq. (4.3) has been worked out in Ref. [229]. Assuming a true population described by $\bar{\lambda}$, the elements of the hyper-Fisher matrix are

$$\Gamma_{\lambda,ij} \equiv -\left\langle \frac{\partial^2 \ln p\left(\{d\}|\lambda\right)}{\partial\lambda^i \partial\lambda^j} \right\rangle \Bigg|_{\lambda=\bar{\lambda}}, \tag{4.6}$$

which we write as a sum of five leading contributions[1]

$$\Gamma_{\lambda,ij} = N\left(\Gamma_{{\rm I},ij} + \Gamma_{{\rm II},ij} + \Gamma_{{\rm III},ij} + \Gamma_{{\rm IV},ij} + \Gamma_{{\rm V},ij}\right), \tag{4.7}$$

where

$$\Gamma_{{\rm I},ij} = -\int \frac{\partial^2 \ln[p_{\rm pop}(\bar{\theta}|\lambda) p_{\rm det}^{-1}(\lambda)]}{\partial\lambda_i \partial\lambda_j}\Bigg|_{\lambda=\bar{\lambda}} p_{\rm det}(\bar{\theta}) p_{\rm pop}(\bar{\theta}|\bar{\lambda}) {\rm d}\bar{\theta}\,, \tag{4.8}$$

$$\Gamma_{{\rm II},ij} = \frac{1}{2}\int \frac{\partial^2 \ln\det(\Gamma_\theta + H)}{\partial\lambda_i \partial\lambda_j}\Bigg|_{\lambda=\bar{\lambda}} p_{\rm det}(\bar{\theta}) p_{\rm pop}(\bar{\theta}|\bar{\lambda}) {\rm d}\bar{\theta}\,, \tag{4.9}$$

$$\Gamma_{{\rm III},ij} = -\frac{1}{2}\int \frac{\partial^2 \left[(\Gamma_\theta + H)^{-1}_{kl}\right]}{\partial\lambda_i \partial\lambda_j}\Bigg|_{\lambda=\bar{\lambda}} \Gamma_{\theta,kl}\, p_{\rm pop}(\bar{\theta}|\bar{\lambda}) {\rm d}\bar{\theta}\,, \tag{4.10}$$

[1]The symbol Γ_λ here indicates the hyper-Fisher matrix and not its single-event contribution as in Ref. [229]. With this notation, the components of Γ_λ^{-1} directly provide the hyperparameters errors and correlations for an astrophysical population of N events. Compared to the derivation presented in Ref. [229], this is equivalent to absorbing a factor of N into the definition of Γ_λ and a factor of $p_{\rm det}^{-1}(\lambda)$ into the definitions of $\Gamma_{\rm I-V}$.

$$\Gamma_{\mathrm{IV},ij} = -\int \frac{\partial^2 \left[P_k (\Gamma_\theta + H)^{-1}_{kl}\right]}{\partial\lambda_i \partial\lambda_j}\bigg|_{\lambda=\bar{\lambda}} D_l \, p_{\mathrm{pop}}(\bar{\theta}|\bar{\lambda}) \mathrm{d}\bar{\theta} \,, \tag{4.11}$$

$$\Gamma_{\mathrm{V},ij} = -\frac{1}{2}\int \frac{\partial^2 \left[P_k (\Gamma_\theta + H)^{-1}_{kl} P_l\right]}{\partial\lambda_i \partial\lambda_j}\bigg|_{\lambda=\bar{\lambda}} p_{\mathrm{det}}(\bar{\theta}) p_{\mathrm{pop}}(\bar{\theta}|\bar{\lambda}) \mathrm{d}\bar{\theta} \,. \tag{4.12}$$

Here, Γ_θ is the single-event Fisher matrix defined in Eq. (4.2) and

$$P_i = \frac{\partial \ln p_{\mathrm{pop}}(\theta|\lambda)}{\partial\theta_i}\bigg|_{\theta=\bar{\theta}} \,, \tag{4.13}$$

$$H_{ij} = -\frac{\partial^2 \ln p_{\mathrm{pop}}(\theta|\lambda)}{\partial\theta_i \partial\theta_j}\bigg|_{\theta=\bar{\theta}} \,, \tag{4.14}$$

$$D_i = \frac{\partial p_{\mathrm{det}}(\theta)}{\partial\theta_i}\bigg|_{\theta=\bar{\theta}} = \frac{\partial p_{\mathrm{det}}}{\partial\rho}\frac{\partial\rho}{\partial\theta_i}\bigg|_{\theta=\bar{\theta}} \,. \tag{4.15}$$

From these expressions, one can compute the covariance matrix Γ_λ^{-1} and thus the errors on the population hyperparameters $(\Gamma_\lambda^{-1})_{ii}^{1/2}$.

In the limit where the uncertainties on the parameters of the individual sources are negligible, one has $\Gamma_\theta \to \infty$ which implies $\Gamma_{\mathrm{II-V}} \to 0$ because $\Gamma_\theta + H \sim \Gamma_\theta$, hence $(\Gamma_\theta + H)^{-1} \to 0$, and $\partial\Gamma_\theta/\partial\lambda = 0$. The population Fisher matrix Γ_λ can thus be approximated with the first term Γ_{I} of Eq. (4.8), which we expect to be the dominant contribution in our forecasting exercise (cf. Ref. [229]).

4.2.3 Interpretation

Equations (4.7–4.12) were computed assuming a Fisher-matrix description for the single-event likelihoods as well as large number of events, namely $N_{\mathrm{det}} \to \infty$ [229]. The appropriate limit for a population Fisher analysis is that of many and well-measured GW signals. In this limit, the population Fisher matrix scales with the size of the catalog as $\Gamma_\lambda \propto 1/\sqrt{N_{\mathrm{det}}}$, cf. Eqs. (4.4) and (4.7) for any given population with efficiency $p_{\mathrm{det}}(\bar{\lambda})$.

Crucially, the expressions above provide the population Fisher matrix only up to additional corrections scaling with inverse powers of N_{det}. This implies that Γ_λ is not

guaranteed to be positive definite, thus casting some issues on the interpretation of its inverse as a covariance matrix. In practice, however, all diagonal terms $(\Gamma_\lambda^{-1})_{ii}$ are positive and those cases where the Fisher matrix Γ_λ is non-positive definite are due to a small but negative eigenvalue. If off-diagonal terms are needed (e.g. when drawing corner plots), we regularize Γ_λ^{-1} by replacing the negative eigenvalue with its absolute value and projecting it back to the original coordinate space.

4.3 Implementation

4.3.1 Approximants and detectors

We implement the population Fisher-matrix formalism described above in the PYTHON programming language, leveraging the infrastructure of the single-event Fisher code GWFAST [104]. All derivatives with respect to λ are performed with automatic differentiation via JAX [234].

We use the IMRPHENOMXPHM [174, 175, 89] waveform model, and perform the single-event Fisher analysis in terms of the parameters $\tilde{\theta} = \{\mathcal{M}_z, \eta, \chi_1, \chi_2, \phi_{JL}, \phi_{12}, \theta_1, \theta_2, d_L, \alpha, \delta, \theta_{\rm JN}, \psi, t_c, \phi_c\}$. These are the detector-frame chirp mass $\mathcal{M}_z$, the symmetric mass ratio η, the dimensionless spin magnitudes $\chi_{1,2}$, tilt angles $\theta_{1,2}$, the azimuthal spin angle $\phi_{1,2}$, the azimuthal angle ϕ_{JL} between the total and orbital angular momenta, the luminosity distance d_L, the sky position coordinates α, δ, the polar angle θ_{JN} between total angular momentum and the line of sight, the polarization angle ψ, the time of coalescence t_c and the phase at coalescence ϕ_c (see Ref. [222]). We assume a flat ΛCDM model with PLANCK18 parameters [235].

We apply Gaussian priors to the spin magnitudes and angular spin parameters; this serves two purposes: avoiding singular matrices $\Gamma_{\tilde{\theta}}$ and incorporating information on the physical range of some parameters. In the Fisher-matrix formalism, applying a prior consists in adding a prior matrix to the Fisher $\Gamma_{\tilde{\theta}}$ [103]. Our prior matrix is diagonal, with

elements chosen as the inverse variances of the Gaussian priors we impose. Specifically, the diagonal elements corresponding to the spin magnitudes, spin orientations, and azimuthal angles are $\{\sigma_{\chi_1}^{-2}, \sigma_{\chi_2}^{-2}, \sigma_{\phi_{JL}}^{-2}, \sigma_{\phi_{12}}^{-2}, \sigma_{\theta_1}^{-2}, \sigma_{\theta_2}^{-2}\} = \{1, 1, (2\pi)^{-2}, (2\pi)^{-2}, \pi^{-2}, \pi^{-2}\}$. All other elements are set to zero.

Note the set of parameters $\tilde{\theta}$ is not equivalent to the set θ entering our population analysis (see below), the difference being in the mass and redshift parameters. We rotate the Fisher matrix $\Gamma_{\tilde{\theta}}$ computed by GWFAST to obtain Γ_θ. In particular, one has $\Gamma_\theta = J^T \Gamma_{\tilde{\theta}} J$ where $J = d\tilde{\theta}/d\theta$ is the Jacobian of the rotation between the $\tilde{\theta}$ and the θ parameters. In practice, this means we are assuming the likelihood is Gaussian also in the θ parameters; this is motivated by the large SNRs delivered by 3G detections.

We compare results obtained assuming a single, triangular shaped ET interferometer against a detector network composed of ET and two CE detectors [223, 228]. Specifically, we use the 10 km noise curve for ET from Ref. [223] assuming a nominal location in Sardinia, Italy, and CE noise curves assuming two interferometers with arm lengths of 40 km and 20 km in the USA [236, 228]. We set a minimum frequency of 2 Hz.

4.3.2 Population models

The results presented in this work rely on a standard parametric population model $p_{\rm pop}(\theta|\lambda)$ borrowed from current LIGO/Virgo/KAGRA analyses [127]. We tackle population inference on source-frame masses, spin tilts, and redshifts, i.e. $\theta = \{m_1, m_2, \chi_1, \chi_2, \theta_1, \theta_2, z, \Omega\}$ We assume that $p_{\rm pop}(\theta|\lambda)$ is separable over the source parameters as follows

$$\begin{aligned} p_{\rm pop}(\theta|\lambda) = & \, p(m_1, m_2|\lambda_m) p(\chi_1|\lambda_\chi) p(\chi_2|\lambda_\chi) \\ & \times p(\theta_1|\lambda_\theta) p(\theta_2|\lambda_\theta) p(z|\lambda_z) p(\Omega)\,, \end{aligned} \tag{4.16}$$

where $\lambda = \{\lambda_m, \lambda_z, \lambda_\chi, \lambda_\theta\}$. The symbol Ω collectively denotes all other parameters entering the GW approximant. Omitting some parameters in population inference is actually equivalent to assuming an astrophysical model where those parameters are distributed as in the Bayesian priors $p(\Omega)$ used in the underlying single-event analyses [96]. However, note that while the parameters Ω factor out in $p_{\rm pop}$, they do enter the detectability function $p_{\rm det}$.

We briefly summarize the population models used in this work as follows, see Table 4.3. Appendix 4.6 provides a more detailed description of these models.

- *Mass distribution.* We use a slightly edited version of the popular Power–Law + Peak model [127, 237, 95] with 9 free hyperparameters

$$\lambda_m = \{\alpha_m, \beta_q, \lambda_{\rm peak}, \mu_m, \sigma_m, m_{\rm min}, m_{\rm max}, \sigma_{\rm l}, \sigma_{\rm h}\}$$

  . The distribution of the primary mass m_1 is described by a power-law with slope $-\alpha_m$ extending between a minimum mass $m_{\rm min}$ and a maximum mass $m_{\rm max}$. A Gaussian feature is included to account for a possible concentration of events just close to the pair-instability supernova mass gap. This Gaussian peak is characterized by weight $\lambda_{\rm peak}$, median μ_m, and standard deviation σ_m. The mass ratio q is modeled by a power-law with spectral index β_q. The distributions of m_1 and q are smoothed over a range of masses $\sigma_{\rm l}$ ($\sigma_{\rm h}$) at the low (high) end of the mass function. Our smoothing prescription differ by that used elsewhere in the literature. In particular, we opted for polynomial smoothing functions for compatibility with the adopted automatic-differentiation strategy and numerical efficiency; see App. 4.6.

- *Redshift distribution.* We use the so-called Madau–Dickinson profile [238] for the star formation rate distribution. This is parameterized by $\lambda_z = \{\alpha_z, \beta_z, z_{\rm p}\}$, namely the redshift $z_{\rm p}$ at which the star formation rate peaks and the power-law

indices α_z and β_z that govern the rise and fall of the star formation rate at low and high redshifts, respectively.

- *Spin distribution.* The spin distribution is modeled after LIGO/Virgo's DEFAULT model where $\lambda_\chi = \{\alpha_\chi, \beta_\chi\}$ and $\lambda_\theta = \{\zeta, \sigma_t\}$ [127, 116]. The component spin magnitudes are independently drawn from a Beta distribution with shape parameters α_χ and β_χ. The distribution of the spin tilt cosines have a uniform component (describing dynamically formed binaries with isotropically oriented spins) and a truncated Gaussian component with mean 1, weight ζ and standard deviation σ_t (modeling BH binaries formed in isolated environment with preferentially aligned spins).

Unless stated otherwise, we adopt fiducial values for the mass and spin hyperparameters $\lambda_m, \lambda_\chi, \lambda_\theta$ using the medians of the marginalized distributions from GWTC-3 [127] and fiducial values of λ_z from Ref. [238]. In particular, we set $\alpha_m = 3.4, \beta_q = 1.1, m_{\rm min} = 9.1\ M_\odot, m_{\rm max} = 87\ M_\odot, \lambda_{\rm peak} = 0.039, \mu_{\rm m} = 34\ M_\odot, \sigma_m = 3.6\ M_\odot, \sigma_{\rm l} = 4\ M_\odot, \sigma_{\rm h} = 20\ M_\odot, \alpha_z = 2.7, \beta_z = 3, z_{\rm p} = 2, \alpha_\chi = 1.6, \beta_\chi = 4.12, \zeta = 0.66, \sigma_t = 1.5$. Note that our choice for the minimum mass $m_{\rm min} = 9.1\ M_\odot$ is different with respect to the median value of $\sim 5\ M_\odot$ reported in Ref. [127]. This is because of the adopted smoothing function (see App. 4.6). In our implementation one has that the population distribution function is zero for $m \leq m_{\rm min} - \sigma_{\rm l}$. As a result, our parameter $m_{\rm min}$ is shifted by $\sigma_{\rm l}$ (which has fiducial value $4\ M_\odot$) compared to that of Ref. [127].

4.3.3 Monte Carlo integrations and regularization

We approximate Eqs. (4.4) and (4.8–4.12) with Monte Carlo integration. In particular, we cast those expressions as

$$I(\bar{\lambda}) = \int X(\bar{\theta}, \bar{\lambda}) p_{\rm draw}(\bar{\theta})\, {\rm d}\bar{\theta} \simeq \frac{1}{N_{\rm draw}} \sum_i^{N_{\rm draw}} X(\bar{\theta}_i, \bar{\lambda})\,, \tag{4.17}$$

where $\bar{\theta}_i \sim p_{\rm draw}(\bar{\theta})$.

The injection distribution $p_{\rm draw}(\bar{\theta})$ is chosen to improve convergence. For instance, it is convenient to use shallower power-laws (with spectral indexes closer to 0) to prevent undersampling in regions of the parameter space where $p_{\rm pop}(\bar{\theta}|\bar{\lambda})$ is low but relevant for a specific element of $\bar{\lambda}$ (notably, this is the case of $m_{\rm max}$). For the mass model, we use a draw population with $\alpha_m = \beta_q = 0.5$ and $\sigma_{\rm h} = 20\ M_\odot$. For all other hyperparamters, it is sufficient to draw samples using the true values $\bar{\lambda}$.

The subdominant terms $\Gamma_{\rm II-V}$ depend on the single-event Fisher matrices Γ_θ evaluated at the draws $\bar{\theta}_i$. If some of these sources have low SNRs and/or are located at the edge of the parameter space, this might result in occasionally large numerical errors on Γ_θ, which in turn make the computation of Γ_λ numerically unstable. As a regularization strategy, we restrict the computation of the Monte Carlo sum in Eq. (4.17) to the inner 95% quantile of the evaluations $X(\bar{\theta}_i, \bar{\lambda})$. At least for the GWFast implementation and using $N_{\rm draw} = \mathcal{O}(10^5)$, this is sufficient to obtain stable results. For additional regularization purposes, we also set $p_{\rm pop}(\theta|\lambda) = 0$ whenever this is $< 10^{-12}$. This helps avoiding numerical instabilities when taking derivatives with respect to the hyperparameters λ and we verified it does not affect our results in any meaningful way.

In the following, the number of significant digits used when presenting the results is determined by the variance of the adopted Monte Carlo estimators.

4.4 Results

We now present our results, starting from a broad overview and then exploring forecasts for mass, redshift, and spins in more details.

Figure 4.1 shows the relative errors σ_λ/λ of the 16 population hyperparameters as a function of the observation time for both the ET and ET+2CE configurations assuming our fiducial model. Results corresponding to $T_{\rm obs} = 10$ yr are also reported in Table 4.1, together with time required to achieve 1% accuracy in each of the hyperparameters. We

	ET		ET+2CE	
	σ_λ/λ $[10^{-3}]$	$T_{1\%}$ [yr]	σ_λ/λ $[10^{-3}]$	$T_{1\%}$ [yr]
α_m	1.5	0.2	1.3	0.2
β_q	7.1	5.1	6.6	4.3
$\lambda_{\rm peak}$	5.8	3.4	5.4	2.9
μ_m	0.7	0.1	0.6	0.04
σ_m	5.7	3.3	5.2	2.7
$m_{\rm min}$	1.1	0.1	0.9	0.1
$m_{\rm max}$	15.7	> 10	15.1	> 10
$\sigma_{\rm l}$	2.7	0.7	2.5	0.6
$\sigma_{\rm h}$	10.3	> 10	9.7	> 10
α_z	6.7	4.5	5.9	1.4
β_z	5.6	3.1	3.8	0.2
$z_{\rm p}$	4.7	2.2	3.7	1.4
α_χ	1.7	0.3	1.5	0.2
β_χ	1.8	0.3	1.5	0.2
ζ_χ	9.8	9.7	8.9	7.9
σ_t	5.3	2.8	5.1	2.6

Table 4.1: Summary results from our fiducial model, see also Fig. 4.1. For ET and ET+2CE, we report relative errors σ_λ/λ after 10 years of observation as well as the required time to achieve a percent-level accuracy $\sigma_\lambda/\lambda = 1\%$. For readability, parameters are divided into four blocks (broad mass features, mass cutoffs, redshift, spins) as in Fig. 4.1.

divide our hyperparameters into four groups, separating those relative to broad mass features (α_m, β_q, μ_m, σ_m, $\lambda_{\rm peak}$), muss cutoffs ($m_{\rm max}$, $m_{\rm min}$, $\sigma_{\rm h}$, $\sigma_{\rm l}$), redshift ($z_{\rm p}$, α_z, β_z) and spins (α_χ, β_χ, ζ, σ_t). We consider a cumulative observation time $T_{\rm obs} \in [0, 10]$ years and compute the corresponding number of detected events $N_{\rm det}$ as described in App. 4.6.

For our fiducial model, an observation time of 10 years corresponds to $\sim 5 \times 10^5$ ($\sim 7 \times 10^5$) detected events and a detection efficiency $p_{\rm det}(\lambda) \simeq 69\%$ (96%) for ET (ET+2CE).

For the case of ET, Fig. 4.2 shows the joint distributions of the mass, spins, and redshift parameters, including their covariances. A similar plot where we compare (with caveats) against current results from LIGO/Virgo is presented in App. 4.7. 3G detectors will be transformative, improving population constraints by orders of magnitudes.

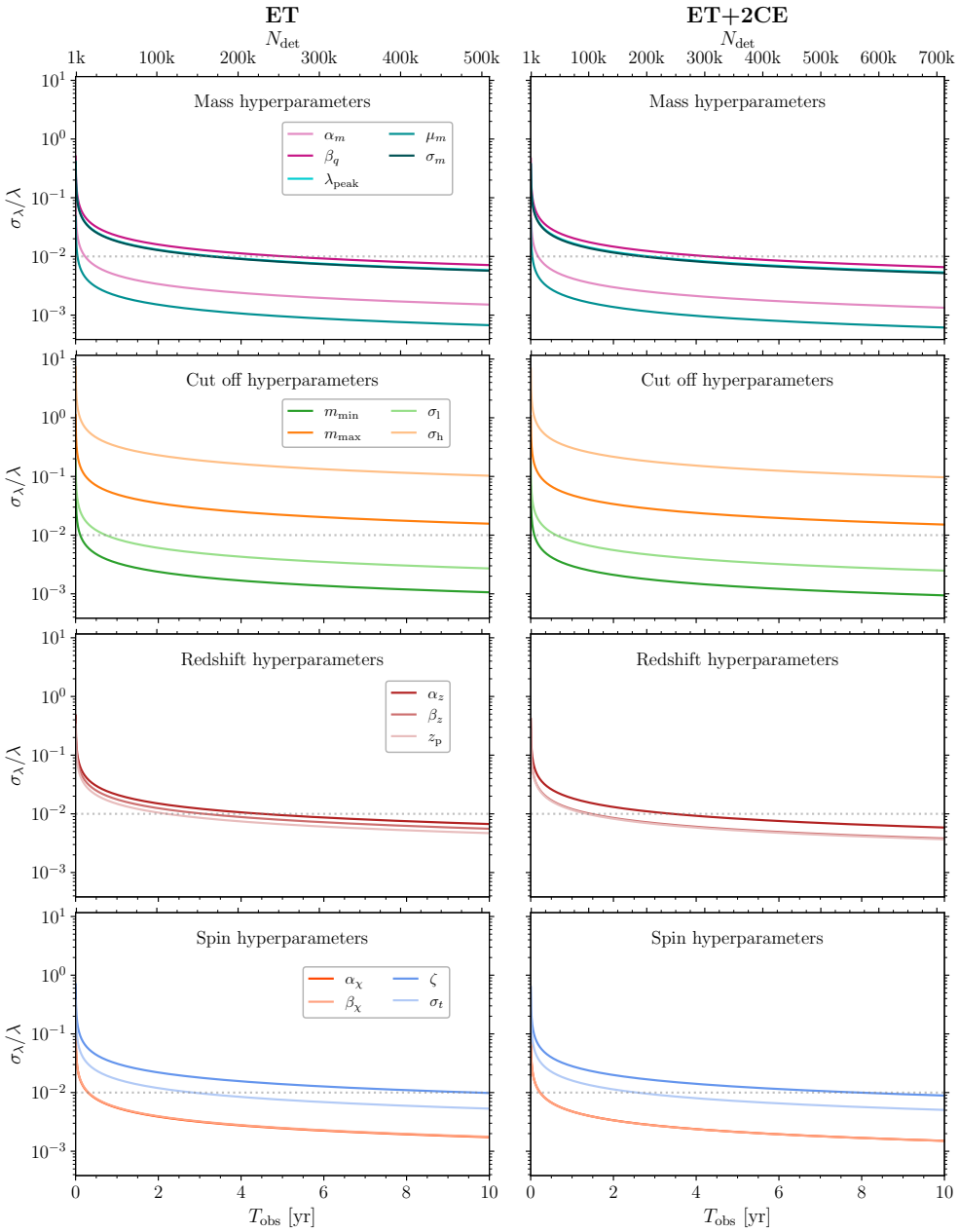

Figure 4.1: Relative errors of the model hyperparameters λ as a function of the observation time (bottom x-axis) for ET (left panels) and ET+2CE (right panels) assuming our fiducial population model. The upper x-axis indicates the corresponding number of detected events, which depends on the chosen detector network. The four rows correspond to the hyperparameters governing broad mass features, mass cutoffs, merger redshift evolution, and BH spins. To guide the eye, the dashed horizontal lines mark the threshold for percent-level accuracy, $\sigma_\lambda/\lambda = 1\%$. Some of these results are also reported in Table 4.1.

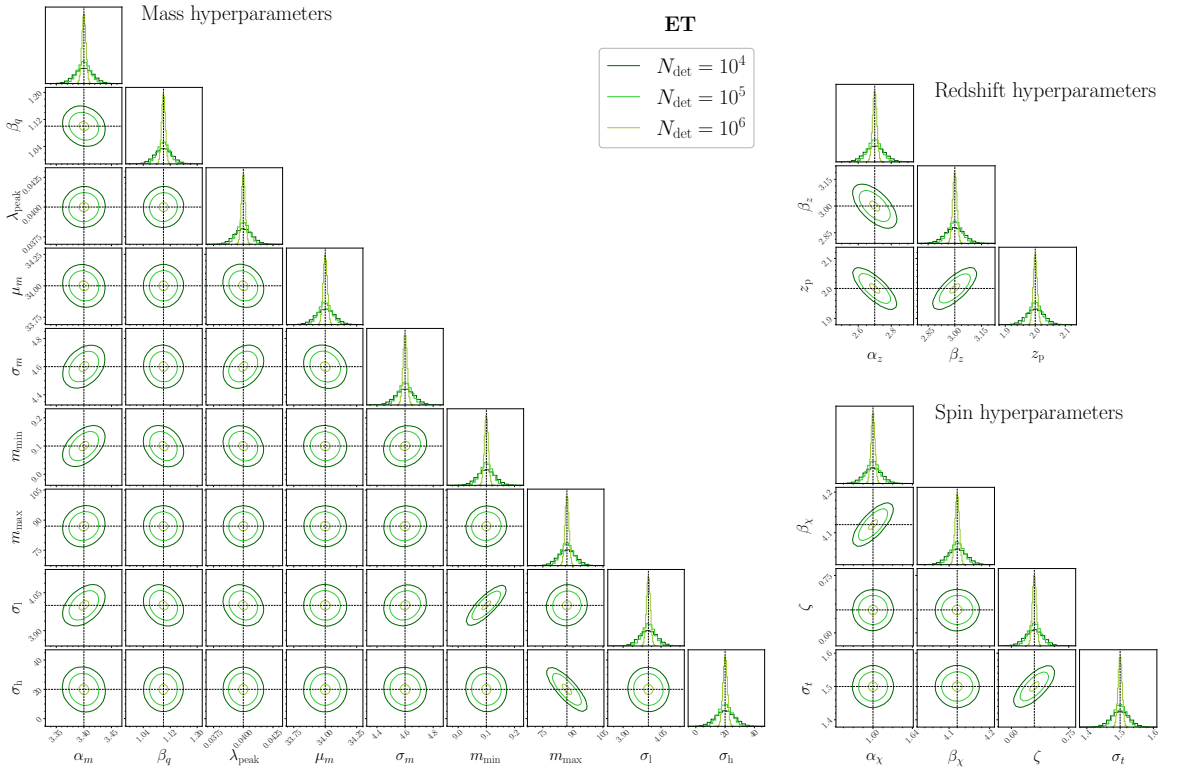

Figure 4.2: Hyperparameter uncertainties and their correlations assuming the ET detector and our fiducial population model. Contours indicate 90% credible intervals and black dashed lines indicate the injected values. We show results for catalogs of $N_{\rm det} = 10^4, 10^5, 10^6$ events (darker to lighter) corresponding to $T_{\rm obs} \simeq 0.2, 2, 19$ years.

We also present some variations around the fiducial model, changing in particular the maximum mass $m_{\rm max}$, the peak of the redshift distribution $z_{\rm p}$, and the aligned/isotropic mixing fraction ζ for the spin directions. These results are presented in Figs. 4.3, 4.4, 4.5, and Table 4.2.

4.4.1 Detector networks

The most evident feature from Fig. 4.1 is the scaling

$$\frac{\sigma_\lambda}{\lambda} \propto \sigma_\lambda \propto \frac{1}{\sqrt{N}} \propto \frac{1}{\sqrt{N_{\rm det}}} \propto \frac{1}{\sqrt{T_{\rm obs}}}\,, \tag{4.18}$$

which is a direct consequence of Eqs. (4.4), (4.7), and (4.27).

From Fig. 4.1 and Table 4.1, we predict that the ET and ET+2CE will measure the population properties of BH binaries equally well, at least qualitatively. While there are small quantitative differences, the overall picture is very similar between the two cases. Adding additional 3G detectors to the network will not be crucial for measuring the intrinsic population properties of merging BHs. This result is largely expected, for a few reasons:

(i) The overall scale of the problem is set by the number of detections. In the expressions above, this is indicated by the weight $p_{\rm det}(\bar{\theta})p_{\rm pop}(\bar{\theta}|\bar{\lambda})$ in the dominant $\Gamma_{\rm I}$ integral of Eq. (4.8). For a given intrinsic number of events N (or equivalently $T_{\rm obs}$), upgrading from ET to ET+2CE improves population constraints by as little as

$$1-\sqrt{\frac{p_{\rm det,ET}(\bar{\lambda})}{p_{\rm det,ET+2CE}(\bar{\lambda})}}\sim 15\%\,. \tag{4.19}$$

This is only an approximate estimate because the detectable population $p_{\rm det}(\bar{\theta})p_{\rm pop}(\bar{\theta}|\bar{\lambda})$ is integrated over an additional derivative term in Eq. (4.8) and additional subdominant terms $\Gamma_{\rm II-V}$ are present.

(ii) While the individual-event measurement accuracies Γ_θ improve with multiple detectors, they enter the population Fisher matrix only through the subdominant terms $\Gamma_{\rm II-V}$.

(iii) In this work and as currently done in most of the literature (but see Ref. [239] for a notable exception) we only consider the population of masses, spins, and redshift. As is already the case for LIGO and Virgo, the accuracy of these parameters is largely dominated by the single most sensitive instrument in the network. On the other hand, it is well known that expanding the network greatly pays off in terms of source localization.

(iv) We restrict our study to the *global* population of sources and assume the BH merger rate roughly follows the star formation rate, i.e. most of the events have $z \lesssim 2$. At these redshifts, the coverage provided by even a single ET instrument of triangular shape is essentially complete. This conclusion cannot be straightforwardly applied to putative sub-populations of sources at very high redshifts, e.g. those coming from population-3 stars [240] or primordial BHs [241]. While the ET+2CE network has a larger detection horizon compared to ET alone, detecting binaries at cosmological distances does not imply an accurate inference of their source parameters [242, 243].

4.4.2 Mass distribution

The primary mass distribution of binary BHs inferred from GWTC-3 spans from a few solar masses up to around 90 $M_\odot$. It shows a power-law behavior, with a Gaussian component at $\sim 35\ M_\odot$, which is below the pair-instability supernova mass gap that is predicted by stellar evolution models for BHs with masses in the range $\sim 40 - 120\ M_\odot$ [244, 245, 246, 247]. The astrophysical origin of this excess is uncertain. The primary mass spectrum reveals another notable peak at $\sim 10\ M_\odot$, which has been suggested as an indicator of a significant contribution from isolated binary evolution [131, 129, 248]. The absence of a sharp cut-off for masses $\gtrsim 40\ M_\odot$ might point to evolutionary paths in dynamical environments, including the occurrence of hierarchical mergers and/or formation in environments like galactic nuclei [249, 250, 194].

Figure 4.1 and Table 4.1 show that hyperparameters such as the minimum mass $m_{\rm min}$, the spectral index α_m, and the position of the Gaussian peak μ_m in the primary mass spectrum are measured with high precision, reaching 1% accuracy in just a few months of 3G operation. All of these are parameters that impact the population where the event rate is sufficiently large (i.e. close to the "top" of the underlying power-law structure). Measuring $m_{\rm min}$ will test the occurrence of the putative mass

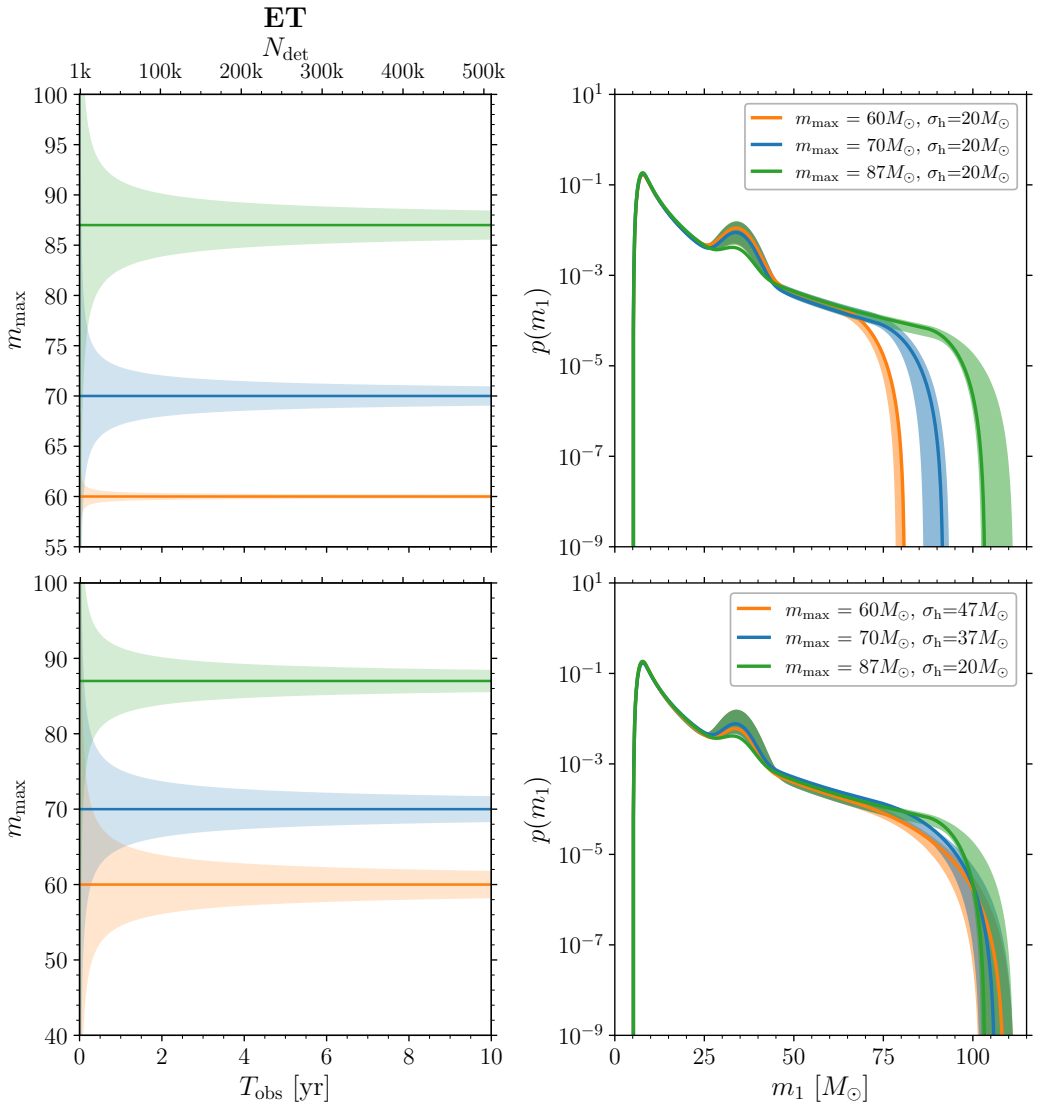

Figure 4.3: Measurability of upper edge of the mass spectrum as a function of the observation time $T_{\rm obs}$ (bottom x) axis for a single ET instrument. The corresponding number of detections is reported in the top x-axis. Top (bottom) panels show model variations around the fiducial model where we vary the maximum mass $m_{\rm max}$ and fix (vary) the related smoothing parameter $\sigma_{\rm h}$. In both panels, the fiducial model is represented by the green curve with $m_{\rm max} = 87\ M_\odot$ and $\sigma_{\rm h} = 20\ M_\odot$. Left panels show the injected values (horizontal lines) and the related 1σ errors (shaded areas). For each model, the solid lines in the right panels represent the corresponding primary mass spectrum with the shaded region showing the 90% credible interval.. Related results are presented in Table 4.2.

gap between BHs and neutron stars, measuring α_m will indicate whether the BH mass spectrum is shallower or steeper than that of stars, and measuring μ_m will delineate the boundary between BHs formed through supernova explosion and those limited by pair instabilities, shedding light on the underlying physics of core collapse. Note that the hyperparameter μ_m can also have other astrophysical interpretation and allow us to explore alternative formation channels, such as hierarchical mergers, direct collapse, and primordial BHs, shedding light on the diverse processes driving core collapse. The secondary-mass spectral index β_q is measured less accurately than that of the primary mass α_m. This is due to both m_1 being easier to constrain than m_2, but also to the specific functional form of $p_{\rm pop}(\theta|\lambda)$ used here, which is written as two independent contributions $p(m_1)p(m_2|m_1)$.

In contrast, parameters such as the maximum mass $m_{\rm max}$ and the smoothing length $\sigma_{\rm h}$ are harder to measure because they affect the low-rate region of the population (at the "bottom" end of the underlying power-law). The accuracy of both these parameters does not go below 1% accuracy even after 10 years of 3G observations.

The corner plot of Fig. 4.2 shows that, as expected, the cutoff parameters at both the lower and upper ends of the mass spectrum exhibit distinct correlations with their respective smoothing hyperparameters. Specifically, the minimum mass $m_{\rm min}$ shows a positive correlation with $\sigma_{\rm l}$, indicating that an increase in $m_{\rm min}$ can be compensated by a simultaneous increase in $\sigma_{\rm l}$ while still fitting the same population model. Conversely, the maximum mass $m_{\rm max}$ displays a negative correlation with $\sigma_{\rm h}$; thus, an increase in $m_{\rm max}$ necessitates a corresponding decrease in $\sigma_{\rm h}$ to maintain the same probability density function $p_{\rm pop}(\theta|\lambda)$.

We also observe a positive correlation between $m_{\rm min}$ (and thus $\sigma_{\rm l}$) and the spectral index α_m. This is because both a steeper power-law (larger α_m) and a more extended mass spectrum (lower $m_{\rm min}$) predict a larger number of low-mass detections. We do not observe a similar correlation between α_m and $m_{\rm max}$ because there are significantly fewer events at the upper end of the mass spectrum. A similar but weaker behavior is

observed also for the secondary-mass spectral index β_q.

The minimum mass $m_{\rm min}$ (and consequently $\sigma_{\rm l}$) is also correlated with the fraction of sources in the Gaussian component $\lambda_{\rm peak}$. As $m_{\rm min}$ and $\sigma_{\rm l}$ increase, the contribution provided by the power-law component decreases (the power-law is "shorter") and thus the weight of the Gaussian peak must increase. We further observe a negative correlation between the spectral indices α_m and β_q of the primary and secondary mass distribution, respectively. This can be understood as follows: a higher value for the primary mass spectral index, i.e. a steeper power-law, results in a larger number of light events, consequently flattening the mass ratio distribution, and thus lowering the value of β_q.

In Fig. 4.3 and Table 4.2, we present two model variation to study the measurability of the upper cutoff $m_{\rm max}$. The impact of these model variations on $p_{\rm det}(\lambda)$ is of $\mathcal{O}(10^{-4})$, such that a given value of $T_{\rm obs}$ corresponds to GW catalogs of essentially the same size $N_{\rm det}$.

(i) First, we vary the maximum mass $m_{\rm max} = 60\ M_\odot$, $70\ M_\odot$, $87\ M_\odot$ while keeping all other hyperparameters fixed to their fiducial values, including the high-mass smoothing $\sigma_{\rm h}$. We find that the uncertainties on the maximum mass increase with the value of the maximum mass itself. After $T_{\rm obs} = 10$ years, ET will be able to constrain the maximum mass with absolute errors of $\pm[1.159, 8.780, 13.623] \times 10^{-1}\ M_\odot$ for $m_{\rm max} = [60, 70, 87]\ M_\odot$. Lowering the value of $m_{\rm max}$ shifts the cutoff closer to the region of the mass spectrum where the event rate is higher. Consequently, the greater density of observed events provides more statistical information, allowing for more precise population constraints. A lower value of the maximum mass also decreases the errors on the smoothing parameter $\sigma_{\rm h}$; this might be due to the correlation between $\sigma_{\rm h}$ and $m_{\rm max}$ highlighted above. The errors on the other mass hyperparameters remain approximately constant under these model variations.

(ii) Second, we simultaneously vary the maximum mass $m_{\rm max} = 60\ M_\odot$, $70\ M_\odot$, $87\ M_\odot$

and the smoothing length $\sigma_{\rm h} = 47\ M_\odot, 37\ M_\odot, 20\ M_\odot$ such that their sum $m_{\rm max} + \sigma_{\rm h} = 107\ M_\odot$ remains constant. Contrary to the previous case, here we find that errors on the maximum mass and that of the smoothing parameter $\sigma_{\rm h}$ decrease with the value of $m_{\rm max}$. In this controlled experiment, lower values of $m_{\rm max}$ imply higher values of $\sigma_{\rm h}$ and thus a smoother high-mass end of the mass spectrum. Such prominent smoothing reduces the importance of the features near the cutoff, making it harder to pinpoint the exact location of $m_{\rm max}$. As a result, the uncertainty on both $m_{\rm max}$ and $\sigma_{\rm h}$ increases with decreasing $m_{\rm max}$.

There is an important statistical caveat that affects some of the results we just described. Errors on broad features scale as the square root of the number of data-points, which is indeed the prediction of Eq. (4.18): $\sigma_\lambda \propto 1/\sqrt{N_{\rm det}}$. This agrees with one's intuition from the central limit theorem. However, the case of cutoff parameters is different and, instead, errors are expected to scale as $1/N_{\rm det}$ (this is the popular "German tank" problem in statistics [251]). This property is not captured by the Fisher matrix formalism, which expands the likelihood around the true value $\bar{\lambda}$ — an expansion that should not be allowed at the edge of a distribution. While mitigated by the smoothing parameters $\sigma_{\rm h}$ and $\sigma_{\rm l}$, this caveat affects our Fisher-based error estimates of $m_{\rm min}$ and $m_{\rm max}$. Our results are therefore conservative: we predict a $1/\sqrt{T_{\rm obs}}$ scaling but this could actually be as rapid as $1/T_{\rm obs}$. This same feature was already noted in single-event parameter estimation when experimenting with truncated waveforms [252].

4.4.3 Redshift distribution

Current GW data provide conclusive evidence that the BH merger rate evolves with redshift. Assuming the merger rate evolves as $R(z) = R_0(1+z)^\kappa$, GWTC-3 data indicate $\kappa \sim 3$ [127], which is consistent with the rise of the star-formation rate with redshift in the local Universe [253, 254, 255, 133]. Modeling the evolution of the merger rate with redshift using a single power-law is not appropriate for 3G detectors

	σ_λ $[10^{-2}]$		
	$m_{\rm max} = 60\ M_\odot$ $\sigma_{\rm h} = 20\ M_\odot$	$m_{\rm max} = 70\ M_\odot$ $\sigma_{\rm h} = 20\ M_\odot$	$\boldsymbol{m_{\rm max} = 87\ M_\odot}$ $\boldsymbol{\sigma_{\rm h} = 20\ M_\odot}$
α_m	0.28	0.25	0.51
β_q	0.71	0.80	0.78
$\lambda_{\rm peak}$	0.02	0.02	0.02
μ_m	2.47	2.52	2.29
σ_m	2.49	2.99	2.05
$m_{\rm min}$	1.02	0.95	0.97
$m_{\rm max}$	11.59	87.80	136.23
$\sigma_{\rm l}$	1.15	1.08	1.08
$\sigma_{\rm h}$	11.77	131.20	205.65
	$m_{\rm max} = 60\ M_\odot$ $\sigma_{\rm h} = 47\ M_\odot$	$m_{\rm max} = 70\ M_\odot$ $\sigma_{\rm h} = 37\ M_\odot$	$\boldsymbol{m_{\rm max} = 87\ M_\odot}$ $\boldsymbol{\sigma_{\rm h} = 20\ M_\odot}$
α_m	0.56	0.54	0.51
β_q	0.82	0.81	0.78
$\lambda_{\rm peak}$	0.02	0.02	0.02
μ_m	2.40	2.40	2.29
σ_m	2.15	2.14	2.05
$m_{\rm min}$	0.94	0.93	0.97
$m_{\rm max}$	169.90	160.62	136.23
$\sigma_{\rm l}$	1.03	1.03	1.08
$\sigma_{\rm h}$	264.31	245.29	205.65
	$z_{\rm p} = 1$	$\boldsymbol{z_{\rm p} = 2}$	$z_{\rm p} = 5$
α_z	3.32	1.81	0.92
β_z	1.35	1.67	2.95
$z_{\rm p}$	0.70	0.94	1.78
	$\zeta = 0.1$	$\boldsymbol{\zeta = 0.66}$	$\zeta = 1$
α_χ	0.27	0.28	0.28
β_χ	0.72	0.73	0.73
ζ	0.32	0.65	0.70
σ_t	20.95	7.99	5.62

Table 4.2: Summary results from a few model variations. We show absolute errors σ_λ after 10 years of observation with ET. Starting from our fiducial model, we vary one or two parameters at a time. In particular, we explore variations of the maximum mass $m_{\rm max}$ (fist block of cells), the maximum mass $m_{\rm max}$ and the smoothing paramenter $\sigma_{\rm h}$ (second block of cells), the peak of the redshift distribution $z_{\rm p}$ (third block of cells), and the fraction of binaries with aligned spins ζ (fourth block of cell). The fiducial model corresponds to $m_{\rm max} = 87\ M_\odot$, $z_{\rm p} = 2$, and $\zeta = 0.66$ as indicated in boldface. Related results are presented in Figs. 4.3, 4.4, 4.5.

because their detection horizon extends far beyond the peak of the star-formation rate. This is indeed the key reason behind choosing the Madau–Dickinson profile instead, cf. Sec. 4.3.2.

The third row of Fig. 4.1 illustrates the relative errors on the redshift hyperparameters for the fiducial model, assuming either ET or ET+2CE. Again, we report qualitatively similar performances for both detector configurations. However, when compared to masses and spins, the redshift distribution is more prominently affected by the expansion of the detector network. This is because a network of well-separated interferometers has better sky localization capabilities, thus breaking the known degeneracy with the distance to the source, and thus with the redshift (let us recall that here we are fixing the cosmology). Improvements saturate as $T_{\rm obs}$ increases and marginal gains require longer times. For our fiducial model, we find that all parameters associated with the redshift evolution achieve 1% accuracy within a range of 2 months to 1.5 years when using ET+2CE, whereas the single ET configuration requires 2 to 4.5 years (Table 4.1).

Figure 4.2 shows that all the redshift hyperparameters are tightly correlated. In particular, there is a positive (negative) correlation between β_z (α_z) and the peak of the star formation rate $z_{\rm p}$. This is largely because we are keeping the local merger rate R_0 fixed. If the redshift peak is higher, the merger rate must be steeper at high redshift and shallower at low redshift.

Finally, Fig. 4.4 shows a model-variation study where we shift the peak of the redshift distribution $z_{\rm p} = 1,\, 2,\, 5$, considering a single ET instrument. We find that the errors decrease with increasing $z_{\rm p}$, which is once more a direct consequence of our rate choice. Fixing the local rate R_0 implies that these different models refer to Universes with a different number of BH mergers N (and thus a different number of detected events $N_{\rm det}$); cf. App. 4.6.2. In this context, assuming a model with a large value of $z_{\rm p}$ corresponds to assuming more merging BHs, which in turn results in better population constraints. An alternative choice would be to fix the total merger rate instead of the local value R_0.

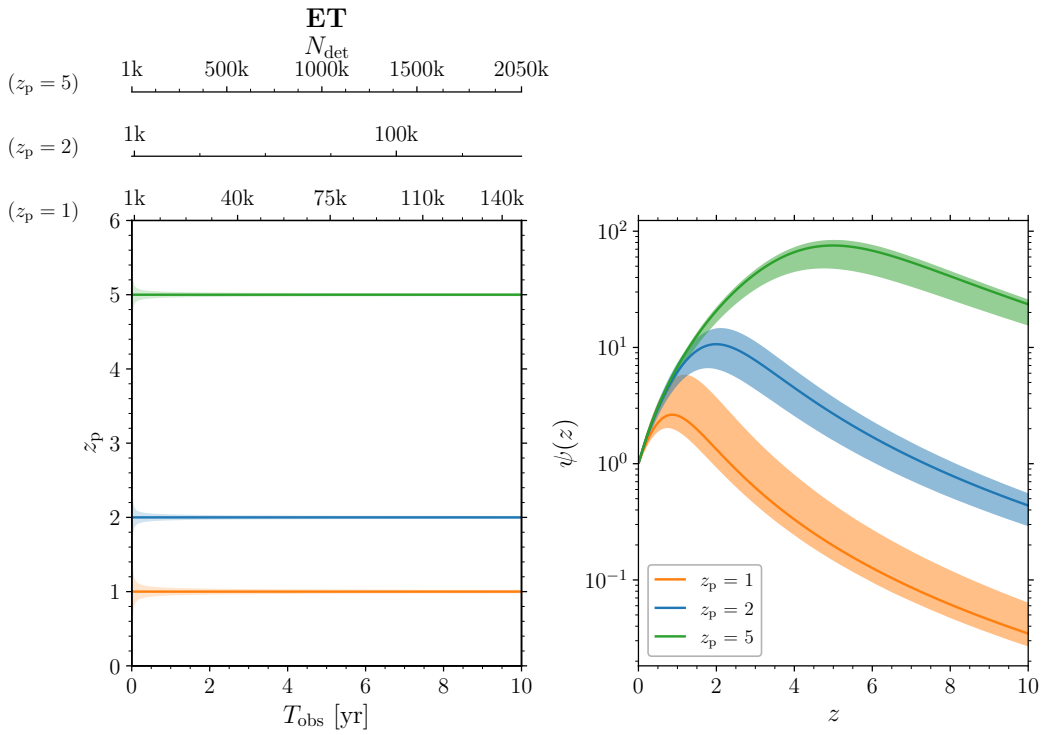

Figure 4.4: Measurability of the redshift at which the merger rate is maximum $z_{\rm p}$ as a function of the observation time $T_{\rm obs}$ (bottom x-axis) for a single ET instrument. We consider three models with $z_{\rm p} = 1$, 2, and 5; the three top x-axes show the corresponding number of detected events $N_{\rm det}$. The fiducial model is represented by the green curve with $z_{\rm p} = 2$. The left panel show the injected values (horizontal lines) and the related 1σ errors (shaded areas). For each model, the corresponding star formation rate $\psi(z)$ is illustrated in the right panels; this is normalized such that $\psi(z = 0) = 1$. The shaded region around each posterior population distribution represent the 90% credible interval. Related results are presented in Table 4.2.

4.4.4 Spin distribution

As for the spin distribution, current GW data reveals evidence of both aligned and misaligned spins. Systems with large spin-orbit misalignment suggest formation in dynamically active environments like globular clusters, where interactions can randomize spin directions; on the other hand, aligned spins suggest formation from isolated binary stars [256, 118, 116, 257]. The BH spin magnitudes contain information on the details

of stellar collapse including the efficiency of angular momentum transfer between stellar cores and envelopes as well as the occurrence of previous mergers [258, 259, 260, 129].

Figure 4.1 and Table 4.1 show that, among the spin hyperparameters, those with tighter constraints are the shape parameters entering the Beta distribution of the spin magnitudes. Both ET and ET+2CE achieve percent-level accuracy for α_χ and β_χ after just a few months of operation. In contrast, the hyperparameters associated with the spin orientations exhibit much longer timescales to reach similar accuracy levels. In particular, the mixing fraction ζ requires nearly 10 years of observations for both configurations to achieve an accuracy of 1%. Note that, for the spin case, the population Fisher terms $\Gamma_{\rm II-V}$ become almost comparable to $\Gamma_{\rm I}$, meaning that the single event errors on the spins are no longer irrelevant for the population analysis.

Our results suggest that the broad program of pinpointing the BH binary formation channel using the spin orientations will face some serious challenges, even in the 3G era. A crucial element on this point will be the achieved low-frequency sensitivity of ET and CE. For the noise curves used here, we have $f \gtrsim 2$ Hz for ET and $f \gtrsim 5$ Hz for CE. Pushing the low-frequency requirement will allow the inspiral phase, where spins are prominent observables, to be better resolved, which impacts both accuracies and event rates.

The corner plot in Fig. 4.2 shows that the hyperparameters modeling the spin magnitudes and those modeling the spin directions are respectively correlated within their own groups. In particular, we find a positive correlation in both $(\alpha_\chi - \beta_\chi)$ and $(\zeta - \sigma_t)$. We do not observe any meaningful cross-correlations between hyperparameters modeling the spin magnitudes and those modeling the spin orientations, which is expected being the two distributions independent.

Finally, in Fig. 4.5 we study the measurability of the relative weight of the isotropic ($\zeta = 0$) and aligned ($\zeta = 1$) component of the spin-direction population, which is meant to capture the fraction of systems originating from dynamical and isolated channels [116]. We vary $\zeta = 0.1, 0.66, 1$ while keeping all other hyperparameters fixed to their fiducial

values. Table 4.2 shows that varying ζ does not influence the absolute errors on either α_χ or β_χ (which agrees with the statement above about these parameters being largely uncorrelated). On the other hand, we observe that the error on σ_t raises (lowers) for smaller (larger) values of ζ. This can be traced to the fact that larger values for ζ predict more events in the Gaussian component, which is then easier to characterize. Finally, we find that the absolute uncertainty on ζ diminishes for lower values of this hyperparameter. As previously remarked, for the spin case single-event parameter reconstruction is found to be relevant through the terms $\Gamma_{\rm II-V}$. For lower values of ζ a larger fraction of the events have precessing spins, in which case tilt angles are easier to reconstruct as compared to $\theta_i \sim \pm\pi/2$.

4.5 Final remarks

We presented a Fisher-matrix implementation to forecast the population properties of merging BH binaries detectable by 3G observatories such as ET and CE. The starting point of this work is the formalism developed by Gair *et al.* [229]. Our results show that future detectors will improve the precision with which we can constrain the population of compact binaries by orders of magnitude compared to LIGO/Virgo, with error bars on the hyperparameters tightening at a rate $\propto 1/\sqrt{T_{\rm obs}}$. This is actually a conservative estimate, as some of the edge parameters could actually improve as fast as $1/T_{\rm obs}$ (the so-called "German tank" problem [251]).

While prospects are exciting, our analysis also reveals that, at least when modeling the population of masses, spins, and redshift, a single ET instrument of triangular shape and a network of ET and two CE's will return qualitatively similar constraints. We expect similar conclusions to hold true also in the case of two L-shaped ET detectors [223]. This point is worth investigating more thoroughly.

We predict that some of the key mass-related hyperparameters, such as the minimum BH mass $m_{\rm min}$ and the spectral index of the primary mass distribution α_m can reach sub-

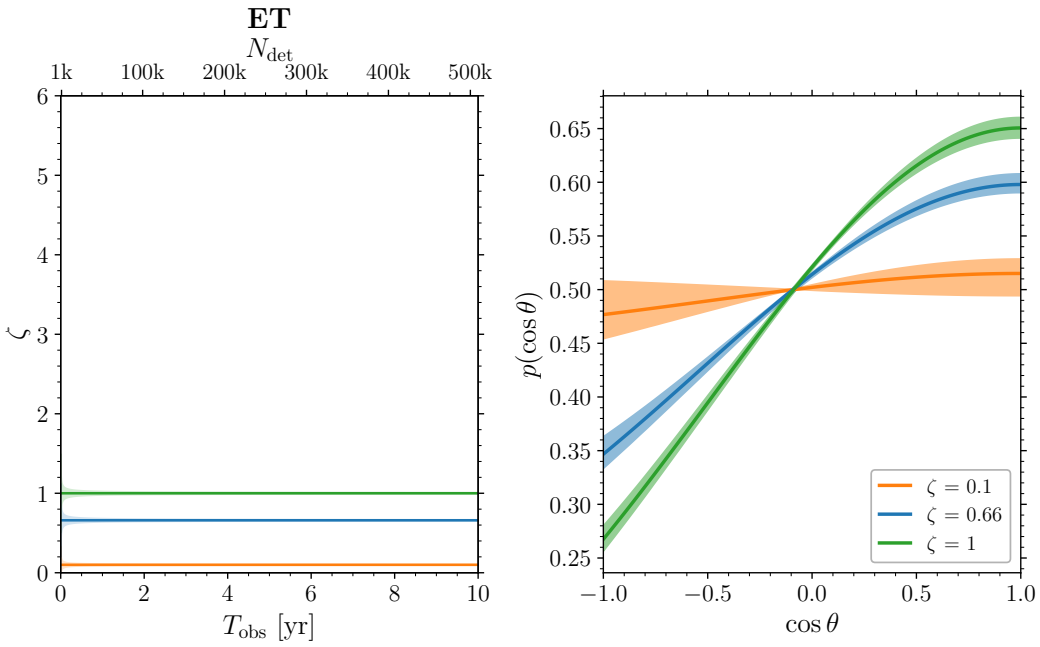

Figure 4.5: Measurability of the mixing fraction ζ between BH binaries with aligned and misaligned spins as a function of the observation time $T_{\rm obs}$ (bottom x-axis) for a single ET instrument. The corresponding number of detections is reported in the top x-axis. We consider three model variations with $\zeta = 0.1$, 0.66, and 1. The fiducial model is represented by the green curve with $\zeta = 0.66$. The left panel shows the injected values (horizontal lines) and the related 1σ errors (shaded areas). For each model, the corresponding distribution of the spin-orbit angle is illustrated in the right panels. The shaded region around each posterior population distribution represent the 90% credible interval. Related results are presented in Table 4.2.

percent accuracy within a few years of observation. The best-measured hyperparameter is μ_m, i.e. the position of the Gaussian peak. This is particularly promising, showing that 3G detectors will be able to deliver precision science on localized features in the population of merging compact objects. On the other hand, we predict that the upper edge of the primary mass spectrum $m_{\rm max}$ as well as the fraction of binaries with aligned vs misaligned spins are significantly harder to measure, reaching an accuracy of 1% only after $\gtrsim 8$ years of operations and several hundreds of thousands of events. This questions the feasibility of their related science case, namely pair-instability physics and formation channels, respectively.

We hope this work will move the signpost from individual events to populations of events in the context of 3G-detector forecasts. This is particularly timely given the current planning activities regarding the construction of such facilities. Beside 3G detectors, our tool could also be useful for current LIGO/Virgo analyses where, even though this Fisher approach is unreliable, it could provide some intuition about the response of specific parametric population models.

Our implementation is built on top of the GWFast code [104] and will be made publicly available in due course. While state-of-the-art for LIGO/Virgo and $\mathcal{O}(100)$ events, the population models implemented here are admittedly too simple for 3G detectors, where the much larger statistics will allow probing far more features in the compact-binary population. Developing forecasts in this direction is left to future work.

Appendix A:

4.6 Details of the population models

In this appendix we provide details about the parameterized population models used in this chapter; a concise list of the relevant hyperparameters is provided in Table 4.3 together with the adopted fiducial values. We stress that, while here we describe only the adopted models, the developed code can trivially be extended to implement different distributions.

4.6.1 Mass population model

For the BH masses we adopt the Power–Law+Peak model first introduced in [128]. The primary (heavier) object is distributed according to

$$\begin{aligned} p(m_1|\lambda_m) \propto\ & [(1-\lambda_{\rm peak}) \\ & \times \mathcal{P}(m_1|-\alpha_m, m_{\rm min}-\sigma_{\rm l}, m_{\rm max}+\sigma_{\rm h}) \\ & + \lambda_{\rm peak}\mathcal{N}(m_1|\mu_m,\sigma_m)]\,\mathcal{S}(m_1|m_{\rm min}-\sigma_{\rm l},\sigma_{\rm l}) \\ & \times [1-\mathcal{S}(m_1|m_{\rm max},\sigma_{\rm h})]\,, \end{aligned} \tag{4.20}$$

where $\mathcal{P}(m_1|-\alpha_m, m_{\rm min}-\sigma_{\rm l}, m_{\rm max}+\sigma_{\rm h})$ is a power-law distribution with spectral index $-\alpha_m$ normalized in the interval $[m_{\rm min}-\sigma_{\rm l}, m_{\rm max}+\sigma_{\rm h}]$, $\mathcal{N}(m_1|\mu_m,\sigma_m)$ is a normalized Gaussian distribution with mean μ_m and standard deviation σ_m. For the smoothing functions $\mathcal{S}$, we design and employ a different prescription compared to that of Ref. [128, 127]. The specific functional form of their smoothing is numerically challenging to treat with automatic differentiation. We also note that their functions only *tend* to zero, without actually reaching the limit value. We use the following polynomial filter, which smoothly evolves from 0 to 1 with vanishing first and second

derivatives at the edges:

$$\mathcal{S}(x|x_0,\sigma_x) = \begin{cases} 0 & x < x_0\,, \\ f(x|x_0,\sigma_x) & x_0 \leq x \leq x_0+\sigma_x\,, \\ 1 & x > x_0+\sigma_x\,, \end{cases} \tag{4.21}$$

$$f(x|x_0,\sigma_x) = \frac{(x-x_0)^3}{\sigma_x^5}[10\sigma_x^2 - 15\sigma_x(x-x_0) + 6(x-x_0)^2]\,, \tag{4.22}$$

where x_0 represents the lower edge, below which the filter vanishes, and σ_x controls the width of the transition. The above definition of the filter makes our implementation numerically stable when computing derivatives, and further allows to analytically compute the normalization of the chosen mass distribution, resulting in a faster numerical evaluation. Within our setting, the filtered mass distribution vanishes at $m_{\rm min} - \sigma_{\rm l}$ and at $m_{\rm max} + \sigma_{\rm h}$, with the onset of the filtering being at $m_{\rm min}$ and $m_{\rm max}$.

As for the mass of the secondary object, the distribution is given by a power-law with the modified smoothing function described above

$$\begin{aligned} p(m_2|\lambda_m, m_1) \propto\ & \mathcal{P}(m_2|\beta_q, m_{\rm min}-\sigma_{\rm l}, m_1) \\ & \times \mathcal{S}(m_2|m_{\rm min}-\sigma_{\rm l}, \sigma_{\rm l}) \\ & \times [1 - \mathcal{S}(m_2|m_{\rm max}, \sigma_{\rm h})]\,, \end{aligned} \tag{4.23}$$

which is then normalized (again analytically) in the interval $[m_{\rm min} - \sigma_{\rm l},\, m_1]$ depending on m_1, ensuring that $\int p(m_1, m_2|\lambda_m)\,{\rm d}m_2 = p(m_1|\lambda_m)$.

4.6.2 Redshift population model

The population Fisher matrix depends on the number N of detectable (not necessarily detected) BH mergers, cf. Eq. (4.7). We assume that the merger rate only depends on

the redshift z, i.e.

$$\frac{\mathrm{d}N}{\mathrm{d}t_s \mathrm{d}V_c} = R_0 \psi(z|\lambda_z)\,, \tag{4.24}$$

where t_s is the source-frame time, V_c is the comoving volume, is R_0 the local merger rate at $z = 0$, and $\psi(z)$ is a function modeling the star-formation rate such that $\psi(z = 0) = 1$. For the latter, we adopt the Madau-Dickinson profile [238], which has the following functional form [261]

$$\psi(z|\lambda_z) = \frac{[(1+z_\mathrm{p})^{\alpha_z+\beta_z}+1]\,(1+z)^{\alpha_z}}{(1+z_\mathrm{p})^{\alpha_z+\beta_z}+(1+z)^{\alpha_z+\beta_z}}\,, \tag{4.25}$$

where α_z (β_z) model the rise (decline) of the function at low (high) redshift, and z_p models the peak of the star-formation rate. The redshift distribution of the events is then given by [262, 237]

$$p(z|\lambda_z) \propto \psi(z|\lambda_z)\frac{\mathrm{d}V_c}{\mathrm{d}z}\frac{1}{1+z}\,. \tag{4.26}$$

The total number of mergers is given by

$$\begin{aligned} N &= \int \frac{\mathrm{d}N}{\mathrm{d}t_s \mathrm{d}V_c}\frac{\mathrm{d}V_c}{\mathrm{d}z}\frac{\mathrm{d}t_s}{\mathrm{d}t_d}\mathrm{d}z\mathrm{d}t_d \\ &= T_\mathrm{obs} R_0 \int_0^\infty \psi(z|\lambda_z)\frac{1}{1+z}\frac{\mathrm{d}V_c}{\mathrm{d}z}\mathrm{d}z\,, \end{aligned} \tag{4.27}$$

where t_d is the time at the detector, $\mathrm{d}t_s/\mathrm{d}t_d = 1/(1+z)$, and $\int \mathrm{d}t_d = T_\mathrm{obs}$ is the data taking period. In Sec. 4.4, we use the expressions above to estimate the time T_obs needed to observe a given number of detected events $N_\mathrm{det} = N p_\mathrm{det}(\lambda)$. The conversion between T_obs and N depends only on the redshift hyperparameters λ_z while the conversion between N and N_det depends on all hyperparameters λ as well as the chosen detector.

In this work, we fix the local merger rate to $R_0 = 17\ \mathrm{Gpc}^{-3}\ \mathrm{yr}^{-1}$ [127]; computing the Fisher-matrix error on R_0 is a natural extension of this work and requires generalizing Eq. (4.3) to the non-marginalized population likelihood (see App. A of Ref. [229]).

4.6.3 Spin population model

For the spins we again follow [128, 127] and adopt their DEFAULT model. This features a Beta distribution for the (independent) spin magnitudes

$$p(\chi_i|\lambda_\chi) = \mathrm{Beta}(\chi_i|\alpha_\chi, \beta_\chi)\,, \qquad i = \{1,\, 2\}\,, \tag{4.28}$$

where α_χ and β_χ are the shape parameters for the distribution. These can be converted into mean and variance, which are used in some of the GW literature to characterize the spin distribution.

The angles θ_i between each spin and the orbital angular momentum of the binary are drawn (again independently) from

$$p(\cos\theta_i|\lambda_\theta) = \zeta\,\mathcal{N}_{\rm tr}(\cos\theta_i|1, \sigma_t) + (1-\zeta)\,\mathcal{I}(\cos\theta_i)\,, \tag{4.29}$$

where $\mathcal{N}_{\rm tr}(\cos\theta_i|1, \sigma_t)$ denotes a truncated Gaussian distribution between $[-1,\, 1]$ centered on 1 and with standard deviation σ_t, and $\mathcal{I}(\cos\theta_i)$ denotes an isotropic distribution in $[-1,\, 1]$.

Appendix B

4.7 LIGO/Virgo vs 3G

In Fig. 4.6 we compare our results against those obtained with current LIGO/Virgo data. For concreteness, we consider our fiducial model and ET with $N_{\rm det} = 10^5$ ($T_{\rm obs} \simeq 2$ yr). We use public LIGO/Virgo/KAGRA data products from GWTC-3 [127], selecting data from their POWER–LAW PLUS PEAK mass distribution, DEFAULT spin model, and POWER–LAW redshift model.

While indicative, there are important caveats to this comparison. As discussed in

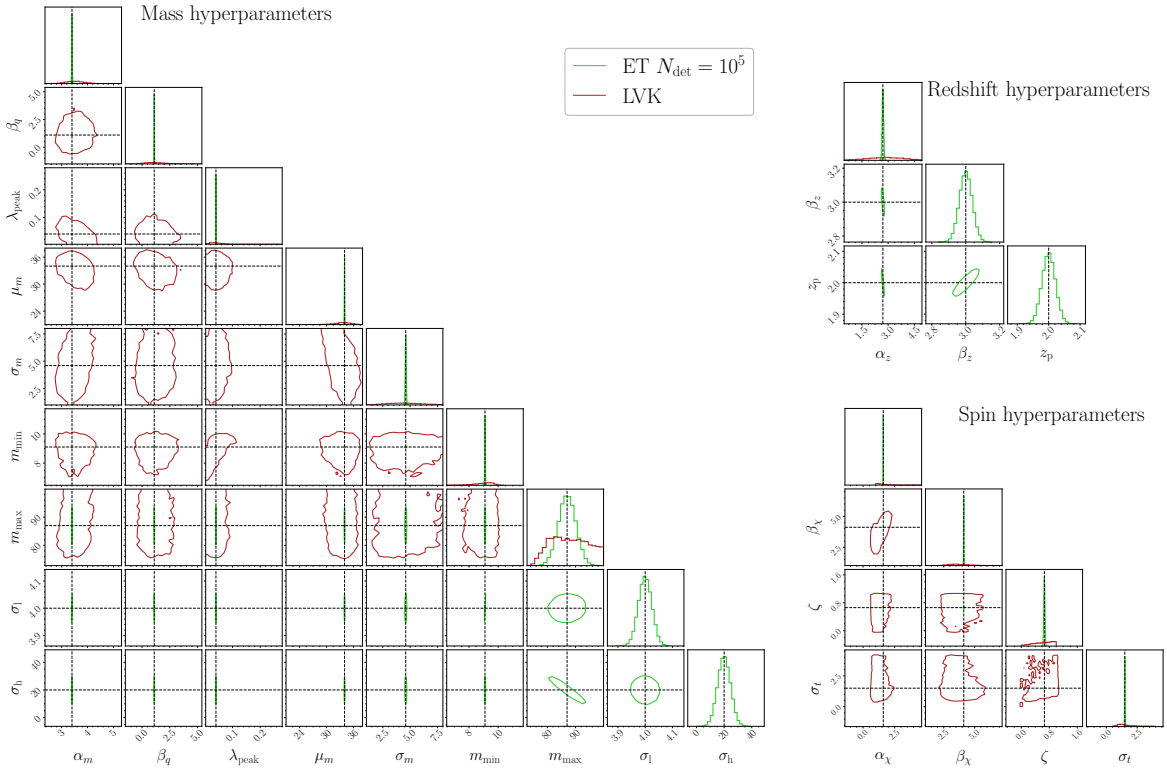

Figure 4.6: Same as Fig. 4.2 but comparing results assuming $N_{\rm det} = 10^5$ detections with ET ($T_{\rm obs} \sim 2$ yr, light curves) against current constraints from LIGO/Virgo data (LVK, dark curves). See App. 4.7 for important caveats.

App. 4.6.1, the smoothing function we adopt differs from that of Ref. [127]. First, we smooth both the lower and upper ends of the mass spectrum while they only smooth the lower end. The parameters $\sigma_{\rm h}$ and $\sigma_{\rm l}$ have different meanings in the two cases, and are thus omitted from Fig. 4.6. Moreover, while in our case the mass distribution vanishes at $m \leq m_{\rm min} - \sigma_{\rm l}$, in the prescription used by LIGO/Virgo/KAGRA it instead vanishes at $m \leq m_{\rm min}$. For a fairer comparison, we artificially shift the GWTC-3 $m_{\rm min}$ posterior by $\sigma_{\rm l} = 4\ M_\odot$. As for the redshift distribution, the Power–Law model used in Ref. [127] is described by a single hyperparameter κ, that can only be approximately compared to α_z from the Madau–Dickinson model.

Parameter	Description	Fiducial
λ_m	Mass model: Power–Law Plus Peak	
α_m	Spectral index for the power-law of the primary mass distribution.	3.4
$m_{\rm min}$	Minimum mass of the power-law component of the primary mass distribution.	9.1 $M_\odot$
$m_{\rm max}$	Maximum mass of the power-law component of the primary mass distribution.	87 $M_\odot$
$\lambda_{\rm peak}$	Fraction of binary BHs in the Gaussian component.	0.039
μ_m	Mean of the Gaussian component in the primary mass distribution.	34
σ_m	Width of the Gaussian component in the primary mass distribution.	3.6
β_q	Spectral index for the power-law of the mass ratio distribution.	1.1
$\sigma_{\rm l}$	Width of mass smoothing at the lower end of the mass distribution.	4.0
$\sigma_{\rm h}$	Width of mass smoothing at the upper end of the mass distribution.	0.5
λ_z	Redshift model: Madau–Dickinson	
α_z	Power-law index governing the rise of the star formation rate at low redshift.	2.7
β_z	Power-law index governing the decline of the star formation rate at high redshift.	3.0
$z_{\rm p}$	Redshift at which the star formation rate peaks.	2.0
$\lambda_\chi, \lambda_\theta$	Spin model: Default	
α_χ	Mean of the Beta distribution of the spin magnitudes.	1.6
β_χ	Standard deviation of the Beta distribution of the spin magnitudes.	4.12
ζ	Mixing fraction of mergers from the truncated Gaussian component for spin orientations.	0.66
σ_t	Width of the truncated Gaussian component for spin orientations.	1.5

Table 4.3: Summary of the population model parameters and their fiducial values.

Chapter 5

Conclusions and future directions

The theory of GR has profoundly transformed our understanding of the universe, reshaping our conception of spacetime and gravity, providing the basis for modern cosmology and explaining the nature of BHs and NSs. GWs, first predicted by Einstein, are a direct consequence of GR and have opened new observational windows into the cosmos. Unlike electromagnetic waves, GWs interact only weakly with matter as they travel through space at the speed of light, allowing us to probe regions of the Universe that would have been otherwise unaccessible. The observation of GWs coming from the earliest BH mergers at high redshift will allow us to explore epochs beyond the reach of electromagnetic observations. Furthermore, for the first time it is possible to test GR in the strong field regime as we will be able to approach regions characterized by a strong gravitational field, like in the vicinity of BHs and NSs. This makes GWs an unparalleled tool for studying high-energy astrophysical phenomena and understanding fundamental aspects of the evolution of our universe.

With a catalog of ~ 90 GW events detected by the LVK collaboration, we are at the forefront of a new era in astronomy. The current generation of detectors has provided unprecedented insights into the mass, spin, and redshift distributions of binary BHs and

NSs. These observations confirmed the existence of intermediate-mass BHs, challenging conventional models of stellar evolution. However, despite these advancements, there are still significant limitations in the precision with which we can measure certain parameters, particularly the spin magnitudes and orientations in merging binaries. The primary reason for this is that spin effects contribute subdominantly to the GW signal, making it harder to resolve them with the current sensitivity. An accurate measurement of spin parameters (and hyperparameters at the population-level) is fundamental to discriminate between possible binary BH formation channels. As pointed out in Chapter 2 and Chapter 3, great improvements in the detectors sensitivity is expected already in the upcoming observing runs. This will enable a deeper study of weak spin effects that were invisible to previous observing runs. But the real revolution in terms of parameter estimation will arrive with the advent of 3G detectors. With their significantly improved sensitivity, they will not only enhance the precision with which individual-event and population-level parameters are measured but, crucially, will also dramatically increase the number of detected events. This is particularly relevant for population-level studies that aim at reconstructing the true population of compact binary systems across cosmic time. In Chapter 4, we highlight that both the ET and ET+2CE configurations will reach percent-level accuracy for the majority of the population parameters within a few years of operation. Despite these breakthroughs, 3G detectors still face challenges in measuring spin orientations with high accuracy. In fact, longer observation times will be needed to achieve the 1% accuracy for the hyperparameters related to spin directions.

In addition to ground-based detectors, space-based missions like LISA will complement the efforts by probing much lower frequencies. LISA will excel in observing supermassive BH binaries and systems in the early stages of the inspiral, allowing for unprecedented accuracy in measuring spin orientations. This is possible due to the longer inspiral timescales at lower frequencies which enable a detailed characterization of spin precession. The ability to track the entire evolution of a binary inspiral phase

will provide invaluable insights into the formation channels of such systems, particularly in environments like AGN disks or globular clusters.

In conclusion, the future of GW astronomy is incredibly bright. We stand on the cusp of answering fundamental questions about the formation of compact objects, the nature of space-time, and the very origins of the universe itself. The unprecedented sensitivity of future detectors promises to open new frontiers in our quest to understand the cosmos.

Bibliography

[1] A. Einstein, in *The Principle of Relativity. Dover Books on Physics. June 1* (1952) pp. 109–164.

[2] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (1973).

[3] S. Husa, General Relativity and Gravitation **41**, 1667 (2009).

[4] J. B. Hartle, *Gravity: an introduction to Einstein's general relativity* (2021).

[5] V. Ferrari, L. Gualtieri, and P. Pani, *General relativity and its applications: black holes, compact stars, and gravitational waves* (2020).

[6] B. P. Abbott et al., Phys. Rev. Lett. **116**, 061102 (2016), arXiv:1602.03837 [gr-qc].

[7] Event Horizon Telescope Collaboration, K. Akiyama, A. Alberdi, W. Alef, K. Asada, R. Azulay, A.-K. Baczko, D. Ball, M. Baloković, J. Barrett, et al., ApJ **875**, L4 (2019), arXiv:1906.11241 [astro-ph.GA].

[8] The Event Horizon Telescope Collaboration, arXiv e-prints , arXiv:2311.08680 (2023), arXiv:2311.08680 [astro-ph.HE].

[9] S. Dodelson and F. Schmidt, *Modern Cosmology* (2020).

[10] A. Friedmann, Zeitschrift fur Physik **10**, 377 (1922).

[11] G. Lemaître, Annales de la Société Scientifique de Bruxelles **47**, 49 (1927).

[12] H. P. Robertson, ApJ **82**, 284 (1935).

[13] A. G. Walker, Proceedings of the London Mathematical Society **42**, 90 (1937).

[14] P. J. E. Peebles, *Principles of Physical Cosmology* (1993).

[15] S. Perlmutter, G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, P. G. Castro, S. Deustua, S. Fabbro, A. Goobar, D. E. Groom, et al., ApJ **517**, 565 (1999), arXiv:astro-ph/9812133 [astro-ph].

[16] A. G. Riess, A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P. M. Garnavich, R. L. Gilliland, C. J. Hogan, S. Jha, R. P. Kirshner, et al., AJ **116**, 1009 (1998), arXiv:astro-ph/9805201 [astro-ph].

[17] A. Einstein, Annalen der Physik **354**, 769 (1916).

[18] A. Einstein, Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften , 688 (1916).

[19] A. Einstein, Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften , 154 (1918).

[20] A. Buonanno, arXiv e-prints , arXiv:0709.4682 (2007), arXiv:0709.4682 [gr-qc].

[21] B. S. Sathyaprakash and B. F. Schutz, Living Reviews in Relativity **12**, 2 (2009), arXiv:0903.0338 [gr-qc].

[22] B. F. Schutz and F. Ricci, arXiv e-prints , arXiv:1005.4735 (2010), arXiv:1005.4735 [gr-qc].

[23] M. Pitkin, S. Reid, S. Rowan, and J. Hough, Living Reviews in Relativity **14**, 5 (2011), arXiv:1102.3355 [astro-ph.IM].

[24] K. Riles, Progress in Particle and Nuclear Physics **68**, 1 (2013), arXiv:1209.0667 [hep-ex].

[25] M. E. Gertsenshteĭn and V. I. Pustovoĭt, Soviet Journal of Experimental and Theoretical Physics **16**, 433 (1963).

[26] R. X. Adhikari, Reviews of Modern Physics **86**, 121 (2014), arXiv:1305.5188 [gr-qc].

[27] P. R. Saulson, Comptes Rendus Physique **14**, 288 (2013).

[28] R. W. P. Drever, J. L. Hall, F. V. Kowalski, J. Hough, G. M. Ford, A. J. Munley, and H. Ward, Applied Physics B: Lasers and Optics **31**, 97 (1983).

[29] B. J. Meers, Phys. Rev. D **38**, 2317 (1988).

[30] P. R. Saulson, *Fundamentals of Interferometric Gravitational Wave Detectors* (1994).

[31] P. R. Saulson, Phys. Rev. D **42**, 2437 (1990).

[32] B. P. Abbott, R. Abbott, R. Adhikari, P. Ajith, B. Allen, G. Allen, R. S. Amin, S. B. Anderson, W. G. Anderson, M. A. Arain, et al., Reports on Progress in Physics **72**, 076901 (2009), arXiv:0711.3041 [gr-qc].

[33] J. Abadie, B. P. Abbott, R. Abbott, M. Abernathy, C. Adams, R. Adhikari, P. Ajith, B. Allen, G. Allen, E. Amador Ceron, et al., Nuclear Instruments and Methods in Physics Research A **624**, 223 (2010), arXiv:1007.3973 [gr-qc].

[34] J. Aasi, J. Abadie, B. P. Abbott, R. Abbott, T. Abbott, M. R. Abernathy, T. Accadia, F. Acernese, C. Adams, T. Adams, et al., Classical and Quantum Gravity **32**, 115012 (2015), arXiv:1410.7764 [gr-qc].

[35] LIGO Scientific Collaboration, J. Aasi, B. P. Abbott, R. Abbott, T. Abbott, M. R. Abernathy, K. Ackley, C. Adams, T. Adams, P. Addesso, et al., Classical and Quantum Gravity **32**, 074001 (2015), arXiv:1411.4547 [gr-qc].

[36] C. M. Caves, Phys. Rev. D **23**, 1693 (1981).

[37] P. R. Saulson, Phys. Rev. D **30**, 732 (1984).

[38] Y. Levin, Phys. Rev. D **57**, 659 (1998), arXiv:gr-qc/9707013 [gr-qc].

[39] D. V. Martynov, E. D. Hall, B. P. Abbott, R. Abbott, T. D. Abbott, C. Adams, R. X. Adhikari, R. A. Anderson, S. B. Anderson, K. Arai, et al., Phys. Rev. D **93**, 112004 (2016), arXiv:1604.00439 [astro-ph.IM].

[40] T. Accadia, F. Acernese, M. Alshourbagy, P. Amico, F. Antonucci, S. Aoudia, N. Arnaud, C. Arnault, K. G. Arun, P. Astone, et al., Journal of Instrumentation **7** (3), 3012.

[41] F. Acernese, M. Agathos, K. Agatsuma, D. Aisa, N. Allemandou, A. Allocca, J. Amarni, P. Astone, G. Balestri, G. Ballardin, et al., Classical and Quantum Gravity **32**, 024001 (2015), arXiv:1408.3978 [gr-qc].

[42] Y. Aso, Y. Michimura, K. Somiya, M. Ando, O. Miyakawa, T. Sekiguchi, D. Tatsumi, and H. Yamamoto, Phys. Rev. D **88**, 043007 (2013), arXiv:1306.6747 [gr-qc].

[43] S. Fairhurst, New Journal of Physics **11**, 123006 (2009), arXiv:0908.2356 [gr-qc].

[44] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, et al., Living Reviews in Relativity **21**, 3 (2018), arXiv:1304.0670 [gr-qc].

[45] M. Punturo, M. Abernathy, F. Acernese, B. Allen, N. Andersson, K. Arun, F. Barone, B. Barr, M. Barsuglia, M. Beker, et al., Classical and Quantum Gravity **27**, 194002 (2010).

[46] D. Reitze, R. X. Adhikari, S. Ballmer, B. Barish, L. Barsotti, G. Billingsley, D. A. Brown, Y. Chen, D. Coyne, R. Eisenstein, et al., in *Bulletin of the American Astronomical Society*, Vol. 51 (2019) p. 35, arXiv:1907.04833 [astro-ph.IM].

[47] P. Amaro-Seoane, H. Audley, S. Babak, J. Baker, E. Barausse, P. Bender, E. Berti, P. Binetruy, M. Born, D. Bortoluzzi, et al., arXiv e-prints , arXiv:1702.00786 (2017), arXiv:1702.00786 [astro-ph.IM].

[48] K. Danzmann and LISA Science Team, Advances in Space Research **32**, 1233 (2003).

[49] P. Amaro-Seoane, S. Aoudia, S. Babak, P. Binétruy, E. Berti, A. Bohé, C. Caprini, M. Colpi, N. J. Cornish, K. Danzmann, et al., GW Notes **6**, 4 (2013), arXiv:1201.3621 [astro-ph.CO].

[50] M. Kramer and D. J. Champion, Classical and Quantum Gravity **30**, 224009 (2013).

[51] M. A. McLaughlin, Classical and Quantum Gravity **30**, 224008 (2013), arXiv:1310.0758 [astro-ph.IM].

[52] R. N. Manchester, G. Hobbs, M. Bailes, W. A. Coles, W. van Straten, M. J. Keith, R. M. Shannon, N. D. R. Bhat, A. Brown, S. G. Burke-Spolaor, et al., PASA **30**, e017 (2013), arXiv:1210.6130 [astro-ph.IM].

[53] G. Agazie, A. Anumarlapudi, A. M. Archibald, Z. Arzoumanian, P. T. Baker, B. Bécsy, L. Blecha, A. Brazier, P. R. Brook, S. Burke-Spolaor, et al., ApJ **951**, L8 (2023), arXiv:2306.16213 [astro-ph.HE].

[54] EPTA Collaboration, J. Antoniadis, S. Babak, A. S. Bak Nielsen, C. G. Bassa, A. Berthereau, M. Bonetti, E. Bortolas, P. R. Brook, M. Burgay, et al., A&A **678**, A48 (2023), arXiv:2306.16224 [astro-ph.HE].

[55] J. Antoniadis, Z. Arzoumanian, S. Babak, M. Bailes, A. S. Bak Nielsen, P. T. Baker, C. G. Bassa, B. Bécsy, A. Berthereau, M. Bonetti, et al., MNRAS **510**, 4873 (2022), arXiv:2201.03980 [astro-ph.HE].

[56] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, et al., ApJ **818**, L22 (2016), arXiv:1602.03846 [astro-ph.HE].

[57] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, et al., Phys. Rev. Lett. **116**, 131102 (2016), arXiv:1602.03847 [gr-qc].

[58] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, et al., Phys. Rev. Lett. **116**, 131103 (2016), arXiv:1602.03838 [gr-qc].

[59] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, et al., ApJ **826**, L13 (2016), arXiv:1602.08492 [astro-ph.HE].

[60] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, et al., Astrophys. J. Lett. **833**, L1 (2016), arXiv:1602.03842 [astro-ph.HE].

[61] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, et al., Physical Review X **6**, 041014 (2016), arXiv:1606.01210 [gr-qc].

[62] B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt, et al., Physical Review X **9**, 031040 (2019), arXiv:1811.12907 [astro-ph.HE].

[63] R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, A. Adams, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt, et al., Physical Review X **11**, 021053 (2021), arXiv:2010.14527 [gr-qc].

[64] R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, N. Adhikari, R. X. Adhikari, V. B. Adya, C. Affeldt, D. Agarwal, et al., Phys. Rev. D **109**, 022001 (2024), arXiv:2108.01045 [gr-qc].

[65] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, et al., Phys. Rev. Lett. **119**, 161101 (2017), arXiv:1710.05832 [gr-qc].

[66] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, et al., ApJ **848**, L13 (2017), arXiv:1710.05834 [astro-ph.HE].

[67] A. von Kienlin, P. Veres, O. J. Roberts, R. Hamburg, E. Bissaldi, M. S. Briggs, E. Burns, A. Goldstein, D. Kocevski, R. D. Preece, et al., ApJ **876**, 89 (2019), arXiv:1901.06158 [astro-ph.HE].

[68] M. C. Díaz, L. M. Macri, D. Garcia Lambas, C. Mendes de Oliveira, J. L. Nilo Castellón, T. Ribeiro, B. Sánchez, W. Schoenell, L. R. Abramo, S. Akras, et al., ApJ **848**, L29 (2017), arXiv:1710.05844 [astro-ph.HE].

[69] S. Valenti, D. J. Sand, S. Yang, E. Cappellaro, L. Tartaglia, A. Corsi, S. W. Jha, D. E. Reichart, J. Haislip, and V. Kouprianov, ApJ **848**, L24 (2017), arXiv:1710.05854 [astro-ph.HE].

[70] R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, N. Adhikari, R. X. Adhikari, V. B. Adya, C. Affeldt, D. Agarwal, et al., Physical Review X **13**, 041039 (2023), arXiv:2111.03606 [gr-qc].

[71] R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Agathos, et al., Phys. Rev. Lett. **125**, 101102 (2020), arXiv:2009.01075 [gr-qc].

[72] O. Barrera and I. Bartos, ApJ **929**, L1 (2022), arXiv:2201.09943 [astro-ph.HE].

[73] A. M. Holgado, A. Ortega, and C. L. Rodriguez, ApJ **909**, L24 (2021), arXiv:2012.09169 [astro-ph.HE].

[74] R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Agathos, et al., ApJ **896**, L44 (2020), arXiv:2006.12611 [astro-ph.HE].

[75] V. Dexheimer, R. O. Gomes, T. Klähn, S. Han, and M. Salinas, Phys. Rev. C **103**, 025808 (2021), arXiv:2007.08493 [astro-ph.HE].

[76] W. Lu, P. Beniamini, and C. Bonnerot, MNRAS **500**, 1817 (2021), arXiv:2009.10082 [astro-ph.HE].

[77] S. Soni, B. K. Berger, D. Davis, F. D. Renzo, A. Effler, T. A. Ferreira, J. Glanzer, E. Goetz, G. González, A. Helmling-Cornell, et al., arXiv e-prints , arXiv:2409.02831 (2024), arXiv:2409.02831 [astro-ph.IM].

[78] R. W. Kiendrebeogo, A. M. Farah, E. M. Foley, A. Gray, N. Kunert, A. Puecher, A. Toivonen, R. O. VandenBerg, S. Anand, T. Ahumada, et al., ApJ **958**, 158 (2023), arXiv:2306.09234 [astro-ph.HE].

[79] J. Creighton and W. Anderson, *Gravitational-Wave Physics and Astronomy: An Introduction to Theory, Experiment and Data Analysis.* (2011).

[80] P. Jaranowski and A. Królak, Living Reviews in Relativity **15**, 4 (2012), arXiv:0711.1115 [gr-qc].

[81] B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, V. B. Adya, C. Affeldt, M. Agathos, et al., Classical and Quantum Gravity **37**, 055002 (2020), arXiv:1908.11170 [gr-qc].

[82] L. Blanchet, Living Reviews in Relativity **5**, 3 (2002), arXiv:gr-qc/0202016 [gr-qc].

[83] A. Buonanno and T. Damour, Phys. Rev. D **59**, 084006 (1999), arXiv:gr-qc/9811091 [gr-qc].

[84] F. Pretorius, Phys. Rev. Lett. **95**, 121101 (2005), arXiv:gr-qc/0507014 [gr-qc].

[85] M. Campanelli, C. O. Lousto, P. Marronetti, and Y. Zlochower, Phys. Rev. Lett. **96**, 111101 (2006), arXiv:gr-qc/0511048 [gr-qc].

[86] T. Damour, in *General Relativity, Cosmology and Astrophysics*, Vol. 177, edited by J. Bičák and T. Ledvinka (2014) p. 111.

[87] M. Hannam, P. Schmidt, A. Bohé, L. Haegel, S. Husa, F. Ohme, G. Pratten, and M. Pürrer, Phys. Rev. Lett. **113**, 151101 (2014), arXiv:1308.3271 [gr-qc].

[88] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. J. Forteza, and A. Bohé, Phys. Rev. D **93**, 044007 (2016), arXiv:1508.07253 [gr-qc].

[89] G. Pratten, C. García-Quirós, M. Colleoni, A. Ramos-Buades, H. Estellés, M. Mateu-Lucena, R. Jaume, M. Haney, D. Keitel, J. E. Thompson, et al., Phys. Rev. D **103**, 104056 (2021), arXiv:2004.06503 [gr-qc].

[90] J. Blackman, S. E. Field, M. A. Scheel, C. R. Galley, D. A. Hemberger, P. Schmidt, and R. Smith, Phys. Rev. D **95**, 104023 (2017), arXiv:1701.00550 [gr-qc].

[91] V. Varma, S. E. Field, M. A. Scheel, J. Blackman, L. E. Kidder, and H. P. Pfeiffer, Phys. Rev. D **99**, 064045 (2019), arXiv:1812.07865 [gr-qc].

[92] D. S. Sivia and J. Skilling (1996).

[93] A. Gelman, J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin, *Bayesian Data Analysis, Third Edition*, Chapman & Hall/CRC Texts in Statistical Science (Taylor & Francis, 2013).

[94] E. Thrane and C. Talbot, PASA **36**, e010 (2019), arXiv:1809.02293 [astro-ph.IM].

[95] C. Talbot and E. Thrane, ApJ **856**, 173 (2018), arXiv:1801.02699 [astro-ph.HE].

[96] S. Vitale, D. Gerosa, W. M. Farr, and S. R. Taylor, in *Handbook of Gravitational Wave Astronomy*, edited by C. Bambi, S. Katsanevas, and K. D. Kokkotas (2022) p. 45.

[97] J. Veitch, V. Raymond, B. Farr, W. Farr, P. Graff, S. Vitale, B. Aylott, K. Blackburn, N. Christensen, M. Coughlin, et al., Phys. Rev. D **91**, 042003 (2015), arXiv:1409.7215 [gr-qc].

[98] G. Ashton, M. Hübner, P. D. Lasky, C. Talbot, K. Ackley, S. Biscoveanu, Q. Chu, A. Divakarla, P. J. Easter, B. Goncharov, et al., ApJS **241**, 27 (2019), arXiv:1811.02042 [astro-ph.IM].

[99] R. J. E. Smith, G. Ashton, A. Vajpeyi, and C. Talbot, MNRAS **498**, 4492 (2020), arXiv:1909.11873 [gr-qc].

[100] M. Dax, S. R. Green, J. Gair, J. H. Macke, A. Buonanno, and B. Schölkopf, Phys. Rev. Lett. **127**, 241103 (2021), arXiv:2106.12594 [gr-qc].

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

[102] J. Lange, R. O'Shaughnessy, and M. Rizzo, arXiv e-prints , arXiv:1805.10457 (2018), arXiv:1805.10457 [gr-qc].

[103] M. Vallisneri, Phys. Rev. D **77**, 042001 (2008), arXiv:gr-qc/0703086 [gr-qc].

[104] F. Iacovelli, M. Mancarella, S. Foffa, and M. Maggiore, Astrophys. J. Supp. S. **263**, 2 (2022), arXiv:2207.06910 [astro-ph.IM].

[105] S. Borhanian, Class. Quantum Grav. **38**, 175014 (2021), arXiv:2010.15202 [gr-qc].

[106] U. Dupletsa, J. Harms, B. Banerjee, M. Branchesi, B. Goncharov, A. Maselli, A. C. S. Oliveira, S. Ronchini, and J. Tissino, Astron. Comput. **42**, 100671 (2023), arXiv:2205.02499 [gr-qc].

[107] K. Krishna, A. Vijaykumar, A. Ganguly, C. Talbot, S. Biscoveanu, R. N. George, N. Williams, and A. Zimmerman, arXiv e-prints , arXiv:2312.06009 (2023), arXiv:2312.06009 [gr-qc].

[108] M. Spera, M. Mapelli, and A. Bressan, MNRAS **451**, 4086 (2015), arXiv:1505.05201 [astro-ph.SR].

[109] M. Mapelli, Frontiers in Astronomy and Space Sciences **7**, 38 (2020), arXiv:2105.12455 [astro-ph.HE].

[110] S. E. Woosley and A. Heger, in *Very Massive Stars in the Local Universe*, Astrophysics and Space Science Library, Vol. 412, edited by J. S. Vink (2015) p. 199, arXiv:1406.5657 [astro-ph.SR].

[111] C. L. Fryer, Classical and Quantum Gravity **20**, S73 (2003).

[112] K. A. Postnov and L. R. Yungelson, Living Reviews in Relativity **17**, 3 (2014), arXiv:1403.4754 [astro-ph.HE].

[113] C. Y. Lam and J. R. Lu, ApJ **955**, 116 (2023), arXiv:2308.03302 [astro-ph.SR].

[114] K. Belczynski, V. Kalogera, and T. Bulik, ApJ **572**, 407 (2002), arXiv:astro-ph/0111452 [astro-ph].

[115] I. Mandel and A. Farmer, Phys. Rep. **955**, 1 (2022), arXiv:1806.05820 [astro-ph.HE].

[116] C. Talbot and E. Thrane, Phys. Rev. D **96**, 023012 (2017), arXiv:1704.08370 [astro-ph.HE].

[117] D. Gerosa, M. Kesden, E. Berti, R. O'Shaughnessy, and U. Sperhake, Phys. Rev. D **87**, 104028 (2013), arXiv:1302.4442 [gr-qc].

[118] C. L. Rodriguez, M. Zevin, C. Pankow, V. Kalogera, and F. A. Rasio, Astrophys. J. Lett. **832**, L2 (2016), arXiv:1609.05916 [astro-ph.HE].

[119] S. Vitale, R. Lynch, R. Sturani, and P. Graff, Classical and Quantum Gravity **34**, 03LT01 (2017), arXiv:1503.04307 [gr-qc].

[120] S. Stevenson, C. P. L. Berry, and I. Mandel, MNRAS **471**, 2801 (2017), arXiv:1703.06873 [astro-ph.HE].

[121] M. Volonteri, Science **337**, 544 (2012), arXiv:1208.1106 [astro-ph.CO].

[122] A. C. Fabian, ARA&A **50**, 455 (2012), arXiv:1204.4114 [astro-ph.CO].

[123] J. Kormendy and L. C. Ho, ARA&A **51**, 511 (2013), arXiv:1304.7762 [astro-ph.CO].

[124] F. Wang, X. Fan, J. Yang, C. Mazzucchelli, X.-B. Wu, J.-T. Li, E. Bañados, E. P. Farina, R. Nanni, Y. Ai, et al., ApJ **908**, 53 (2021), arXiv:2011.12458 [astro-ph.GA].

[125] A. M. Ghez, B. L. Klein, M. Morris, and E. E. Becklin, ApJ **509**, 678 (1998), arXiv:astro-ph/9807210 [astro-ph].

[126] M. Milosavljević and D. Merritt, in *The Astrophysics of Gravitational Wave Sources*, American Institute of Physics Conference Series, Vol. 686, edited by J. M. Centrella (AIP, 2003) pp. 201–210, arXiv:astro-ph/0212270 [astro-ph].

[127] R. Abbott et al., Phys. Rev. X **13**, 011048 (2023), arXiv:2111.03634 [astro-ph.HE].

[128] R. Abbott et al., ApJ **913**, L7 (2021), arXiv:2010.14533 [astro-ph.HE].

[129] K. Belczynski, J. Klencki, C. E. Fields, A. Olejak, E. Berti, G. Meynet, C. L. Fryer, D. E. Holz, R. O'Shaughnessy, D. A. Brown, et al., A&A **636**, A104 (2020), arXiv:1706.07053 [astro-ph.HE].

[130] K. Belczynski, D. E. Holz, T. Bulik, and R. O'Shaughnessy, Nature **534**, 512 (2016), arXiv:1602.04531 [astro-ph.HE].

[131] N. Giacobbo and M. Mapelli, Mon. Not. R. Astron. Soc. **480**, 2011 (2018), arXiv:1806.00001 [astro-ph.HE].

[132] S. Banerjee, MNRAS **503**, 3371 (2021), arXiv:2011.07000 [astro-ph.HE].

[133] F. Santoliquido, M. Mapelli, N. Giacobbo, Y. Bouffanais, and M. C. Artale, MNRAS **502**, 4877 (2021), arXiv:2009.03911 [astro-ph.HE].

[134] V. De Renzis, D. Gerosa, G. Pratten, P. Schmidt, and M. Mould, Phys. Rev. D **106**, 084040 (2022), arXiv:2207.00030 [gr-qc].

[135] V. De Renzis, D. Gerosa, M. Mould, R. Buscicchio, and L. Zanga, Phys. Rev. D **108**, 024024 (2023), arXiv:2304.13063 [gr-qc].

[136] V. De Renzis, F. Iacovelli, D. Gerosa, M. Mancarella, and C. Pacilio, (2024), arXiv:2410.17325 [astro-ph.HE].

[137] D. Gangardt, D. Gerosa, M. Kesden, V. De Renzis, and N. Steinle, Phys. Rev. D **106**, 024019 (2022), arXiv:2204.00026 [gr-qc].

[138] D. Gerosa, G. Fumagalli, M. Mould, G. Cavallotto, D. P. Monroy, D. Gangardt, and V. De Renzis, Phys. Rev. D **108**, 024042 (2023), arXiv:2304.04801 [gr-qc].

[139] G. Fumagalli, I. Romero-Shaw, D. Gerosa, V. De Renzis, K. Kritos, and A. Olejak, Phys. Rev. D **110**, 063012 (2024), arXiv:2405.14945 [astro-ph.HE].

[140] D. Gerosa, V. De Renzis, F. Tettoni, M. Mould, A. Vecchio, and C. Pacilio, arXiv e-prints , arXiv:2409.07519 (2024), arXiv:2409.07519 [gr-qc].

[141] U. von Luxburg, arXiv e-prints , arXiv:0711.0189 (2007), arXiv:0711.0189 [cs.DS].

[142] T. A. Apostolatos, C. Cutler, G. J. Sussman, and K. S. Thorne, Phys. Rev. D **49**, 6274 (1994).

[143] L. E. Kidder, Phys. Rev. D **52**, 821 (1995), arXiv:gr-qc/9506022 [gr-qc].

[144] A. Sesana, E. Barausse, M. Dotti, and E. M. Rossi, ApJ **794**, 104 (2014), arXiv:1402.7088 [astro-ph.CO].

[145] M. Sayeb, L. Blecha, L. Z. Kelley, D. Gerosa, M. Kesden, and J. Thomas, MNRAS **501**, 2531 (2021), arXiv:2006.06647 [astro-ph.GA].

[146] N. Loutrel, T. Tanaka, and N. Yunes, Phys. Rev. D **98**, 064020 (2018), arXiv:1806.07431 [gr-qc].

[147] R. Abbott et al., (2021), arXiv:2108.01045 [gr-qc].

[148] R. Abbott et al., (2021), arXiv:2111.03606 [gr-qc].

[149] M. Hannam, C. Hoy, J. E. Thompson, S. Fairhurst, V. Raymond, m. o. t. LIGO, and Virgo collaborations, (2021), arXiv:2112.11300 [gr-qc].

[150] V. Varma, S. Biscoveanu, T. Islam, F. H. Shaik, C.-J. Haster, M. Isi, W. M. Farr, S. E. Field, and S. Vitale, Phys. Rev. Lett. **128**, 191102 (2022), arXiv:2201.01302 [astro-ph.HE].

[151] E. Payne, S. Hourihane, J. Golomb, R. Udall, D. Davis, and K. Chatziioannou, (2022), arXiv:2206.11932 [gr-qc].

[152] I. Romero-Shaw, P. D. Lasky, E. Thrane, and J. Calderón Bustillo, ApJ **903**, L5 (2020), arXiv:2009.04771 [astro-ph.HE].

[153] R. Abbott et al., (2021), arXiv:2111.03634 [astro-ph.HE].

[154] T. A. Callister, S. J. Miller, K. Chatziioannou, and W. M. Farr, ApJ **937**, L13 (2022), arXiv:2205.08574 [astro-ph.HE].

[155] M. Mould, D. Gerosa, F. S. Broekgaarden, and N. Steinle, (2022), arXiv:2205.12329 [astro-ph.HE].

[156] S. Khan, K. Chatziioannou, M. Hannam, and F. Ohme, Phys. Rev. D **100**, 024059 (2019), arXiv:1809.10113 [gr-qc].

[157] S. Biscoveanu, M. Isi, V. Varma, and S. Vitale, Phys. Rev. D **104**, 103018 (2021), arXiv:2106.06492 [gr-qc].

[158] V. Varma, M. Isi, S. Biscoveanu, W. M. Farr, and S. Vitale, Phys. Rev. D **105**, 024045 (2022), arXiv:2107.09692 [astro-ph.HE].

[159] D. Gerosa, M. Mould, D. Gangardt, P. Schmidt, G. Pratten, and L. M. Thomas, Phys. Rev. D **103**, 064067 (2021), arXiv:2011.11948 [gr-qc].

[160] P. Schmidt, F. Ohme, and M. Hannam, Phys. Rev. D **91**, 024043 (2015), arXiv:1408.1810 [gr-qc].

[161] C. Henshaw, R. O'Shaughnessy, and L. Cadonati, Classical and Quantum Gravity **39**, 125003 (2022), arXiv:2201.05220 [gr-qc].

[162] T. Damour, Phys. Rev. D **64**, 124013 (2001), arXiv:gr-qc/0103018 [gr-qc].

[163] É. Racine, Phys. Rev. D **78**, 044021 (2008), arXiv:0803.1820 [gr-qc].

[164] D. Gerosa, M. Kesden, U. Sperhake, E. Berti, and R. O'Shaughnessy, Phys. Rev. D **92**, 064016 (2015), arXiv:1506.03492 [gr-qc].

[165] L. M. Thomas, P. Schmidt, and G. Pratten, Phys. Rev. D **103**, 083022 (2021), arXiv:2012.02209 [gr-qc].

[166] D. Gangardt, N. Steinle, M. Kesden, D. Gerosa, and E. Stoikos, Phys. Rev. D **103**, 124026 (2021), arXiv:2103.03894 [gr-qc].

[167] S. Fairhurst, R. Green, M. Hannam, and C. Hoy, Phys. Rev. D **102**, 041302 (2020), arXiv:1908.00555 [gr-qc].

[168] M. Kesden, D. Gerosa, R. O'Shaughnessy, E. Berti, and U. Sperhake, Phys. Rev. Lett. **114**, 081103 (2015), arXiv:1411.0674 [gr-qc].

[169] G. Pratten, P. Schmidt, R. Buscicchio, and L. M. Thomas, Physical Review Research **2**, 043096 (2020), arXiv:2006.16153 [gr-qc].

[170] J. Veitch and A. Vecchio, Phys. Rev. D **78**, 022001 (2008), arXiv:0801.4313 [gr-qc].

[171] J. S. Speagle, MNRAS **493**, 3132 (2020), arXiv:1904.02180 [astro-ph.IM].

[172] S. Vitale, D. Gerosa, C.-J. Haster, K. Chatziioannou, and A. Zimmerman, Phys. Rev. Lett. **119**, 251103 (2017), arXiv:1707.04637 [gr-qc].

[173] K. Chatziioannou, G. Lovelace, M. Boyle, M. Giesler, D. A. Hemberger, R. Katebi, L. E. Kidder, H. P. Pfeiffer, M. A. Scheel, and B. Szilágyi, Phys. Rev. D **98**, 044028 (2018), arXiv:1804.03704 [gr-qc].

[174] G. Pratten, S. Husa, C. García-Quirós, M. Colleoni, A. Ramos-Buades, H. Estellés, and R. Jaume, Phys. Rev. D **102**, 064001 (2020), arXiv:2001.11412 [gr-qc].

[175] C. García-Quirós, M. Colleoni, S. Husa, H. Estellés, G. Pratten, A. Ramos-Buades, M. Mateu-Lucena, and R. Jaume, Phys. Rev. D **102**, 064002 (2020), arXiv:2001.10914 [gr-qc].

[176] H. Estellés, M. Colleoni, C. García-Quirós, S. Husa, D. Keitel, M. Mateu-Lucena, M. d. L. Planas, and A. Ramos-Buades, Phys. Rev. D **105**, 084040 (2022), arXiv:2105.05872 [gr-qc].

[177] V. Varma, S. E. Field, M. A. Scheel, J. Blackman, D. Gerosa, L. C. Stein, L. E. Kidder, and H. P. Pfeiffer, Physical Review Research **1**, 033015 (2019), arXiv:1905.09300 [gr-qc].

[178] Ž. Ivezić, A. J. Connelly, J. T. Vanderplas, and A. Gray, *Statistics, Data Mining, and Machine Learning in Astronomy* (Princeton University Press, 2019).

[179] J. Chambers, W. Cleveland, B. Kleiner, and P. Tukey, *Graphical Methods for Data Analysis* (Wadsworth International Group, 1983).

[180] S. Vitale, R. Lynch, J. Veitch, V. Raymond, and R. Sturani, Phys. Rev. Lett. **112**, 251101 (2014), arXiv:1403.0129 [gr-qc].

[181] M. Pürrer, M. Hannam, and F. Ohme, Phys. Rev. D **93**, 084042 (2016), arXiv:1512.04955 [gr-qc].

[182] K. Chatziioannou, A. Klein, N. Yunes, and N. Cornish, Phys. Rev. D **95**, 104004 (2017), arXiv:1703.03967 [gr-qc].

[183] S. Marsat and J. G. Baker, (2018), arXiv:1806.10734 [gr-qc].

[184] E. Hamilton, L. London, J. E. Thompson, E. Fauchon-Jones, M. Hannam, C. Kalaghatgi, S. Khan, F. Pannarale, and A. Vano-Vinuales, Phys. Rev. D **104**, 124027 (2021), arXiv:2107.08876 [gr-qc].

[185] V. De Renzis and D. Gerosa, github.com/ViolaDeRenzis/twoprecessingspins, doi.org/10.5281/zenodo.6777952 (2022).

[186] M. Mould and D. Gerosa, Phys. Rev. D **105**, 024076 (2022), arXiv:2110.05507 [astro-ph.HE].

[187] V. Baibhav, E. Berti, D. Gerosa, M. Mapelli, N. Giacobbo, Y. Bouffanais, and U. N. Di Carlo, Phys. Rev. D **100**, 064060 (2019), arXiv:1906.04197 [gr-qc].

[188] M. Mapelli, in *Handbook of Gravitational Wave Astronomy* (Springer, 2021) p. 16.

[189] D. Gerosa, M. Kesden, R. O'Shaughnessy, A. Klein, E. Berti, U. Sperhake, and D. Trifirò, Phys. Rev. Lett. **115**, 141102 (2015), arXiv:1506.09116 [gr-qc].

[190] C. O. Lousto and J. Healy, Phys. Rev. D **93**, 124074 (2016), arXiv:1601.05086 [gr-qc].

[191] N. K. Johnson-McDaniel, S. Kulkarni, and A. Gupta, Phys. Rev. D **106**, 023001 (2022), arXiv:2107.11902 [astro-ph.HE].

[192] V. Varma, M. Mould, D. Gerosa, M. A. Scheel, L. E. Kidder, and H. P. Pfeiffer, Phys. Rev. D **103**, 064003 (2021), arXiv:2012.07147 [gr-qc].

[193] B. McKernan, K. E. S. Ford, R. O'Shaugnessy, and D. Wysocki, Mon. Not. R. Astron. Soc. **494**, 1203 (2020), arXiv:1907.04356 [astro-ph.HE].

[194] D. Gerosa and M. Fishbach, Nature Astronomy **5**, 749 (2021), arXiv:2105.03439 [astro-ph.HE].

[195] J. M. Bardeen and J. A. Petterson, ApJ **195**, L65 (1975).

[196] M. Mould and D. Gerosa, Phys. Rev. D **101**, 124037 (2020), arXiv:2003.02281 [gr-qc].

[197] H. Jeffreys and R. B. Lindsay, Physics Today **16**, 68 (1963).

[198] W. D. Penny and G. R. Ridgway, PLoS ONE **8**, e59655 (2013).

[199] L. Kelley, J. Open Source Softw. **6**, 2784 (2021).

[200] D. Gerosa and M. Kesden, Phys. Rev. D **93**, 124066 (2016), arXiv:1605.01067 [astro-ph.HE].

[201] D. Gerosa, G. Fumagalli, M. Mould, G. Cavallotto, D. Padilla Monroy, D. Gangardt, and V. De Renzis, (2023), arXiv:2304.04801 [gr-qc].

[202] E. Payne, C. Talbot, and E. Thrane, Phys. Rev. D **100**, 123017 (2019), arXiv:1905.05477 [astro-ph.IM].

[203] C. Mills and S. Fairhurst, Phys. Rev. D **103**, 024042 (2021), arXiv:2007.04313 [gr-qc].

[204] R. Cotesta, A. Buonanno, A. Bohé, A. Taracchini, I. Hinder, and S. Ossokine, Phys. Rev. D **98**, 084028 (2018), arXiv:1803.10701 [gr-qc].

[205] V. De Renzis and D. Gerosa, github.com/ViolaDeRenzis/updowninjections, doi.org/10.5281/zenodo.7974556 (2023).

[206] M. C. Miller and J. H. Krolik, ApJ **774**, 43 (2013), arXiv:1307.6569 [astro-ph.HE].

[207] N. Steinle and D. Gerosa, MNRAS **519**, 5031 (2023), arXiv:2211.00044 [astro-ph.HE].

[208] K. Chatziioannou, N. Cornish, A. Klein, and N. Yunes, Phys. Rev. D **89**, 104023 (2014), arXiv:1404.3180 [gr-qc].

[209] B. P. Abbott et al., Astrophys. J. Lett. **882**, L24 (2019), arXiv:1811.12940 [astro-ph.HE].

[210] T. A. Callister, C.-J. Haster, K. K. Y. Ng, S. Vitale, and W. M. Farr, Astrophys. J. Lett. **922**, L5 (2021), arXiv:2106.00521 [astro-ph.HE].

[211] C. Adamcewicz and E. Thrane, Mon. Not. R. Astron. Soc. **517**, 3928 (2022), arXiv:2208.03405 [astro-ph.HE].

[212] C. Adamcewicz, P. D. Lasky, and E. Thrane, Astrophys. J. **958**, 13 (2023), arXiv:2307.15278 [astro-ph.HE].

[213] J. Heinzel, S. Vitale, and S. Biscoveanu, Phys. Rev. D **109**, 103006 (2024), arXiv:2312.00993 [astro-ph.HE].

[214] S. Rinaldi, W. Del Pozzo, M. Mapelli, A. Lorenzo-Medina, and T. Dent, Astron. Astrophys. **684**, A204 (2024), arXiv:2310.03074 [astro-ph.HE].

[215] G. Pierra, S. Mastrogiovanni, and S. Perriès, (2024), arXiv:2406.01679 [gr-qc].

[216] J. Heinzel, M. Mould, and S. Vitale, (2024), arXiv:2406.16844 [astro-ph.HE].

[217] I. Mandel, W. M. Farr, and J. R. Gair, Mon. Not. R. Astron. Soc. **486**, 1086 (2019), arXiv:1809.02063 [physics.data-an].

[218] V. Tiwari, Class. Quantum Grav. **35**, 145009 (2018), arXiv:1712.00482 [astro-ph.HE].

[219] W. M. Farr, Res. Notes AAS **3**, 66 (2019), arXiv:1904.10879 [astro-ph.IM].

[220] C. Talbot and J. Golomb, Mon. Not. R. Astron. Soc. **526**, 3495 (2023), arXiv:2304.06138 [astro-ph.IM].

[221] M. Punturo et al., Class. Quantum Grav. **27**, 194002 (2010).

[222] F. Iacovelli, M. Mancarella, S. Foffa, and M. Maggiore, Astrophys. J. **941**, 208 (2022), arXiv:2207.02771 [gr-qc].

[223] M. Branchesi, M. Maggiore, et al., J. Cosmology Astropart. Phys. **2023**, 068 (2023), arXiv:2303.15923 [gr-qc].

[224] Y. Li, I. S. Heng, M. L. Chan, C. Messenger, and X. Fan, Phys. Rev. D **105**, 043010 (2022), arXiv:2109.07389 [astro-ph.IM].

[225] M. Pieroni, A. Ricciardone, and E. Barausse, Sci. Rep. **12**, 17940 (2022), arXiv:2203.12586 [astro-ph.CO].

[226] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, K. Ackley, C. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, C. Affeldt, et al., Class. Quantum Grav. **34**, 044001 (2017), arXiv:1607.08697 [astro-ph.IM].

[227] M. Maggiore, C. Van Den Broeck, et al., J. Cosmology Astropart. Phys. **2020**, 050 (2020), arXiv:1912.02622 [astro-ph.CO].

[228] M. Evans, R. X. Adhikari, C. Afle, S. W. Ballmer, S. Biscoveanu, S. Borhanian, D. A. Brown, Y. Chen, R. Eisenstein, A. Gruson, et al., (2021), arXiv:2109.09882 [astro-ph.IM].

[229] J. R. Gair, A. Antonelli, and R. Barbieri, Mon. Not. R. Astron. Soc. **519**, 2736 (2023), arXiv:2205.07893 [gr-qc].

[230] M. Maggiore, *Gravitational Waves. Vol. 1: Theory and Experiments* (Oxford, 2007).

[231] E. Pizzati, S. Sachdev, A. Gupta, and B. S. Sathyaprakash, Phys. Rev. D **105**, 104016 (2022), arXiv:2102.07692 [gr-qc].

[232] J. Janquart, T. Baka, A. Samajdar, T. Dietrich, and C. Van Den Broeck, Mon. Not. R. Astron. Soc. **523**, 1699 (2023).

[233] D. Gerosa and M. Bellotti, Class. Quantum Grav. **41**, 125002 (2024), arXiv:2404.16930 [astro-ph.HE].

[234] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, et al., github.com/google/jax (2018).

[235] N. Aghanim et al., Astron. Astrophys. **641**, A6 (2020), arXiv:1807.06209 [astro-ph.CO].

[236] V. Srivastava, D. Davis, K. Kuns, P. Landry, S. Ballmer, M. Evans, E. D. Hall, J. Read, and B. S. Sathyaprakash, Astrophys. J. **931**, 22 (2022), arXiv:2201.10668 [gr-qc].

[237] M. Fishbach, D. E. Holz, and W. M. Farr, Astrophys. J. Lett. **863**, L41 (2018), arXiv:1805.10270 [astro-ph.HE].

[238] P. Madau and M. Dickinson, ARA&A **52**, 415 (2014), arXiv:1403.0007 [astro-ph.CO].

[239] E. Payne, S. Banagiri, P. D. Lasky, and E. Thrane, Phys. Rev. D **102**, 102004 (2020), arXiv:2006.11957 [astro-ph.CO].

[240] T. Kinugawa, K. Inayoshi, K. Hotokezaka, D. Nakauchi, and T. Nakamura, Mon. Not. R. Astron. Soc. **442**, 2963 (2014), arXiv:1402.6672 [astro-ph.HE].

[241] M. Sasaki, T. Suyama, T. Tanaka, and S. Yokoyama, Class. Quantum Grav. **35**, 063001 (2018), arXiv:1801.05235 [astro-ph.CO].

[242] K. K. Y. Ng, S. Chen, B. Goncharov, U. Dupletsa, S. Borhanian, M. Branchesi, J. Harms, M. Maggiore, B. S. Sathyaprakash, and S. Vitale, Astrophys. J. Lett. **931**, L12 (2022), arXiv:2108.07276 [astro-ph.CO].

[243] M. Mancarella, F. Iacovelli, and D. Gerosa, Phys. Rev. D **107**, L101302 (2023), arXiv:2303.16323 [gr-qc].

[244] K. Belczynski, A. Heger, W. Gladysz, A. J. Ruiter, S. Woosley, G. Wiktorowicz, H. Y. Chen, T. Bulik, R. O'Shaughnessy, D. E. Holz, et al., Astron. Astrophys. **594**, A97 (2016), arXiv:1607.03116 [astro-ph.HE].

[245] P. Marchant, M. Renzo, R. Farmer, K. M. W. Pappas, R. E. Taam, S. E. de Mink, and V. Kalogera, Astrophys. J. **882**, 36 (2019), arXiv:1810.13412 [astro-ph.HE].

[246] S. Stevenson, M. Sampson, J. Powell, A. Vigna-Gómez, C. J. Neijssel, D. Szécsi, and I. Mandel, Astrophys. J. **882**, 121 (2019), arXiv:1904.02821 [astro-ph.HE].

[247] S. E. Woosley, Astrophys. J. **836**, 244 (2017), arXiv:1608.08939 [astro-ph.HE].

[248] L. A. C. van Son, S. E. de Mink, M. Renzo, S. Justham, E. Zapartas, K. Breivik, T. Callister, W. M. Farr, and C. Conroy, Astrophys. J. **940**, 184 (2022), arXiv:2209.13609 [astro-ph.HE].

[249] F. Antonini and F. A. Rasio, Astrophys. J. **831**, 187 (2016), arXiv:1606.04889 [astro-ph.HE].

[250] H. Tagawa, Z. Haiman, and B. Kocsis, Astrophys. J. **898**, 25 (2020), arXiv:1912.08218 [astro-ph.GA].

[251] G. Clark, A. Gonye, and S. J. Miller, (2021), arXiv:2101.08162 [stat.OT].

[252] I. Mandel, C. P. L. Berry, F. Ohme, S. Fairhurst, and W. M. Farr, Classical and Quantum Gravity **31**, 155005 (2014), arXiv:1404.2382 [gr-qc].

[253] M. Fishbach and V. Kalogera, Astrophys. J. Lett. **914**, L30 (2021), arXiv:2105.06491 [astro-ph.HE].

[254] L. A. C. van Son, S. E. de Mink, T. Callister, S. Justham, M. Renzo, T. Wagg, F. S. Broekgaarden, F. Kummer, R. Pakmor, and I. Mandel, Astrophys. J. **931**, 17 (2022), arXiv:2110.01634 [astro-ph.HE].

[255] C. J. Neijssel, A. Vigna-Gómez, S. Stevenson, J. W. Barrett, S. M. Gaebel, F. S. Broekgaarden, S. E. de Mink, D. Szécsi, S. Vinciguerra, and I. Mandel, Mon. Not. R. Astron. Soc. **490**, 3740 (2019), arXiv:1906.08136 [astro-ph.SR].

[256] I. Mandel and R. O'Shaughnessy, Class. Quantum Grav. **27**, 114007 (2010), arXiv:0912.1074 [astro-ph.HE].

[257] D. Gerosa, E. Berti, R. O'Shaughnessy, K. Belczynski, M. Kesden, D. Wysocki, and W. Gladysz, Phys. Rev. D **98**, 084036 (2018), arXiv:1808.02491 [astro-ph.HE].

[258] D. Gerosa and E. Berti, Phys. Rev. D **95**, 124046 (2017), arXiv:1703.06223 [gr-qc].

[259] M. Fishbach, D. E. Holz, and B. Farr, ApJ **840**, L24 (2017), arXiv:1703.06869 [astro-ph.HE].

[260] J. Fuller and L. Ma, Astrophys. J. Lett. **881**, L1 (2019), arXiv:1907.03714 [astro-ph.SR].

[261] P. Madau and T. Fragos, Astrophys. J. **840**, 39 (2017), arXiv:1606.07887 [astro-ph.GA].

[262] M. Dominik, E. Berti, R. O'Shaughnessy, I. Mandel, K. Belczynski, C. Fryer, D. E. Holz, T. Bulik, and F. Pannarale, Astrophys. J. **806**, 263 (2015), arXiv:1405.7016 [astro-ph.HE].

Acknowledgments

Aknowledgments

I thank my supervisor, Davide Gerosa, for his constant support and dedication throughout my research. His passion and guidance have been a constant source of inspiration for me. I also thank Daria Gangardt, Nathan Steinle, Floor Broekgaarden, Roberto Cotesta, Colm Talbot, Isobel Romero-Shaw, Chris Moore, Salvatore Vitale, Neil Cornish, Sylvia Biscoveanu, Vijay Varma, Max Isi, Andrea Antonelli, Ssohrab Borhanian, Rodrigo Tenorio, Arianna Renzini, and Christopher Moore for discussions.

Funding

I am supported by ERC Starting Grant No. 945155–GWmining, Cariplo Foundation Grant No. 2021-0555, MUR PRIN Grant No. 2022-Z9X4XS, MUR Grant "Progetto Dipartimenti di Eccellenza 2023-2027" (BiCoQ), and the ICSC National Research Centre funded by NextGenerationEU, and by an "Exchange Extra-EU" scholarship of the University of Milano-Bicocca. Computational work was performed at CINECA with allocations through INFN and Bicocca.