UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA
Department of Physics “Giuseppe Occhialini”
Ph.D. program in Physics and Astronomy - Cycle XXXVIII
Astrophysics Curriculum

Methods for exploiting gravitational-wave data from massive black-hole binaries

Alice Spadaro
Registration number: 814890

Supervisor: Prof. Davide Gerosa

Co-supervisor: Dr. Riccardo Buscicchio

Tutor: Prof. Monica Colpi

Coordinator: Prof. Laura D’Alfonso

Academic Year 2024/2025

Gutta cavat lapidem
Lucrezio

Abstract

Over the past decade, gravitational-wave astronomy has evolved from a largely theoretical field into a mature observational science. This transition was marked by the first direct detection of gravitational waves from a binary black-hole merger in 2015. Since then, continued observations have revealed a steadily growing population of compact-object mergers. Looking ahead, the scientific potential of gravitational-wave astronomy is expected to expand substantially with the advent of future observatories, which will extend observations to new frequency bands and enable the detection of previously inaccessible astrophysical sources. In this context, the future LISA mission will play a central role by surveying the millihertz regime, where mergers of massive black-hole binaries represent some of the most promising and information-rich sources. In this Thesis, we present advances in the statistical data-analysis framework for massive black-hole binaries, ranging from population-based studies to the characterization of individual sources, while also addressing key challenges relevant for the scientific exploitation of the mission. In particular, we focus on: (i) investigating the generation of hardening processes within a model-selection framework aimed at distinguishing between different astrophysical formation channels; (ii) the impact of instrumental glitches on parameter inference; and (iii) the characterization of single-source parameters using simulation-based inference. Together, these results yield a significant contribution to an accurate and robust inference of massive black-hole binaries in the LISA observational regime.

Riassunto

Con la prima rivelazione diretta delle onde gravitazionali, avvenuta nel 2015 e associata alla fusione di una binaria di buchi neri, l'astronomia gravitazionale si é affermata come scienza osservativa. Le rivelazioni successive hanno progressivamente arricchito il catalogo di eventi, aprendo una nuova finestra sull'astrofisica degli oggetti compatti. L'avvento di osservatori di nuova generazione estenderá le osservazioni a bande di frequenza finora inesplorate, aprendo l'accesso a nuove classi di sorgenti astrofisiche. In questo scenario, la futura missione spaziale LISA avrá un ruolo centrale nello studio della banda millihertz, in cui le fusioni di binarie di buchi neri massicci costituiscono una delle sorgenti di maggiore interesse per l'astrofisica. Questa Tesi si inserisce in tale contesto contribuendo allo sviluppo di metodologie avanzate di analisi dei dati per lo studio di queste sorgenti, affrontando temi legati sia all'inferenza a livello di popolazione sia alla caratterizzazione dettagliata di singoli eventi. Il lavoro si concentra: (i) sull'inferenza dei canali di formazione attraverso modelli astrofisici di hardening; (ii) sull'analisi dell'impatto di transienti strumentali sull'inferenza dei parametri fisici; (iii) e sulla caratterizzazione delle sorgenti usando techniche di machine learning. Nel loro insieme, questi risultati contribuiscono a rafforzare le basi metodologiche necessarie per un'interpretazione accurata e affidabile delle osservazioni di binarie di buchi neri massicci nel regime osservativo di LISA.

Acronyms

BH	black-hole
GR	General Relativity
GW	gravitational-wave
KL	Kullback-Leibler
LISA	Laser Interferometer Space Antenna
LPF	LISA Pathfinder
MCMC	Markov Chain Monte Carlo
NPE	neural posterior estimation
PSD	power spectral density
SBI	simulation-based inference
SNR	signal-to-noise ratio
TDI	time-delay interferometry
TM	test mass

Notation

Constants and units

Throughout the Thesis we adopt the CGS units system.

The speed of light and the Newton constant are normally written explicitly and set, respectively, to $c = 3 \times 10^{10}\,\mathrm{cm} \cdot \mathrm{s}^{-1}$ and $G = 6.67 \times 10^{-8}\,\mathrm{g}^{-1} \cdot \mathrm{cm}^3 \cdot \mathrm{s}^{-2}$. Occasionally there are Sections where we use unit $c = 1$, but in this case the use of this unit is always stated within the section.

The solar mass is set to $\mathrm{M}_\odot = 2 \times 10^{33}\,\mathrm{g}$ and the typical astrophysical distances are set in units of parsec ($1\,\mathrm{pc} = 3.09 \times 10^{18}\,\mathrm{cm}$).

Contents

1 Introduction **2**

1.1 Executive summary . . . 2
1.2 Theoretical foundations of gravitational waves . . . 2
1.3 From ground to space: gravitational-wave observatories . . . 4
1.4 The Laser Interferometer Space Antenna . . . 5
1.4.1 Massive black-hole binaries in the LISA landscape . . . 5
1.4.2 Mission concept and instrument design . . . 7
1.5 Gravitational-wave data analysis . . . 10
1.5.1 Signal extraction . . . 11
1.5.2 Bayesian statistical inference . . . 11
1.5.3 Future challenges: the LISA case . . . 13
1.6 Computational approaches for Bayesian inference . . . 15
1.6.1 Stochastic sampling . . . 15
1.6.2 Simulation-based inference . . . 18
1.7 Research summary . . . 22

2 Stars or gas? Constraining the hardening processes of massive black-hole binaries with LISA **25**

Overview . . . 25
2.1 Executive summary . . . 26
2.2 Astrophysical models . . . 26
2.2.1 Galaxy evolution model . . . 26
2.2.2 Black-hole growth . . . 26
2.2.3 Binary hardening . . . 28
2.2.4 Fiducial population . . . 28
2.3 Mock LISA catalogs . . . 29
2.4 Statistical inference . . . 32
2.5 Results & Discussion . . . 34
2.5.1 Inference . . . 34
2.5.2 Model selection . . . 35
2.5.3 Source parameters . . . 36
2.5.4 Importance of the BH spins . . . 37

3 Glitch systematics on the observation of massive black-hole binaries with LISA **40**

Overview . . . 40
3.1 Executive summary . . . 41
3.2 LISA Pathfinder . . . 41
3.2.1 LPF glitches . . . 41

3.2.2 Differential TM acceleration . . . 42
3.3 Transients modeling . . . 43
3.3.1 Acceleration transients . . . 44
3.3.2 OMS Transients . . . 45
3.3.3 GW Transients . . . 45
3.4 Acceleration TDIs . . . 48
3.5 Inference . . . 52
3.6 Results . . . 53
3.6.1 GW + Glitch . . . 55
3.6.2 Glitch only . . . 60

4 Accurate and efficient simulation-based inference for massive black-hole binaries with LISA 65
Overview . . . 65
4.1 Executive summary . . . 66
4.2 Methods . . . 66
4.2.1 LISA response . . . 66
4.2.2 Dingo setup . . . 67
4.3 Results . . . 69
4.3.1 Model validation . . . 69
4.3.2 Benchmarking . . . 70
4.3.3 Comparison with previous work . . . 79
4.4 Outlook . . . 80

5 Conclusions and outlook 84

Bibliography 87

List of Figures 98

List of Tables 99

1 Introduction

1.1 Executive summary

This Chapter provides the background framework for my research. In Sec. 1.2, we review the theoretical foundations of gravitational waves (following closely Ref. [114]) and summarize the key observational milestones that marked the beginning of gravitational-wave (GW) astronomy (Sec. 1.3). In Sec. 1.4.2, we introduce the future Laser Interferometer Space Antenna (LISA) and discuss the properties of massive black-hole (BH) binaries (Sec. 1.4.1), which constitute the main focus of this Thesis. Sec. 1.5 presents the Bayesian data-analysis methods currently employed in GW astronomy, while Sec. 1.5.3 discusses future challenges for LISA. In Sec. 1.6 we outline the main computational approaches for Bayesian inference, with a focus on stochastic sampling methods (Sec. 1.6.1) and simulation-based inference techniques (Sec. 1.6.2). Finally, Sec. 1.7 summarizes the work presented in the remaining Chapters of this Thesis.

1.2 Theoretical foundations of gravitational waves

Within the framework of the theory of General Relativity (GR), formulated by Albert Einstein in 1915 [73, 74, 157], the gravitational interaction is described by a set of field equations that relate the geometry of spacetime, encoded in the metric $g_{\mu\nu}$, to the distribution of matter and energy, described by the stress-energy tensor $T_{\mu\nu}$:

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = \frac{8\pi G}{c^4} T_{\mu\nu}. \tag{1.1}$$

Here, $R_{\mu\nu}$ and R are the Ricci tensor and the Ricci scalar, respectively.

In the weak-field regime, the spacetime metric $g_{\mu\nu}$ can be expressed as a small perturbation $h_{\mu\nu}$ of the flat Minkowski metric $\eta_{\mu\nu}$ as

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \tag{1.2}$$

with $|h_{\mu\nu}| \ll 1$. By expanding the Einstein's equations (Eq. 1.1) to linear order in the metric perturbations $\bar{h}_{\mu\nu}$ and adopting the Lorentz gauge, one obtains

$$\Box \bar{h}_{\mu\nu} = -\frac{16\pi G}{c^4} T_{\mu\nu}, \tag{1.3}$$

where $\Box = \eta^{\mu\nu}\partial_\mu\partial_\nu$ is the d'Alembert operator in flat spacetime, and $\bar{h}_{\mu\nu} \equiv h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}h$ denotes the trace-reversed metric perturbation, with $h = \eta^{\mu\nu}h_{\mu\nu}$. In vacuum ($T_{\mu\nu} = 0$), Eq. 1.3 admits propagating wave solutions for the metric perturbations. These solutions are known as *gravitational waves*.

In the linearised regime, gravitational waves propagate at the speed of light and can be interpreted as ripples in the fabric of spacetime, corresponding to small perturbations of an otherwise flat Minkowski geometry. Eq. 1.3 further shows that gravitational waves are sourced by time-varying distributions of matter and energy through the stress-energy tensor $T_{\mu\nu}$. However, the factor $G/c^4 \sim 8 \times 10^{-50}\,\mathrm{s}^2\,\mathrm{g}^{-1}\,\mathrm{cm}^{-1}$ sets the scale of the emitted wave, indicating that GW effects are intrinsically weak.

Adopting the transverse-traceless gauge in the radiation zone, GWs are characterized by only two independent quadrupolar polarization states, denoted by h_+ and $h_\times$, both transverse to the direction of propagation. To illustrate their physical effect, let us consider a ring of freely falling test masses (TMs) lying in the xy plane and a gravitational wave propagating along the z direction. If the wave is linearly polarized with its principal axes aligned with the coordinate axes, the corresponding polarization state is referred to as the *plus* polarization.

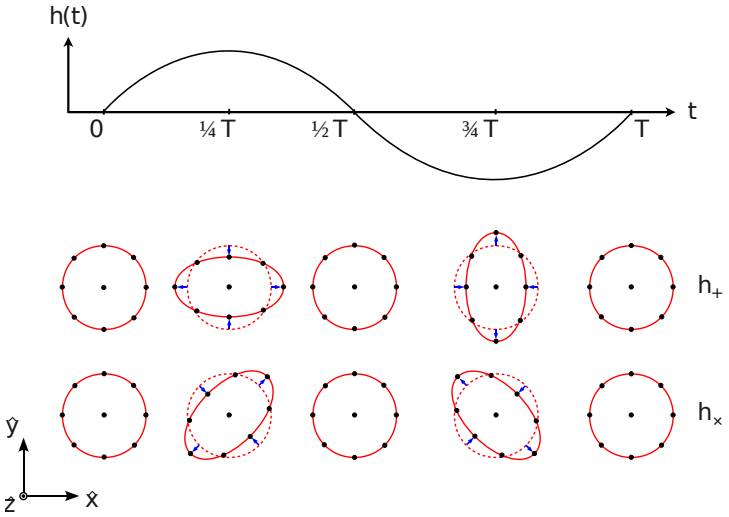

Figure 1.1: Monochromatic gravitational wave of frequency $\omega = 2\pi/T$ (top) propagating along the $\hat{z}$ direction. The characteristic quadrupolar pattern of the gravitational radiation is shown in both the polarizations h_+ and h_x (bottom). Figure adapted from Ref. [165].

If instead the wave is linearly polarized along directions rotated by 45° with respect to the x axis in the xy plane, the corresponding polarization state is referred to as the *cross* polarization. As illustrated in Fig. 1.1, a gravitational wave with plus polarization

h_+ alternately stretches and squeezes proper distances along the x and y axes: during one half-cycle, distances along the x axis increase while those along the y axis decrease, and the opposite occurs during the subsequent half-cycle. For the cross polarization $h_\times$, the same deformation pattern is present but rotated by 45° in the xy plane. Gravitational waves therefore induce oscillatory deformations of objects, quantified by a dimensionless measure known as the *strain*. For the plus polarization, the strain can be expressed as

$$h = \frac{\Delta(L_x - L_y)}{L}, \tag{1.4}$$

where L denotes the proper length.

1.3 From ground to space: gravitational-wave observatories

The direct detection of gravitational waves represents a remarkable experimental challenge, owing to the extremely small amplitudes of the signals produced by astrophysical sources. For instance, stellar-mass compact binaries observed by ground-based interferometers emit gravitational waves in the kilohertz frequency range, with characteristic strain amplitudes of order 10^{-21}. Reaching the sensitivity required to observe such signals demanded decades of advances in interferometry, precision measurement techniques, and noise mitigation strategies.

As a result, nearly a century of theoretical, experimental, and technological developments was required before the first direct detection of gravitational waves could be achieved. In 2015, gravitational waves were finally observed for the first time from the merger of a binary BH system [6], marking the beginning of GW astronomy [2–5, 7]. The field rapidly expanded with the detection, in 2017, of a binary neutron-star merger observed jointly through gravitational waves [1] and electromagnetic radiation [53, 65, 130, 152, 167, 169], inaugurating the era of multimessenger astronomy [8]. In the subsequent years, observations performed by the global network of ground-based interferometers—comprising the two Advanced LIGO detectors in the United States [104, 122], Advanced Virgo in Italy [15, 16], and KAGRA in Japan [24]—have led to the detection of approximately 300 GW events to date, enabling population studies of compact binaries and providing the first observational insights into the mass, spin, and mass-ratio distribution of stellar-mass black holes [9, 12, 13, 162]. Among these detections, particularly remarkable events include exceptionally massive mergers consistent with the existence of intermediate-mass black holes (e.g. GW190521 [10]) and highly asymmetric systems involving objects in the putative lower mass gap (e.g. GW190814 [11]).

More recently, in 2023, pulsar timing array experiments reported compelling evidence

for a stochastic GW background in the nanohertz frequency band [17, 21, 142, 175]. These observations provide a novel probe of the assembly and cosmic evolution of supermassive BH binaries with typical masses of 10^8–$10^9\,M_\odot$. While ground-based interferometers

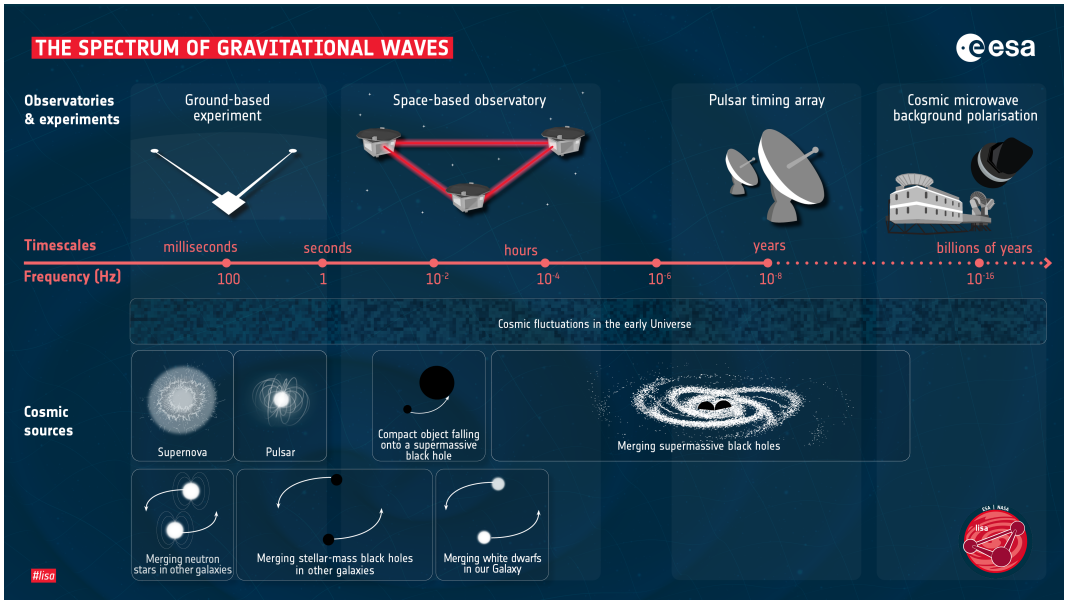

Figure 1.2: GW spectrum and the corresponding observational landscape. *Top*: GW observatories, including ground-based interferometers, space-based detectors, pulsar timing arrays, and cosmological probes. *Middle*: GW frequency bands accessed by each class of observatory. *Bottom*: GW sources populating each frequency band, ranging from stellar-mass compact binaries to massive and supermassive BH binaries. Figure adapted from the European Space Agency.

and pulsar timing arrays currently operate at the high- and low-frequency ends of the gravitational-wave spectrum, respectively (Fig. 1.2), the intermediate millihertz band remains largely unexplored. This frequency range will be accessed by the future LISA mission [49].

1.4 The Laser Interferometer Space Antenna

LISA is a space-based GW mission formally adopted by the European Space Agency in 2024 and currently planned for launch in the mid-2030s. Sec 1.4.1 outlines the scientific case for LISA, with particular emphasis on massive BHs binaries, while Sec. 1.4.2 provides an overview of the mission concept and the principal components of the LISA instrument.

1.4.1 Massive black-hole binaries in the LISA landscape

The millihertz band of the GW spectrum is expected to host a rich variety of astrophysical sources [49]. An overview of these sources is shown in Fig. 1.3, where they are represented

in the GW frequency–amplitude plane in terms of their characteristic strain. This population includes massive BH binaries (multi-coloured tracks in Fig. 1.3), stellar-origin black holes (blue and purple tracks), extreme mass-ratio inspirals (orange tracks), and compact binary stars within our Galaxy (light-blue points). The unresolved component of Galactic compact binaries gives rise to a stochastic foreground that dominates the low-frequency band (solid teal line in Fig. 1.3).

Among these sources, massive BH binaries constitute one of the primary scientific targets of the LISA mission. They will be observed at cosmological distances, with total masses spanning $\sim 10^4$–$10^7\,\mathrm{M}_\odot$ and typical signal-to-noise ratios (SNRs) ranging from approximately ten to several thousand. From a spectral perspective, these systems are transient sources whose signals sweep across the millihertz band. Depending on the system parameters, the inspiral phase can last from hours to several months, ultimately culminating in a merger within the LISA band.

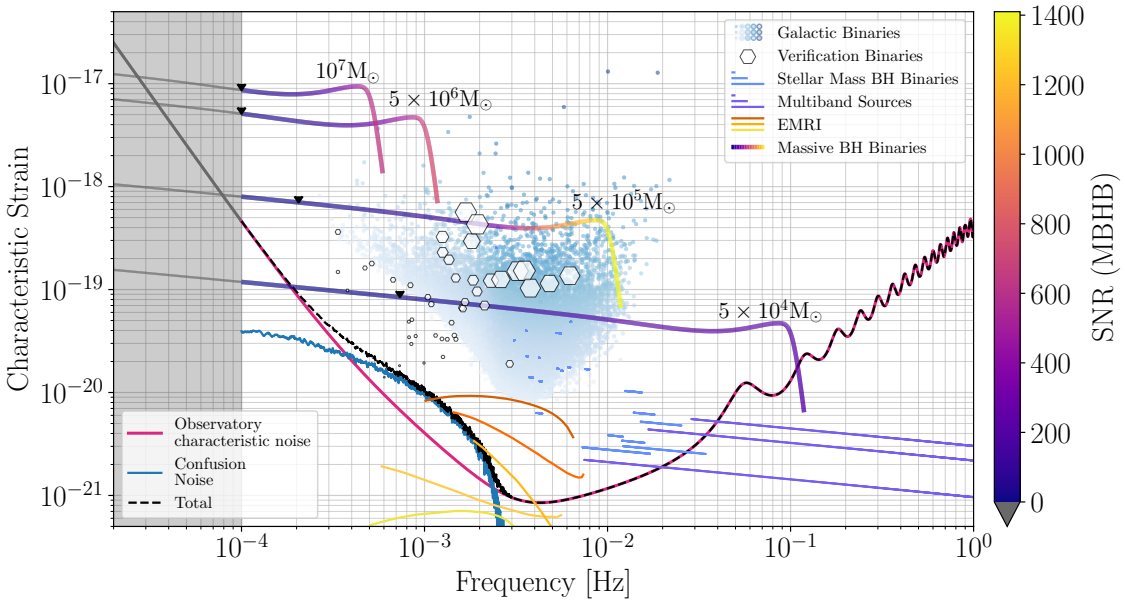

Figure 1.3: Illustration of the variety of LISA sources in the GW frequency–amplitude plane. The figure shows merging massive BH binaries and an extreme mass-ratio inspiral at moderate redshift; stellar-mass black holes, including potential multiband sources, at low redshift; and Galactic binaries, including verification binaries, within the Milky Way. Solid teal, solid blue, and dashed black curves indicate the sensitivity limits due to instrumental noise alone, the unresolved GW foreground, and their combined contribution, respectively. The grey shaded region denotes an extrapolation of the LISA instrumental noise below 0.1 mHz. All quantities are expressed as characteristic strain;the noise curves correspond to the noise amplitude $h_n(f) = \sqrt{f\,S_n(f)}$, where $S_n(f)$ is the power spectral density (PSD). Figure adapted from Ref. [49].

From an astrophysical point of view, massive BH binaries are a natural outcome of hierarchical galaxy formation [171]. Following a galaxy merger [32], the massive

BHs initially sink toward the center of the merger remnant through dynamical friction against the surrounding matter distribution, eventually forming a gravitationally bound binary. At later stages, the astrophysical environment surrounding the binary plays a crucial role in shaping its orbital evolution and promoting the final coalescence [19]. In particular, mergers can only be explained if dissipation mechanisms of astrophysical nature are at play in the early inspiral, before binaries enter their GW-driven regime. Broadly speaking, hardening mechanisms can be divided into *gas hardening* and *stellar hardening* [68]. The former, which dominates in gas-rich environments, is driven by the interaction of the massive BH binary with a gaseous circumbinary disk [56, 113]. Conversely, the latter takes place primarily in gas-poor environments, where the binary orbital separation shrinks due to three-body interactions with individual stars [141].

Galaxy formation and evolution are governed by coupled, highly non-linear phenomena [48], including gravitational dynamics, gas hydrodynamics, and radiative cooling and feedback, which cannot be treated self-consistently within a purely analytical framework. This complexity is further amplified by the enormous dynamical range involved, spanning from galactic scales of kiloparsecs down to milliparsec separations, where GW emission becomes dominant. Modeling these processes therefore remains an active area of research, as different physical assumptions can lead to significantly different predictions for both the properties and merger rates of massive black holes [100, 149]. In this context, numerical simulations and semi-analytical models represent essential tools for investigating the formation, evolution, and final coalescence of massive BH binaries.

By reverse engeneering the problem, LISA observations of coalescing massive BHs can provide a powerful probe of these processes. The physical information encoded in the observed GW signals can in fact be used to constrain the evolutionary pathways that lead to the formation and merger of massive black hole binaries. This constitutes the central scientific motivation of Chapter 2.

1.4.2 Mission concept and instrument design

LISA consists of three identical spacecraft arranged in a triangular configuration, following a heliocentric, Earth-trailing orbit at a distance of approximately 50-65 Mkm from Earth and with a mean inter-spacecraft separation of 2.5 Mkm. The resulting constellation is inclined by 60° with respect to the ecliptic plane and undergoes a cartwheeling motion, completing one rotation per orbit around the Sun. Each spacecraft operates in drag-free mode, with the orbital configuration passively maintaining an approximately equilateral triangular formation with minimal need for orbital corrections [174]. As illustrated in Fig. 1.4, this configuration remains quasi-stable over the nominal mission

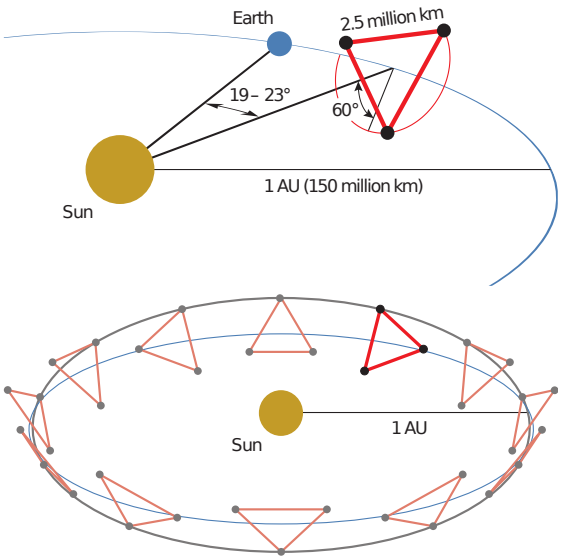

Figure 1.4: Schematic representation of the LISA orbit. *Top*: the 2.5 Mkm triangular constellation trails the Earth by about 20° and is inclined by 60° with respect to the ecliptic plane. *Bottom*: illustration of the cartwheeling motion of the LISA constellation around the Sun. The blue line represents the ecliptic plane, while the gray solid line shows the trajectory of one of the three spacecraft. Figure adapted from Ref. [18].

lifetime.

Each spacecraft hosts two free-falling TMs, realized as 46 mm platinum–gold cubes with a mass of 2 kg. These TMs define the endpoints of the interferometric arms and enable the measurement of variations in the inter-spacecraft optical pathlengths. Direct reflection of laser light between spacecraft is not feasible due to the large arm lengths[1]. Instead, the spacecraft are connected by a continuous interferometric scheme, providing two laser links per arm and thus two independent phase measurements at each spacecraft. The effective LISA observables are constructed in post-processing through linear combinations of appropriately time-delayed single-link measurements. This technique, known as time-delay interferometry (TDI) [166], suppresses the otherwise dominant laser phase noise and yields interferometric observables sensitive to gravitational waves. Over time, TDI has been extended to increasingly sophisticated formulations, including implementations that account for realistic LISA orbits with unequal and time-varying arm lengths. For the purposes of this Thesis, and in the analyses presented in the following Chapters, we restrict our attention to the original TDI combinations, derived under the assumption of equal and constant arm lengths.

[1] For a laser power of 2 W, beam divergence over baselines of order 10^6 km results in only ~ 100 pW of received power [18].

Figure 1.5 illustrates the overall geometric configuration of the LISA detector. The

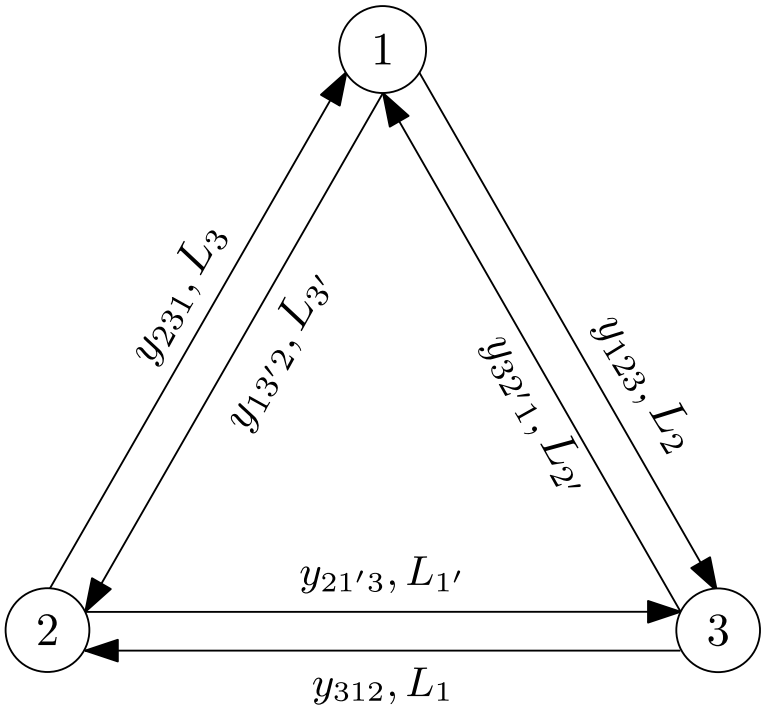

Figure 1.5: Schematic of single laser links conventions. The constellation is made of three satellites (white circles), each connected to the other two by four laser links. A phase measurement is denoted by y_{ijk} or $y_{ij'k}$ according to the following convention: the signal is emitted by the i-th spacecraft and received by the k-th spacecraft, propagating along arm L_j or $L_{j'}$. The indices j and j' denote cyclic and anticyclic permutations of $(1, 2, 3)$, respectively. This figure establishes the notation and geometric conventions used in the following Chapters.

spacecraft, labelled 1, 2, and 3, are ordered counter-clockwise, and their separations are denoted by L_j, with $j \in \mathbb{Z}_3$ indicating the spacecraft opposite to the corresponding arm. Six independent phase measurements, one for each laser link, enter any TDI combination and are denoted by y_{ijk}, where $i, k \in \mathbb{Z}_3$ label the emitting and receiving spacecraft, respectively. Laser beams propagating clockwise are associated with even permutations of (ijk), while counter-clockwise propagation corresponds to odd permutations; in the latter case, the phase measurement is denoted by $y_{ij'k}$ to emphasize the opposite direction along the arm.

As an explicit example, y_{231} denotes the phase measurement of a laser beam emitted by spacecraft 2 and received by spacecraft 1, propagating along arm L_3. Similarly, $y_{13'2}$ represents the phase measurement of the beam travelling in the opposite direction along arm $L_{3'}$, i.e. from spacecraft 1 to spacecraft 2. We also introduce a convenient notation for time-delayed measurements,

$$\mathcal{D}[f(t)] \equiv f(t - L), \tag{1.5}$$

where $\mathcal{D}$ denotes the time-delay operator and units with $c = 1$ are assumed.

We now illustrate the construction of the original TDI Michelson-like combination centered on spacecraft 1, commonly referred to as the M_X channel:

$$M_X = y_{231} + \mathcal{D}y_{13'2} - y_{32'1} - \mathcal{D}y_{123}. \tag{1.6}$$

The sum $y_{231} + \mathcal{D}y_{13'2}$ defines a two-way interferometric variable associated with the arm pair $L_{3'} - L_3$, while $y_{32'1} + \mathcal{D}y_{123}$ defines the corresponding variable associated with $L_2 - L_{2'}$. The M_X combination synthetically interferes these two paths at spacecraft 1, thereby cancelling laser phase noise to first order. The remaining TDI channels M_Y and M_Z are obtained by cyclic permutations of the indices and are centred on spacecraft 2 and 3, respectively:

$$M_Y = y_{312} + \mathcal{D}y_{21'3} - y_{13'2} - \mathcal{D}y_{231} \tag{1.7}$$

and

$$M_Z = y_{123} + \mathcal{D}y_{32'1} - y_{21'3} - \mathcal{D}y_{312}. \tag{1.8}$$

While the TDI observables are designed to optimally capture the GW signal, non-stationary transient noise sources can also couple to the detector and contaminate the measurements. Such disturbances are commonly referred to as *glitches*. During the LISA Pathfinder (LPF) mission [22], launched by the European Space Agency in 2015 to validate key technologies for space-based GW detection, glitches were observed at a rate of approximately one per day. Given the increased complexity of both the spacecraft and payload design in LISA compared to LPF, similar transient disturbances are expected to occur. This, in turn, motivates dedicated studies aimed at assessing their impact on the scientific performance of the LISA mission, which constitutes the main focus of Chapter 3.

1.5 Gravitational-wave data analysis

Careful data analysis is crucial to fully exploit the scientific content of GW observations. In Secs. 1.5.1 and 1.5.2, we review the basic concepts of GW data analysis as implemented in current pipelines, while Sec. 1.5.3 introduces the data-analysis framework specific to LISA.

1.5.1 Signal extraction

Under the assumption of well-calibrated instruments and stationary noise, the output of a GW detector can be modeled as

$$d(t) = h(t;\theta) + n(t), \tag{1.9}$$

where $h(t;\theta)$ denotes the GW signal, which depends on the source parameters θ, and $n(t)$ represents the instrumental noise, characterized by its PSD $S_n(f)$.

Despite extensive noise-mitigation efforts, GW signals are typically buried in detector noise and cannot be identified by direct inspection of the data. Their detection therefore relies on dedicated statistical methods. In current searches, signal extraction is commonly performed using matched filtering, which consists of correlating the data with a bank of template waveforms spanning the expected parameter space.

For each template, a matched-filter signal-to-noise ratio (SNR) time series is computed. The detection statistic is defined as the maximum SNR over the template bank,

$$\rho = \max_{\theta} \frac{\langle d(t) \mid h(t;\theta)\rangle}{\sqrt{\langle h(t;\theta) \mid h(t;\theta)\rangle}}, \tag{1.10}$$

which is compared to a predefined threshold. Events exceeding this threshold are identified as candidate GW signals and subsequently subjected to further validation and parameter-estimation analyses.

The inner product $\langle \cdot \mid \cdot \rangle$ is defined as the noise-weighted cross-correlation between two time series $a(t)$ and $b(t)$,

$$\langle a \mid b\rangle = 4\,\mathrm{Re}\int_0^\infty \frac{\tilde{a}^*(f)\,\tilde{b}(f)}{S_n(f)}\,\mathrm{d}f, \tag{1.11}$$

where $\tilde{a}(f)$ and $\tilde{b}(f)$ denote the Fourier transforms of $a(t)$ and $b(t)$, respectively, and $S_n(f)$ is the one-sided noise PSD.

1.5.2 Bayesian statistical inference

Candidate signals identified in the search stage are subsequently characterized through dedicated parameter-estimation pipelines. In GW astronomy, this task is commonly performed within a Bayesian framework, which provides a principled statistical approach to infer the physical properties of astrophysical sources from detector data.

Given the observed data d, Bayesian inference aims to compute the posterior proba-

bility distribution of the source parameters θ,

$$p(\theta|d) = \frac{\mathcal{L}(d|\theta)\, p(\theta)}{\mathcal{Z}}, \tag{1.12}$$

where the likelihood $\mathcal{L}(d|\theta)$ describes the data-generating process, $p(\theta)$ denotes the prior probability distribution encoding information available before analyzing the data, and $\mathcal{Z}$ is the Bayesian evidence, which ensures proper normalization of the posterior.

The likelihood function encodes the statistical properties of the detector noise. Assuming Gaussian and stationary noise, the probability density associated with a specific noise realization $n_0(t)$ can be written as

$$p(n_0) = \mathcal{N} \exp\left[-\frac{1}{2} 4 \int_0^\infty \frac{|\tilde{n}_0(f)|^2}{S_n(f)}\, \mathrm{d}f\right], \tag{1.13}$$

where $\mathcal{N}$ is a normalization constant that depends on the noise PSD. Using the noise-weighted inner product defined in Eq. 1.11, this expression can be equivalently written as

$$p(n_0) = \mathcal{N}\ \exp\left[-\frac{1}{2}(n_0 \mid n_0)\right]. \tag{1.14}$$

If a GW signal described by a template $h(t;\theta)$ is present in the data, the detector output can be written as

$$d(t) = h(t;\theta_\mathrm{t}) + n_0(t), \tag{1.15}$$

where θ_t denotes the (unknown) true source parameters. The likelihood function for the observed data $d(t)$ is then obtained by substituting $n_0 = d - h(\theta)$ into Eq. 1.14, yielding

$$\mathcal{L}(d \mid \theta) = \mathcal{N} \exp\left[-\frac{1}{2}\big(d - h(\theta) \,\big|\, d - h(\theta)\big)\right]. \tag{1.16}$$

In other words, the probability $p(d|\theta)$ of measuring the data $d(t)$ given a signal $h(t;\theta)$ is equal to the probability $p(d' \mid 0)$ of observing the residual $d' = d - h(t;\theta)$ under the hypothesis that no signal is present.

In the specific case of LISA, as discussed in Sec. 1.4.2, the detector output consists of a vector of data streams—either in the time or frequency domain—corresponding to the three TDI channels. Bayesian inference is therefore performed through a coherent analysis of the full set of LISA TDI observables. In practice, this analysis is typically carried out using the three noise-orthogonal TDI channels (A, E, T) [166]. These channels are obtained by diagonalizing the noise correlation matrix of the TDI combinations and form a basis in which the instrumental noise is approximately uncorrelated. This representation is computationally advantageous, as it allows the likelihood to be evaluated by summing

the contributions from the individual channels while neglecting cross-correlation terms. In this basis, the log-likelihood of the LISA data d given the source parameters θ can be written as

$$\log \mathcal{L}(d \mid \theta) = -\frac{1}{2} \sum_{k \in \{A,E,T\}} \Big(d_k - h_k(\theta) \,\Big|\, d_k - h_k(\theta) \Big)_k + \text{const}, \tag{1.17}$$

where the subscript $k = A, E, T$.

As already introduced in Eq. 1.12, the quantity $\mathcal{Z}$ denotes the Bayesian evidence, a normalization constant that is independent of θ and ensures that the posterior distribution is properly normalized. It is defined as

$$\mathcal{Z} = \int \mathcal{L}(d|\theta)\, \pi(\theta)\, \mathrm{d}\theta. \tag{1.18}$$

While the evidence plays no role in parameter estimation, it becomes a central quantity in Bayesian model comparison. Model selection addresses the question of which competing model is statistically preferred by the data, and by how much. The relative support for one model M_1 over another M_2 is quantified by the odds ratio,

$$\frac{P(M_1|d)}{P(M_2|d)} = BF_{12}\, \frac{\pi(M_1)}{\pi(M_2)}, \tag{1.19}$$

which is given by the Bayes factor BF_{12} multiplied by the ratio of the model priors. The Bayes factor is defined as

$$BF_{12} = \frac{\mathcal{Z}_{M_1}}{\mathcal{Z}_{M_2}}, \tag{1.20}$$

where $\mathcal{Z}_{M_1}$ and $\mathcal{Z}_{M_2}$ denote the evidences associated with the two competing models.

1.5.3 Future challenges: the LISA case

As outlined in Sec. 1.4.1, LISA will observe a wide range of astrophysical sources (Fig. 1.3). A large fraction of these signals will be simultaneously present in the data, giving rise to a signal-dominated regime over the whole LISA lifetime. This observational context poses challenges that are not addressed by data-analysis pipelines developed for current ground-based detectors, and calls for a framework capable of jointly inferring the tens of thousands of resolvable sources overlapping in the data stream, together with the instrumental noise.

The development of such pipelines represents a major and ongoing effort within the LISA community. Early studies addressing the LISA global-fit problem explored trans-dimensional Bayesian inference techniques, including reversible-jump MCMC [84],

designed to simultaneously infer the number of signals and their parameters [51, 55, 108, 109]. Building on these developments, more recent efforts have focused on scalable and modular analysis pipelines tailored to the complexity of the LISA data stream. In these frameworks, the inference is organized through a so-called global-fit "wheel", in which dedicated modules operate in parallel and iteratively exchange information within a cyclic scheme, each targeting a specific source class or instrumental component. Several prototype implementations of this strategy have already been described in the literature and made publicly available [63, 98, 110, 159]. While these pipelines differ in their computational strategies, workflow architectures, and resource requirements, they share a common conceptual structure, as shown in Fig. 1.6.

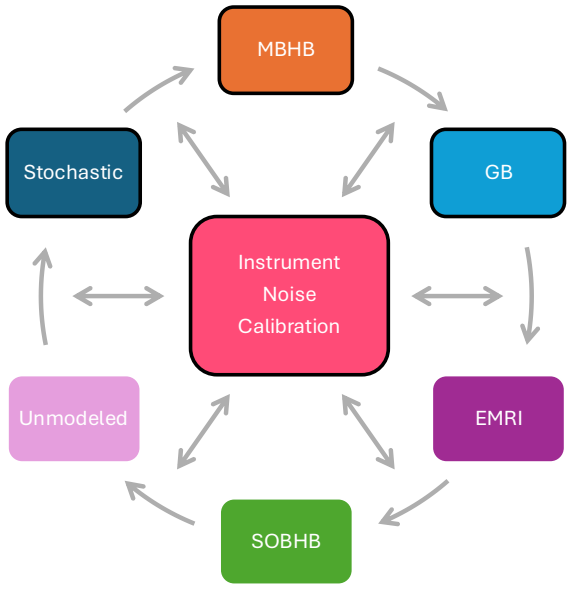

Figure 1.6: Schematic overview of a global-fit "wheel" for LISA data analysis. Dedicated modules targeting different GW sources are coupled through an iterative clockwise inference cycle. Both resolved and unresolved components are inferred jointly within a unified global inference scheme, while instrumental noise and calibration information are updated throughout the process. Figure adapted from [98].

In practice, the global-fit procedure is implemented as an iterative process. Before identifying individual astrophysical signals, an initial estimate of the instrumental noise and unresolved foreground is obtained to initialize the global analysis. As new data are transmitted to the ground, the analysis proceeds by identifying and modeling resolvable GW sources that can be reliably detected given the current characterization of the noise and foreground, while accounting for possible instrumental non-stationarities. Once incorporated into the global model, these sources are removed from the data, yielding residuals that are subsequently analyzed to identify additional classes of signals, while

the noise and foreground models are updated. This procedure is repeated sequentially across different source populations, with parameter estimates progressively refined as information from other data components is incorporated. At a higher level, the whole process is formulated as an iterative looping structure, in which successive global-fit cycles lead to an increasingly accurate characterization of both the astrophysical signals and the instrumental noise.

In the global-fit context, Chapter 3 presents an approach aimed at the simultaneous characterization of GW signals and noise transients, and assesses its impact on inference in the LISA observational regime.

1.6 Computational approaches for Bayesian inference

The connection between theoretical models and empirical observations can be approached in two complementary ways: forward modeling, which predicts data from a given model, and inverse modeling, which infers model parameters from observations. While the forward problem is typically well defined, the inverse problem is inherently more challenging.

In the following, we outline the main computational strategies commonly adopted to tackle this problem, namely standard stochastic sampling methods (Sec. 1.6.1) and simulation-based inference techniques (Sec. 1.6.2).

1.6.1 Stochastic sampling

To illustrate the point highlighted above, let us consider the computation of the posterior distribution for the ~ 15 parameters describing a quasi-circular, precessing massive BH binary signal. A naive exploration of the parameter space could be attempted by constructing a grid according to the prior distribution, with, for instance, ten bins along each parameter dimension, and evaluating the likelihood at every grid point. Even at this extremely coarse resolution, such an approach would require $\sim 10^{15}$ likelihood evaluations. Since waveform generation typically takes of the order of milliseconds per template, the total computational cost of an analysis with only ten points per parameter would amount to $10^{15} \times 10^{-3}\,\mathrm{s} \sim 10^{4}\,\mathrm{yr}$, rendering this strategy entirely impractical. The situation worsens rapidly as the dimensionality of the parameter space increases or a finer resolution is required.

An effective strategy to mitigate the curse of dimensionality relies on stochastic sampling algorithms, which extract samples from the posterior distribution through a stochastic exploration of parameter space guided by the likelihood. Common sam-

pling approaches can be broadly divided into two main classes: Markov Chain Monte Carlo (MCMC) and nested sampling. MCMC methods sample the posterior distribution by constructing a Markov chain whose stationary distribution coincides with the target posterior [87, 123]. Despite their robustness and broad applicability, conventional MCMC algorithms can struggle to efficiently explore complex posterior landscapes, particularly in the presence of strong parameter correlations or multiple separated modes. In addition, they do not directly provide estimates of the Bayesian evidence, which must be computed through additional techniques. To address these limitations, several extensions and complementary strategies have been developed to improve sampling efficiency. Among the most widely used approaches, Hamiltonian Monte Carlo exploits gradient information of the target distribution to guide the exploration of parameter space, allowing the sampler to propose distant moves with high acceptance probability and significantly improving performance in high-dimensional settings [128]. Sequential Monte Carlo methods, instead, evolve a population of particles through a sequence of intermediate distributions, progressively approximating the target posterior and providing a flexible framework capable of capturing complex or multi-modal structures [70]. Parallel tempering offer another strategy to address multi-modality by running multiple chains at different temperatures and periodically exchanging their states, thereby facilitating transitions between separated modes [81, 161, 170].

A further challenge arises when the dimensionality of the parameter space is itself unknown. This situation naturally occurs in signal-dominated analyses such as the LISA global-fit problem (Sec. 1.5.3), where the number of signals present in the data must be inferred simultaneously with their parameters. In this setting, trans-dimensional MCMC methods provide a natural extension of standard MCMC techniques. In particular, reversible-jump MCMC [84] enables the sampler to move between models with different dimensionalities, allowing simultaneous parameter estimation and model selection over the number of signals.

While the methods discussed above extend the MCMC framework to tackle increasingly complex inference problems, a conceptually different approach is provided by nested sampling. Unlike standard MCMC methods, nested sampling algorithms are specifically designed to compute the Bayesian evidence as their primary objective [151], with posterior samples obtained as a by-product of the evidence estimation. Given their central role in the following chapters, the basic principles underlying nested sampling algorithms are briefly summarized below.

Computing the Bayesian evidence (Eq. 1.18) generally involves evaluating an integral over a high-dimensional parameter space, which is a computationally demanding task. The key advantage of nested sampling is that it recasts this multidimensional integral

into a one-dimensional integral that is significantly easier to evaluate. The algorithm begins by drawing a set of N_{live} live points from the prior distribution $p(\theta)$, initially spanning the full parameter space. At each iteration i, the live point with the lowest likelihood value $\mathcal{L}_i$ is removed from the ensemble and recorded as a dead point. This point is then replaced by a new live point sampled from the prior under the constraint $\mathcal{L}(\theta) > \mathcal{L}_i$. The fraction of prior mass enclosed within an iso-likelihood contour $\mathcal{L}(\theta) = \mathcal{L}$ defines the prior volume,

$$X(\mathcal{L}) = \int_{\mathcal{L}(\theta)>\mathcal{L}} \pi(\theta)\,\mathrm{d}\theta, \tag{1.21}$$

which decreases monotonically as the likelihood threshold is increased. As the nested

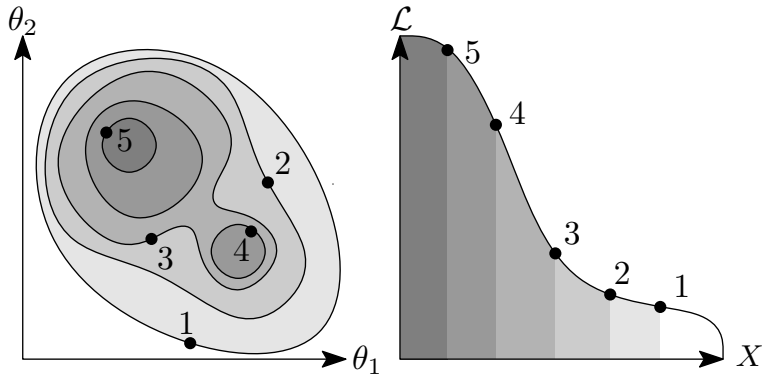

Figure 1.7: The nested sampling volume transformation. *Left*: two-dimensional multimodal posterior distribution with five iso-likelihood contours. Each contour encloses some fraction of the prior volume X, indicated by color. *Right*: the likelihood $\mathcal{L}$ as a function of the volume X enclosed by contour. The evidence is the area under the curve. Plots adapted from Ref. [86].

sampling run progresses, the prior volume is progressively compressed and the ensemble of live points migrates toward regions of higher likelihood, eventually concentrating around the dominant posterior mode(s). The algorithm thus generates an ordered sequence of dead points, each associated with a parameter set θ_i, a likelihood value $\mathcal{L}_i$, and an estimate of the corresponding prior volume X_i. Using the prior-volume parametrization introduced above, the Bayesian evidence can be rewritten as a one-dimensional integral over the unit interval,

$$\mathcal{Z} = \int_0^1 \mathcal{L}(X)\,\mathrm{d}X, \tag{1.22}$$

where $\mathcal{L}(X)$ denotes the likelihood expressed as a function of the enclosed prior volume X. By construction, $\mathcal{L}(X)$ is a positive and monotonically decreasing function, as illustrated in Fig. 1.7.

In practice, the evidence is estimated numerically using the ordered sequence of N_{dead}

dead points produced by the nested sampling algorithm. Approximating the integral by numerical quadrature yields

$$\mathcal{Z} \simeq \sum_{i=1}^{N_{\text{dead}}} \alpha_i, \tag{1.23}$$

where α_i denotes the quadrature weight associated with the i-th dead point. Using a rectangular or trapezoidal integration scheme, the quadrature weights can be written as $\alpha_i = \mathcal{L}_i (X_{i-1} - X_i)$ or $\alpha_i = \frac{1}{2} (\mathcal{L}_{i-1} + \mathcal{L}_i)(X_{i-1} - X_i)$, respectively. The step in prior volume, $X_{i-1} - X_i$, is determined by the number of live points N_{live}, with the typical scaling $X_{i-1} - X_i \sim X_{i-1}/N_{\text{live}}$ [151]. Posterior samples can be obtained as a by-product of the nested sampling run by assigning each dead point an importance weight proportional to its contribution to the evidence. The normalized posterior weight of the i-th sample is then given by

$$p_i = \frac{\alpha_i}{\mathcal{Z}}. \tag{1.24}$$

1.6.2 Simulation-based inference

Compared to stochastic sampling techniques, simulation-based inference (SBI) adopts a fundamentally different strategy for performing Bayesian inference. Rather than relying on explicit evaluations of the likelihood function, SBI methods require only the ability to generate joint samples $(d, \theta) \sim p(d, \theta) = p(d|\theta)p(\theta)$ by drawing parameters from the prior and evaluating a forward simulator that emulates the data-generating process. This approach makes SBI particularly well suited to regimes where traditional likelihood-based methods become impractical or infeasible, including computationally prohibitive inference problems and cases involving analytically or computationally intractable likelihoods. Recent advances in deep learning and generative modeling have made SBI increasingly practical [54], allowing complex posterior distributions to be learned efficiently from simulations. This flexibility, however, comes at the cost of substantial training and validation efforts, as well as the need for careful control of simulation fidelity and prior coverage. SBI methods can be broadly classified according to the Bayesian quantity they aim to approximate, namely neural posterior estimation (NPE) [133], neural likelihood estimation [134], and neural ratio estimation [89]:

- Neural posterior estimation
  The posterior distribution is learned directly from simulated data using neural density estimation techniques.

- Neural likelihood estimation
  The likelihood function associated with the data-generating process is approxi-

mated using neural density estimators.

- Neural ratio estimation
  Likelihood ratios, such as the likelihood-to-evidence or posterior-to-prior ratio, are learned using classification-based models.

NPE constitutes the primary focus of Chapter 4, therefore we briefly outline its methodological framework below.

The core idea of NPE is to train a conditional density estimator $q_\phi(\theta|d)$, parameterized by a neural network with weights ϕ, to approximate the Bayesian posterior distribution of the parameters θ given the observed data d,

$$q_\phi(\theta|d) \approx p(\theta|d). \tag{1.25}$$

The training objective is therefore to minimize a suitable measure of discrepancy between the learned distribution $q_\phi(\theta|d)$ and the true posterior $p(\theta|d)$, averaged over the joint distribution of simulated data and parameters. A widely used measure of dissimilarity between two probability distributions $p(x)$ and $q(x)$ is the Kullback-Leibler (KL) divergence [102], defined as

$$D_{\mathrm{KL}}(p(x) \,\|\, q(x)) = \int \mathrm{d}x \, p(x) \log\left(\frac{p(x)}{q(x)}\right). \tag{1.26}$$

The KL divergence is asymmetric: $D_{\mathrm{KL}}(p \,\|\, q)$ and $D_{\mathrm{KL}}(q \,\|\, p)$ are commonly referred to

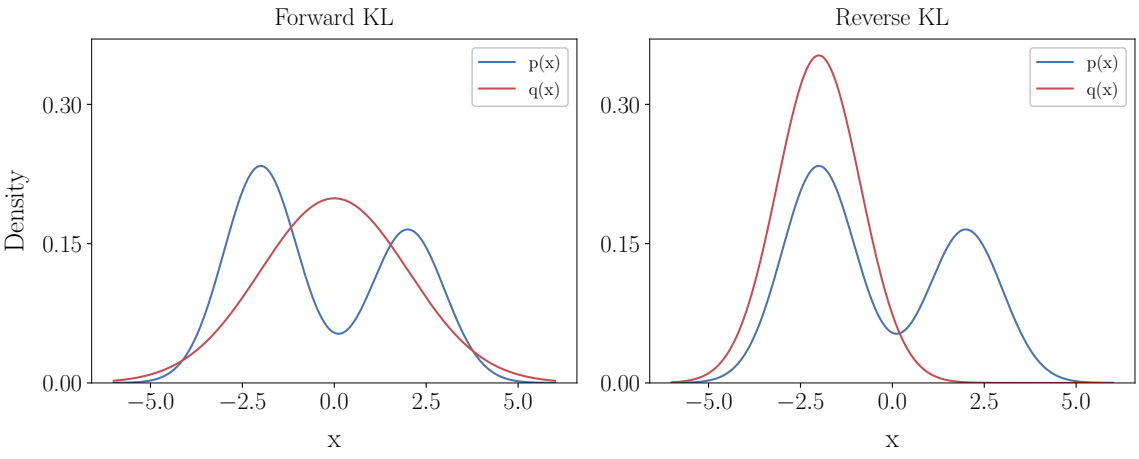

Figure 1.8: Illustration of forward (left panel) and reverse (right panel) KL divergence for a bimodal target distribution $p(x)$ (blue) approximated by a unimodal Gaussian distribution $q(x)$ (red).

as the forward and reverse KL divergences, respectively, and their minimization leads

to different behaviors. To illustrate this difference, consider a target distribution with two distinct modes, approximated by a unimodal Gaussian distribution $q(x)$ (Fig. 1.8). Minimizing the forward KL divergence penalizes regions where $p(x)$ has high probability density but $q(x)$ assigns little or no mass. As a result, the optimization encourages $q(x)$ to cover the full support of $p(x)$, even at the expense of assigning probability mass to regions where $p(x) \approx 0$. This behavior is commonly referred to as *mode covering*. In contrast, minimizing the reverse KL divergence penalizes assigning high probability density to regions where the target distribution has low probability, while being comparatively insensitive to regions where $q(x) \approx 0$ but $p(x) > 0$. This leads to a *mode-seeking* behavior, in which the approximation concentrates on fitting a single mode of $p(x)$ with high accuracy, potentially ignoring other modes entirely. The choice between forward and reverse KL divergence therefore depends on the desired properties of the approximation. In the context of posterior inference, where capturing all plausible regions of parameter space is essential, a mode-covering behavior is generally preferred; this is the choice adopted throughout Chapter 4.

Assuming the forward KL divergence, Eq. 1.26 can be written as

$$D_{\mathrm{KL}}(p \parallel q_\phi) = \int d\theta \, p(\theta|d) \, \log\left(\frac{p(\theta|d)}{q_\phi(\theta|d)}\right) = \mathbb{E}_{\theta\sim p(\theta|d)}\left[\log\frac{p(\theta|d)}{q_\phi(\theta|d)}\right] \tag{1.27}$$

Since the expectation value is taken with respect to the true posterior distribution, this quantity cannot be evaluated directly in practice. However, in SBI one has access to samples drawn from the joint generative distribution$p(d, \theta) = p(d|\theta)\, p(\theta)$, which enables an alternative formulation. By taking the expectation of the forward KL divergence over simulated data realizations $d \sim p(d)$, the expected forward KL divergence becomes

$$\mathbb{E}_{d\sim p(d)}\left[D_{KL}(p \parallel q_\phi)\right] = \mathbb{E}_{\theta,d\sim p(d,\theta)}\left[\log\frac{p(\theta|d)}{q_\phi(\theta|d)}\right] \tag{1.28}$$

$$= -\mathbb{E}_{d,\theta\sim p(d,\theta)} \log q_\phi(\theta|d) + \mathbb{E}_{d,\theta\sim p(d,\theta)} \log p(\theta|d). \tag{1.29}$$

The second term does not depend on the network parameters ϕ and therefore does not contribute to the optimization. The first term corresponds to the expected log-probability assigned by the approximate posterior $q_\phi(\theta|d)$ to (d, θ) pairs drawn from the true generative model, and quantifies how well the learned distribution captures the true posterior structure. As a result, training an NPE model amounts to minimizing the following loss function,

$$\mathcal{L}_{\mathrm{NPE}}(\phi) = -\mathbb{E}_{d,\theta\sim p(d,\theta)} \log q_\phi(\theta|d), \tag{1.30}$$

which corresponds to maximizing the conditional log-likelihood of the parameters under the learned posterior model, averaged over simulated data.

The density estimator $q_\phi(\theta|d)$ plays a central role in NPE, as it defines the approximate posterior distribution used for inference. In this context, normalizing flows have emerged as a standard choice in NPE owing to their high expressive power and tractable density evaluation. They construct flexible probability densities by mapping a simple base distribution, typically a standard normal, to a more complex target distribution through a sequence of invertible and differentiable transformations.

Given a base variable $z_0 \sim p(z_0)$ and a sequence of K invertible transformations f_k, with $k = 1, \dots, K$, we define intermediate variables recursively as $z_k = f_k(z_{k-1})$. Samples from the approximate posterior are then obtained by applying the chain of transformations

$$\theta = z_K = f_K \circ \cdots \circ f_1(z_0). \tag{1.31}$$

The corresponding model density is obtained by applying the change-of-variables formula to the inverse flow,

$$\log q_\phi(\theta|d) = \log p(z_0) - \sum_{k=1}^{K} \log \left| \det J_{f_k}(z_{k-1}) \right|, \tag{1.32}$$

where $J_{f_k}(z_{k-1})$ denotes the Jacobian matrix of the transformation f_k evaluated at z_{k-1}.

When parameterized by neural networks, normalizing flows provide flexible and tractable conditional density estimators that can be efficiently trained by minimizing the NPE loss in Eq. 1.30, making them a natural choice for complex high-dimensional posterior inference.

1.7 Research summary

The work presented in this Thesis is devoted to the study of massive BH systems as anticipated GW sources for the future space-based mission LISA. Overall, the Thesis investigates this topic at different levels, from the statistical analysis of astrophysical populations to the characterization of individual sources, as well as addressing some of the key data-analysis challenges relevant for the scientific exploitation of the mission. The main contributions presented in this Thesis can be summarised as follows.

- In Chapter 2, we examine how different hardening processes imprint distinctive signatures on the population of massive BH binaries, and we assess their statistical distinguishability using simulated observations from LISA.

  The main body of this Chapter is based on Ref. [154]. I led this study in collaboration with Riccardo Buscicchio, David Izquierdo-Villalba, Davide Gerosa, Antoine Klein, and Geraint Pratten. The astrophysical catalogs employed in the analysis were provided by David Izquierdo-Villalba. I constructed the mock LISA catalogs and led the parameter-estimation campaign on the TheCatalog sample used for the statistical inference, with contributions from Riccardo Buscicchio. Antoine Klein and Geraint Pratten made major contributions to the modules of the Balrog code used in this campaign. The numerical implementation of the statistical analysis was developed by myself under the guidance of Riccardo Buscicchio and Davide Gerosa. Data supporting this paper, including all mock LISA catalogs and posterior samples, are publicly released at Ref. [153].

- In Chapter 3, we present, for the first time in the literature, a systematic study of the impact of glitches on the inference of individual massive BH binary signals.

  The main body of this Chapter is based on Ref. [155]. I led this study in collaboration with Riccardo Buscicchio, Daniele Vetrugno, Antoine Klein, Davide Gerosa, Stefano Vitale, Rita Dolesi, William Joseph Weber, and Monica Colpi. Daniele Vetrugno, Stefano Vitale, Rita Dolesi, and William Joseph Weber highlighted the advantages of performing the analysis using acceleration-based TDI observables. The analytical derivation of the LISA response to acceleration and displacement artifacts was carried out by myself under the guidance of Riccardo Buscicchio. At the implementation level, this work builds on the Balrog framework, within which I contributed to the development the glitch-analysis module under the guidance of Riccardo Buscicchio and with valuable inputs from Antoine Klein.. I performed all the parameter-estimation runs presented in this work.

- In Chapter 4, we propose a simulation-based inference framework for Bayesian parameter estimation of individual massive BH binary signals. This Chapter constitutes the main body of a publication currently in preparation. I led this study in collaboration with Jonathan Gair, Davide Gerosa, Stephen Green, Riccardo Buscicchio, Nihar Gupte, Rodrigo Tenorio, Samuel Clyne, Michael Püerrer, and Natalia Korsakova. This work builds on the Dingo software, whose extension to the LISA frequency band was developed by myself as a substantial component of my PhD research. Jonathan Gair, Davide Gerosa, and Stephen Green provided guidance and valuable suggestions throughout the project. The LISA version of the Dingo code is under active development, and progress can be tracked at a dedicated branch of the main Dingo repository: github.com/AliSword/dingo-lisa. I carried out the neural-network training, as well as all validation and benchmarking analyses. Riccardo Buscicchio contributed to the development of the statistical framework used to benchmark the algorithm against standard inference methods and provided valuable input throughout the project, while Nihar Gupte provided technical input on the code development. Rodrigo Tenorio provided valuable help during the final stages of the project.

Beyond my core research activities, my PhD also included supervision, dissemination, research visits, tutoring, coursework, and outreach. In particular, I supervised a Bachelor's student in the context of the work presented in Chapter 3. Within this project, we estimated the probability of temporal overlap between massive BH binary signals and glitches. The analysis considered three astrophysical population scenarios spanning different expected detection rates over the nominal four-year LISA mission.

Throughout my doctoral studies, I presented my work at international conferences within the LISA community, as well as at broader meetings of the GW community. As part of my research activities, I completed a six-month research visit at the Max Planck Institute for Gravitational Physics in Potsdam, where I initiated the research presented in Chapter 4 and gave departmental seminars in the "LISA meeting" series on the work described in Chapter 3 and Chapter 4. I was also invited to present the results of Chapter 3 in a departmental seminar at the Institute for Gravitational Wave Astronomy in Birmingham, with which I collaborated during my PhD. In parallel with my research activities, I was actively involved throughout my PhD in outreach initiatives related to astrophysics, GW astronomy, and cosmic-ray science.

2 Stars or gas? Constraining the hardening processes of massive black-hole binaries with LISA

Overview

Massive BH binaries will be the loudest sources detectable by LISA. These systems are predicted to form during the hierarchical assembly of cosmic structures and coalesce by interacting with the surrounding environment. The hardening phase of their orbit is driven by either stars or gas and encodes distinctive features into the binary black holes that can potentially be reconstructed with gravitational-wave observations. We present a Bayesian framework to assess the likelihood of massive mergers being hardened by either gaseous or stellar interactions. We use state-of-the-art astrophysical models tracking the cosmological evolution of massive BH binaries and construct a large number of simulated catalogs of sources detectable by LISA. From these, we select a representative catalog and run both parameter estimation assuming a realistic LISA response as well as model comparison capturing selection effects. Our results suggest that, at least within the context of the adopted models, future LISA observations can confidently constrain whether stars or gas are responsible for the binary hardening. We stress that accurate astrophysical modeling of the BH spins and the inclusion of subdominant emission modes in the adopted signal might be crucial to avoid systematic biases.

2.1 Executive summary

This Chapter is organized as follows. In Sec. 2.2, we illustrate the adopted astrophysical models. In Sec. 2.3, we present the construction of the LISA catalogs. In Sec. 2.4, we describe the parameter-estimation and model-selection strategies adopted to perform the analysis. Finally, Sec. 2.5, presents and discusses the results. Throughout this work, we use units where $c = 1$. The material presented in this Chapter is based on Ref [154].

2.2 Astrophysical models

2.2.1 Galaxy evolution model

We use the L-Galaxies semi-analytical model as presented in Refs. [88, 93, 156]. In particular, the model is built on top of merger trees from the Millenium-II [41] cosmological simulation, which follows the evolution of 2160^3 dark matter particles with a mass of $6.885 \times 10^6 \mathrm{M}_\odot/h$ within a comoving box of side 100 Mpc/h, later rescaled [20] to match the Planck cosmology [138] (here $h = 10^{-2} H_0$ Mpc s/km and H_0 is the Hubble constant).

L-Galaxies associates a fraction of baryonic matter to each newly resolved DM halo in the form of a diffuse, spherical, and quasi-static hot gas atmosphere. As the gas cools down, it settles into a disk [160]; this facilitates episodes of star formation, which in turn results in the assembly of a stellar disk component. Galaxies are also allowed to assemble an overdensity of stars in their central regions (i.e. their bulges). This is triggered by either internal processes, namely non-axisymmetric instabilities that redistribute the stellar matter, or galaxy mergers.

2.2.2 Black-hole growth

For each newly resolved DM halo, L-Galaxies assess whether physical conditions are favorable for the formation of a massive BH seed. The model incorporates for different processes, including direct collapse of pristine gas clouds, runaway stellar mergers, and PopIII star formation [156]. As a result, the initial BH mass function arises as a combination of different formation channels, with masses ranging from $\sim 10^2$ to $\sim 10^5$ $\mathrm{M}_\odot$ [156]. The initial dimensionless spin parameter χ is set uniformly in $[0, 0.998]$ [163].

The evolution of masses and spins is influenced by the accretion of surrounding gas into the BHs as well as BH-binary coalescences. In particular, gas accretion is triggered by both galaxy mergers and disk instabilities [28, 34, 76], and predominantly drives the mass growth [71]. The cold gas available for accretion settles around the BH in a reservoir

of mass $M_{\rm res}$, proportional to the cold gas mass of the galaxy [94]. The gas reservoir is then progressively consumed through a sequence of transient accretion disks [92, 93]. As for the BH spin, its value is set by the frequency of accretion events consumed in prograde or retrograde orbits, which in L-Galaxies is linked to the coherence of the bulge kinematics [67, 150]. The contribution of binary coealescences to the BH mass and spin evolution is important at low redshifts ($z < 2$), where galaxies have mostly depleted their gas reservoirs through star formation and feedback processes.

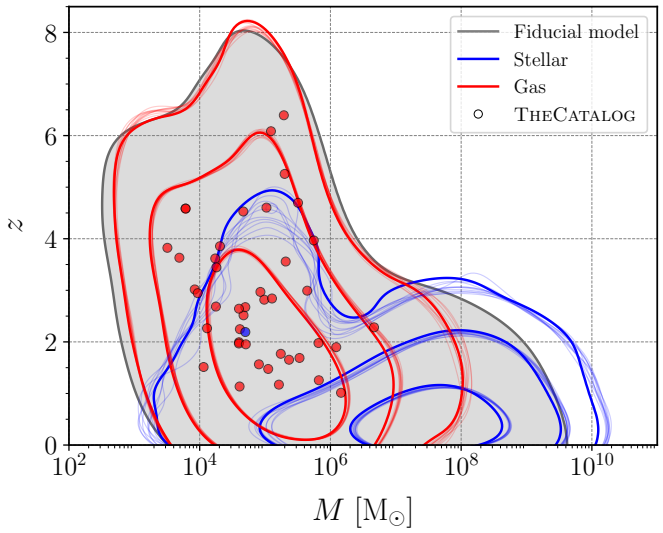

Figure 2.1: Density distribution of the total mass M and redshift z of massive BH binary mergers. Our fiducial population is shown in gray, where the contour encloses 99% of the estimated source density. The red (blue) distribution corresponds to the sub-population of mergers hardened by gas (stars); contours contain 50%, 90%, and 99% of the estimated source density. Thin contours refer to additional realizations to capture the statistical fluctuations in the underlying astrophysical model. Circles denote the 44 sources from TheCatalog used in our analysis, with colors indicating their respective $f_{\rm gas}$ values. Note the two pairs of overlapping sources, c.f. Fig. 2.4.

The post-merger BH masses and spins are computed using fits to numerical-relativity simulations [30, 164]. During the BH evolution, L-Galaxies tracks the evolution of the spin magnitude χ but not the spin direction, which is however important for LISA observations. In this work, we construct this property for each binary component using the gas fraction of the environment $f_{\rm gas} = M_{\rm res}/(M_{\rm res} + M)$, where M is the total mass of the BH binary. If $f_{\rm gas} > 0.5$, coalescence occurs in a gas-rich environment and the two spins are assumed to align to the binary orbital angular momentum via the Bardeen-Petterson effect [31, 69, 80]. If instead $f_{\rm gas} < 0.5$, the environment is gas-poor and the BH spins are assumed to be isotropically distributed [39, 71]. This prescription assumes a sharp transition between gas- and stellar-driven evolution at $f_{\rm gas} = 0.5$, effectively treating the two regimes as mutually exclusive. In realistic galactic environments, however, both

processes may act simultaneously, and a more refined model would account for their combined contribution [40].

2.2.3 Binary hardening

The dynamical pathway of massive BH binaries can be divided into three phases: pairing, hardening, and GW inspiral [32]. Following a galaxy merger, the pairing phase reduces the BH-BH separation from ~kpc to ~pc through dynamical friction. This is implemented in L-Galaxies as in Ref. [35]. As the binary forms, different physical processes contribute to the shrinking of the binary semi-major axis depending on the gas fraction $f_{\rm gas}$. For $f_{\rm gas} > 0.5$, gas hardening dominates and the decay of the binary orbit is set by accretion torques from the circumbinary disk. In gas-poor environments with $f_{\rm gas} < 0.5$, instead, the binary evolution is driven by capture and ejection of stars. Gravitational-radiation reaction becomes dominant at roughly milliparsec scales and drives the system to the merger. For details on the L-Galaxies implementation see Ref. [93].

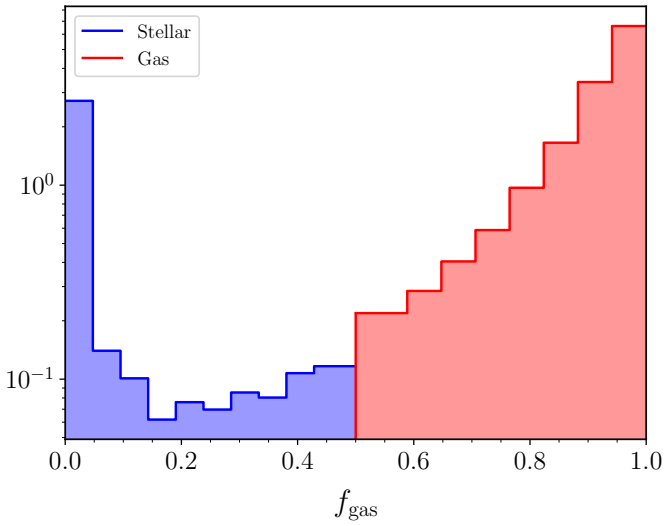

Figure 2.2: Distribution of the gas fraction $f_{\rm gas}$ of the environment across the fiducial population. Sources in blue (red) with $f_{\rm gas} < 0.5$ ($f_{\rm gas} > 0.5$) are classified as mergers occurring in a gas-poor (gas-rich) environment. The two sub-population distributions are jointly normalized to the full population.

2.2.4 Fiducial population

The model described in the previous subsections provides the astrophysical population of BH mergers we refer to as "fiducial", which is depicted in grey in Fig. 2.1 and presented in Ref. [94]. This population comprises two distinct sub-populations with peaks at

$f_{\rm gas} \sim 0$ (stellar hardening) and $f_{\rm gas} \sim 1$ (gas hardening) parameter (Fig. 2.2). The latter dominates the overall merger rate by a factor of ~ 5. Figure 2.1 shows the respective mass and redshift distributions for each sub-populations. Systems evolving in gaseous environments merge on average at higher redshifts $z \lesssim 8$ and have lower masses $M \sim 10^4 - 10^6\ \mathrm{M}_\odot$ compared to binaries hardened by stellar interactions which instead are more prominent at $z \lesssim 3$ and have $M \sim 10^5 - 10^9\ \mathrm{M}_\odot$. Nonetheless, the two populations overlap significantly in the $M - z$ parameter space, making their distinguishability with LISA non trivial.

2.3 Mock LISA catalogs

We build mock LISA observation catalogs on top of the L-Galaxies simulations. The latter provides the number density of massive BH binary mergers, $\mathrm{d}^2N/\mathrm{d}z\,\mathrm{d}V_c$, per unit redshift z and comoving volume V_c. We proceed as follows:

(i) We compute the total merger rate of massive BH mergers predicted by the fiducial model [29],

$$\frac{\mathrm{d}N}{\mathrm{d}t_{\rm obs}} = \int \frac{\mathrm{d}N}{\mathrm{d}z}\, w(z)\, \mathrm{d}z, \tag{2.1}$$

where

$$w(z) = 4\pi \left[\frac{d_{\rm L}}{1+z}\right]^2 \frac{\mathrm{d}z}{\mathrm{d}V_c}. \tag{2.2}$$

Here, $d_{\rm L}$ is the luminosity distance and $t_{\rm obs}$ denotes the time measured at the detector.

(ii) We draw the number of sources $N_{\rm astro}$ from a Poisson distribution with mean $\lambda = T\,\mathrm{d}N/\mathrm{d}t_{\rm obs}$, where T is the duration of the observing period. Massive BH binary signals in LISA typically last from hours to months and gradually accumulate signal-to-noise ratio (SNR) during their inspiral. We conservatively set $T = 5\,\mathrm{yr}$, which also accounts for sources merging after the nominal mission duration $T_{\rm LISA} = 4\,\mathrm{yr}$ [49].

(iii) We extract $N_{\rm astro}$ sources from the astrophysical population, assigning each source a relative weight $w_i = w(z_i)$. Each source is characterized by the binary component masses $m_{1,2}$, mass ratio $q = m_2/m_1$, redshift z, and the aligned dimensionless spins $\chi_{1,2} = \vec{\chi}_{1,2} \cdot \hat{L}$, where $\vec{\chi}_{1,2}$ are the dimensionless BH spins and $\hat{L}$ is the unit vector along the direction of the orbital angular momentum.

(iv) From this population, we further select sources satisfying $f_{\rm cut} > 10^{-4}\,\mathrm{Hz}$ and $q > 0.01$. We define the cutoff frequency as $f_{\rm cut} = 5 f_{\rm ISCO}$, where $f_{\rm ISCO}$ is the

detector-frame GW frequency at the innermost stable circular orbit. This ensures that the dominant harmonic lies within the LISA sensitivity band and that the adopted waveform model is reliable.

(v) We model the GW signal using the IMRPhenomXHM waveform approximant [78], which describes the full coalescence of quasi-circular, non-precessing BH binaries. The harmonic modes $h_{\ell m}$ are calibrated to numerical relativity and include $(\ell, |m|) = \{(2,2),(2,1),(3,3),(3,2),(4,4)\}$. The implementation of the LISA response for this waveform in the Balrog code is described in Ref. [140].

(vi) For each source, we sample the extrinsic parameters assuming uniform distributions: the time to merger $t_m \in [0, 5\,\mathrm{yr}]$, initial phase $\phi_0 \in [0, 2\pi]$, polarization angle $\psi \in [-\pi/2, \pi/2]$, ecliptic sine latitude $\sin\beta \in [-1, 1]$, ecliptic longitude $\lambda \in [0, 2\pi]$, and cosine of the inclination $\cos\iota \in [-1, 1]$.

(vii) The SNR of each source, given its parameters θ, is evaluated using the three noise-orthogonal TDI channels h_k with $k = A, E, T$,

$$\mathrm{SNR}^2 = \sum_{k=A,E,T} \left\langle h_k(\theta) \middle| h_k(\theta) \right\rangle_k. \tag{2.3}$$

We assume the constant and equal-arm-length approximation [166]. The noise-weighted inner product is defined as

$$\langle a|b\rangle_k = 2 \int_{f_{\min}}^{f_{\max}} \frac{\tilde{a}(f)\tilde{b}^*(f) + \tilde{a}^*(f)\tilde{b}(f)}{S_k(f)}\, \mathrm{d}f, \tag{2.4}$$

where $\tilde{a}(f)$ denotes the Fourier transform of the time series $a(t)$ and $S_k(f)$ is the noise power spectral density of the k-th TDI channel. We adopt the semi-analytical noise model of Ref. [25], which accounts for stationary instrumental noise and astrophysical confusion noise from unresolved Galactic binaries. We use a low-frequency cutoff $f_{\min} = 0.1\,\mathrm{mHz}$ [49] and integrate up to $f_{\max} = 1\,\mathrm{Hz}$, well above the maximum frequency of all simulated GW signals. Sources are considered detectable if they exceed the threshold $\mathrm{SNR} = 10$ [49].

We iterate the above procedure 1000 times, resulting in a distribution of LISA catalogs. The corresponding distributions for the number of merger events is shown in Fig. 2.3. Specifically, steps (i)–(iii) generate the light-gray histogram in the left panel, depicting the distribution of astrophysical mergers predicted by the fiducial model. The dark-gray histogram is obtained by applying steps (iv)–(vii) to the astrophysical catalogs and shows the distribution of massive BH merger events detectable by LISA assuming the dominant

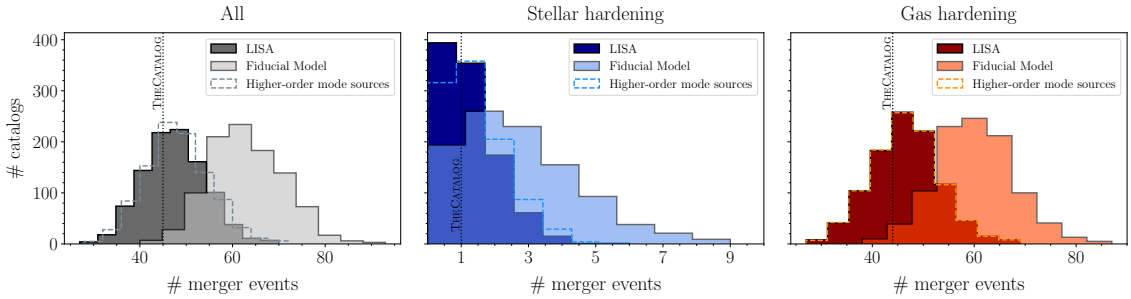

Figure 2.3: Distribution of massive BH binary mergers across the generated catalogs. The left (gray), middle (blue), and right (red) panel shows the full population of BHs, the sub-population of stellar-hardened sources, and the sub-population of gas-hardened sources, respectively. Light histograms show the astrophysical distributions [steps (i)–(iii) in Sec. 2.3] while dark histograms show the distributions of merger events detectable by LISA during an observation time $T = 5\text{yr}$ [steps (iv)-(vii) in Sec. 2.3]. The dashed distributions are a variation of the latter where we include higher-order harmonics and remove the condition on $f_{\rm cut}$ [step (iv)]. The vertical dotted line denotes the selected catalog used to perform the statistical analysis of Sec. 2.4.

$l = 2, |m| = 2$ emission mode. The blue (red) histograms in the middle (right) panel show the subpopulations of mergers that evolved via stellar (gas)-hardening. Overall, the LISA catalogs have about one order of magnitude fewer events compared to the astrophysical catalogs, which is due to the detection criteria described above. The observable catalogs predominantly feature massive BH binary mergers occurring in gas-rich environments (46 on average) compared to systems evolving in gas-poor environments (1 on average). The averaged count of detectable sources mildly increases to 50 when we relax the assumption on the mass ratio, allowing q to be as small as 0.001. These sources, and those at even lower mass ratio, border the extreme mass-ratio inspiral regime and will need to be treated differently; we leave this to future work.

The dashed distributions in all the three panels of Fig. 2.3 show the detectable events modeled using the additional modes $(l, |m|) = \{(2, 1), (3, 3), (3, 2)(4, 4)\}$ in the waveform approximant and removing the condition on $f_{\rm cut}$ [step (iv)]. We find higher-order modes marginally increase the number of detectable sources (see Ref. [158]). This is more significant for massive binaries with $M \gtrsim 10^8 \mathrm{M}_\odot$, where the (2,2) mode frequency falls out of band. As shown in Fig. 2.3, this preferentially impacts binaries hardened via stellar processes: including higher-order modes in LISA analyses might have important repercussions on the astrophysical interpretation of the data.

In the following, we study in detail one specific realization of our simulated LISA catalogs, which we refer to as TheCatalog. This specific catalog is close to the medians of the distributions in Fig. 2.3 and can thus be taken as representative. The 45 detectable sources of TheCatalog are shown as colored circles in the right panel of Fig. 2.4 together

with the astrophysical population from which they are extracted. The left panel puts these sources into context by showing the common "waterfall plot", i.e. the averaged SNR in the mass-redshift space [49]. LISA is expected to detect massive BH binary systems in the mass range of $M_{\rm tot} \sim 10^4 - 10^7$ $\mathrm{M}_\odot$ and out to $z \sim 7$, covering an SNR range from ~ 10 to ~ 3000. The 13 sources marked with white triangles in Fig. 2.4 are undetectable by LISA. Of these, 9 sources with $M > 10^4$ $\mathrm{M}_\odot$ merge after the LISA mission nominal duration, thereby limiting their SNR growth, cf. step (ii) in Sec. 2.3. Conversely, the 4 sources with $M < 10^4$ $\mathrm{M}_\odot$ merge within the mission lifetime $T_{\rm LISA}$ but remain with SNR < 10 nonetheless. These undetectable sources are excluded from the analysis presented in Sec. 2.4.

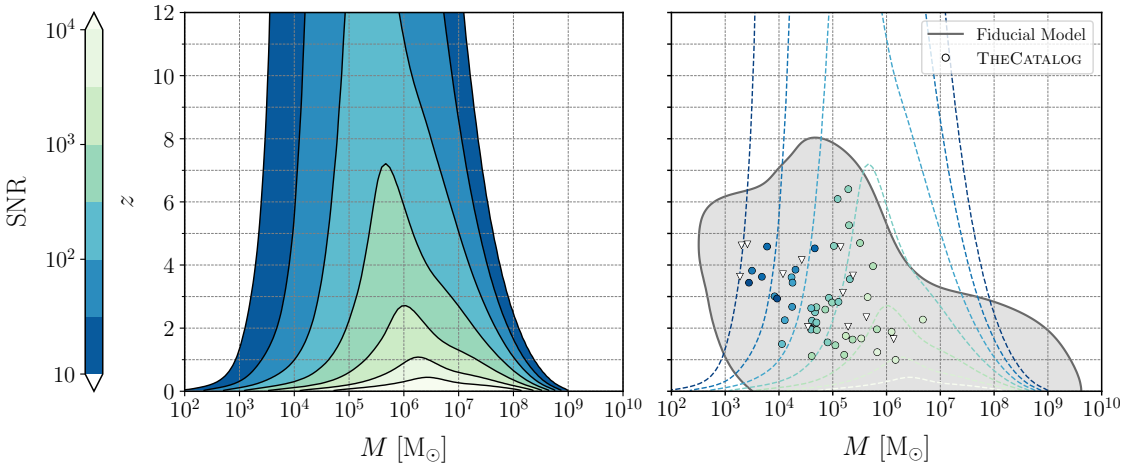

Figure 2.4: *Left panel.* Contour lines of constant SNR from massive BH binaries ($q = 0.5$, $\chi_{1,2} = 0.2$) detectable by LISA, adapted from Ref. [49]. The SNR shown on the color scale has been averaged over sky location, polarization and inclination assuming the dominant quadrupole emission mode. *Right panel.* The gray contour encompasses 99% of the estimated astrophysical BH population simulated with L-Galaxies. Circles indicate the 45 detectable sources of TheCatalog, colored according to their SNR [see step (vii) in Sec. 2.3]. Of note, two pairs of sources overlap in this plot: one, located at $z \sim 4.5$ with $M \sim 6 \times 10^3 \mathrm{M}_\odot$, have the same weights w_i [i.e. identical source parameters, see step (iii) in Sec. 2.3]; the other pair, located at $z \sim 2$ with $M \sim 4 \times 10^4 \mathrm{M}_\odot$, are characterized by very similar parameter values. White triangles indicate sources with SNRs below the detection threshold. Note that undetectable sources with high masses ($M > 10^4 \mathrm{M}_\odot$) are affected by the cut-off on the time to merger [step (ii) in Sec. 2.3]. The dashed contour lines of constant SNR are the same of the left panel.

2.4 Statistical inference

We compare a single GW event against distributions of simulated sources using the Bayesian formalism spelled out in Ref. [124]; see also Refs. [116, 168]. Specifically, we

quantify the relative degree of consistency between individual detections from THE-CATALOG and the two sub-population models of binary hardening: gas (G) and stellar (S).

Given the posterior distribution $p(\theta|d,\mathrm{U})$ from each individual GW event obtained with some uninformative priors $p(\theta|\mathrm{U})$, we compute the Bayes factor between models G and S as follows:

$$\mathcal{B}_{\mathrm{G/S}} = \frac{\mathcal{B}_{\mathrm{G/U}}}{\mathcal{B}_{\mathrm{S/U}}} = \frac{\int p(\theta|d,\mathrm{U})\frac{p(\theta|\mathrm{G})}{p(\theta|\mathrm{U})}d\theta}{\int p(\theta|d,\mathrm{U})\frac{p(\theta|\mathrm{S})}{p(\theta|\mathrm{U})}d\theta}\,. \tag{2.5}$$

The astrophysical probability densities $p(\theta|\mathrm{G})$ and $p(\theta|\mathrm{S})$ act as new informative priors on the targeted parameters which we reconstruct from the simulated sources using Gaussian kernel density estimates (KDEs) [148].

Parameter estimation is performed through BALROG on $\theta = \{\theta_\alpha, \theta_\zeta\}$, where $\theta_\alpha = \{\mathcal{M}_c, \delta\mu, \chi_1, \chi_2, d_\mathrm{L}\}$ represent the quantities used to perform the astrophysical model selection [Eq. (2.6)] and $\theta_\zeta = \{t_m, \phi_0, \psi, \sin\beta, \lambda, \cos\iota\}$ are the additional extrinsic parameters characterizing the modeled signals. In particular, we use the redshifted chirp mass $\mathcal{M}_c$ and the dimensionless mass difference $\delta\mu = (m_1 - m_2)/(m_1 + m_2)$ as they are the two mass parameters that enter the Post-Newtonian (PN) evolution at the leading- and next-to-leading order, respectively. We run full Bayesian inference on simulated data using the nested sampling algorithm [151] as implemented in NESSAI [172]. All the injections are in zero noise and we choose uniform priors $p(\theta|\mathrm{U})$ on each parameter over either its entire definition domain or a range that is sufficiently large to enclose the entire posterior. In particular, we classify sources into three SNR intervals: low $(10-100)$, moderate $(100-1000)$, and high (>1000). For each of these, we estimate prior limits by calibrating against a fiducial source selected at the lower boundary of that range. Therefore, the terms $p(\theta_\alpha|\mathrm{U})$, $p(\theta_\zeta|\mathrm{U})$, $p(\theta_\zeta|\mathrm{G})$, $p(\theta_\zeta|\mathrm{S})$ [see step (vi) of Sec. 2.3] are constant, we can factor them out the integrals of Eq. (2.5) and evaluate the KDEs exclusively on θ_α. This approach significantly improves the computational efficiency. Equation (2.5) thus simplifies to

$$\mathcal{B}_{G/S} = \frac{\int p(\theta_\alpha|d,\mathrm{U})p(\theta_\alpha|\mathrm{G})d\theta_\alpha}{\int p(\theta_\alpha|d,\mathrm{U})p(\theta_\alpha|\mathrm{S})d\theta_\alpha}\,. \tag{2.6}$$

Note that our detection statistic is approximated as GW detectability depends on the data realization (for detailed discussions on this point see e.g. Refs. [75, 125]). We verified that all of our posterior samples have SNR > 10 and $q > 0.01$, which ensure consistency

with the adopted detection criterion. Furthermore, it is important to note that posterior samples constrained to a parameter space θ_α that is not supported by the gas [stellar] probability density, i.e. $p(\theta_\alpha|\mathrm{G}) = 0$ [$p(\theta_\alpha|\mathrm{S}) = 0$], result in a formally null [infinite] value for the Bayes factor in Eq. (2.6). As a concrete example, see the discussion on the spin distributions in Sec.2.5.

Crucially, one needs to account for selection effects. The detectability-conditioned Bayes factor is given by [124]

$$\mathcal{D}_{\mathrm{G/S}} = \frac{P(\mathrm{det}|\mathrm{S})}{P(\mathrm{det}|\mathrm{G})}\mathcal{B}_{\mathrm{G/S}}, \tag{2.7}$$

where $P(\mathrm{det}|\mathrm{S})$ and $P(\mathrm{det}|\mathrm{G})$ are the fraction of sources from the sub-populations S and G, respectively, that can be detected *a priori* given adopted detection statistic SNR > 10 and $q > 0.01$. Our numerical implementation follows that described in Ref. [124].

2.5 Results & Discussion

2.5.1 Inference

We apply the formalism from Sec. 2.4 to the 45 mergers in TheCatalog, which LISA is expected to detect. Our parameter-estimation pipeline successfully recovers 44 sources, with posterior distributions on source parameters θ well confined within physically unbounded priors. However, for one source with SNR $\sim$ 13, the stochastic algorithm fails to converge because the data provide limited information on the source parameters, leaving the posterior distribution close to the prior and therefore challenging to sample efficiently. Therefore, we choose to exclude this source from the population analysis. Additionally, sources with SNRs in the range $10 - 100$ frequently exhibit non-linear correlations and multimodalities on θ_ζ; however both are sufficiently decoupled from the reconstructed θ_α's, which are the target parameters of our astrophysical model selection. In the following, we quote parameter estimates at 90% confidence interval.

Among the target parameters, $\mathcal{M}_c$ is the most precisely measured. Specifically, we constrain it with a relative precision $\Delta\mathcal{M}_c/\mathcal{M}_c$ of $\sim 10^{-6} - 10^{-4}$ for sources with moderate and high SNRs. For low SNRs, the relative precision is $\sim 10^{-4} - 10^{-3}$. The spin components χ_1 and χ_2 are measured with a relative precision of $0.006\% - 1\%$ at high and moderate SNRs, and 1% to 65% at low SNRs.

Six systems in the low–SNR range exhibit mild biases on $\delta\mu$ due to the broad posteriors ($\Delta\delta\mu/\delta\mu \sim 6\% - 160\%$ at 90% credible level) and projection effects associated to non-linear parameter correlations. Of these, three systems have posterior distributions that include

the injected values only in their 99.9% confidence inteval. For the other three sources, the true injected value lies within the 98.5% credible interval. Finally, we emphasize that LISA will be able to measure the source luminosity distance with a relative precision $\Delta d_{\rm L}/d_{\rm L}$ of $\sim 0.3\% - 2\%$ for high SNRs, $\sim 1\% - 45\%$ for moderate SNRs, and $\sim 10\% - 115\%$ for low SNRs. Uncertainties related to sky localization and host galaxy identification, which are crucial for multimessenger detections of gas-rich mergers [117], will be addressed in future work.

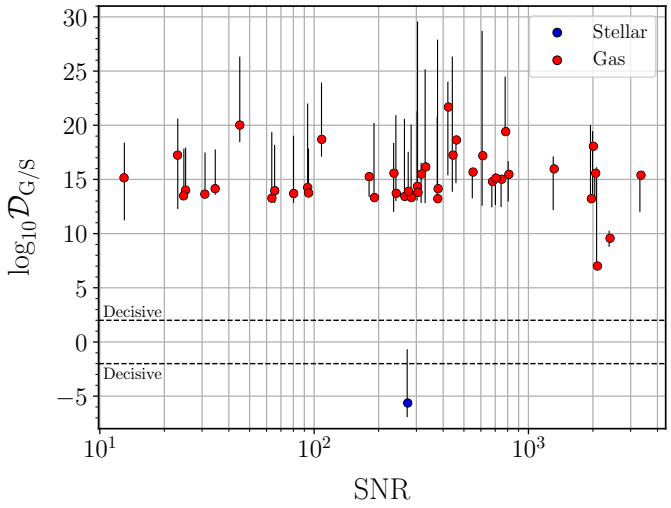

Figure 2.5: Logarithmic Bayes factor between the gas- and stellar- hardening hypotheses as a function of the source SNR. Circles represent the 44 sources from TheCatalog used in the population analysis, with colors indicating their respective $f_{\rm gas}$ values. Horizontal dashed lines denote the threshold values of the Bayes factor according to the Jeffrey scale. Black error bars refer to the statistical (but not systematic) fluctuations associated to the underlying astrophysical model.

2.5.2 Model selection

We use the posteriors described above to evaluate the detection-conditioned Bayes factors $\mathcal{D}_{\rm G/S}$ of Eq. (2.7). Results are presented in Fig. 2.5, where sources are ordered by their SNR and colored according to their true astrophysical sub-population. For all the sources we find "decisive" evidence [95] in favor of the gas (stellar) sub-population model, with $\log_{10}\mathcal{D}_{\rm G/S} > 2$ ($\log_{10}\mathcal{D}_{\rm G/S} < -2$). Additionally, no significant correlation is observed between $\mathcal{D}_{\rm G/S}$ and the SNR of each source. This is largely due to the highly precise measurement of the target parameters, as described above. Even the wider posterior distributions are well localized within the sub-population parameter space, so increasing the SNR does not lead to improved discrimination between the two astrophysical models.

We observe that most of the Bayes factors fall within the range $10 \lesssim |\log_{10}\mathcal{D}_{\mathrm{G/S}}| \lesssim 25$, with three exceptions that exhibit lower values. Among theese, two sources with SNR ~ 2000 and $7 \lesssim |\log_{10}\mathcal{D}_{\mathrm{G/S}}| \lesssim 10$ evolved trough a gas-hardening phase ($f_{\mathrm{gas}} \sim 0.72$). Conversely, the lowest Bayes factor ($|\log_{10}\mathcal{D}_{\mathrm{G/S}}| = 5.63$) is associated with the only source in TheCatalog that hardened through stellar interactions ($f_{\mathrm{gas}} \sim 0.14$).

We further test the analysis pipeline by generating ten additional realizations of our astropyhysical model to evaluate the impact of simulation uncertainties on the results. The error bars in Fig. 2.5 indicate the resulting variability associated with each log-Bayes factor. Even when accounting for these statistical fluctuations, model G is confidently favored for all the G sources, thus validating the previous results. For the one S source, model S remains favored, though to a lesser extent. Crucially, these error bars capture the statistical fluctuations in the underlying astrophysical model and not the systematic effects due to the many assumptions entering the model itself.

2.5.3 Source parameters

Based on these results, we now investigate how the astrophysical properties of massive BH binaries might influence the distinguishability between models G and S. In Fig. 2.1, we show the 44 analyzed systems from TheCatalog together with the astrophysical sub-population distributions marginalized on the total mass and redshift. Notably, we observe that the gas-hardened sources are predominantly supported by the gas-hardening sub-population, with most falling within (outside) the 90% contour level of the gas (stellar) distribution. Conversely, the stellar-hardened source, hereafter referred to as s_1, falls outside the 90% contour level of the stellar-hardening sub-population but is located close to the peak (i.e. within the 50% contour level) of the gas-hardening sub-population. This suggest that additional parameters may be impacting its $\log_{10}\mathcal{D}_{\mathrm{G/S}}$ value, favoring the stellar-hardening model.

To investigate further, we analyze the one-dimensional distribution of the aligned component spins χ_1 and χ_2, see Fig. 2.6. For the source s_1 considered so far, the primary component spin $\chi_1 = 0.49^{+0.08}_{-0.08}$ lies below the 92^{nd} and 1.5^{th} percentiles of the stellar- and gas-hardening distributions, respectively, thus placing it confidently in the former. Additionally, the secondary component spin $\chi_2 = 0.39^{+0.13}_{-0.13}$ is well below the 90^{th} percentile of the stellar-hardening distribution and lies within only the 0.3^{rd} percentile of the gas-hardening distribution. We find that the recovered $\delta\mu$ parameter does not provide significant support for model S, with the full posterior information falling below the 14^{th} percentile of the stellar-hardening distribution but within the 50^{th} percentile of the gas-hardening sub-population.

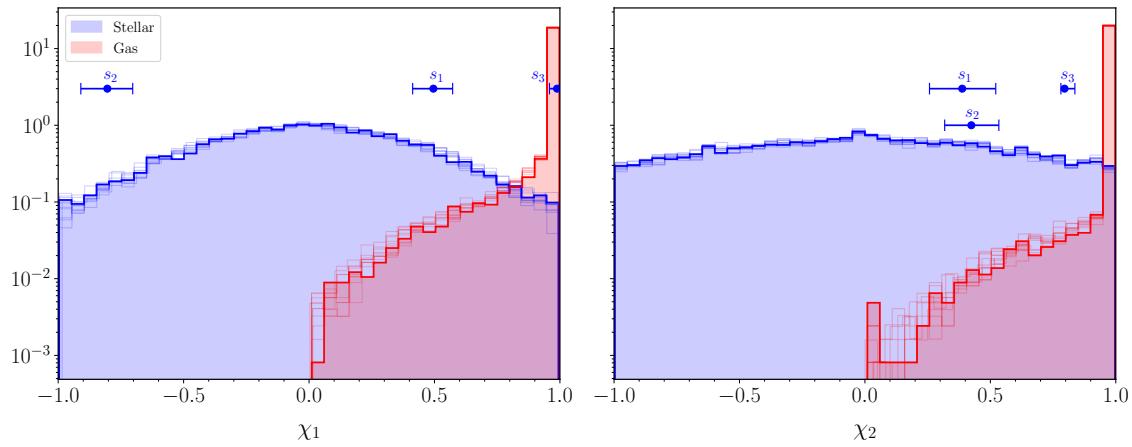

Figure 2.6: Marginal distribution of the aligned spin component of the primary BH χ_1 (left panel) and the secondary BH χ_2 (right panel) for mergers in gaseous (red) and stellar (blue) environments. Thick histograms indicate our fiducial model; thin histograms indicate additional realizations. Blue dots denote the median value of source s_1 from TheCatalog as well as the two additional sources $s_{2,3}$, c.f. Sec. 2.5. The error bars indicate the 90% confidence intervals. The two sub-population distributions are jointly normalized to the full population.

2.5.4 Importance of the BH spins

From the discussion above, we conclude that, at least for the astrophysical models considered here, the component spins are the crucial parameters for inferring the hardening processes of massive BH binaries. We stress that the employed models describe a simplified scenario. Massive BH binaries evolving in gas-rich environments experience accretion torques that might align their component spins with the orbital angular momentum of the binary. The models used here assume that the spin alignment process occurs rapidly, typically within the time-scale of the merger, thus leading to high and positive spin configuration at merger [39] ($\chi_{1,2} \sim 0.98$ on average, as indicated by the red distributions in Fig. 2.6). In contrast, binaries that evolve in gas-poor environments do not experience significant gas accretion and likely enter the GW dominated phase with roughly isotropic spin orientations, as reflected by the blue distributions in Fig. 2.6. Nevertheless, as pointed out in the literature, the overall picture is more complex due to the internal properties of the disk. Indeed, the effectiveness of disk alignment may be influenced by the gas temperature [69], accretion rate [79], as well as the disk viscosity [111] and its initial angle of misalignment, which can lead to critical disk-breaking configurations [131]. Accurate astrophysical models accounting for these properties are crucial for addressing such scenarios in future study.

Finally, we extend the analysis from TheCatalog by including two additional sources, s_2 and s_3, selected from the LISA stellar-hardening distribution shown in Fig. 2.3. In

order to provide a representative sample of LISA observations, we specifically select sources with SNR in the range $100 - 1000$, which is most densely populated interval for detectable massive BH binaries in LISA (see Fig. 2.4). This region is populated by sources with $M \sim 10^4 - 10^6 \mathrm{M}_\odot$ at redshift $z \sim 2 - 4$. Additionally, we select the two sources such that their primary spin components fall in the tails of the stellar-hardening sub-population from Fig. 2.6. This allow us to further investigate the role of spins in the model comparison.

First, as shown in Fig. 2.6, the primary component spin $\chi_1 = -0.8^{+0.1}_{-0.1}$ for source s_2 is entirely constrained within a region where $p(\chi_1|\mathrm{G}) = 0$, thus making the KDE evaluation unnecessary, as pointed out in Sec. 2.4. Therefore, we can definitely favor model S over model G ($\log_{10}\boldsymbol{\mathcal{D}}_{\mathrm{G/S}} \ll -2$) without even considering the influence of the other binary properties. For source s_3, the analysis indicates a preference for the stellar-hardening scenario with $\log_{10}\boldsymbol{\mathcal{D}}_{\mathrm{G/S}} = -4.65$. This case is particularly challenging because the gas-hardening model has also strongly support over the source posterior. Specifically, s_3 has a mass $M \sim 7 \times 10^4 M_\odot$ and is located at a redshift $z \sim 1.6$, positioning it at the peak of the $M - z$ distribution for the gas-hardening sub-population (see Fig. 2.1). Additionally, it has a measured primary spin component of $\chi_1 = 0.98^{+0.01}_{-0.03}$. This value is notably high and exceeds the 99.5^{th} percentile of the stellar-hardening distribution and the 7^{th} percentile of the gas-hardening distribution. As illustrated in Fig. 2.6, s_3 is situated in a high-density region for gas-hardened sources but in a relative sparse region for stellar-hardened sources. Furthermore, the recovered value of $\delta\mu$ falls below the 9^{th} (38^{th}) percentile for the stellar (gas) sub-population, providing no further support for model S with respect model G. On the other hand, the secondary component spin strongly favors model S. In fact, the recovered value $\chi_2 = 0.79^{+0.04}_{-0.01}$ falls below the 95^{th} (1.2^{nd}) percentile for the stellar- (gas-) hardening sub-population.

This further emphasizes that, at least in the astrophysical setup considered here, component spins play a crucial role in constraining the hardening mechanisms responsible for the final coalescence of massive BH binaries, even in ambiguous but rare cases (less than 1% of the stellar-hardening population sources detectable by LISA). However, simulation-induced fluctuations reduce the log-Bayes factor of source s_3 to as low as 0.35, indicating only a weak preference for model S. Incorporating spin precession into the population analysis may be crucial to enhance model discrimination and gain further insights into the galactic environments where massive BH binary mergers occur. This is left to future work.

3 Glitch systematics on the observation of massive black-hole binaries with LISA

Overview

Detecting and coherently characterizing thousands of GW signals is a core data-analysis challenge for LISA. Transient artifacts, or "glitches", with disparate morphologies are expected to be present in the data, potentially affecting the scientific return of the mission. We present the first joint reconstruction of short-lived astrophysical signals and noise artifacts. Our analysis is inspired by glitches observed by the LPF mission, including both acceleration and fast displacement transients. We perform full Bayesian inference using LISA TDI data and gravitational waveforms describing mergers of massive black holes. We focus on a representative binary with a detector-frame total mass of $6 \times 10^7 \mathrm{M}_\odot$ *at redshift* 5*, yielding a signal lasting* ~ 30 h *in the LISA sensitivity band. We explore two glitch models of different flexibility, namely a fixed parametric family and a shapelet decomposition. In the most challenging scenario, we report a complete loss of the gravitational-wave signal if the glitch is ignored; more modest glitches induce biases on the black-hole parameters. On the other hand, a joint inference approach fully sanitizes the reconstruction of both the astrophysical and the glitch signal. We also inject a variety of glitch morphologies in isolation, without a superimposed gravitational signal, and show we can identify the correct transient model. Our analysis is an important stepping stone toward a realistic treatment of LISA data in the context of the highly sought-after "global fit".*

3.1 Executive summary

This Chapter is organized as follows. In Sec. 3.2, we introduce the phenomenology of the expected instrumental artifacts. In Sec. 3.3, we present our glitch models and provide a brief summary of the fiducial GW-source and glitch parameters. In Sec. 3.4, we derive an alternative set of TDI variables suitable for the simultaneous treatment of glitches and GW signals. In Sec. 3.5, we provide definitions of relevant statistical quantities and details on our parameter-estimation runs. Finally, in Sec. 3.6, we present our inference results. Throughout this paper, we use units where $c = 1$. The material presented in this Chapter is based on Ref [155].

3.2 LPF glitches in LISA data

3.2.1 Phenomenology of LPF glitches

Glitches are observed as additional signals in the data stream. They can be thus modeled and subtracted from the data as such. The strategy here is to (i) get a consistent estimate of the power spectral density (PSD) of the underlying quasi-stationary noise over the entire data stream and thus (ii) improve the astrophysical signal inference by making it robust against glitch-induced biases. The latter constitutes a key element of the LISA data processing pipeline in view of the targeted "global fit" [51,110]. The properties of glitches, namely amplitude, duration, and time morphology, depend both on the measurement system and the originating physical process.

LPF observed two main kinds of glitches: a first class treated as an effective displacement-measurement artifact in the optical metrology chain and another class due to spurious forces acting on the TMs. Displacement glitches have been rarely observed in nominal conditions, have a typical duration comparable with the LISA sampling cadence, and carry negligible impulse per unit of mass as compared to the typical forces acting on the TMs [23]. As a consequence, fast, low-impulse glitches could be expected to affect the geodesic motion of the LISA constellation only mildly. On the contrary, force events result in impulse-carrying glitches lasting from tens of seconds to several hours, have a significant impact on the noise performance, and can potentially contaminate GW detection and parameter estimation.

During its ordinary runs, LPF observed 102 impulse-carrying glitches and 81 of these were visible in the data stream as a sharp, positive offset of the residual force-per-unit-mass (henceforth loosely referred to as "acceleration") [23]. These acceleration glitches correspond to the two TMs moving toward each other along the sensitive axis of the pair,

i.e. the direction joining their respective centers of mass. The rate of these events has been estimated to be about 1 per day and compatible with a Poisson distribution [23]. Several possible physical origins for glitches have been vetoed by extensive cross-checking and correlation analysis on LPF data, with the most plausible explanation pointing to either gas outbursts or virtual leaks in the vacuum chamber and the material surrounding the TMs. Dedicated experimental studies are underway to corroborate this hypothesis [23].

3.2.2 Guiding principles for LISA differential acceleration measurements

We now list a few guiding principles behind our modeling choices:

- Long-lived glitches related to force phenomena such as those observed by LPF are the most relevant for LISA; For these, we adopt a phenomenological parameterization suitable to describe their temporal evolution in terms of differential TM accelerations;
- Constructing the corresponding signal model for fractional phase observables in the frequency domain is more complex, although doable;
- Long-lived transients present in a displacement (optical phase) or velocity (optical frequency) observable disappear in an acceleration observable, with the signal disturbance limited to the duration of the external force transient. Likewise, glitch parameters related to the initial conditions – position and velocity – are eliminated with an acceleration observable;
- In a realistic operational setup, systematic errors arising from force disturbances (e.g. stiffness coupling) could be subtracted directly in acceleration. Thus, our fitting model does not require any additional integration or whitening filter;
- When the effective glitch "signal" has spectral content mainly near the low-frequency end of the LISA sensitivity range, differentiation is numerically safer than integration. In this regime, data correction from systematics in the displacement variables is still viable;
- The corresponding TDI variables written in acceleration allow for a straightforward inclusion of LPF glitches in a Bayesian inference framework;
- GW signal models can be easily rewritten as effective accelerations by differentiating those already available in phase or fractional frequency.

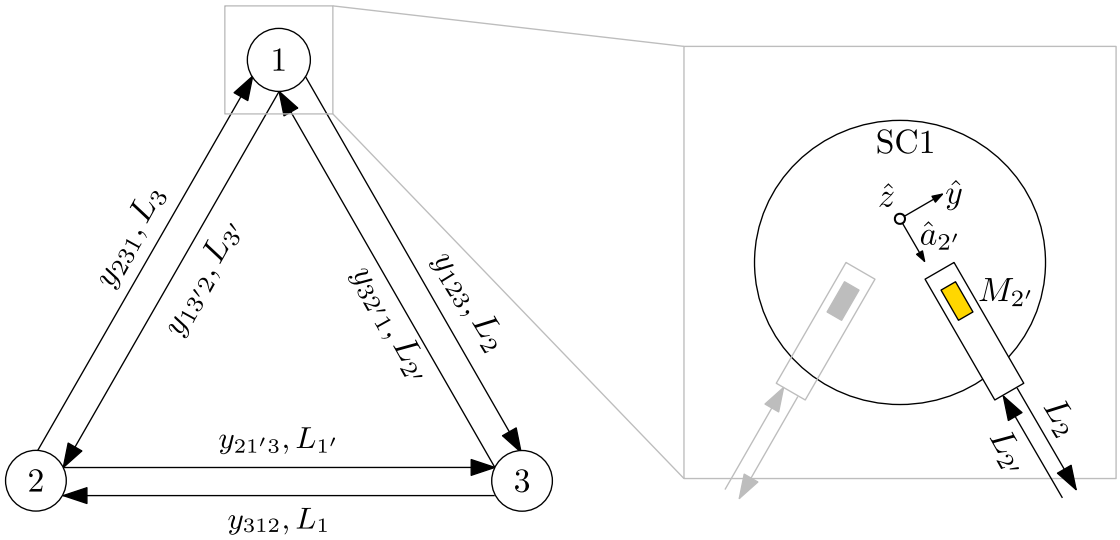

Figure 3.1: Schematic of single laser links and glitch reference system conventions. The left panel reproduces the single-link geometry introduced in Fig. 1.5, while the right panel adds the local reference system used to model acceleration glitches acting on the TMs (right inset, yellow and gray boxes). Unit vectors $\hat{a}_j$ parametrize the glitch component along the incoming (outgoing) link $L_{j'}$ (L_j) associated with the TM $M_{j'}$. On satellite 1, generic acceleration glitches acting on TM $M_{2'}$ and M_3 are described by the components $a_{2'}$ and a_3, respectively. The former [latter] affects link $y_{32'1}(t)$ [$y_{231}(t)$] at reception and link $y_{123}(t-L)$ [$y_{13'2}(t-L)$] at emission.

These broad considerations are mostly inspired by the observational equivalence between GWs and tidal forces accelerating TMs relative to their local inertial frames [50]. We thus opt to implement our joint inference for glitches and GWs with suitable acceleration TDI variables.

3.3 Transients modeling

The fundamental observable in LISA is the phase evolution $\Delta\phi$ of a one-way propagating laser along each of the six links connecting the satellites. This can be equivalently written as an optical pathlength

$$L = \frac{\Delta\phi}{\omega_l}\,, \tag{3.1}$$

where ω_l is the central frequency of the laser signal, which is assumed to be constant.

We now focus on three different mechanisms perturbing the phase readout.

3.3.1 Acceleration transients

The two TMs housed in each of the LISA satellites are expected to independently exchange momentum with their surrounding environment (see Fig. 3.1 for a schematic representation). We model the resulting transient acceleration profile $\vec{a}_i$ of the i-th TM as in Ref. [26]. We use a two-damped exponential model inspired by glitches observed in LPF, namely

$$g(t; A, \beta_1, \beta_2, \tau) = \frac{A}{\beta_1 - \beta_2}\left(e^{-\frac{t-\tau}{\beta_1}} - e^{-\frac{t-\tau}{\beta_2}}\right)\Theta(t-\tau), \tag{3.2}$$

which we refer to as Model A1. Equation (3.2) integrates to the net transferred momentum per unit mass:

$$\int_{-\infty}^{+\infty} g(t; A, \beta_1, \beta_2, \tau)\, \mathrm{d}t = A\,. \tag{3.3}$$

The parameters β_1, β_2 describe the typical timescales of the two exponentials while τ is the glitch onset time entering the Heaviside step function Θ. The corresponding Fourier-domain representation is

$$\tilde{g}(\omega; A, \beta_1, \beta_2, \tau) = -A\frac{e^{-i\tau\omega}}{(\beta_1\omega - i)(\beta_2\omega - i)}\,. \tag{3.4}$$

Accommodating glitches of unknown shape requires a more flexible model. We construct this using a superposition of S Gabor-Morlet shapelets

$$g(t) = \sum_i^S \sigma\left(t; A_i, \tau_i, \beta_i, n_i\right), \tag{3.5}$$

where

$$\sigma\left(t; A, \tau, \beta, n\right) = c_n \psi_n\left(\frac{t-\tau}{\beta}\right), \tag{3.6}$$

$$\psi_n\left(t\right) = \frac{2t}{n} e^{-t/n} L_{n-1}^{(1)}\left(\frac{2t}{n}\right)\Theta\left(t\right), \tag{3.7}$$

$$c_n = (-1)^{n-1}\frac{A}{2\beta n^2}, \tag{3.8}$$

and $L_n^{(\alpha)}(t)$ is the $n-$th generalized Laguerre polynomial [14]. We refer to these expressions as Model A2. Comparing to Ref. [26], we use a different normalization c_n for the individual shapelets such that

$$\int_{-\infty}^{+\infty} \sigma(t; A, \tau, \beta, n)\, \mathrm{d}t = A\,, \qquad \forall n \in \mathbb{N}\,. \tag{3.9}$$

In the frequency domain Eq. (3.6) reads

$$\tilde{\sigma}(\omega; A, \tau, \beta, n) = (-1)^n e^{-i\omega\tau} A \frac{(n\beta\omega + i)^{n-1}}{(n\beta\omega - i)^{n+1}}. \tag{3.10}$$

Shapelets in this parametric family are quasi-orthogonal, i.e.

$$\int_{-\infty}^{+\infty} \tilde{\sigma}(\omega; A, \tau, \beta, n)\tilde{\sigma}^*(\omega; A', \tau, \beta, m)\,\mathrm{d}\omega = \delta_{nm}\frac{\pi AA'}{2n\beta}, \tag{3.11}$$

$$\int_{-\infty}^{+\infty} \tilde{\sigma}(\omega; A, \tau, \beta, n)\tilde{\sigma}^*(\omega; A', \tau', \beta, n)\mathrm{d}\omega = \frac{\pi AA'}{2n^2\beta^2} e^{-\frac{|\tau-\tau'|}{n\beta}}(n\beta + |\tau - \tau'|). \tag{3.12}$$

From Eqs. (3.4) and (3.10) it is immediate to show that Model A1 tends to Model A2 with $n = 1$ in the limit where $\beta_1 \to \beta_2$.

3.3.2 Displacement transients

The interferometer readout system is also expected to generate transient phase fluctuations. From Eq. (3.1), we model these as effective displacement transients with the same agnostic shapelet parameterization used in Eq. (3.5). We use a superposition of S shapelets

$$\Delta L(t) = \sum_i^S \sigma\left(t; D_i, \tau_i, \beta_i, n_i\right), \tag{3.13}$$

where

$$\int_{-\infty}^{+\infty} \mathrm{d}t \Delta L(t) = \sum_i^S D_i \tag{3.14}$$

is the net integrated displacement experienced by the TM before returning asymptotically to its free-fall condition. We refer to this parametric family of glitches as Model D. The frequency domain representation follows from Eq. (3.10) and reads

$$\tilde{\sigma}(\omega; D, \tau, \beta, n) = (-1)^n e^{-i\omega\tau} D \frac{(n\beta\omega + i)^{n-1}}{(n\beta\omega - i)^{n+1}}. \tag{3.15}$$

3.3.3 GW transients

Among the large variety of typical sources populating the LISA sensitivity band, the most massive binary systems detectable produce hours to years-long transient signals.

To leading-order, the binary time to merger t_m from a reference frequency $f_{\rm ref}$ is [36, 137]

$$t_m \sim \left(\frac{3}{4\eta}\right)\left(\frac{f_{\rm ref}}{0.1\ {\rm mHz}}\right)^{-\frac{8}{3}}\left(\frac{M_z}{10^7 M_\odot}\right)^{-\frac{5}{3}} {\rm days}\,, \tag{3.16}$$

where $\eta \equiv m_1 m_2/(m_1+m_2)^2$ is the symmetric mass ratio and $M_z = (1+z)(m_1+m_2)$ is the solar-system barycenter frame total mass for a source of component masses m_1 and m_2. By contrast, glitches observed by LPF have typical durations of seconds to hours and are positively correlated with the transferred momentum per unit mass ranging from 10^{-2} to 10^3 pm/s [23]. Their broadband, short-lived morphology makes them the most likely to impact parameter estimation for GW transient sources of comparable duration.

We select three fiducial noise transients and superimpose them on a short-lived ($t_m = 30$ hours) high-mass ($M_z = 6\times10^7\,M_\odot, \eta = 3/16$) MBHB at redshift $z = 5$. We assume zero sensitivity below 0.1 mHz [140]. We consider a short-duration Model D glitch ($\beta = 5$ s), a moderate-duration Model A2 ($\beta = 40$ s), and a long-duration Model A1 glitch with $\beta_1 + \beta_2 = 3300$ s.

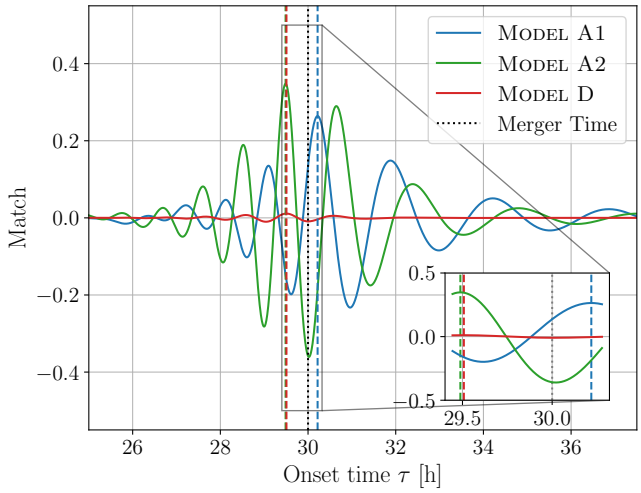

Figure 3.2: Match between the GW and glitch signals as a function of onset time. The blue, green, and red solid curve corresponds to the Model A1 (A2, D) glitch shown in Fig. 3.3. The GW signal is fixed to that of our fiducial MBHB. Dashed vertical lines with matching color denote the onset time that maximizes the match. The black dotted line denotes the GW source nominal merger time. The inset shows a 40-minute interval zoom-in around the merger and glitch onset times.

All three glitches have peak amplitudes close to the merger time of the GW source, as shown in Fig. 3.3. For a conservative approach, we fine-tune the glitch onset times to maximally impact the reconstruction of GW source parameters. This is done by

maximizing the match between the glitch and GW waveforms as shown in Fig. 3.2 (see Sec. 3.5 for more details).

We model the GW signal with the IMRPhenomXHM [78, 139] waveform approximant which captures the full coalescence of a quasi-circular, non-precessing black-hole binary. The implementation of the LISA response to this GW signal in the Balrog code has been presented in Ref. [140].

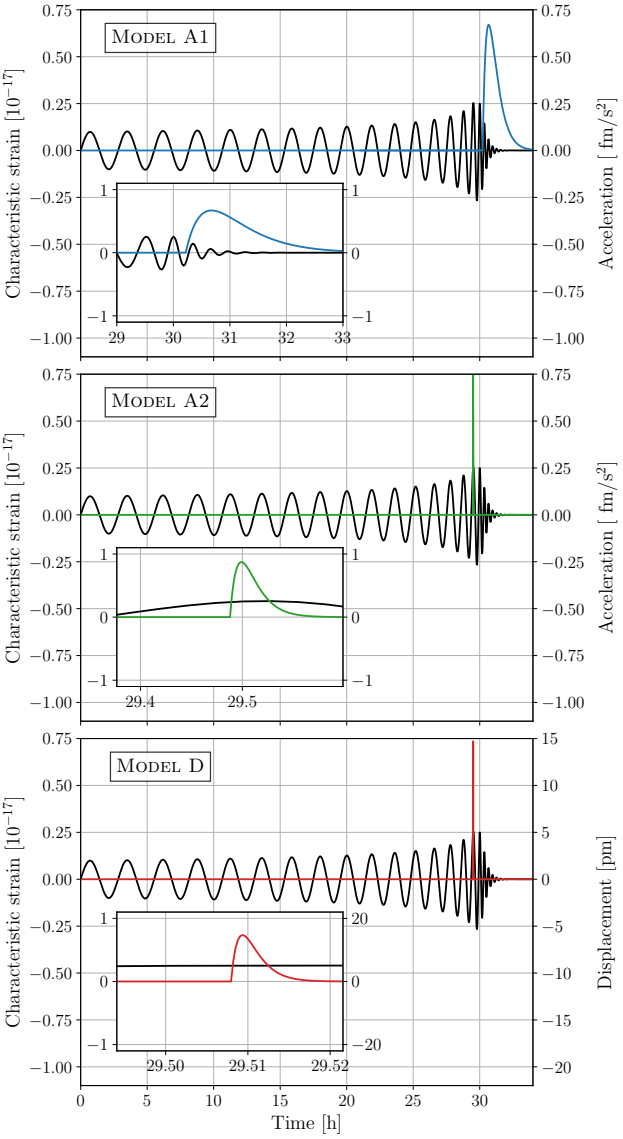

Figure 3.3: Fiducial waveforms for our parameter-estimation runs. Black solid curves show the MBHB signal we consider ($M_z = 6 \times 10^7\, M_\odot$ and $z = 5$), which is identical across the three panels. Colored curves in the top, middle, and bottom panels describe the Model A1, Model A2, Model D glitch amplitudes, respectively. Signals shown in the three panels correspond to injections in runs 9, 10, and 11 and exemplify glitches lasting hours, minutes, and seconds, respectively (cf. Table 3.1). The parameters of the injected signals are shown in Tables 3.3, 3.4, and 3.5.

We choose to parametrize the GW signal injected as follows: $m_{1z,2z}$ and $\chi_{1,2}$ denote the binary component redshifted masses and aligned dimensionless spin, respectively; t_m, ϕ_0, ψ denote the time to merger introduced in Eq. (3.16), initial phase and polarization, respectively; $\sin\beta, \lambda$ denote the (sine-)ecliptic latitude and longitude; d_L and ι denote the source luminosity distance and inclination. Tables 3.3, 3.4 and 3.5 list the parameter

values of our fiducial GW source which has an SNR of 187 and is common across all of our runs.

3.4 Acceleration TDIs

We use Eqs. (3.4), (3.10), and (3.15) to model the TDI variables [166] $\tilde{s}_k(f;\boldsymbol{\theta})$ entering the likelihood, cf. Sec. 3.5. As described in Sec. 1.4.2, the analysis is performed within the constant equal-armlength approximation.

In this framework, incorporating Model A1 and Model A2 signals into Eqs. (1.6), (1.7), and (1.8) requires integrating the single-link differential accelerations twice. However, any non-zero total transferred momentum necessitate artificial regularization or ad-hoc approximations to construct a Fourier-domain representation of the signal. We solve this problem by introducing a set of "acceleration TDIs" $G_{X,Y,Z}$ which are trivially related to Eqs. (1.6), (1.7), and (1.8) by double differentiation. In the frequency domain one has

$$\mathcal{F}\left[G_X\right] = (2\pi f)^2 \mathcal{F}\left[M_X\right] \tag{3.17}$$

$$G_X = g_{231} + \mathcal{D}g_{13'2} - g_{32'1} - \mathcal{D}g_{123}, \tag{3.18}$$

where $\mathcal{F}$ denotes the Fourier transform operator and

$$g_{ijk}(t) = \frac{d}{dt^2}\left[y_{ijk}(t)\right]. \tag{3.19}$$

Similar definitions hold for G_Y, G_Z upon cyclic permutation of indices.

The key advantage of introducing a new set of TDIs lies in its instrumental robustness. Equation (3.17) also allows us to conveniently recycles signal models available in fractional displacement by including both Model D glitches and GW signals. Furthermore, Eq. (3.19) does not require a transfer function to model acceleration glitches.

Following the conventions shown in Fig. 3.1, the single-link perturbation $g_{ijk}(t)$ is obtained from the instantaneous accelerations $\vec{g}_i(t)$ and $\vec{g}_k(t-L)$ which are experienced by sender i and receiver k along the link j, and projected along the unit-vectors $\hat{a}_j(t-L)$ and $\hat{a}_{j'}(t)$, respectively. We associate a unit vector $\hat{a}_j$ to each TM M_j pointing in the direction opposite to L_j. For simplicity, we denote the associated vector components a_j. Given the choice of the local reference system, a positive value a_i corresponds to a negative displacement ΔL_i. The three TDI observables in terms of the individual TM

accelerations are

$$G_X = (1 + \mathcal{D}^2)(a_{2'} - a_3) + 2\mathcal{D}(a_2 - a_{3'}), \tag{3.20}$$

$$G_Y = (1 + \mathcal{D}^2)(a_{3'} - a_1) + 2\mathcal{D}(a_3 - a_{1'}), \tag{3.21}$$

$$G_Z = (1 + \mathcal{D}^2)(a_{1'} - a_2) + 2\mathcal{D}(a_1 - a_{2'}). \tag{3.22}$$

It is importante to note how the acceleration TDI variable G_X (G_Y, G_Z) is insensitive to glitches acting on links L_1 and L_1' (L_2 and L_2', L_3 and L_3'). This would no longer be true if a single glitch affects more than one TM (or more optical phase measurements); further modeling on this point will be presented elsewhere. Following the standard procedure [166], we combine G_X, G_Y, and G_Z into three noise-orthogonal variables

$$G_A = \frac{G_Z - G_X}{\sqrt{2}}, \tag{3.23}$$

$$G_E = \frac{G_X - 2G_Y + G_Z}{\sqrt{6}}, \tag{3.24}$$

$$G_T = \frac{G_X + G_Y + G_Z}{\sqrt{3}}. \tag{3.25}$$

Equations (3.23), (3.24), and (3.25) define the data pieces entering our inference pipeline.

ID	Injection		Recovery		$\log_{10}\mathcal{Z}$	Figure	Table
	Glitch	GW	Glitch	GW			
1	✗	✓ (acc. TDI)	✗	✓ (acc. TDI)	−35.27	✗	✗
2	✗	✓ (disp. TDI)	✗	✓ (disp. TDI)	−35.19	✗	✗
3	A1	✗	A1	✗	−14.0	✗	✗
4	A2	✗	A2	✗	−9.1	✗	✗
5	D	✗	D	✗	−8.8	✗	✗
6	A1	✓	✗	✓	−14537.8	Fig. 3.4	Tab. 3.3
7	A2	✓	✗	✓	−296.5	Fig. 3.5	Tab. 3.4
8	D	✓	✗	✓	−48.8	Fig. 3.6	Tab. 3.5
9	A1	✓	A1	✓	−46.8	Fig. 3.4	Tab. 3.3
10	A2	✓	A2	✓	−43.9	Fig. 3.5	Tab. 3.4
11	D	✓	D	✓	−40.8	Fig. 3.6	Tab. 3.5
12	✗	✓	A1	✓	−75.0	✗	✗
13	✗	✓	A2	✓	−44.1	✗	✗
14	✗	✓	D	✓	−52.2	✗	✗

Table 3.1: Summary of our runs containing single glitches and the fiducial GW signal. Rows highlighted in teal denote runs where the recovery signal model matches that of the injection. We first perform disjoint parameter estimation on our fiducial GW source and three glitch models (IDs 1–5). We then generate signals from the superposition of a GW signal with single glitches and study them both ignoring (IDs 6–8) and including (IDs 9–11) the glitch in the data model. Finally, we generate GW signals and perform parameter estimation on them including glitches in the data model (IDs 12–14).

ID	Injection				Recovery				$\log_{10} \mathcal{Z}$	Figure	Table
	Glitch1	Glitch2	Glitch3	Glitch4	Glitch1	Glitch2	Glitch3	Glitch4			
15	D(1,1)	D(1,2)	✗	✗	D(1,1)	D(1,2)	✗	✗	−16.1	✗	Tab. 3.7
16	✗	✗	✗	✗	D(1,1)	D(1,3)	✗	✗	−18.0	✗	✗
17	✗	✗	✗	✗	D(1,2)	D(1,3)	✗	✗	−20.1	✗	✗
18	✗	✗	✗	✗	D(1,1)	✗	✗	✗	−22.9	✗	✗
19	✗	✗	✗	✗	D(1,2)	✗	✗	✗	−23.9	✗	✗
20	✗	✗	✗	✗	D(1,3)	✗	✗	✗	−34.4	✗	✗
21	✗	✗	✗	✗	D(1,1)	D(1,2)	D(1,3)	✗	−17.0	✗	✗
22	D(1,1)	✗	✗	✗	D(1,1)	✗	✗	✗	−15.2	✗	Tab. 3.7
23	✗	✗	✗	✗	D(2,1)	✗	✗	✗	−3650.2	✗	✗
24	✗	✗	✗	✗	D(3,1)	✗	✗	✗	−3650.18	✗	✗
25	✗	✗	✗	✗	D(1′,1)	✗	✗	✗	−224.1	✗	✗
26	✗	✗	✗	✗	D(2′,1)	✗	✗	✗	−3628.6	✗	✗
27	✗	✗	✗	✗	D(3′,1)	✗	✗	✗	−3640.8	✗	✗
28	D(1,2)	D(1,3)	D(3,1)	D(3,2)	D(1,2)	D(1,3)	D(3,1)	D(3,2)	−34.8	Fig. 3.9	Tab. 3.7
29	A2(1,1)	✗	A2(2′,2)	✗	A2(1,1)	✗	A2(2′,2)	✗	−15.9	Fig. 3.8	Tab. 3.6
30	A1S(1)	✗	✗	✗	A1S(1)	✗	✗	✗	−13.6	Fig. 3.7	Tab. 3.6
31	A1M(1)	✗	✗	✗	A1M(1)	✗	✗	✗	−18.0	Fig. 3.7	Tab. 3.6
32	A1L(1)	✗	✗	✗	A1L(1)	✗	✗	✗	−16.6	Fig. 3.7	Tab. 3.6

Table 3.2: Summary of a large set of injected glitches and associated recoveries. Injected glitches are labeled by X(i,n), with X, n, and i describing the glitch model, the injection point, and the shapelet order (when applicable), respectively. We explore the number of components and shapelet order (IDs 16-20), the number of glitches (ID 21), and the potential misidentification of the injection point (IDs 22-27). Additionally, we simulate data from glitches of increasing complexity (IDs 28, 29) and consider three representative glitches inspired by LPF data (IDs 30-32). These are a short duration and small amplitude glitch (A1S), a medium duration and amplitude glitch (A1M), and a long duration and large amplitude glitch (A1L). Runs with same injected signals are grouped by horizontal lines.

3.5 Inference

The initial search of a GW in noisy data is achieved through matched-filtering techniques [129] which provide initial guesses on the signal parameters. If glitches are present, their preliminary detection and subtraction might not be sufficient to provide data that are sufficiently cleaned to accurately infer the parameters of the astrophysical source [26]. Previous studies presented a matching-pursuit algorithm for an automated and systematic glitch detection [115] showing that, while the search grid on the damping parameter is too coarse to accurately obtain the best-fit glitch, it provides a reliable initial guess. For practical purposes, here we assume that such guess has been identified from the data and can be used to inform our subsequent analyses.

We perform a joint parameter estimation, fitting simultaneously for GW signals and noise artifacts. We construct posteriors on parameters $\boldsymbol{\theta}$

$$p(\boldsymbol{\theta}|d) \propto \mathcal{L}(d|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \tag{3.26}$$

through stochastic sampling of the likelihood $\mathcal{L}(d|\boldsymbol{\theta})$ under a prior $\pi(\boldsymbol{\theta})$. We employ a coherent analysis on the three noise-orthogonal TDI channels $d = \{d_k; k = M_A, M_E, M_T\}$ when considering displacement variables and $d = \{d_k; k = G_A, G_E, G_T\}$ when considering acceleration variables. We use a Gaussian likelihood [58]

$$\ln\mathcal{L}(d|\boldsymbol{\theta}) = -\sum_k \frac{(d_k - s_k(\boldsymbol{\theta})|d_k - s_k(\boldsymbol{\theta}))_k}{2} + \text{const.}, \tag{3.27}$$

where $\tilde{s}_k$ is the k-th TDI output frequency series associated to the injected signal $\tilde{s}(f;\boldsymbol{\theta})$. The output $\tilde{s}_k$ represent either acceleration or fractional displacements depending on the chosen TDI variable set, thus containing acceleration glitches, displacement glitches, GW transients, or a combination of these (cf. Sec. 3.3). The noise-weighted inner product is defined as

$$(a \mid b)_k = 4\mathfrak{R}\int_{f_{\min}}^{f_{\max}} \frac{\tilde{a}^*(f)\tilde{b}(f)}{S_k(f)} df, \tag{3.28}$$

where $\mathfrak{R}$ denotes the real part, $\tilde{a}(f)$ is the Fourier transform of the time series $a(t)$, and $S_k(f)$ is the one-sided noise spectral density of the k-th TDI channel. We use the match between two signals

$$M(a,b) = \frac{(a \mid b)}{(a \mid a)^{1/2}(b \mid b)^{1/2}} \tag{3.29}$$

to optimize the onset time of the injected glitches as discussed in Sec. 3.3.3. Model

selection is performed using log-Bayes factors

$$\log_{10} \mathcal{B}_i^j = \log_{10} \mathcal{Z}_i - \log_{10} \mathcal{Z}_j, \tag{3.30}$$

where i and j are labels identifying the competing models, and

$$\mathcal{Z}(d) = \int d\boldsymbol{\theta} \mathcal{L}(d \mid \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) \tag{3.31}$$

is the evidence of each parameter estimation.

We consider a LISA mission lifetime of $T_{\rm LISA} = 4$ years, roughly equivalent to a calendar observation time of 4.5 years with an effective duty cycle of 82%. Our frequency resolution is therefore $\Delta f \approx 1/T_{\rm LISA} = 1.7 \times 10^{-8}\,\mathrm{Hz}$. We set $f_{\rm min} = 0.1\,\mathrm{mHz}$ and $f_{\rm max} = 4\,\mathrm{mHz}$, which is well above the fiducial GW and the maximum frequencies of all glitch signals. We use a semi-analytical noise spectral density model $S_k(f)$ [107] describing the superposition of LISA stationary instrumental noise and astrophysical confusion noise from unresolved Galactic binaries [25]. In order to reduce the computational cost, we evaluate inner products from Eq. (3.28) using a Clenshaw-Curtis integration algorithm [47], see e.g Ref. [43] for a summary of its application to LISA data.

Parameter estimation is performed with the Balrog code, which is designed to work with different stochastic samplers. In particular, in this paper we use the nested sampling algorithm [151] as implemented in Nessai [172]. We choose uniform priors on each parameter over either its entire definition domain or a range that is sufficiently large to enclose the entire posterior.

3.6 Results

We perform two sets of parameter-estimation runs:

(i) Joint inference runs on both GW signal and glitches (Sec. 3.6.1), listed with IDs 1 to 14 in Table 3.1;

(ii) Inference runs where we inject and recover glitches without GW signal (Sec. 3.6.2), listed with IDs 15 to 32 in Table 3.2.

	MBHB											Model A1			
ID	m_{1z} [$10^7 M_\odot$]	m_{2z} [$10^7 M_\odot$]	t_m [h]	χ_1	χ_2	d_L [Gpc]	ι [rad]	β [rad]	λ [rad]	ϕ [rad]	ψ [rad]	A [pm/s]	β_1 [s]	β_2 [s]	τ [h]
	4.5	1.5	30.0	0.4	0.3	47.6	0.6	0.30	2.0	1.0	1.7	3.0	1500.0	1800.0	30.21
9	$4.5^{+0.2}_{-0.2}$	$1.5^{+0.3}_{-0.3}$	$30.00^{+0.10}_{-0.08}$	$0.4^{+0.1}_{-0.1}$	$0.3^{+0.6}_{-1.0}$	44^{+15}_{-15}	$0.8^{+0.3}_{-0.6}$	$0.3^{+0.6}_{-0.1}$	$2.0^{+0.3}_{-0.1}$	$1.6^{+1.4}_{-1.3}$	$1.6^{+1.3}_{-1.4}$	$3.0^{+0.2}_{-0.1}$	1637^{+594}_{-417}	1647^{+584}_{-426}	$30.219^{+0.006}_{-0.006}$
6	$4.25^{+0.01}_{-0.01}$	$0.610^{+0.007}_{-0.007}$	$29.447^{+0.007}_{-0.002}$	$-0.308^{+0.005}_{-0.005}$	$-0.51^{+0.09}_{-0.09}$	$10.010^{+0.032}_{-0.009}$	$0.04^{+0.05}_{-0.03}$	$0.150^{+0.005}_{-0.005}$	$1.792^{+0.005}_{-0.005}$	$1.6^{+1.4}_{-1.4}$	$1.6^{+1.4}_{-1.4}$	✗	✗	✗	✗
	4.5	1.5	30.0	0.4	0.3	47.6	0.6	0.30	2.0	1.0	1.7	✗	✗	✗	✗
1	$4.5^{+0.2}_{-0.2}$	$1.5^{+0.3}_{-0.3}$	$30.01^{+0.09}_{-0.08}$	$0.4^{+0.1}_{-0.1}$	$0.3^{+0.6}_{-1.0}$	44^{+15}_{-14}	$0.8^{+0.3}_{-0.6}$	$0.3^{+0.6}_{-0.1}$	$2.0^{+0.2}_{-0.1}$	$1.6^{+1.5}_{-1.3}$	$1.6^{+1.3}_{-1.4}$	✗	✗	✗	✗

Table 3.3: Parameter estimation results for a GW signal contaminated by a Model A1 glitch. The injected parameters are listed in the white rows. Medians and 90% credible intervals for the recovered posteriors are listed in the two rows highlighted in teal. While accounting for the presence of a glitch (ID 9) allows for joint unbiased reconstruction of all parameters, ignoring its potential occurrence (ID 6) yields large systematic biases. Ignoring the presence of a glitch is disfavored with $\log_{10} \mathcal{B}^6_9 = -14491$. Joint posterior distributions for both these runs are shown in Fig. 3.4. For comparison, the bottom row shows our reference run where we only inject the GW source (ID 1). The subset of parameters common across runs 1 and 9 does not show appreciable differences.

	MBHB											Model A2		
ID	m_{1z} [$10^7 M_\odot$]	m_{2z} [$10^7 M_\odot$]	t_m [h]	χ_1	χ_2	d_L [Gpc]	ι [rad]	β [rad]	λ [rad]	ϕ [rad]	ψ [rad]	A [pm/s]	β [s]	τ [h]
	4.5	1.5	30.0	0.4	0.3	47.6	0.6	0.30	2.0	1.0	1.7	0.0948	40.0	29.48
10	$4.5^{+0.2}_{-0.2}$	$1.5^{+0.3}_{-0.3}$	$30.01^{+0.10}_{-0.08}$	$0.4^{+0.1}_{-0.1}$	$0.3^{+0.6}_{-1.0}$	44^{+15}_{-15}	$0.8^{+0.3}_{-0.6}$	$0.3^{+0.6}_{-0.1}$	$2.0^{+0.2}_{-0.1}$	$1.6^{+1.4}_{-1.3}$	$1.6^{+1.3}_{-1.4}$	$0.095^{+0.007}_{-0.007}$	41^{+22}_{-23}	$29.48^{+0.01}_{-0.01}$
7	$4.4^{+0.3}_{-0.1}$	$1.4^{+0.2}_{-0.5}$	$29.85^{+0.06}_{-0.06}$	$0.26^{+0.16}_{-0.04}$	$0.95^{+0.05}_{-1.77}$	36^{+12}_{-12}	$0.8^{+0.3}_{-0.6}$	$0.3^{+0.4}_{-0.1}$	$1.95^{+0.29}_{-0.07}$	$1.5^{+1.4}_{-1.3}$	$1.6^{+1.3}_{-1.4}$	✗	✗	✗
	4.5	1.5	30.0	0.4	0.3	47.6	0.64	0.30	2.0	1.0	1.7	✗	✗	✗
1	$4.5^{+0.2}_{-0.2}$	$1.5^{+0.3}_{-0.3}$	$30.01^{+0.09}_{-0.08}$	$0.4^{+0.1}_{-0.1}$	$0.3^{+0.6}_{-1.0}$	44^{+15}_{-14}	$0.8^{+0.3}_{-0.6}$	$0.3^{+0.6}_{-0.1}$	$2.0^{+0.2}_{-0.1}$	$1.6^{+1.5}_{-1.3}$	$1.6^{+1.3}_{-1.4}$	✗	✗	✗

Table 3.4: Parameter estimation results for a GW signal contaminated by a Model A2 glitch. Results are organized as in Table 3.3. This glitch, if present in data and ignored upon inference, introduces milder biases when compared to run with ID 6: this is due to its shorter duration resulting in a smaller match with the GW waveform. Joint posterior distributions for both runs are shown in Fig. 3.5.

	MBHB											Model D		
ID	m_{1z} [$10^7 M_\odot$]	m_{2z} [$10^7 M_\odot$]	t_m [h]	χ_1	χ_2	d_L [Gpc]	ι [rad]	β [rad]	λ [rad]	ϕ [rad]	ψ [rad]	D [pm·s]	β [s]	τ [h]
	4.5	1.5	30.0	0.4	0.3	47.6	0.6	0.3	2.0	1.0	1.7	200.0	5.0	29.507
11	$4.5^{+0.2}_{-0.2}$	$1.5^{+0.3}_{-0.3}$	$30.01^{+0.09}_{-0.08}$	$0.4^{+0.1}_{-0.1}$	$0.3^{+0.6}_{-1.0}$	44^{+15}_{-14}	$0.8^{+0.3}_{-0.6}$	$0.3^{+0.5}_{-0.1}$	$1.99^{+0.25}_{-0.09}$	$1.6^{+1.4}_{-1.3}$	$1.6^{+1.3}_{-1.4}$	216^{+83}_{-73}	16^{+13}_{-14}	$29.502^{+0.009}_{-0.008}$
8	$4.5^{+0.2}_{-0.2}$	$1.5^{+0.3}_{-0.3}$	$30.01^{+0.09}_{-0.08}$	$0.4^{+0.1}_{-0.1}$	$0.3^{+0.6}_{-1.0}$	44^{+15}_{-15}	$0.8^{+0.3}_{-0.6}$	$0.3^{+0.5}_{-0.1}$	$1.99^{+0.25}_{-0.10}$	$1.5^{+1.5}_{-1.4}$	$1.6^{+1.3}_{-1.4}$	✗	✗	✗
	4.5	1.5	30.0	0.4	0.3	47.6	0.6	0.3	2.0	1.0	1.7	✗	✗	✗
2	$4.5^{+0.2}_{-0.2}$	$1.5^{+0.3}_{-0.3}$	$30.01^{+0.09}_{-0.08}$	$0.4^{+0.1}_{-0.1}$	$0.3^{+0.6}_{-1.0}$	44^{+15}_{-14}	$0.8^{+0.3}_{-0.6}$	$0.3^{+0.5}_{-0.1}$	$1.99^{+0.25}_{-0.09}$	$1.5^{+1.5}_{-1.3}$	$1.6^{+1.3}_{-1.4}$	✗	✗	✗

Table 3.5: Parameter estimation results for a GW signal contaminated by a Model D glitch. Results are organized as in Table 3.3. This glitch, if present in data and ignored upon inference, introduces negligible biases when compared to runs with IDs 6 and 7. This is due to its very short duration, which superimposes with the GW signal only for a few seconds yielding a low match. Joint posterior distributions for both runs are shown in Fig. 3.6.

3.6.1 Joint inference with glitches and GWs

If a preliminary search fails to identify and remove a glitch from the data, it is important to assess its impact on the parameters of the overlapping GW source. We thus tackle the following cases for each of the three signals illustrated in Fig. 3.3:

- Parameter estimation in the absence of glitch in the data ("reference" runs, with IDs 1 and 2);
- Parameter estimation ignoring a glitch when present in the data ("glitch-ignorant" runs, with IDs 6-8);
- Parameter estimation including in the signal model a glitch that is present in the data ("glitch-complete" runs, with IDs 9-11).

Bayesian evidence for each run is listed Tab. 3.1. We report $\log_{10}\mathcal{B}^{6}_{9}$, $\log_{10}\mathcal{B}^{7}_{10}$, and $\log_{10}\mathcal{B}^{8}_{11}$ much greater than 2, indicating a "decisive" evidence [95] in favor of a glitch being present in the data.

Summaries are provided in Tables 3.3, 3.4, and 3.5. We find no appreciable differences in the posterior distribution of the GW-source parameters when comparing reference runs and glitch-complete runs, which is encouraging for LISA science. Individual parameters are well reconstructed, which is expected given the brightness of the source (SNR $\simeq$ 187). In particular, the MBHB component masses, the primary aligned spin component, and time to merger are measured with an accuracy of $\Delta m_i/m_i \approx 8-40\%$, $\Delta\chi_1 \approx 0.2$, and $\Delta t_m \approx$ 600 s (where we quote the 90% credible interval of the marginal posterior distributions). Figures 3.4, 3.5, and 3.6 show the posterior distribution for the fiducial MBHB of each glitch-complete run. Similarly, we do not report any appreciable difference with either fractional displacement or acceleration TDIs to model the same GW signal (see runs 1 and 2).

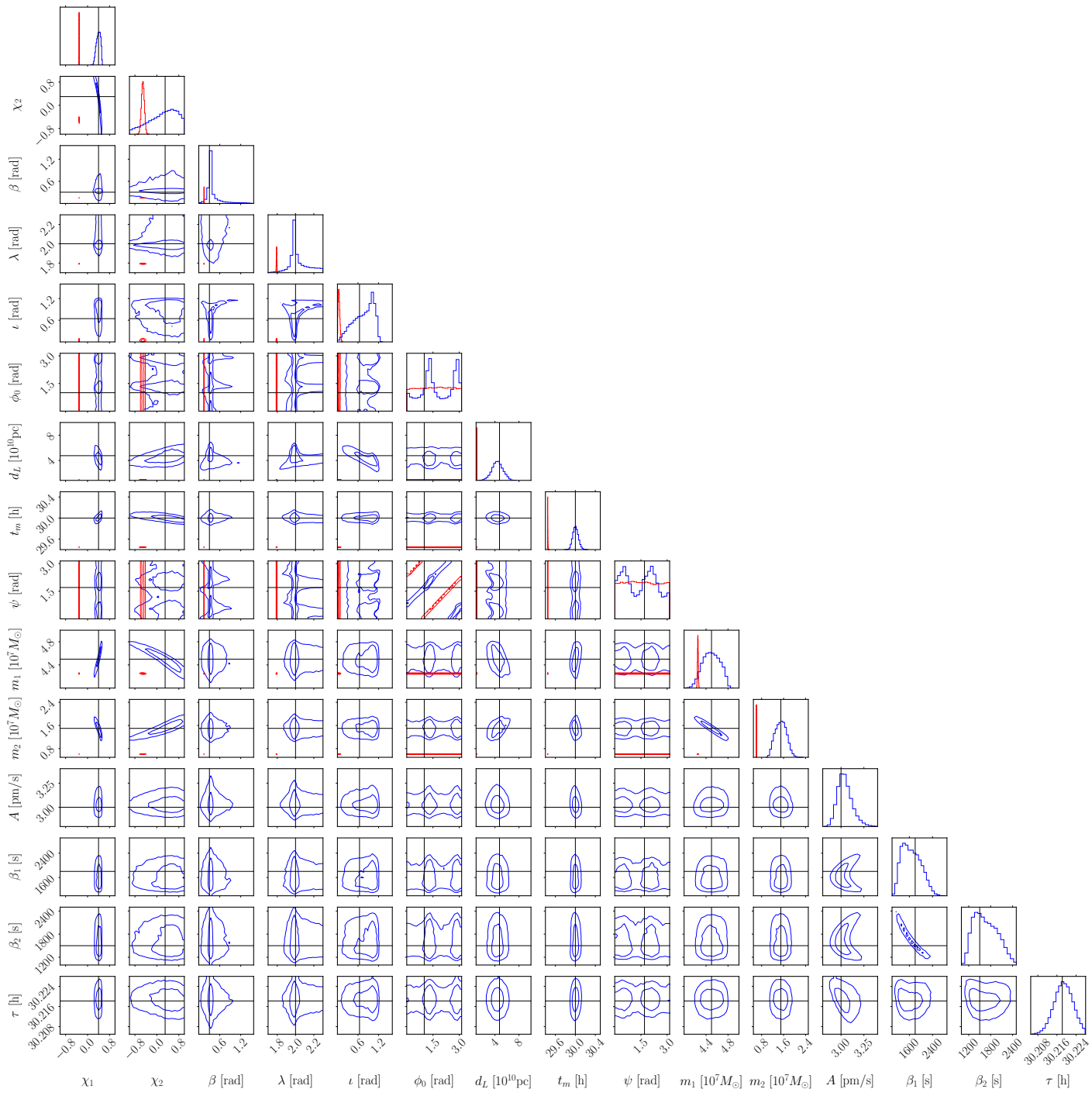

Figure 3.4: Posterior distribution in blue (red) corresponding to run ID 9 (6) where a Model A1 glitch is (not) included in the recovery process. Contours indicate the 50% and 90% credible regions; solid black lines indicate the injected values as listed in Table 3.3. When the glitch is included in the inference, each model injected parameter is recovered within the 90% one-dimensional credible region. We do not report notable correlations between glitch and GW parameters. If the glitch is excluded, all MBHB parameters except the initial phase ϕ_0 and the polarization angle ψ are systematically biased. In particular, the posterior on the luminosity distance d_L rails heavily against the prior lower bound.

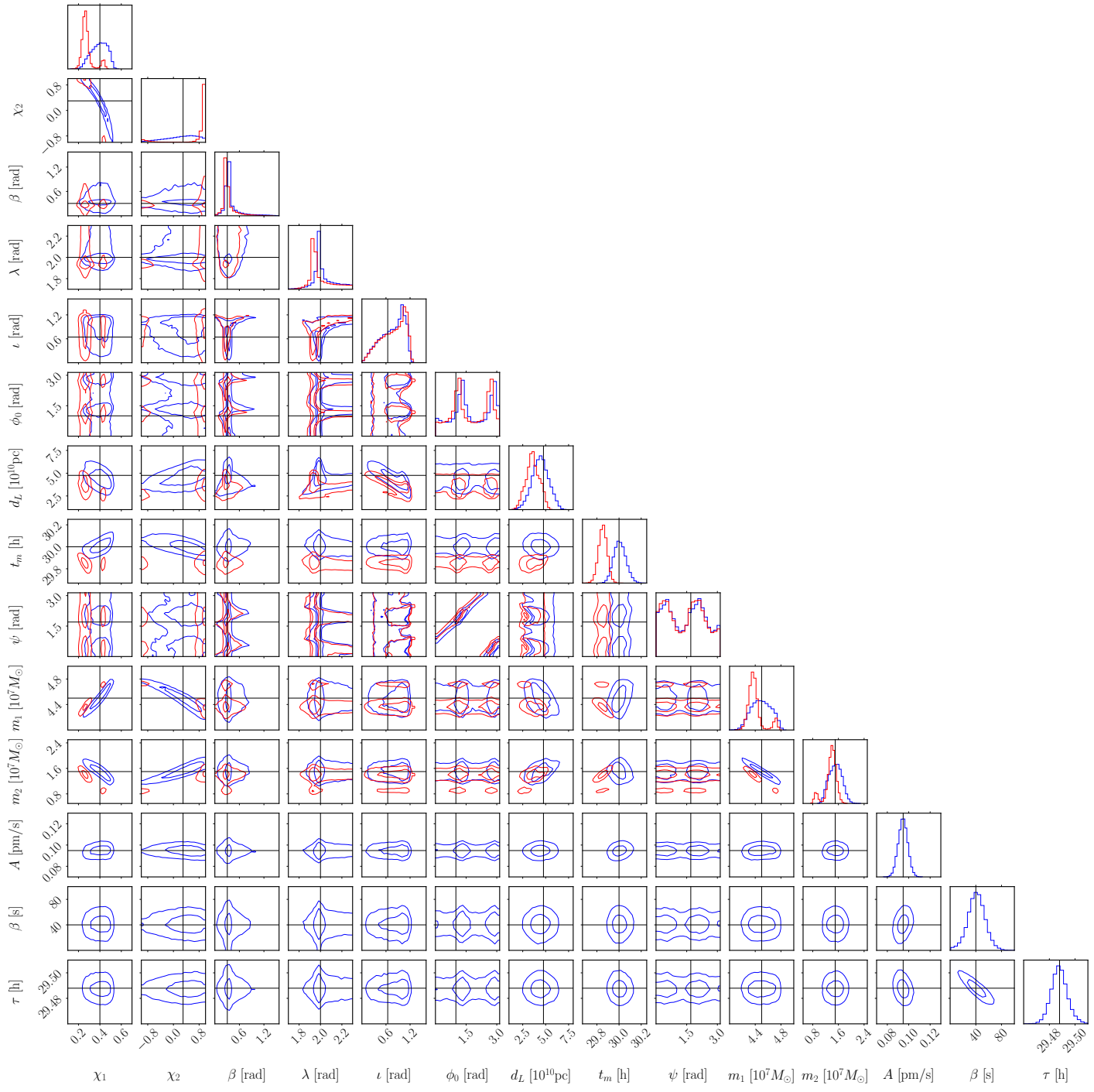

Figure 3.5: Posterior distribution in blue (red) corresponding to run ID 10 (7) where a Model A2 glitch is (not) included in the recovery process. Contours indicate the 50% and 90% credible regions; solid black lines indicate the injected values as listed in Table 3.4. When the glitch is ignored, the MBHB parameters are somewhat biased, in particular the black-hole masses and spins. When the glitch is included in the recovery process, all model parameters are recovered within their 90% one-dimensional credible regions. We do not report notable correlations between glitch and GW parameters.

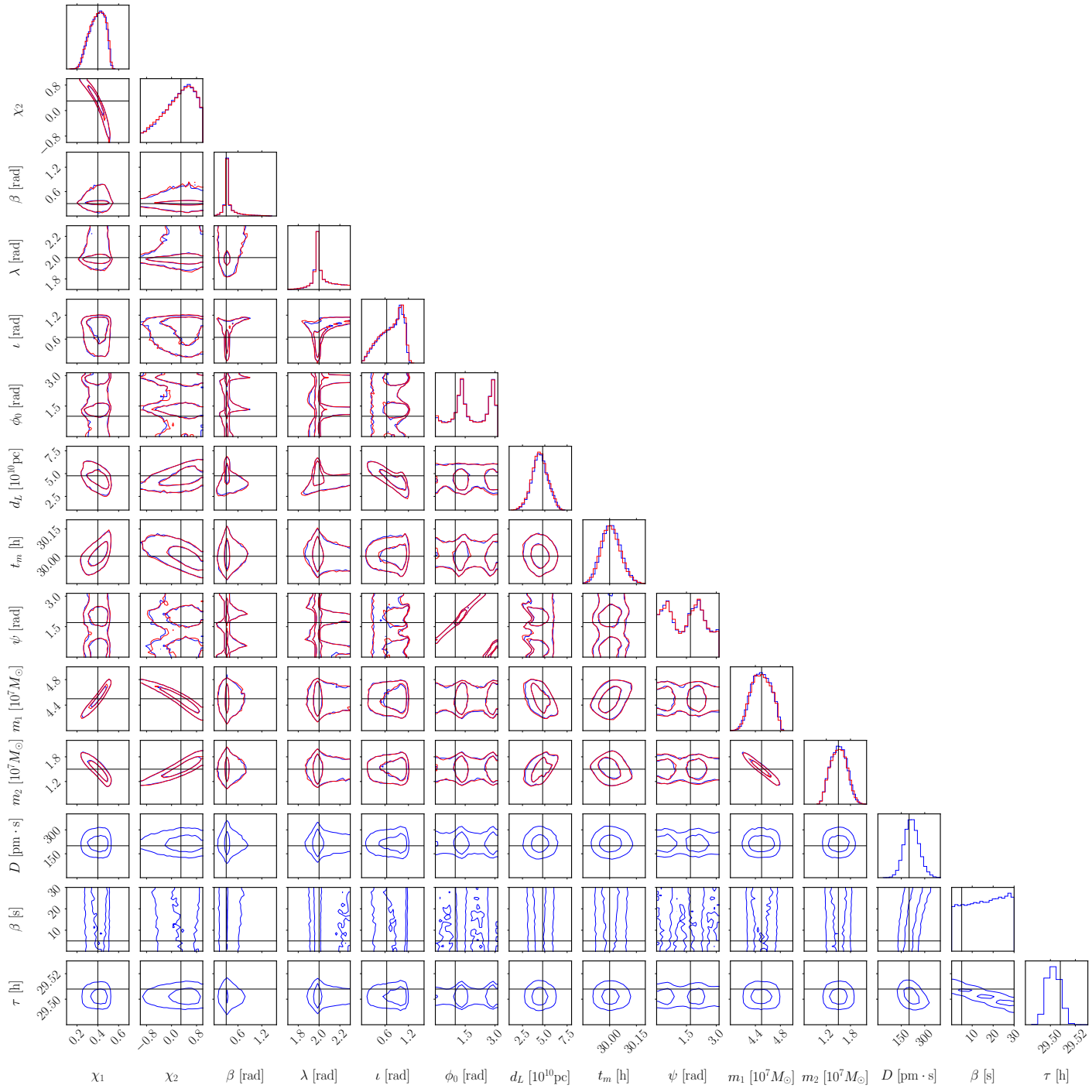

Figure 3.6: Posterior distribution in blue (red) corresponding to run ID 11 (8) where a Model D glitch is (not) included in the recovery process. Contours indicate the 50% and 90% credible regions; solid black lines indicate the injected values as listed in Table 3.5. When the glitch is ignored, the MBHB parameters are only very mildly biased. In both cases, all model parameters are recovered within their 90% one-dimensional credible regions.

On the contrary, glitch-ignorant runs point to a different conclusion. The resulting posterior depends on the chosen duration and amplitude of each transient (see runs 6, 7,

and 8). We find a long-duration, small-amplitude Model A1 glitch massively contaminates the reconstruction of the GW parameters, to a point that the signal cannot be recovered at all. This is shown in Fig. 3.4, where the glitch-ignorant distribution (red) shows evident issues in the underlying stochastic-sampling procedure. This has to be contrasted with the regularity of the glitch-complete posterior distribution (blue), where instead the parameters of both GW signal and noise transient are successfully recovered. In particular, when the glitch is ignored we find that the posterior on the luminosity distance rails heavily against the lower bound of its prior, thus making the GW source reconstruction highly biased, even in a parameter space that largely encloses the posterior of the glitch-complete run.

As shown in Fig. 3.5, a Model A2 glitch with moderate duration and amplitude induces milder biases. Although the posterior support is far from the prior boundaries, the injected values lie outside the 99% credible interval for both mass and spin parameters. For the merger time, the true value lies on the 97% confidence interval of the corresponding marginalized posterior distribution. The injected values of polarization, initial phase, inclination, and source position are within their one-dimensional 90% confidence interval.

Equivalent runs for a Model D glitch are shown in Fig. 3.6. This is a noise transient that overlaps with the GW signal only for a small fraction of a cycle. As expected, we find such a glitch does not significantly impact the measurement of the GW parameters.

Finally, we note that glitch-complete runs do not exhibit significant cross-correlations between the glitch and GW parameters, thus effectively decoupling the inference on the two signals.

	Model A1				Model A2					
ID	A [pm/s]	β_1 [s]	β_2 [s]	τ [h]	A_0 [pm/s]	β_0 [s]	τ_0 [h]	A_1 [pm/s]	β_1 [s]	τ_1 [h]
	✗	✗	✗	✗	1.48	3600.0	11.94	3.72	3600.0	36.94
29	✗	✗	✗	✗	$1.6^{+0.6}_{-0.4}$	3735^{+770}_{-543}	$11.94^{+0.03}_{-0.03}$	$4.2^{+2.4}_{-1.4}$	3848^{+993}_{-719}	$36.93^{+0.04}_{-0.04}$
	0.3	21.0	20.0	12.0	✗	✗	✗	✗	✗	✗
30	$0.300^{+0.004}_{-0.004}$	20^{+13}_{-17}	20^{+13}_{-17}	$12.001^{+0.003}_{-0.003}$	✗	✗	✗	✗	✗	✗
	2.0	900.0	400.0	12.0	✗	✗	✗	✗	✗	✗
31	$2.00^{+0.02}_{-0.02}$	439^{+485}_{-55}	848^{+75}_{-465}	$12.000^{+0.002}_{-0.002}$	✗	✗	✗	✗	✗	✗
	100.0	7500.0	7400.0	12.0	✗	✗	✗	✗	✗	✗
32	102^{+5}_{-4}	7453^{+2211}_{-1417}	7453^{+2201}_{-1402}	$12.000^{+0.002}_{-0.002}$	✗	✗	✗	✗	✗	✗

Table 3.6: Parameter-estimation results on Model A1 (ID 30-32) and Model A2 (ID 29) glitches. In particular, the former corresponds to glitches inspired by LPF observations, with varying duration and amplitudes. White rows show the injected values and teal rows show the recovered median and 90% confidence interval. The posterior distribution for these runs is provided in Fig. 3.7 and 3.8.

	Glitch 1						Glitch 2					
	Component 1			Component 2			Component 1			Component 2		
ID	D_0 [pm · s]	β_0 [s]	τ_0 [h]	D_1 [pm · s]	β_1 [s]	τ_1 [h]	D_2 [pm · s]	β_2 [s]	τ_2 [h]	D_3 [pm · s]	β_3 [s]	τ_3 [h]
	2480.0	20.0	12.0	✗	✗	✗	✗	✗	✗	✗	✗	✗
22	2481^{+64}_{-64}	$20.0^{+0.9}_{-0.9}$	$12.0000^{+0.0003}_{-0.0004}$	✗	✗	✗	✗	✗	✗	✗	✗	✗
	542.0	40.0	12.0	1420.0	80.0	12.0	✗	✗	✗	✗	✗	✗
15	672^{+536}_{-278}	69^{+60}_{-29}	$11.997^{+0.005}_{-0.007}$	1336^{+690}_{-578}	77^{+16}_{-15}	$12.015^{+0.010}_{-0.012}$	✗	✗	✗	✗	✗	✗
	5000.0	100.0	11.111	1000.0	10.0	11.111	5000.0	40.0	13.89	20000.0	120.0	13.89
28	5021^{+1461}_{-1355}	101^{+25}_{-26}	$11.111^{+0.004}_{-0.003}$	986^{+489}_{-333}	$10.2^{+2.2}_{-1.9}$	$11.111^{+0.003}_{-0.004}$	4975^{+820}_{-821}	$40.3^{+3.8}_{-3.5}$	$13.889^{+0.001}_{-0.002}$	19822^{+3942}_{-3998}	$120.2^{+8.4}_{-7.6}$	$13.89^{+0.02}_{-0.02}$

Table 3.7: Parameter estimation results assuming Model D glitches of increasing complexity. White rows show the injected values and teal rows show the recovered median and 90% confidence interval. In particular, we consider a single-component glitch (ID 22), a glitch with two components (ID 15), and two glitches separated by 200 seconds with two components each (ID 28). The posterior distribution for the latter, most complex case is shown in Fig 3.9.

3.6.2 Inference with glitches alone, without GWs

We consider all three glitch models presented in Sec. 3.3 and inject them separately in the LISA data stream. Results are shown in Figs. 3.7, 3.8, and 3.9 as well as Tables 3.6 and 3.7.

We perform model selection with different (i) number and order of shapelet components, (ii) number of glitches, and (iii) injection point. In particular, in Tab. 3.2 we report "strong" evidence in favor of the correct noise-transient model for the selection of the number and order of shapelets; these are discrete parameters we can confidently identify

using $\log_{10} \mathcal{B}^{j}_{15}$ with $j = 16, \ldots, 20$. We obtain a "substantial" evidence $\log_{10} \mathcal{B}^{21}_{15} = 0.9$ for selecting the correct number of glitches. Injection points are selected with a "decisive" evidence given by $\mathcal{B}^{n}_{22}$ with $n = 23, \ldots, 27$.

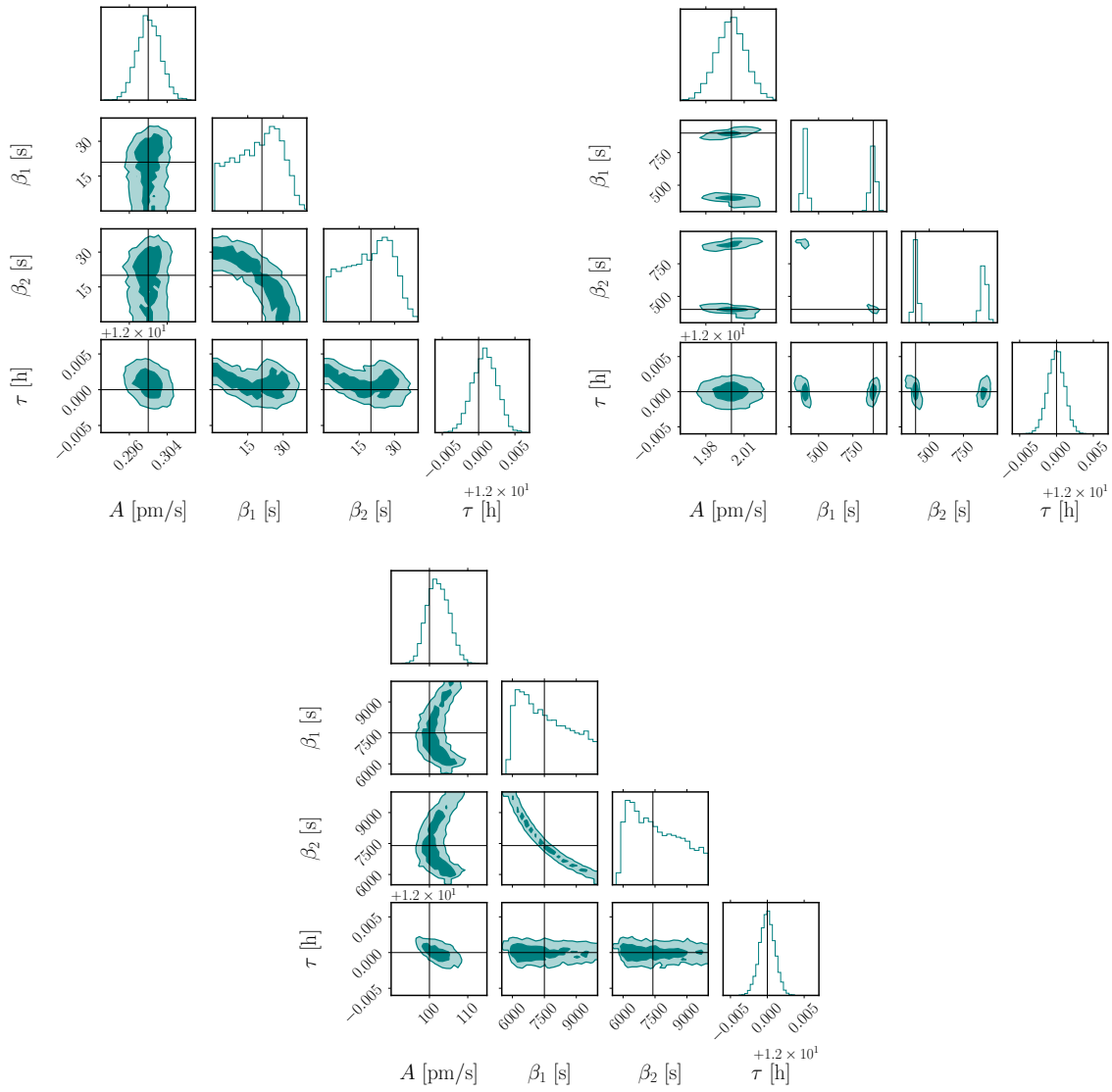

Figure 3.7: *Top Left*: Posterior distribution for short Model A1 glitch (ID 30). Darker (lighter) shaded areas indicate 90% (50%) credible regions and solid black lines denote the injected values. The correlation between the fall time β_1 and the rise time β_2 reflects their intrinsic degeneracy (see Eqs. (3.2) and (3.4)). Injected values and some posterior summary statistics are listed in Table 3.6. *Top Right*: Posterior distribution for medium Model A1 glitch (ID 31). The larger separation between the injected value of $\beta_1 - \beta_2 = 500$s partially breaks it into a strong multimodality. Injected values and some posterior summary statistics are listed in Table 3.6. *Bottom*: Posterior distribution for a long Model A1 glitch (ID 32). Injected values and some posterior summary statistics are listed in Table 3.6.

All runs point to the same, encouraging result: glitch parameters are confidently

reconstructed. In particular, we recover amplitudes across all models (i.e. A, $A_{0,1}$, $D_{0,1,2,3}$) with accuracies of $1\% - 30\%$ at 90% credible level. Glitch-onset times are recovered with fractional accuracy $\lesssim 0.1\%$. The parameters β_i's in Model D glitches are recovered with an accuracy of 20%. On the other hand, Model A1 glitches exhibit correlation and multimodalities for the joint posterior on β_1 and β_2. This is expected given the waveform degeneracy upon exchange of these two parameters, cf. Eqs. (3.2) and Eq. (3.4).

Finally, the joint inference strategy adopted here has also been explored using trans-dimensional stochastic samplers. In particular, Ref. [127] performs joint inference of GW signals and glitches within a reversible-jump MCMC framework that allows the number of glitches to vary, demonstrating the feasibility of this strategy within the broader LISA global-fit problem. An alternative strategy is to employ heavy-tailed likelihood functions, which provide robustness against non-Gaussian noise by mitigating the impact of large residuals that may otherwise bias the signal reconstruction [146, 147]. Rather than explicitly modeling glitches, heavy-tailed likelihoods account for data outliers within the statistical noise model and can be viewed as complementary to global-fit approaches.

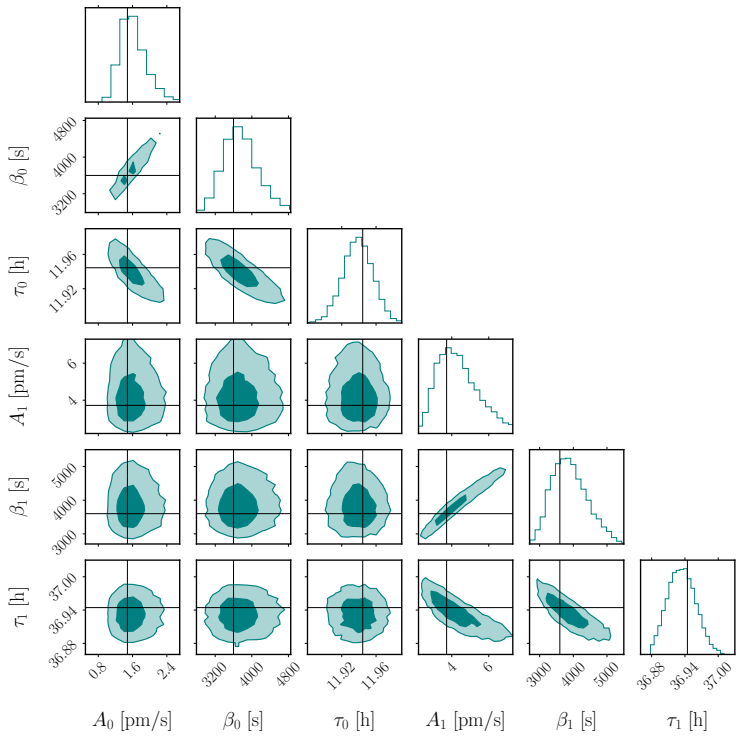

Figure 3.8: Posterior distributions for two Model A2 glitches (run ID 29). Injected values and some posterior summary statistics are listed in Table 3.6. Darker (lighter) shaded areas indicate 90% (50%) credible regions and solid black lines indicate the injected values. The lower-left panels show the joint distribution between parameters describing the two glitches, which do not present significant correlations.

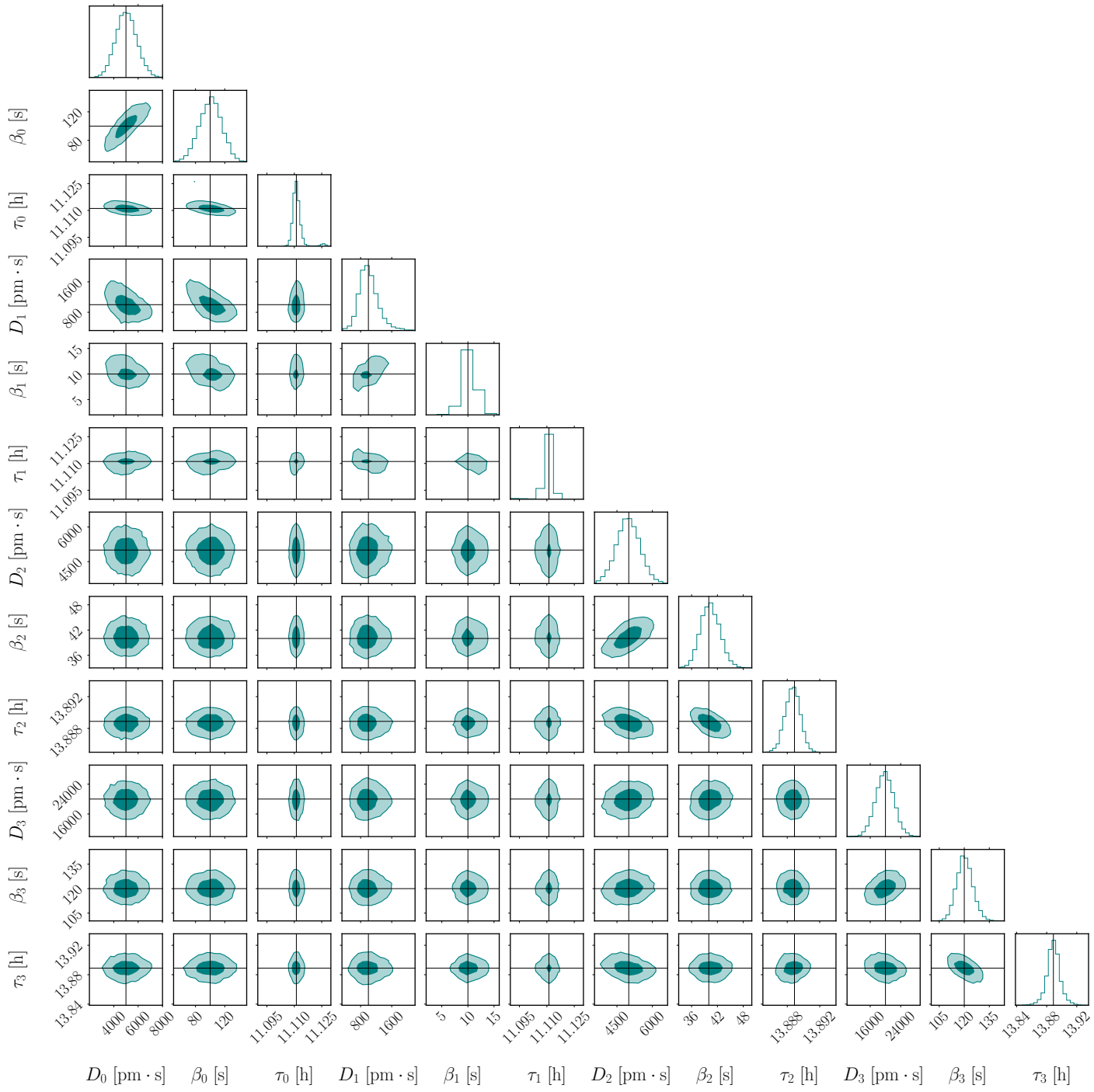

Figure 3.9: Posterior distributions for two Model D glitches (run ID 28). Injected values and selected posterior summary statistics are listed in Table 3.7. Each glitch is composed of two components, with injected values $\tau_0 = \tau_1$ and $\tau_2 = \tau_3$ for the first and second glitch, respectively. Darker (lighter) shaded areas indicate 90% (50%) credible regions, and solid black lines denote the injected values. Glitch parameters are successfully recovered, and cross-glitch correlations are negligible.

4 Accurate and efficient simulation-based inference for massive black-hole binaries with LISA

Overview

We develop an accurate simulation-based inference framework for high-mass ($\gtrsim 10^7 M_\odot$) black-hole binaries observable by LISA. The method is implemented within the Dingo *gravitational-wave parameter-estimation code, extending its application from ground-based detectors to the LISA band. We train a normalizing-flow model using aligned-spin higher-mode waveform models and a low-frequency approximation of the detector response. After sampling, we importance-sample to the true posterior. We validate performance on simulated signals spanning the signal-to-noise regimes relevant for LISA observations and benchmark our new* Dingo *implementation against standard methods. We report robust agreement in the inferred posterior distributions up to signal-to-noise ratios of* ~ 500*. At higher signal-to-noise ratios of* ~ 1000*, we observe a reduction in sampling efficiency, while still yielding unbiased and tightly localized posteriors that can be used as a starting point for follow-up with traditional methods. The trained flow can generate 20 thousand posterior samples in less than a minute, establishing* Dingo *as a promising neural inference framework for rapid full-parameter estimation of massive black-hole binaries in the LISA band. The likelihood-free nature of this approach allows for straightforward generalizations, including a time-dependent detector response, non-stationary noise artifacts such as gaps and glitches, and low-latency parameter estimations.*

4.1 Executive summary

This Chapter is organized as follows. In Sec. 4.2, we present the LISA targeted sources, describe the adopted detector response, and detail the neural posterior estimation methodology used in this analysis. In Sec. 4.3, we present and discuss our results, and briefly compare them with previous work. Finally, in Sec. 4.4, we examine the implications of our findings and outline the main challenges and limitations of the proposed approach.

The LISA version of the Dingo code is under active development. Progress can be tracked at a dedicated branch of the main Dingo repository: github.com/AliSword/dingo-gw/dingo/tree/dingo-lisa.

4.2 Methods

Among SBI-based approaches, the Dingo software has been extensively validated in GW astronomy. Most notably, Ref. [60] demonstrated fast and accurate parameter inference for compact binary signals observed by ground-based detectors. Subsequent works have extended the framework to neutron-star binaries [59], explored its performance for third-generation detectors [145], and more recently generalized it through transformer-based architectures [101]. In this work, we adapt Dingo to perform neural posterior estimation on quasi-circular, non-precessing massive BH binaries observed by LISA.

In particular, our analysis targets the most massive BH binary systems observable by LISA, with redshifted total masses $M \gtrsim 10^7 \mathrm{M}_\odot$ and signal-to-noise-ratios (SNRs) ranging from ~ 10 to ~ 1000. Owing to the mission low-frequency cut-off at $f_{\mathrm{low}} = 0.1$ mHz [49], these binaries sweep through the LISA band over relatively short timescales, $\tau \sim \mathcal{O}(\mathrm{hours} - \mathrm{days})$. As a consequence, LISA is expected to observe only the late inspiral, merger, and ringdown, spanning the final few to few tens of GW cycles, $N_{\mathrm{GW}} \sim \tau f_{\mathrm{low}}$. The resulting signals exhibit a morphology similar to that of the transient events detected by the current ground-based observatories LIGO and Virgo. As such, the analysis of massive BH binaries observed by LISA can be naturally framed within the existing Dingo framework. In the following, we describe the LISA response model implemented within Dingo (Sec. 4.2.1) and outline the corresponding training setup (Sec. 4.2.2).

4.2.1 LISA response

The observables of space-based GW interferometers are obtained through time-delay interferometry (TDI), which is a post-processing technique that reduce the overwhelming laser frequency noise in the phase measurements [166]. In this work, we model the LISA

response within the low-frequency approximation [57], which provides a simplified yet well-motivated description for the high-mass binary signals produced by the sources considered here. Accordingly, LISA is modeled as a rigid triangular constellation with equal arm lengths, neglecting both the time and frequency dependence of the response [52]. Under these assumptions, the constellation is effectively treated as fixed along its orbit and the TDI variables are approximated by two noise orthogonal data channels (commonly denoted A and E), whose GW response is equivalent to that of two co-located, motionless Michelson interferometers rotated with respect to each other by an angle of $\pi/4$.

We assume the noise in the two channels to be identical, stationary, and Gaussian, with statistical properties fully described by a power spectral density (PSD) $S_n(f)$. We adopt the SciRDv1 noise model [107] simulated using `lisatools` [96], while neglecting the confusion noise arising from the superposition of unresolved Galactic binaries (see e.g. Ref. [44] and references therein). The inclusion of this additional noise component, characterized by non-Gaussian and non-stationary features, is a natural direction for future SBI studies. The instrument response is implemented within the Dingo framework in close correspondence with the formalism of Ref. [57].

4.2.2 Dingo setup

We perform Bayesian neural posterior estimation in the frequency domain between $f_{\min} = 0.1\,\mathrm{mHz}$ and $f_{\max} = 4\,\mathrm{mHz}$, with a resolution of $\Delta f = 10^{-5}\,\mathrm{Hz}$. As shown in Ref. [120], the frequency dependence of the response vanishes in the limit $f \ll f_L \sim 0.19\,\mathrm{Hz}$, where $f_L = c/(2\pi L)$ is the detector transfer frequency and L the LISA arm length. Although departures from the low-frequency approximation become noticeable around 2 mHz, most of the signal content for the systems considered here lies at lower frequencies, so that the approximation remains accurate over the frequency range analyzed. We perform inference over the full 11-dimensional parameter space of quasi-circular spin-aligned compact binaries

$$\theta = \{\mathcal{M}_c, q, \chi_{1,2}, d_L, \theta_{JN}, \theta_S, \phi_S, \psi, t_c, \phi_c\},$$

where $\mathcal{M}_c$ is the redshifted chirp mass, q is the mass ratio, $\chi_{1,2}$ are the dimensionless spin components aligned with the orbital angular momentum, d_L is the luminosity distance, θ_{JN} is the inclination angle between the total angular momentum and the line of sight, θ_S is the ecliptic colatitude, ϕ_S is the ecliptic longitude, ψ is the polarization angle, t_c is the coalescence time, and ϕ_c is the coalescence phase. All extrinsic parameters are defined in the frame of the Solar System barycenter.

For density estimation, we employ a conditional normalizing flow based on a rational-quadratic spline coupling transforms [72], adopting the same network architecture as in Ref. [60]. In addition to the flow network, DINGO incorporates an embedding network that compresses the input strain data into a 128-dimensional feature vector for each detector channel. Together, these two networks comprise approximately 361 millions learnable parameters.

We train the neural network on simulated data generated with the IMRPHENOMXHM waveform approximant [78], including all available emission modes and additive stationary Gaussian noise realizations sampled from the fixed PSD. The training dataset includes 10^7 waveforms covering the intrinsic-parameter space, while the extrinsic parameters are generated on-the-fly. Parameters θ are drawn as follows: chirp mass $\mathcal{M}_c \in [8 \times 10^6, 2 \times 10^7]\,\mathrm{M}_\odot$ and mass ratio $q \in [0.2, 1]$ are derived from components masses sampled uniformly in the range $m_1, m_2 \in [7 \times 10^6, 3 \times 10^7]\,\mathrm{M}_\odot$, subject to the constraint $m_1 > m_2$. This choice ensures that the signal remains within the validity range of the low-frequency response model discussed above. Following Ref. [85], we adopt a uniform prior in luminosity distance, $d_L \in [15, 50]$ Gpc, which improves training performance, while ensuring coverage of the target SNR range across the mass interval considered. Dimensionless spin magnitudes are drawn from a uniform distribution

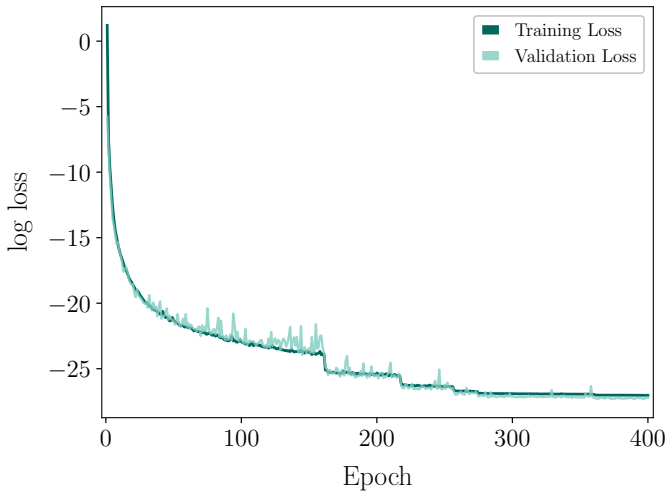

Figure 4.1: Evolution of the training (green) and validation (teal) losses as a function of epoch. The drops at epochs ∼160, 220, 255, 275, and 360 are due to learning rate reductions triggered by the `ReduceLROnPlateau` scheduler [135].

$a_{1,2} \in [0, 0.99]$, while spin orientations are assumed to be isotropic and marginalized over the planar components. The resulting prior on the aligned dimensionless spins $\chi_{1,2}$ peaks at zero, as shown in Fig. 4.6. Standard uninformative priors are adopted for all remaining angular parameters. All training signals are evaluated at a fixed reference time $t_{\mathrm{ref}} = 1\,\mathrm{yr}$,

so that the neural network does not need to learn the modulation induced by LISA's heliocentric motion. As discussed in Ref. [103], correcting for detector motion would be straightforward. Finally, we adopt a uniform prior on the coalescence time t_c within 30 minutes around the reference time $t_{\rm ref}$, implicitly assuming that a sufficiently accurate trigger time is supplied by a preceding low-latency search pipeline. The short runtime of our method (see Sec. 4.3.2 for details) allows continuous processing of consecutive data segments as an alternative to a trigger-based approach, provided the model remains uninformative in the absence of a signal.

The network is trained for 400 epochs with a batch size of 4096. We use the Adam optimizer [99] with an initial learning rate of 2×10^{-4}, which is reduced by a factor of 0.4 when the validation loss stagnates for 15 epochs. As shown in Fig. 4.1, the training and validation losses closely track each other throughout optimization, indicating stable convergence and no evidence of overfitting. Training was performed on a single 40 GB NVIDIA A100 GPU and required approximately 10 days.

4.3 Results

We validate our inference code both internally, using a large number of injections drawn from the prior distribution (Sec. 4.3.1), and externally, by comparison with a standard stochastic sampler (Sec. 4.3.2) and previous works (Sec. 4.3.3).

4.3.1 Model validation

Once trained, we employ the flow-based model for posterior inference, enabling rapid sampling for data consistent with the training priors and the assumed PSD of the noise. Residual discrepancies in the learned posterior approximation—arising from limited training or finite network expressivity—are mitigated via importance sampling (IS) [61] using the flow posterior as a proposal distribution. We quantify the quality of the reweighted samples through the sampling efficiency

$$\epsilon = \frac{\left(\sum_i^N w_i\right)^2}{N \sum_i^N w_i^2} \in (0, 100]\% \,, \tag{4.1}$$

which depends on the distribution of the importance weights w_i associated with the N samples θ_i drawn from the flow. The importance weights are defined as

$$w_i = \frac{\mathcal{L}(d|\theta_i)\,\pi(\theta_i)}{q(\theta_i)} \,, \tag{4.2}$$

where $\mathcal{L}(d|\theta)$ denotes the likelihood of the data d, $\pi(\theta)$ the prior distribution, and $q(\theta)$ the proposal distribution defined by the flow. The corresponding effective sample size is then given by $N_{\text{eff}} = \epsilon \times N$. We assess the statistical calibration of the model using an ensemble

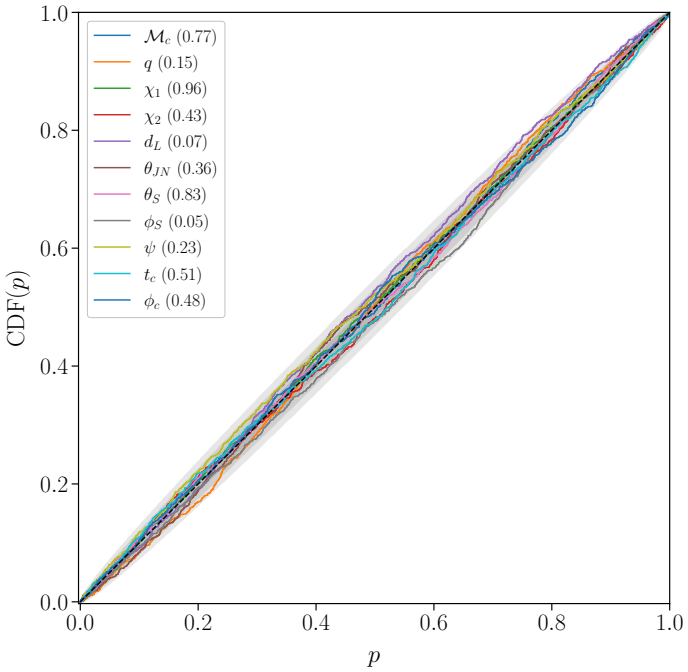

Figure 4.2: Probability-probability (p–p) plot for 1000 simulated injections obtained with DINGO-IS. Colored curves show results for each of the 11 marginal distributions. The dashed line indicates a uniform distribution, and grey shaded regions denote the expected 1σ, 2σ, 3σ confidence intervals. KS test p-values are provided in the legend.

of 1000 simulated signals drawn from the prior and injected into stationary Gaussian noise. The results are summarized in the p–p plot shown in Fig. 4.2, which illustrates the percentiles of the cumulative distribution function (CDF) of the marginal posterior at which the injected parameter values are recovered. For all parameters, the percentile distributions are consistent with uniform distributions, indicating well-calibrated results. This is further supported by a Kolmogorov-Smirnov (KS) test, which does not reject the null hypothesis of uniformity at the 95% significance level and yields a combined p–value of 0.27 using Fisher's method [77]. Across the analyzed events, we obtain a median sampling efficiency of $\epsilon \sim 22\%$, consistent with previous application of DINGO-IS to LIGO-Virgo-KAGRA observations [61].

4.3.2 Benchmarking

We benchmark DINGO against posterior samples obtained with the nested sampling algorithm [151] as implemented in NESSAI [173]. To ensure a consistent comparison,

Table 4.1: Injected parameter values for the three representative sources at low, moderate, and high SNR used in this work.

Parameter	Low SNR SNR = 87	Moderate SNR SNR = 500	High SNR SNR = 1000
$\mathcal{M}_c\,[10^7 M_\odot]$	1.61	1.30	0.90
q	0.82	0.54	0.47
χ_1	−0.47	−0.09	0.34
χ_2	−0.11	0.66	0.28
d_L [Gpc]	44.14	21.91	16.96
θ_{JN} [rad]	1.77	2.33	0.73
θ_S [rad]	0.14	1.62	2.94
ϕ_S [rad]	3.09	4.72	3.09
t_c [s]	820.0	395.0	−568.0
ψ [rad]	1.13	0.98	2.51
ϕ_c [rad]	3.36	2.18	2.74

both methods employ the same waveform model, prior distributions, analysis settings, and noise PSD described in Sec. 4.2. We consider three representative signal injections spanning the targeted SNR range, corresponding to low, moderate, and high SNRs of approximately 87, 500, and 1000, respectively. The injected parameters are reported in Table 4.1.

For each injection, we draw 2×10^4 samples from the flow posterior. Initial sampling requires less than two seconds on a single 40 GB NVIDIA A100 GPU, while IS adds an additional ten to fifteen seconds on 20 CPU cores, resulting in a full-parameter inference time well below one minute per event. By comparison, the corresponding Nessai runs were performed using 4000 live points and without acceleration or optimization techniques. The runtime ranged from roughly ten days for the low-SNR case to about forty days for the high-SNR scenario. These values likely represent a pessimistic estimate, as GPU-accelerated analyses combined with heterodyned likelihood techniques (see e.g. Ref. [97]) have been shown to reduce the computational cost of stochastic sampling methods to tens of minutes. Nevertheless, the inference approach presented in this work remains significantly less computationally demanding.

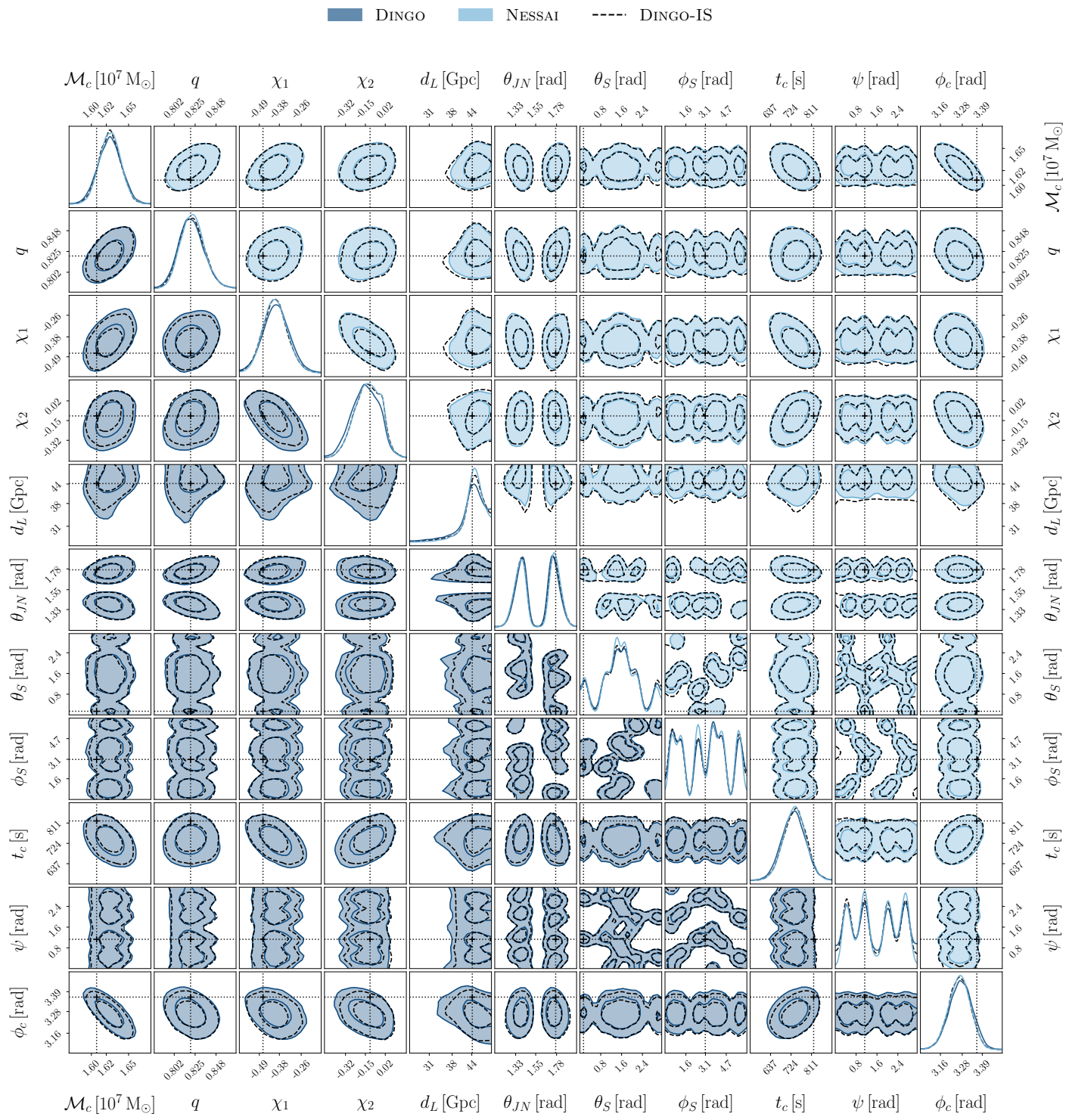

Figure 4.3: Posterior distributions for our representative source with low SNR (∼87) obtained with different sampling methods. The lower left triangle compares results from DINGO (dark-blue filled contours) and DINGO-IS (dashed black contours), while the upper right triangle compares NESSAI (light-blue filled contours) and DINGO-IS. Contours indicate the 50% and 90% credible regions. The true injected values are indicated by dotted lines.

In Fig. 4.3, we show the posterior distributions for the low-SNR injection, comparing samples obtained from the raw DINGO output, after IS correction (DINGO-IS), and from the reference NESSAI analysis. DINGO and NESSAI posteriors are in very close agreement, with the corresponding DINGO-IS results fully overlapping both. This indicates that IS introduces only negligible corrections in this regime, a conclusion further supported by the recovered sample efficiency, $\epsilon = 50\%$, which corresponds to an effective sample size of $N_{\text{eff}} = 10^4$. Beyond the overall agreement, DINGO successfully captures the non-trivial structure of the posterior, including correlations between luminosity distance d_L, inclination θ_{JN}, and sky location (θ_S, ϕ_S). In particular, the characteristic multimodal sky structure induced by the low-frequency response of LISA [120] is consistently recovered.

Figures 4.4 and 4.5 show the posterior distributions for the moderate- and high-SNR sources, respectively. At moderate SNR, the raw DINGO posteriors display small but noticeable deviations from the NESSAI reference. These residual discrepancies are effectively compensated by IS, bringing the resulting DINGO-IS posteriors into close agreement with NESSAI. In this case, we report a sample efficiency of $\epsilon = 12\%$ and an effective sample size of $N_{\text{eff}} = 2.4 \times 10^3$.

For the high-SNR injection, the DINGO posteriors exhibit broader support than the NESSAI reference, resulting in limited overlap between the two distributions. In this case, the learned proposal distribution assigns non-negligible probability mass to regions of low true posterior density, leading to highly variable importance weights (Eq. 4.2) and a sampling efficiency of $\epsilon = 0.05\%$. The resulting posterior estimate obtained with DINGO-IS remains dominated by statistical fluctuations, as illustrated by the one- and two-dimensional marginal distributions in Fig. 4.5, due to the extremely small effective sample size ($N_{\text{eff}} = 10 \ll N$). Increasing the number of flow samples might improve the stability and smoothness of the reweighted posterior. Despite this, it is worth noting that the raw DINGO posterior correctly localizes the source within a restricted region of the parameter space, reducing the original prior volume by approximately a factor of 10^{13} at the 90% credible level. The corresponding one- and two-dimensional marginal distributions also provide good coverage of the target distribution, offering a reliable starting point for subsequent refinement with conventional stochastic sampling methods. The sky location and polarization parameters remain weakly constrained owing to intrinsic degeneracies in the adopted LISA response. As in the low- and moderate-SNR injections, the posteriors are highly multimodal and extend over most of the prior range. Nonetheless, the model captures the characteristic sky pattern expected in the low-frequency regime.

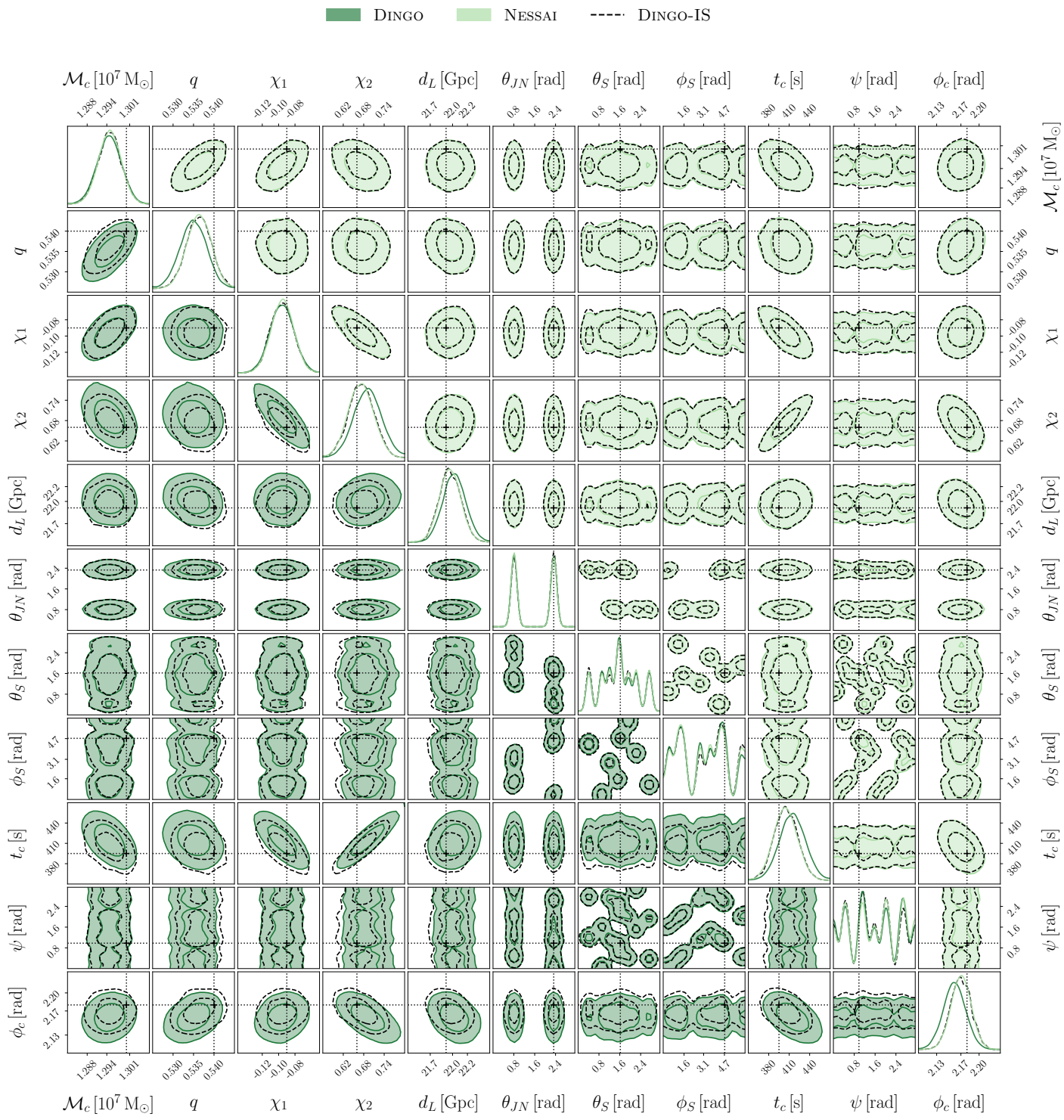

Figure 4.4: Posterior distributions for our representative source with moderate SNR (∼500) obtained with different sampling methods. The lower left triangle compares results from DINGO (dark-green filled contours) and DINGO-IS (dashed black contours), while the upper right triangle compares NESSAI (light-green filled contours) and DINGO-IS. Contours indicate the 50% and 90% credible regions. The true injected values are indicated by dotted lines.

To shed light on the origin of the low sampling efficiency observed for the high-SNR injection, we investigate how it varies across the source parameter space. This is illustrated in Fig. 4.6, which shows the prior distribution used to construct the training dataset, together with the injected parameter values of the three representative BH binaries described above. The visualization is restricted to parameters with non-uniform priors, since only these imprint a non-trivial structure on the sampling density. We draw 2×10^4 sources from this prior and evaluate their posterior distributions using our trained model. The resulting SNR and sampling efficiency ϵ are reported in Fig. 4.6 as a function of the source parameters.

Crucially, we observe a strong anti-correlation between SNR and sampling efficiency. This behaviour can be understood as a consequence of the mass-covering property of the training objective used in neural posterior estimation. Specifically, the loss penalizes the density estimator when it assigns insufficient probability mass to regions supported by the true posterior, thereby favoring approximations that cover its full support. In the presence of residual modeling inaccuracies, this property typically results in overly diffuse posterior distributions. While avoiding posterior undercoverage is generally desirable in statistical inference, this tendency becomes particularly problematic in the high-SNR regime, where the true posterior is sharply concentrated. In this situation, even a mild overestimation of the posterior support causes a substantial fraction of samples to fall in low-likelihood regions, resulting in highly variable importance weights and a corresponding reduction in the effective sample size.

An additional factor contributing to the observed reduction in sampling efficiency is the coverage of the training prior. As illustrated in Fig. 4.6, the high-SNR source lies in a sparsely represented region of this distribution, with the injected chirp mass $\mathcal{M}_c = 9 \times 10^6\ \mathrm{M}_\odot$ falling below the 3.5^{th} percentile. Consequently, low chirp-mass systems are rarely encountered during training and therefore contribute less to the optimization objective, leading to a poorer approximation of the posterior and a corresponding decrease in sampling efficiency. A similar effect is observed for the primary spin. Although the injected value, $\chi_1 = 0.34$, lies well within the allowed range, the peaked structure of its prior places it above the 85^{th} percentile of the training prior.

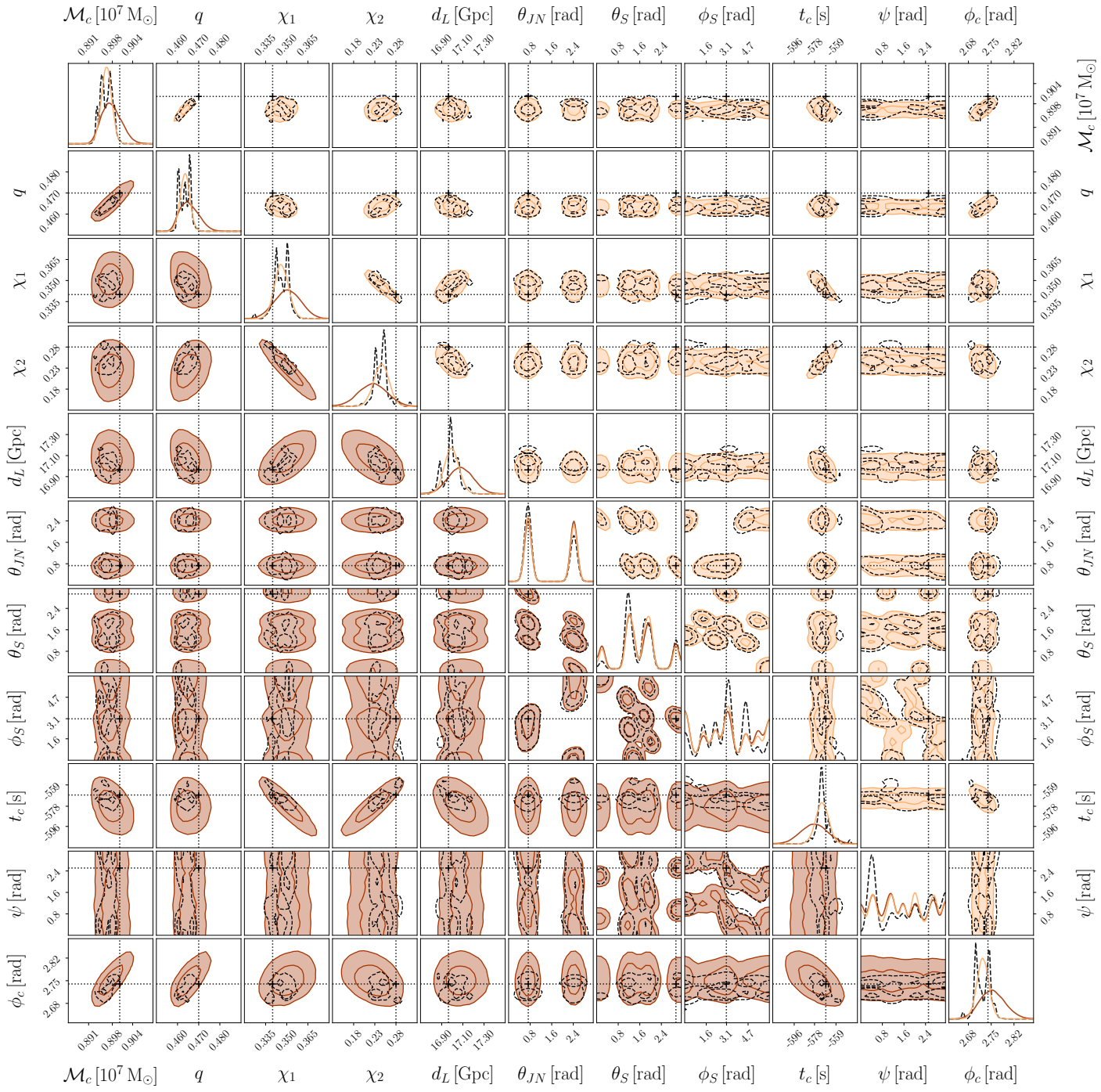

Figure 4.5: Posterior distributions for our representative source with high SNR (∼ 1000) obtained with different sampling methods. The lower left triangle compares results from Dingo (dark-orange filled contours) and Dingo-IS (dashed black contours), while the upper right triangle compares Nessai (light-orange filled contours) and Dingo-IS. Contours indicate the 50% and 90% credible regions. The true injected values are indicated by solid dotted lines.

To explore this dependence, we vary the chirp mass and primary spin values while keeping all other parameters fixed to the reference values listed in Tab. 4.1. The luminosity

distance d_L and source inclination θ_{JN} are adjusted to maintain a constant SNR. Increasing the chirp mass to $1.3 \times 10^7\,\mathrm{M}_\odot$ shifts the system toward a more densely sampled region of the training prior and raises the sampling efficiency to 1.5%. A similar improvement is obtained by setting the primary spin to $\chi_1 = 0$, which increases the efficiency to 0.5%. By contrast, varying the sky colatitude θ_S has little effect on the sampling efficiency. Even when moved toward more densely sampled regions of the training prior, the efficiency remains close to the reference value of $\epsilon = 0.05\%$. These results indicate that underrepresentation in the training prior does not affect all parameters equally. Waveforms with low chirp masses and high primary spins exhibit more oscillatory structure and are therefore more difficult to model accurately, making sparse training coverage in these regions particularly detrimental to the performance of the density estimator.

We emphasize that limitations in prior coverage are not exclusive to high-SNR sources. Indeed, among the 1000 injections used to construct the p–p plot presented in Sec. 4.3.1, we identify a small subset of sources (~ 10) with SNR $\in [250, 500]$ and sample efficiency below 1%. These systems typically have parameters located near the boundaries of the training prior, most commonly featuring either high primary spins above the 95^{th} percentile or low chirp masses below the 5^{th} percentile. Two sources differ from this pattern: one has a mass ratio above the 95^{th} percentile together with an anti-aligned primary spin below the 5^{th} percentile, while the other exhibits strongly anti-aligned spins, with χ_1 and χ_2 below the 0.1^{th} and 4^{th} percentiles, respectively.

Overall, our results show that sampling efficiency systematically decreases at high SNR. Extending the training does not significantly improve the sampling efficiency, suggesting that insufficient optimization is unlikely to be the primary source of the observed mismatch. Further improvements may instead require increased model capacity or more flexible conditional density parameterizations. In addition, the sensitivity of the sampling efficiency to peripheral regions of parameter space indicates that the adopted training prior plays a non-negligible role. This dependence could be mitigated by broadening the prior support or by adopting a distribution that provides more uniform coverage of the most informative parameters. A systematic exploration of the full LISA mass spectrum and its implications for prior coverage is left for future work.

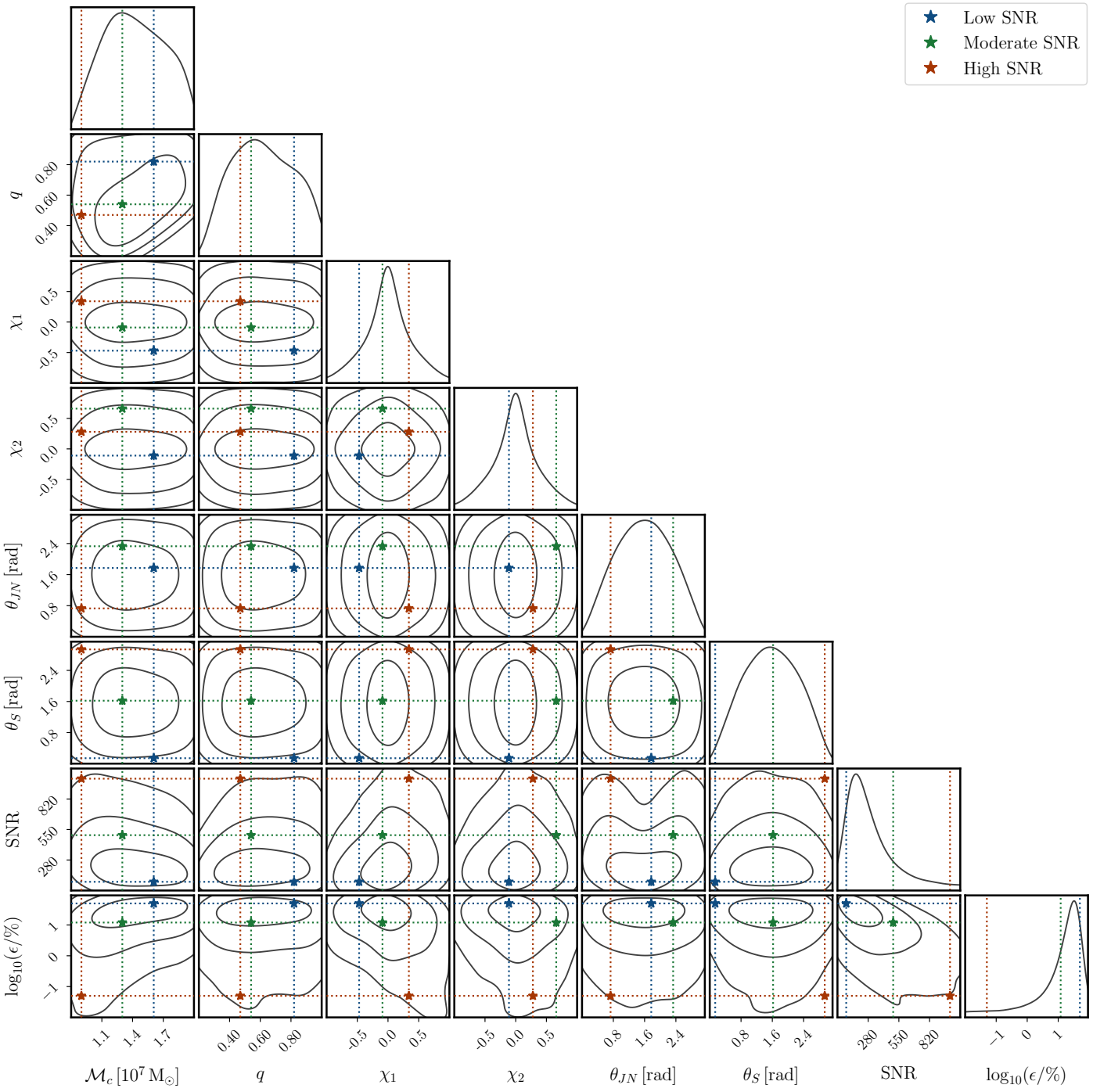

Figure 4.6: Prior distributions used to construct the target training set. Injected values for the low- (blue), moderate- (green), and high-SNR (orange) cases are shown as dotted lines, with colored stars marking their locations. Contours contain 50%, 90%, and 99% of the estimated prior density. Of note, the bottom right section shows the joint and marginal distributions of the SNR and sample efficiency ϵ across the ensemble of simulated injections.

Finally, as done in previous works [60, 61, 145], we use the Jensen-Shannon (JS) divergence [105] to quantitatively compare the marginal posterior distributions obtained with DINGO and NESSAI across 100 independent injections for each representative source considered in this study. As discussed above, posterior recovery in the high-SNR case already shows qualitative differences, and we therefore leave a quantitative assessment of this regime to future work. The injections are generated using the same noise realization for each representative source, thereby isolating sampling errors. The JS divergence ranges from 0 bit to 1 bit, corresponding to identical and maximally distinct distributions, respectively. Across all parameters, we find median divergencies of $D_{JS}^{L} \in [6.3\times10^{-4}, 2\times10^{-2}]$ for the low-SNR injections and $D_{JS}^{M} \in [1.2 \times 10^{-4}, 0.1]$ for the moderate-SNR case. Applying IS to the DINGO posteriors leads to a systematic reduction of the divergence in both regimes, yielding $D_{JS}^{L} \in [1.9 \times 10^{-4}, 1 \times 10^{-2}]$ and $D_{JS}^{M} \in [1.7 \times 10^{-4}, 7.3 \times 10^{-4}]$, respectively. For a reference scale, let us consider the JS divergence between two univariate Gaussian distributions with mean $\mu_{1,2}$ and standard deviation $\sigma_{1,2}$. The largest median divergence observed in our analysis ($D_{JS}^{M} = 0.1$, corresponding to the phase parameter of the moderate-SNR source) is comparable either to a mean offset of $|\mu_1 - \mu_2| \simeq 0.8\,\sigma$ for a case where $\sigma = \sigma_1 = \sigma_2$, or to a standard-deviation ratio $\sigma_1/\sigma_2 \simeq 1.75$ for a case with $\mu_1 = \mu_2$. For the low-SNR case, the largest median divergence ($D_{JS}^{M} = 2 \times 10^{-2}$) is obtained for the luminosity distance parameter. When mapped onto the Gaussian reference, this corresponds either to a mean offset of $|\mu_1 - \mu_2| \simeq 0.3\,\sigma$ for equal variances, or to a standard-deviation ratio $\sigma_1/\sigma_2 \simeq 1.3$ for equal means. After IS, a substantial improvement is observed, particularly in the moderate-SNR case, where the largest median JS divergence corresponds to $|\mu_1 - \mu_2| \simeq 0.06\,\sigma$ and $\sigma_1/\sigma_2 \simeq 1.04$. Simultaneous differences in mean and variances smaller than the values provided yield smaller JS divergences.

4.3.3 Comparison with previous work

Previous works have applied SBI techniques to the analysis of massive BH binaries. Although these studies primarily focused on lower-mass systems than those considered here, in the following we present a qualitative comparison of the underlying methodological approaches.

In contrast to the present work, Ref. [121] adopts a neural likelihood estimation strategy, training neural networks to construct a surrogate of the likelihood function. This method achieves a substantial reduction in the number of waveform evaluations relative to traditional stochastic samplers, while yielding posterior distributions that are qualitatively consistent with standard likelihood-based results. However, a comprehensive quantitative validation against established methods is still lacking, preventing a robust assessment

of its overall performance. Inference also relies on a dedicated training process for each individual source, with reported computational costs of the order of days, thereby limiting the applicability of this method to time-critical analyses.

By comparison, Ref. [103] adopts a neural posterior-estimation framework more closely aligned with the approach presented in this work, based on continuous normalizing flows [46] trained using the flow matching algorithm [106]. The inferred posterior distributions show qualitative agreement with those obtained using nested sampling; however, systematic deviations persist across multiple parameters, indicating that further methodological refinement may be required to achieve inference accuracy comparable to that achieved in this work. In particular, model performance is evaluated under different noise settings: Galactic foreground noise is included in the test set but not during training, while the nested-sampling analysis explicitly accounts for it. Consequently, the two approaches are not assessed under matched conditions, and direct comparisons should be treated with caution. From a computational perspective, their network features approximately 200 million learnable parameters and trains in about two days, whereas the discrete normalizing-flow model considered here has roughly 350 million parameters and requires approximately 10 days of training. However, per-event inference typically requires several minutes to generate $\sim 1.4 \times 10^4$ posterior samples, compared to a few seconds in this work.

4.4 Outlook

The successful demonstration of Dingo as an accurate and efficient framework for full parameter estimation of massive BH binary signals enables a broad range of applications for LISA data analysis. Among these, low-latency pipelines are particularly compelling, as they are crucial for enabling coordinated observations of the diverse electromagnetic signatures expected to accompany massive BH binaries [38,62,66,68]. The sky-localization uncertainties anticipated for LISA observations of such sources are expected to encompass numerous potential host galaxies, posing a significant challenge for unambiguous host identification [112]. Rapid inference of additional source properties —most notably the mass and luminosity distance— is therefore essential to provide independent constraints on the host environment and to efficiently reduce the potential host-galaxy candidates. In this context, developing a Dingo network trained on pre-merger signals (see Ref. [59] for an application to binary neutron-star systems) and integrating it within low-latency search pipelines would be a promising step forward toward LISA parameter estimation in realistic early-warning scenarios.

Beyond low-latency applications, the fast inference capabilities of Dingo can be

leveraged within the LISA global fit to substantially reduce the computational cost associated with massive BH binary analyses. More specifically, this can be achieved by integrating the proposed SBI approach within a blocked Gibbs framework [82,126,143]. In this hybrid scheme, Dingo could generate samples for individual sources at each global-fit iteration, either replacing Markov Chain Monte Carlo sampling [83] when the learned posterior approximation is sufficiently accurate, or acting as a proposal distribution to accelerate convergence otherwise. This will facilitate the adoption of more accurate waveform models, whose computational expense would otherwise be prohibitive for fully stochastic inference frameworks. Extending this approach to self-consistent global-fit implementations would require relaxing the fixed-PSD assumption adopted in this work and training the network on a representative ensemble of instrumental noise and Galactic foreground realizations, thereby accounting for event-to-event variations in the noise model. In the signal-dominated regime expected for LISA, the noise PSD cannot be estimated independently of the GW signals and must instead be inferred as part of the global fit. In this context the PSD estimated at a previous stage of the analysis could be used to condition Dingo at inference time, ensuring consistency between the source parameter estimation and the evolving noise model.

A further challenge common to both low-latency and global-fit pipelines is the presence of non-stationary features in the data, such as glitches and data gaps. In global-fit analyses, glitches are either incorporated directly into the data model and inferred jointly with the astrophysical signals [127, 144, 155], or mitigated by effectively treating them as data gaps through direct masking [45]. In parallel, machine-learning–based approaches have been developed for the rapid identification and mitigation of noise transients in low-latency settings [90]. However, these methods are not designed to deliver full end-to-end parameter estimation. Data gaps have also been addressed through a variety of techniques, including apodization [64], inpainting [37], data augmentation [27, 118], and non-diagonal likelihood formulations [42]. These approaches often come at the cost of additional noise-model approximations, potential biases introduced by gap-filling procedures, or significant computational overhead. More recently, Ref. [136] proposed an efficient wavelet-domain augmentation scheme; however, its applicability relies on the assumption of locally stationary noise.

Finally, future work will explore the performance of Dingo on precessing and potentially eccentric signals, as well as its extension to lower-mass systems emitting GW signals that typically last weeks to months. In this regime, the breakdown of the low-frequency approximation requires the full time- and frequency-dependent LISA response. These effects substantially increase the effective input dimensionality and the computational cost of training, motivating the investigation of data compression techniques, such as

multibanding [59], alongside the development of more flexible model architectures.

Taken together, these considerations highlight the need for fast, preliminary source-characterization methods that remain effective in the presence of realistic, time-varying noise and physics-rich waveform models. SBI is the obvious answer here, as noise artifacts can be included in the training data without the need to specify a likelihood function (note, however, that IS —used both in this paper and in previous Dingo work— *does* require evaluating the likelihood). In this context, Ref. [119] has demonstrated that SBI provides a flexible and efficient framework capable of handling data gaps. This capability is particularly promising for time-critical analyses and may also facilitate the initialization of global-fit pipelines by improving convergence and robustness in complex noise environments. The recent transformer-based extension of Dingo [101], designed to accommodate variations in data-analysis settings, represents a natural next step that builds on the approach presented in this work toward addressing key challenges for LISA, such as data gaps and source overlap (see also Ref. [33, 91, 132]).

5 Conclusions and outlook

The direct detection of GWs represents a major experimental milestone for GR, providing direct access to the dynamics of spacetime in the strong-field regime. By complementing electromagnetic observations, GW astronomy enables the study of compact-object mergers that would otherwise remain inaccessible.

Nevertheless, the field is still in its infancy, with current detectors probing only a limited portion of the GW spectrum. Future observatories will significantly broaden the accessible frequency range, opening new windows onto previously unexplored astrophysical sources and physical regimes.

In this landscape, LISA will play a central role in unveiling a rich population of GW sources in the millihertz band. Among these, mergers of massive BH binaries stand out as particularly promising targets. Understanding the formation, growth, and cosmic evolution of these systems remains one of the central open challenges in contemporary astrophysics. Current theoretical predictions about these sources are affected by substantial uncertainties, reflecting the complex interplay of highly non-linear processes acting across widely separated scales, including galaxy mergers, gas dynamics, accretion, and stellar interactions. As a consequence, different physical assumptions can yield markedly different predictions for merger rates and source properties.

In this context, GW observations can offer a fundamentally data-driven avenue to investigate the evolution of coalescing massive BH binaries. In this Thesis, we presented a Bayesian framework to assess the likelihood that putative massive BH binaries observed by LISA have evolved in either gas or stellar environments. Within the state-of-the-art of astrophysical models, our results indicated that the majority of sources are expected to evolve in gas-rich environments. Bayesian inference and model selection further showed that the expected precision of LISA measurements can place meaningful constraints on the dominant hardening channel that drives massive binaries toward final coalescence. Crucially, we found that accurately modeling BH spins is essential to avoid significant systematic biases in LISA astrophysical inference. In addition, our results suggest that subdominant waveform modes may preferentially impact sources formed through specific astrophysical pathways, highlighting the importance of combining accurate waveform modeling with population-informed inference

From a data-analysis perspective, a key challenge for LISA arises from non-stationary transient features in the data stream, such as glitches and data gaps, which can directly affect parameter estimation. Assessing their impact is therefore paramount to ensuring reliable inference of the source parameters. Our results showed that neglecting glitches

in the data may introduce significant systematic biases in the inferred source parameters, with the glitch duration emerging as a particularly critical factor. At the same time, we demonstrated that a joint inference framework enables accurate reconstruction of the binary parameters, making the approach naturally suited to the LISA global-fit context. While the present analysis focuses on single-source scenarios, a broader exploration of source and noise parameters, including configurations with multiple overlapping sources, will be required to robustly assess the impact of noise transients on future LISA catalogs.

In addition to probing BH formation and evolution over cosmic time, massive BH binaries may also be associated with electromagnetic emission. Joint GW and electromagnetic observations therefore offer a powerful avenue to characterize the physical environments in which these systems form and evolve. In this perspective, developing data-analysis frameworks capable of supporting multi-messenger studies is essential to fully exploit the astrophysical potential of the LISA mission. Within this context, the results presented in this Thesis demonstrate that Dingo provides an accurate and computationally efficient framework for full parameter estimation of massive BH binaries, enabling a broad range of applications in LISA data analysis. In particular, its efficiency makes it well suited to time-critical scenarios, including low-latency analyses of pre-merger signals relevant for coordinated multi-messenger follow-up efforts. More generally, the reduced computational cost renders the proposed approach naturally compatible with the LISA global-fit framework. This scalability opens the door to the routine adoption of more accurate waveform models and supports the construction of reliable source catalogs for astrophysical, cosmological, and fundamental-physics investigations.

Bibliography

[1] B. Abbott et al. GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral. Phys. Rev. Lett., 119(16):161101, October 2017.

[2] B. P. Abbott et al. Astrophysical Implications of the Binary Black-hole Merger GW150914. ApJ, 818(2):L22, February 2016.

[3] B. P. Abbott et al. GW150914: Implications for the Stochastic Gravitational-Wave Background from Binary Black Holes. Phys. Rev. Lett., 116(13):131102, April 2016.

[4] B. P. Abbott et al. GW150914: The Advanced LIGO Detectors in the Era of First Discoveries. Phys. Rev. Lett., 116(13):131103, April 2016.

[5] B. P. Abbott et al. Localization and Broadband Follow-up of the Gravitational-wave Transient GW150914. ApJ, 826(1):L13, July 2016.

[6] B. P. Abbott et al. Observation of Gravitational Waves from a Binary Black Hole Merger. *prl*, 116(6):061102, February 2016.

[7] B. P. Abbott et al. The Rate of Binary Black Hole Mergers Inferred from Advanced LIGO Observations Surrounding GW150914. ApJ, 833(1):L1, December 2016.

[8] B. P. Abbott et al. Gravitational Waves and Gamma-Rays from a Binary Neutron Star Merger: GW170817 and GRB 170817A. ApJ, 848(2):L13, October 2017.

[9] B. P. Abbott et al. GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs. *Physical Review X*, 9(3):031040, July 2019.

[10] R. Abbott et al. GW190521: A Binary Black Hole Merger with a Total Mass of 150 M⊙. Phys. Rev. Lett., 125(10):101102, September 2020.

[11] R. Abbott et al. GW190814: Gravitational Waves from the Coalescence of a 23 Solar Mass Black Hole with a 2.6 Solar Mass Compact Object. ApJ, 896(2):L44, June 2020.

[12] R. Abbott et al. Population Properties of Compact Objects from the Second LIGO-Virgo Gravitational-Wave Transient Catalog. ApJ, 913(1):L7, May 2021.

[13] R. Abbott et al. GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo during the Second Part of the Third Observing Run. *Physical Review X*, 13(4):041039, October 2023.

[14] Milton Abramowitz and Irene A. Stegun. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. Dover, 1965.

[15] T. Accadia et al. Virgo: a laser interferometer to detect gravitational waves. *Journal of Instrumentation*, 7(3):3012, March 2012.

[16] F. Acernese et al. Advanced Virgo: a second-generation interferometric gravitational wave detector. *Classical and Quantum Gravity*, 32(2):024001, January 2015.

[17] Gabriella Agazie et al. The NANOGrav 15 yr Data Set: Evidence for a Gravitational-wave Background. *Astrophys. J. Lett.*, 951(1):L8, July 2023.

[18] Pau Amaro-Seoane et al. Laser Interferometer Space Antenna. , page arXiv:1702.00786, February 2017.

[19] Pau Amaro-Seoane et al. Astrophysics with the Laser Interferometer Space Antenna. *lrr*, 26(1):2, December 2023.

[20] R. E. Angulo and S. D. M. White. One simulation to fit them all - changing the background parameters of a cosmological N-body simulation. *Mon. Not. R. Astron. Soc.*, 405(1):143–154, June 2010.

[21] J. Antoniadis et al. The second data release from the European Pulsar Timing Array. III. Search for gravitational wave signals. *Astron. Astrophys.*, 678:A50, October 2023.

[22] M. Armano et al. Beyond the required lisa free-fall performance: New lisa pathfinder results down to 20 μHz. *Phys. Rev. Lett.*, 120:061101, Feb 2018.

[23] M. Armano et al. Transient acceleration events in LISA Pathfinder data: Properties and possible physical origin. *prd*, 106(6):062001, September 2022.

[24] Yoichi Aso, Yuta Michimura, Kentaro Somiya, Masaki Ando, Osamu Miyakawa, Takanori Sekiguchi, Daisuke Tatsumi, and Hiroaki Yamamoto. Interferometer design of the KAGRA gravitational wave detector. Phys. Rev. D, 88(4):043007, August 2013.

[25] Stanislav Babak, Jonathan Gair, Alberto Sesana, Enrico Barausse, Carlos F. Sopuerta, Christopher P. L. Berry, Emanuele Berti, Pau Amaro-Seoane, Antoine Petiteau, and Antoine Klein. Science with the space-based interferometer LISA. V. Extreme mass-ratio inspirals. *prd*, 95(10):103012, May 2017.

[26] Quentin Baghi, Natalia Korsakova, Jacob Slutsky, Eleonora Castelli, Nikolaos Karnesis, and Jean-Baptiste Bayle. Detection and characterization of instrumental transients in lisa pathfinder and their projection to LISA. *Physical Review D*, 105(4):042002, February 2022.

[27] Quentin Baghi, James Ira Thorpe, Jacob Slutsky, John Baker, Tito Dal Canton, Natalia Korsakova, and Nikos Karnesis. Gravitational-wave parameter estimation with gaps in LISA: A Bayesian data augmentation method. *Phys. Rev. D*, 100(2):022003, July 2019.

[28] Enrico Barausse. The evolution of massive black holes and their spins in their galactic hosts. *Mon. Not. R. Astron. Soc.*, 423(3):2533–2557, July 2012.

[29] Enrico Barausse, Irina Dvorkin, Michael Tremmel, Marta Volonteri, and Matteo Bonetti. Massive Black Hole Merger Rates: The Effect of Kiloparsec Separation Wandering and Supernova Feedback. *Astrophys. J.*, 904(1):16, November 2020.

[30] Enrico Barausse and Luciano Rezzolla. Predicting the Direction of the Final Spin from the Coalescence of Two Black Holes. *Astrophys. J. Lett.*, 704(1):L40–L44, October 2009.

[31] James M. Bardeen and Jacobus A. Petterson. The Lense-Thirring Effect and Accretion Disks around Kerr Black Holes. *Astrophys. J. Lett.*, 195:L65, January 1975.

[32] M. C. Begelman, R. D. Blandford, and M. J. Rees. Massive black hole binaries in active galactic nuclei. *Nature*, 287(5780):307–309, September 1980.

[33] Martin Benedikt and Ippocratis D. Saltas. GraviBERT: Transformer-based inference for gravitational-wave time series. , December 2025.

[34] Emanuele Berti and Marta Volonteri. Cosmological Black Hole Spin Evolution by Mergers and Accretion. *Astrophys. J.*, 684(2):822–828, September 2008.

[35] James Binney and Scott Tremaine. *Galactic Dynamics: Second Edition*. 2008.

[36] Luc Blanchet. Gravitational Radiation from Post-Newtonian Sources and Inspiralling Compact Binaries. *lrr*, 17(1):2, December 2014.

[37] Aurore Blelly, Jérôme Bobin, and Hervé Moutarde. Sparse data inpainting for the recovery of Galactic-binary gravitational wave signals from gapped data. *Mon. Not. R. Astron. Soc.*, 509(4):5902–5917, February 2022.

[38] Tamara Bogdanović, M. Coleman Miller, and Laura Blecha. Electromagnetic counterparts to massive black-hole mergers. *Living Rev. Relativ.*, 25(1):3, December 2022.

[39] Tamara Bogdanović, Christopher S. Reynolds, and M. Coleman Miller. Alignment of the Spins of Supermassive Black Holes Prior to Coalescence. *Astrophys. J. Lett.*, 661(2):L147–L150, June 2007.

[40] Elisa Bortolas, Alessia Franchini, Matteo Bonetti, and Alberto Sesana. The Competing Effect of Gas and Stars in the Evolution of Massive Black Hole Binaries. ApJ, 918(1):L15, September 2021.

[41] Michael Boylan-Kolchin, Volker Springel, Simon D. M. White, Adrian Jenkins, and Gerard Lemson. Resolving cosmic structure formation with the Millennium-II Simulation. *Mon. Not. R. Astron. Soc.*, 398(3):1150–1164, September 2009.

[42] Ollie Burke, Sylvain Marsat, Jonathan R. Gair, and Michael L. Katz. Addressing data gaps and assessing noise mismodeling in LISA. *Phys. Rev. D*, 111(12):124053, June 2025.

[43] Riccardo Buscicchio, Antoine Klein, Elinore Roebber, Christopher J. Moore, Davide Gerosa, Eliot Finch, and Alberto Vecchio. Bayesian parameter estimation of stellar-mass black-hole binaries with LISA. *prd*, 104(4):044065, August 2021.

[44] Riccardo Buscicchio, Federico Pozzoli, Daniele Chirico, and Alberto Sesana. The first year of LISA Galactic foreground. , November 2025.

[45] Eleonora Castelli, Quentin Baghi, John G. Baker, Jacob Slutsky, Jérôme Bobin, Nikolaos Karnesis, Antoine Petiteau, Orion Sauter, Peter Wass, and William J. Weber. Extracting gravitational wave signals from LISA data in the presence of artifacts. *Class. Quantum Grav.*, 42(6):065018, March 2025.

[46] Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. Neural Ordinary Differential Equations. , June 2018.

[47] C.W. Clenshaw and A.R. Curtis. A method for numerical integration on an automatic computer. *Numer. Math.*, 2:197–205, 1960.

[48] Monica Colpi. Massive binary black holes in galactic nuclei and their path to coalescence. *Space Science Reviews*, 183(1):189–221, 2014.

[49] Monica Colpi et al. LISA Definition Study Report. , February 2024.

[50] Giuseppe Congedo, Rita Dolesi, Mauro Hueller, Stefano Vitale, and William J. Weber. Space-borne gravitational-wave detectors as time-delayed differential dynamometers. *Phys. Rev. D*, 88:082003, Oct 2013.

[51] Neil J. Cornish and Jeff Crowder. LISA data analysis using Markov chain Monte Carlo methods. *prd*, 72(4):043005, August 2005.

[52] Neil J. Cornish and Louis J. Rubbo. Publisher's Note: LISA response function [Phys. Rev. D 67, 022001 (2003)]. *Phys. Rev. D*, 67(2):029905, January 2003.

[53] P. S. Cowperthwaite et al. The Electromagnetic Counterpart of the Binary Neutron Star Merger LIGO/Virgo GW170817. II. UV, Optical, and Near-infrared Light Curves and Comparison to Kilonova Models. ApJ, 848(2):L17, October 2017.

[54] Kyle Cranmer, Johann Brehmer, and Gilles Louppe. The frontier of simulation-based inference. *Proceedings of the National Academy of Science*, 117(48):30055–30062, December 2020.

[55] Jeff Crowder and Neil J. Cornish. Extracting galactic binary signals from the first round of Mock LISA Data Challenges. *Classical and Quantum Gravity*, 24(19):S575–S585, October 2007.

[56] J. Cuadra, P. J. Armitage, R. D. Alexander, and M. C. Begelman. Massive black hole binary mergers within subparsec scale gas discs. *Mon. Not. R. Astron. Soc.*, 393(4):1423–1432, March 2009.

[57] Curt Cutler. Angular resolution of the LISA gravitational wave detector. *Phys. Rev. D*, 57(12):7089–7102, June 1998.

[58] Curt Cutler and Éanna E. Flanagan. Gravitational waves from merging compact binaries: How accurately can one extract the binary's parameters from the inspiral waveform\? *prd*, 49(6):2658–2697, March 1994.

[59] Maximilian Dax, Stephen R. Green, Jonathan Gair, Nihar Gupte, Michael Pürrer, Vivien Raymond, Jonas Wildberger, Jakob H. Macke, Buonanno, et al. Real-time inference for binary neutron star mergers using machine learning. *Nature*, 639(8053):49–53, March 2025.

[60] Maximilian Dax, Stephen R. Green, Jonathan Gair, Jakob H. Macke, Alessandra Buonanno, and Bernhard Schölkopf. Real-Time Gravitational Wave Science with Neural Posterior Estimation. *Phys. Rev. Lett.*, 127(24):241103, December 2021.

[61] Maximilian Dax, Stephen R. Green, Jonathan Gair, Michael Pürrer, Jonas Wildberger, Jakob H. Macke, Alessandra Buonanno, and Bernhard Schölkopf. Neural Importance Sampling for Rapid and Reliable Gravitational-Wave Inference. *Phys. Rev. Lett.*, 130(17):171403, April 2023.

[62] Alessandra De Rosa et al. The quest for dual and binary supermassive black holes: A multi-messenger view. *New Astron. Rev.*, 86:101525, December 2019.

[63] Senwen Deng, Stanislav Babak, Maude Le Jeune, Sylvain Marsat, Éric Plagnol, and Andrea Sartirana. Modular global-fit pipeline for LISA data analysis. *Phys. Rev. D*, 111(10):103014, May 2025.

[64] Kallol Dey, Nikolaos Karnesis, Alexandre Toubiana, Enrico Barausse, Natalia Korsakova, Quentin Baghi, and Soumen Basak. Effect of data gaps on the detectability and parameter estimation of massive black hole binaries with LISA. *Phys. Rev. D*, 104(4):044035, August 2021.

[65] M. C. Díaz et al. Observations of the First Electromagnetic Counterpart to a Gravitational-wave Source by the TOROS Collaboration. ApJ, 848(2):L29, October 2017.

[66] Daniel J. D'Orazio and Maria Charisi. Observational Signatures of Supermassive Black Hole Binaries. In *Black Holes in the Era of Gravitational Wave Astronomy*. Elsevier, October 2023.

[67] M. Dotti, M. Colpi, S. Pallini, A. Perego, and M. Volonteri. On the Orientation and Magnitude of the Black Hole Spin in Galactic Nuclei. *Astrophys. J.*, 762(2):68, January 2013.

[68] M. Dotti, A. Sesana, and R. Decarli. Massive Black Hole Binaries: Dynamical Evolution and Observational Signatures. *Advances in Astronomy*, 2012:940568, January 2012.

[69] M. Dotti, M. Volonteri, A. Perego, M. Colpi, M. Ruszkowski, and F. Haardt. Dual black holes in merger remnants - II. Spin evolution and gravitational recoil. *Mon. Not. R. Astron. Soc.*, 402(1):682–690, February 2010.

[70] Arnaud Doucet, Nando De Freitas, and Neil Gordon. An introduction to sequential monte carlo methods. In *Sequential Monte Carlo methods in practice*, pages 3–14. Springer, 2001.

[71] Yohan Dubois, Marta Volonteri, and Joseph Silk. Black hole evolution - III. Statistical properties of mass growth and spin evolution using large-scale hydrodynamical cosmological simulations. *Mon. Not. R. Astron. Soc.*, 440(2):1590–1606, May 2014.

[72] Conor Durkan, Artur Bekasov, Iain Murray, and George Papamakarios. Neural Spline Flows. *Neural Information Processing Systems*, 33(675):7511, June 2019.

[73] Albert Einstein. Die grundlagen der allgemeinen. *Relativitats- teorie, Annale der Physic*, 49:769, 1916.

[74] Albert Einstein. Näherungsweise integration der feldgleichungen der gravitation. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften*, pages 688–696, 1916.

[75] Reed Essick and Maya Fishbach. Ensuring Consistency between Noise and Detection in Hierarchical Bayesian Inference. *Astrophys. J.*, 962(2):169, February 2024.

[76] N. Fanidakis, C. M. Baugh, A. J. Benson, R. G. Bower, S. Cole, C. Done, and C. S. Frenk. Grand unification of AGN activity in the ΛCDM cosmology. *Mon. Not. R. Astron. Soc.*, 410(1):53–74, January 2011.

[77] Ronald Aylmer Fisher. Statistical methods for research workers. In *Breakthroughs in statistics: Methodology and distribution*, pages 66–70. Springer, 1970.

[78] Cecilio García-Quirós, Marta Colleoni, Sascha Husa, Héctor Estellés, Geraint Pratten, Antoni Ramos-Buades, Maite Mateu-Lucena, and Rafel Jaume. Multimode frequency-domain model for the gravitational wave signal from nonprecessing black-hole binaries. *Phys. Rev. D*, 102(6):064002, September 2020.

[79] D. Gerosa, B. Veronesi, G. Lodato, and G. Rosotti. Spin alignment and differential accretion in merging black hole binaries. *Mon. Not. R. Astron. Soc.*, 451(4):3941–3954, August 2015.

[80] Davide Gerosa, Giovanni Rosotti, and Riccardo Barbieri. The Bardeen-Petterson effect in accreting supermassive black hole binaries: a systematic approach. *Mon. Not. R. Astron. Soc.*, 496(3):3060–3075, August 2020.

[81] Charles J Geyer. Practical markov chain monte carlo. *Statistical science*, pages 473–483, 1992.

[82] Walter R Gilks and Pascal Wild. Adaptive rejection sampling for gibbs sampling. *J. R. Stat. Soc. C*, 41(2):337–348, 1992.

[83] Jonathan Goodman and Jonathan Weare. Ensemble samplers with affine invariance. *Comm. App. Math. and Comp. Sci.*, 5(1):65–80, January 2010.

[84] Peter J Green. Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, 82(4):711–732, 1995.

[85] Stephen R. Green and Jonathan Gair. Complete parameter inference for GW150914 using deep learning. , August 2020.

[86] W. J. Handley, M. P. Hobson, and A. N. Lasenby. POLYCHORD: next-generation nested sampling. *mnras*, 453(4):4384–4398, November 2015.

[87] W. Keith Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[88] Bruno M. B. Henriques, Simon D. M. White, Peter A. Thomas, Raul Angulo, Qi Guo, Gerard Lemson, Volker Springel, and Roderik Overzier. Galaxy formation in the Planck cosmology - I. Matching the observed evolution of star formation rates, colours and stellar masses. *Mon. Not. R. Astron. Soc.*, 451(3):2663–2680, August 2015.

[89] Joeri Hermans, Volodimir Begy, and Gilles Louppe. Likelihood-free MCMC with Amortized Approximate Ratio Estimators. *arXiv e-prints*, page arXiv:1903.04057, March 2019.

[90] Niklas Houba, Luigi Ferraioli, and Domenico Giardini. Detection and mitigation of glitches in LISA data: A machine learning approach. *Phys. Rev. D*, 109(8):083027, April 2024.

[91] Niklas Houba, Giovanni Giarda, and Lorenzo Speri. SlotFlow: Amortized Trans-Dimensional Inference with Slot-Based Normalizing Flows. , November 2025.

[92] David Izquierdo-Villalba, Silvia Bonoli, Massimo Dotti, Alberto Sesana, Yetli Rosas-Guevara, and Daniele Spinoso. From galactic nuclei to the halo outskirts: tracing supermassive black holes across cosmic history and environments. *Mon. Not. R. Astron. Soc.*, 495(4):4681–4706, July 2020.

[93] David Izquierdo-Villalba, Alberto Sesana, Silvia Bonoli, and Monica Colpi. Massive black hole evolution models confronting the n-Hz amplitude of the stochastic gravitational wave background. *Mon. Not. R. Astron. Soc.*, 509(3):3488–3503, January 2022.

[94] David Izquierdo-Villalba, Alberto Sesana, Monica Colpi, Daniele Spinoso, Matteo Bonetti, Silvia Bonoli, and Rosa Valiante. Connecting low-redshift LISA massive black hole mergers to the nHz stochastic gravitational wave background. *Astron. Astrophys.*, 686:A183, June 2024.

[95] Robert E. Kass and Adrian E. Raftery. Bayes factors. *J. Am. Stat. Assoc.*, 90(430):773–795, 1995.

[96] Michael Katz et al. Lisaanalysistools. *Zenodo 10930980 github.com/mikekatz04/LISAanalysistools*, 2024.

[97] Michael L. Katz. Fully automated end-to-end pipeline for massive black hole binary signal extraction from LISA data. *Phys. Rev. D*, 105(4):044055, February 2022.

[98] Michael L. Katz, Nikolaos Karnesis, Natalia Korsakova, Jonathan R. Gair, and Nikolaos Stergioulas. Efficient GPU-accelerated multisource global fit pipeline for LISA data analysis. Phys. Rev. D, 111(2):024060, January 2025.

[99] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *International Conference for Learning Representations*, 2017.

[100] Antoine Klein, Enrico Barausse, Alberto Sesana, Antoine Petiteau, Emanuele Berti, Stanislav Babak, Jonathan Gair, Sofiane Aoudia, Ian Hinder, Frank Ohme, and Barry Wardell. Science with the space-based interferometer eLISA: Supermassive black hole binaries. *Phys. Rev. D*, 93(2):024003, January 2016.

[101] Annalena Kofler, Maximilian Dax, Stephen R. Green, Jonas Wildberger, Nihar Gupte, Jakob H. Macke, Jonathan Gair, Alessandra Buonanno, and Bernhard Schölkopf. Flexible Gravitational-Wave Parameter Estimation with Transformers. , December 2025.

[102] Solomon Kullback and Richard A Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86, 1951.

[103] Bo Liang, Minghui Du, He Wang, Yuxiang Xu, Chang Liu, Xiaotong Wei, Peng Xu, Li-e. Qiang, and Ziren Luo. Rapid parameter estimation for merging massive black hole binaries using continuous normalizing flows. *Mach. Learn. Sci. Tech.*, 5(4):045040, December 2024.

[104] LIGO Scientific Collaboration et al. Advanced LIGO. *Classical and Quantum Gravity*, 32(7):074001, April 2015.

[105] J. Lin. Divergence measures based on the shannon entropy. *IEEE Transactions on Information Theory*, 37(1):145–151, 1991.

[106] Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow Matching for Generative Modeling. , October 2022.

[107] LISA Science Study Team. LISA Science Requirements Document, ESA-L3-EST-SCI-RS-001, May 2018.

[108] Tyson B. Littenberg. Detection pipeline for Galactic binaries in LISA data. Phys. Rev. D, 84(6):063009, September 2011.

[109] Tyson B. Littenberg and Neil J. Cornish. Bayesian approach to the detection problem in gravitational wave astronomy. Phys. Rev. D, 80(6):063007, September 2009.

[110] Tyson B. Littenberg and Neil J. Cornish. Prototype global analysis of LISA data with multiple source types. *prd*, 107(6):063004, March 2023.

[111] G. Lodato and D. Gerosa. Black hole mergers: do gas discs lead to spin alignment? *Mon. Not. R. Astron. Soc.*, 429:L30–L34, February 2013.

[112] Gaia Lops, David Izquierdo-Villalba, Monica Colpi, Silvia Bonoli, Alberto Sesana, and Alberto Mangiagli. Galaxy fields of LISA massive black hole mergers in a simulated universe. *Mon. Not. R. Astron. Soc.*, 519(4):5962–5986, March 2023.

[113] Andrew I. MacFadyen and Miloš Milosavljević. An Eccentric Circumbinary Accretion Disk and the Detection of Binary Massive Black Holes. *Astrophys. J.*, 672(1):83–93, January 2008.

[114] M. Maggiore. *Gravitational waves - Volume 1: Theory and experiments.* Oxford University Press, 2008.

[115] S. G. Mallat and Zhifeng Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Process.*, 41(12):3397–3415, December 1993.

[116] Ilya Mandel, Will M. Farr, and Jonathan R. Gair. Extracting distribution parameters from multiple uncertain observations with selection biases. *Mon. Not. R. Astron. Soc.*, 486(1):1086–1093, June 2019.

[117] Alberto Mangiagli, Chiara Caprini, Marta Volonteri, Sylvain Marsat, Susanna Vergani, Nicola Tamanini, and Henri Inchauspé. Massive black hole binaries in LISA: Multimessenger prospects and electromagnetic counterparts. *Phys. Rev. D*, 106(10):103017, November 2022.

[118] Ruiting Mao, Jeong Eun Lee, and Matthew C. Edwards. Novel stacked hybrid autoencoder for imputing LISA data gaps. *Phys. Rev. D*, 111(2):024067, January 2025.

[119] Ruiting Mao, Jeong Eun Lee, and Matthew C. Edwards. Robust and scalable simulation-based inference for gravitational wave signals with gaps. , December 2025.

[120] Sylvain Marsat, John G. Baker, and Tito Dal Canton. Exploring the Bayesian parameter estimation of binary black holes with LISA. *Phys. Rev. D*, 103(8):083011, April 2021.

[121] Iván Martín Vílchez and Carlos F. Sopuerta. Efficient Massive Black Hole Binary parameter estimation for LISA using Sequential Neural Likelihood. *J. Cosmology Astropart. Phys.*, 2025(4):022, April 2025.

[122] D. V. Martynov et al. Sensitivity of the Advanced LIGO detectors at the beginning of gravitational wave astronomy. Phys. Rev. D, 93(11):112004, June 2016.

[123] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of State Calculations by Fast Computing Machines. J. Chem. Phys., 21(6):1087–1092, June 1953.

[124] Matthew Mould, Davide Gerosa, Marco Dall'Amico, and Michela Mapelli. One to many: comparing single gravitational-wave events to astrophysical populations. *Mon. Not. R. Astron. Soc.*, 525(3):3986–3997, November 2023.

[125] Matthew Mould, Christopher J. Moore, and Davide Gerosa. Calibrating signal-to-noise ratio detection thresholds using gravitational-wave catalogs. *Phys. Rev. D*, 109(6):063013, March 2024.

[126] Peter Muller. A genetic approach to posterior integration and gibbs sampling. *Department of Statistics, Purdue University*, 1991.

[127] Martina Muratore, Jonathan Gair, Olaf Hartwig, Michael L. Katz, and Alexandre Toubiana. Pipeline for searching and fitting instrumental glitches in LISA data. *Phys. Rev. D*, 112(6):063041, September 2025.

[128] Radford M. Neal. Mcmc using hamiltonian dynamics. In Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng, editors, *Handbook of Markov Chain Monte Carlo*, pages 113–162. Chapman and Hall/CRC, 2011.

[129] J. Neyman and E. S. Pearson. On the Problem of the Most Efficient Tests of Statistical Hypotheses. *Philos. Trans. R. Soc.*, 231:289–337, January 1933.

[130] M. Nicholl et al. The Electromagnetic Counterpart of the Binary Neutron Star Merger LIGO/Virgo GW170817. III. Optical and UV Spectra of a Blue Kilonova from Fast Polar Ejecta. ApJ, 848(2):L18, October 2017.

[131] Chris Nixon, Andrew King, and Daniel Price. Tearing up the disc: misaligned accretion on to a binary. *Mon. Not. R. Astron. Soc.*, 434(3):1946–1954, September 2013.

[132] Lucia Papalini, Federico De Santi, Massimiliano Razzano, Ik Siong Heng, and Elena Cuoco. Can Transformers help us perform parameter estimation of overlapping signals in gravitational wave detectors? *Class. Quantum Grav.*, 42(18):185012, December 2025.

[133] George Papamakarios and Iain Murray. Fast ϵ-free Inference of Simulation Models with Bayesian Conditional Density Estimation. *arXiv e-prints*, page arXiv:1605.06376, May 2016.

[134] George Papamakarios, David C. Sterratt, and Iain Murray. Sequential Neural Likelihood: Fast Likelihood-free Inference with Autoregressive Flows. *arXiv e-prints*, page arXiv:1805.07226, May 2018.

[135] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An Imperative Style, High-Performance Deep Learning Library. , December 2019.

[136] Noah Pearson and Neil J. Cornish. Handling Data Gaps for the Next Generation of Gravitational-Wave Observatories. , September 2025.

[137] P. C. Peters and J. Mathews. Gravitational Radiation from Point Masses in a Keplerian Orbit. *Phys. Rev.*, 131(1):435–440, July 1963.

[138] Planck Collaboration, P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, C. Baccigalupi, A. J. Banday, and et al. Planck 2013 results. XVI. Cosmological parameters. *Astron. Astrophys.*, 571:A16, November 2014.

[139] Geraint Pratten, Sascha Husa, Cecilio García-Quirós, Marta Colleoni, Antoni Ramos-Buades, Héctor Estellés, and Rafel Jaume. Setting the cornerstone for a family of models for gravitational waves from compact binaries: The dominant harmonic for nonprecessing quasicircular black holes. *prd*, 102(6):064001, September 2020.

[140] Geraint Pratten, Antoine Klein, Christopher J. Moore, Hannah Middleton, Nathan Steinle, Patricia Schmidt, and Alberto Vecchio. On the LISA science performance in observations of short-lived signals from massive black hole binary coalescences. , 107(12):123026, December 2022.

[141] Gerald D. Quinlan. The dynamical evolution of massive black hole binaries I. Hardening in a fixed stellar background. *New Astron.*, 1(1):35–56, July 1996.

[142] Daniel J. Reardon et al. Search for an Isotropic Gravitational-wave Background with the Parkes Pulsar Timing Array. *Astrophys. J. Lett.*, 951(1):L6, July 2023.

[143] Christian Ritter and Martin A Tanner. Facilitating the gibbs sampler: the gibbs stopper and the griddy-gibbs sampler. *J. Am. Stat. Assoc.*, 87(419):861–868, 1992.

[144] Travis Robson and Neil J. Cornish. Detecting gravitational wave bursts with lisa in the presence of instrumental glitches. *prd*, 99(2):024019, January 2019.

[145] Filippo Santoliquido, Jacopo Tissino, Ulyana Dupletsa, Marica Branchesi, Jan Harms, Manuel Arca Sedda, Maximilian Dax, Annalena Kofler, Stephen R. Green, et al. Fast and accurate parameter estimation of high-redshift sources with the Einstein Telescope. *Phys. Rev. D*, 112(10):103015, November 2025.

[146] Argyro Sasli, Minas Karamanis, Nikolaos Karnesis, Michael W. Coughlin, Vuk Mandic, Uroš Seljak, and Nikolaos Stergioulas. Beyond Gaussian Assumptions: A new robust statistical framework for gravitational-wave data analysis. *arXiv e-prints*, page arXiv:2602.22074, February 2026.

[147] Argyro Sasli, Nikolaos Karnesis, and Nikolaos Stergioulas. Heavy-tailed likelihoods for robustness against data outliers: Applications to the analysis of gravitational wave data. Phys. Rev. D, 108(10):103005, November 2023.

[148] D. W. Scott. *Multivariate Density Estimation: Theory, Practice, and Visualization.* 2015.

[149] A. Sesana. Gravitational wave emission from binary supermassive black holes. *Class. Quantum Grav.*, 30(24):244009, December 2013.

[150] A. Sesana, E. Barausse, M. Dotti, and E. M. Rossi. Linking the Spin Evolution of Massive Black Holes to Galaxy Kinematics. *Astrophys. J.*, 794(2):104, October 2014.

[151] John Skilling. Nested sampling for general Bayesian computation. *Bayesian Analysis*, 1(4):833–859, 2006.

[152] M. Soares-Santos et al. The Electromagnetic Counterpart of the Binary Neutron Star Merger LIGO/Virgo GW170817. I. Discovery of the Optical Counterpart Using the Dark Energy Camera. ApJ, 848(2):L16, October 2017.

[153] Alice Spadaro and Riccardo Buscicchio. *github.com/RiccardoBuscicchio/lisa-mbhb-cats-and-samps, doi.org/10.5281/zenodo.13787675*, 2024.

[154] Alice Spadaro, Riccardo Buscicchio, David Izquierdo–Villalba, Davide Gerosa, Antoine Klein, and Geraint Pratten. Stars or gas? Constraining the hardening processes of massive black-hole binaries with LISA. Phys. Rev. D, 111(2):023004, February 2025.

[155] Alice Spadaro, Riccardo Buscicchio, Daniele Vetrugno, Antoine Klein, Davide Gerosa, Stefano Vitale, Rita Dolesi, William Joseph Weber, and Monica. Colpi. Glitch systematics on the observation of massive black-hole binaries with LISA. *Phys. Rev. D*, 108(12):123029, December 2023.

[156] D. Spinoso, S. Bonoli, R. Valiante, R. Schneider, and D. Izquierdo-Villalba. Multiflavour SMBH seeding and evolution in cosmological environments. *Mon. Not. R. Astron. Soc.*, 518(3):4672–4692, January 2023.

[157] Wolfgang Steinicke. Einstein and the gravitational waves. *Astronomische Nachrichten*, 326(7):640–641, 2005.

[158] Nathan Steinle, Hannah Middleton, Christopher J. Moore, Siyuan Chen, Antoine Klein, Geraint Pratten, Riccardo Buscicchio, Eliot Finch, and Alberto Vecchio. Implications of pulsar timing array observations for LISA detections of massive black hole binaries. *Mon. Not. R. Astron. Soc.*, 525(2):2851–2863, October 2023.

[159] Stefan H. Strub, Luigi Ferraioli, Cédric Schmelzbach, Simon C. Stähler, and Domenico Giardini. Global analysis of LISA data with Galactic binaries and massive black hole binaries. *Phys. Rev. D*, 110(2):024005, July 2024.

[160] Ralph S. Sutherland and M. A. Dopita. Cooling Functions for Low-Density Astrophysical Plasmas. *Astrophys. J. Supp. S.*, 88:253, September 1993.

[161] Robert H. Swendsen and Jian-Sheng Wang. Replica Monte Carlo simulation of spin glasses. Phys. Rev. Lett., 57(21):2607–2609, November 1986.

[162] The LIGO Scientific Collaboration, the Virgo Collaboration, the KAGRA Collaboration, et al. GWTC-4.0: Updating the Gravitational-Wave Transient Catalog with Observations from the First Part of the Fourth LIGO-Virgo-KAGRA Observing Run. *arXiv e-prints*, page arXiv:2508.18082, August 2025.

[163] Kip S. Thorne. Disk-Accretion onto a Black Hole. II. Evolution of the Hole. *Astrophys. J.*, 191:507–520, July 1974.

[164] Wolfgang Tichy and Pedro Marronetti. Final mass and spin of black-hole mergers. *Phys. Rev. D*, 78(8):081501(R), October 2008.

[165] Alexandre Le Tiec and Jérôme Novak. Theory of gravitational waves. In *An Overview of Gravitational Waves*, pages 1–41. WORLD SCIENTIFIC, February 2017.

[166] Massimo Tinto and Sanjeev V. Dhurandhar. Time-delay interferometry. *Living Reviews in Relativity*, 24(1):1, December 2021.

[167] Stefano Valenti, David J. Sand, Sheng Yang, Enrico Cappellaro, Leonardo Tartaglia, Alessandra Corsi, Saurabh W. Jha, Daniel E. Reichart, Joshua Haislip, and Vladimir Kouprianov. The Discovery of the Electromagnetic Counterpart of GW170817: Kilonova AT 2017gfo/DLT17ck. ApJ, 848(2):L24, October 2017.

[168] Salvatore Vitale, Davide Gerosa, Will M. Farr, and Stephen R. Taylor. Inferring the Properties of a Population of Compact Binaries in Presence of Selection Effects. In *Handbook of Gravitational Wave Astronomy*, page 45. 2022.

[169] A. von Kienlin, P. Veres, O. J. Roberts, R. Hamburg, E. Bissaldi, M. S. Briggs, E. Burns, A. Goldstein, D. Kocevski, R. D. Preece, C. A. Wilson-Hodge, C. M. Hui, B. Mailyan, and C. Malacaria. Fermi-GBM GRBs with Characteristics Similar to GRB 170817A. ApJ, 876(1):89, May 2019.

[170] W. D. Vousden, W. M. Farr, and I. Mandel. Dynamic temperature selection for parallel tempering in Markov chain Monte Carlo simulations. MNRAS, 455(2):1919–1937, January 2016.

[171] S. D. M. White and M. J. Rees. Core condensation in heavy halos: a two-stage theory for galaxy formation and clustering. *Mon. Not. R. Astron. Soc.*, 183:341–358, May 1978.

[172] Michael J. Williams, John Veitch, and Chris Messenger. Nested sampling with normalizing flows for gravitational-wave inference. *Phys. Rev. D*, 103(10):103006, May 2021.

[173] Michael J. Williams, John Veitch, and Chris Messenger. Nested sampling with normalizing flows for gravitational-wave inference. *Phys. Rev. D*, 103(10):103006, May 2021.

[174] Yan Xia, Guangyu Li, Gerhard Heinzel, Albrecht Rudiger, and Yongjie Luo. Orbit design for the Laser Interferometer Space Antenna (LISA). *Sci. China Phys. Mech. Astron.*, 53:179–186, 2010.

[175] Heng Xu et al. Searching for the Nano-Hertz Stochastic Gravitational Wave Background with the Chinese Pulsar Timing Array Data Release I. *Research in Astronomy and Astrophysics*, 23(7):075024, July 2023.

List of Figures

1.1 Characteristic quadrupolar pattern of the gravitational radiation. 3
1.2 GW spectrum and the corresponding observational landscape. 5
1.3 LISA astrophysical landscape. 6
1.4 Orbital configuration of the LISA mission. 8
1.5 Schematic of the single laser links 9
1.6 Schematic overview of a global-fit for LISA data analysis. 14
1.7 Nested-sampling volume transformation. 17
1.8 Forward and reverse KL divergence. 19

2.1 Density distribution of the total mass M and redshift z of massive BH binary mergers. 27
2.2 Distribution of the gas fraction $f_{\rm gas}$ of the environment across the fiducial population. 28
2.3 Distribution of massive BH binary mergers across the generated catalogs. 31
2.4 LISA waterfall plot. 32
2.5 Logarithmic Bayes factor between the gas- and stellar- hardening hypotheses as a function of the source SNR. 35
2.6 Marginal distribution of the aligned spin components. 37

3.1 Schematic of single laser links and glitch reference system conventions. . 43
3.2 Match between the GW and glitch signals as a function of onset time. . . 46
3.3 Fiducial waveforms for our parameter-estimation runs. 47
3.4 Corner plot Model A1 glitch and MBHB. 56
3.5 Corner plot Model A2 glitch and MBHB. 57
3.6 Corner plot Model D glitch and MBHB. 58
3.7 Corner plots for single glitch-only analyses. 61
3.8 Corner plot for two Model A2 glitches. 62
3.9 Corner plot for two Model D glitches. 63

4.1 Training and validation losses. 68
4.2 p–p plot for 1000 simulated injections. 70
4.3 Comparative corner plot for the low-SNR source. 72
4.4 Comparative corner plot for the moderate-SNR source. 74
4.5 Comparative corner plot for the high-SNR source. 76
4.6 Prior distribution of the training set. 78

List of Tables

3.1 Summary of the runs containing single glitches and the fiducial GW signal. 50
3.2 Summary of a large set of injected glitches and associated recoveries. . . 51
3.3 Parameter estimation results for a GW signal contaminated by a Model A1 glitch. 54
3.4 Parameter estimation results for a GW signal contaminated by a Model A2 glitch. 54
3.5 Parameter estimation results for a GW signal contaminated by a Model D glitch. 54
3.6 Parameter-estimation results for Model A1 and Model A2 glitches. . . . 60
3.7 Parameter-estimation results for Model D glitches. 60

4.1 Injected parameter values for the three representative sources. 71