A Glitch and the Matrix: Advances in gravitational-wave glitch mitigation and acceleration of pulsar timing analyses

Thesis by

Sophie Rose Hourihane

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2025
Defended May 12th 2025

Sophie Rose Hourihane
ORCID: 0000-0002-9152-0719

ACKNOWLEDGEMENTS

I would first like to thank my advisor, Katerina Chatziioannou. Being your first graduate student was likely not always easy; I appreciated your availability to talk through my (usually last-minute) concerns, for pushing me to actually finish projects that I would have left 90% completed and unpublished, and for allowing me the freedom granted to join and contribute to multiple gravitational-wave collaborations.

Thank you to my committee, Alan Weinstein, Michele Vallisneri, and Saul Teukolsky.

I would also like to thank my collaborators: The BayesWave and BayesWave C++ team, particularly Howard Deshong, Meg Millhouse, Tyson Littenberg, Katerina Chatziioannou, James Clark, Aaron Johnson, Sophie Bini, and Sudarshan Ghonge.

Thank you to my SURF students: Cailin Plunkett, Julien Kearns, Maven Holst, and Seth Moriarty for patience while I figured out how to mentor a project.

Thank you to the other LIGO Caltech graduate students and postdocs with whom I have shared a lot of time, frustrations, science, and discoveries. Shout out especially to Simona Miller, Jacob Golomb, Isaac Legred, Rhiannon Udall, Ian MacMillan, Alvin Li, Ethan Payne, Taylor Knapp, Alex Ramirez, Rico Lo, Anchal Gupta, Shruti Maliakal, and Radhika Bhatt as well as to the postdocs Patrick Meyers, Derek Davis, Colm Talbot, Ryan Magee, Eliot Finch, Jane Glanzer, Lucy Thomas, Sophie Bini, Aaron Johnson, Arianna Renzini, and also honorarily Sylvia Biscoveanu.

Helping to organize the Caltech Graduate Student and Postdoc Union (now UAW local 2478) has been one of the most meaningful endeavors I have been a part of. Together we won one of the strongest contracts in higher education, particularly around protections from harassment and abuse, setting a new industry standard. Thank you to Piril Nergis and Simona Miller for getting me involved in the first place. Thank you to the communications team, especially Aditi Narayanan, Jasmine Emtage, Piril Nergis, and Rohan Shenoy for the late nights and deep statements. Thanks also to Patrick Meyers, Dee Dunne, Sam Ponnada, Jakob Faber, George Zhou, Mike McKeon, Kyle Nelli, Harshda

Saxena, Jack Wilding, Ruth Moormon, Ian Anderson, Aaron Weaver, Richard Horak, Dan Pollak, Nikos Mynhier, Jess Griffiths, Varun Wadia, Desi Ascencio, Marina Lecoeuche, David Abramovitch, Angela Gao, Tom Naragon, Korbinian Thalhammer, Tessa Rusch, and Emma Lenz for helping build and strengthen a union that will (hopefully) outlast us all.

Thank you also to my co-organizers in the gravitational-wave solidarity group, particularly Amanda Farah.

Thank you also to my Caltech RISE mentee, Sofia Barrios, for being a light in my week and to Liz Jackman for putting us together.

Thank you to my parents for supporting me to become *this* kind of doctor, and to my five siblings—Mimi, Eoin, Max, Naomi, and Ali—for developing in me the mental fortitude to finish a PhD. Mom, I hope you are not still afraid of black holes.

Shout out to Patrick Meyers who made both me and my time here better.

Thank you to Piril Nergis who fundamentally altered my life path.

Thank you to Simona Miller, without whom I do not believe I could have finished my degree.

SH was supported by the National Science Foundation Graduate Research Fellowship under Grant DGE-1745301.

This material is based upon work supported by NSF's LIGO Laboratory which is a major facility fully funded by the National Science Foundation. LIGO was constructed by the California Institute of Technology and Massachusetts Institute of Technology with funding from the National Science Foundation, and operates under cooperative agreement PHY-2309200. The authors are grateful for computational resources provided by the LIGO Laboratory and supported by National Science Foundation Grants PHY-0757058 and PHY-0823459.

ABSTRACT

Since the first detection of gravitational-waves in 2015, the field of gravitational-wave astronomy has developed rapidly. Today, there are more than 300 transient gravitational-wave event candidates from stellar-mass sources and we have found evidence for a stochastic background of supermassive black-holes. In this thesis I present work addressing two significant challenges on analyzing these data. The first: mitigating transient, non-Gaussian noise in gravitational-wave detectors, or "glitches", that can bias our estimates of physical properties of compact objects. The second: introducing a faster method to analyze pulsar-timing data containing a stochastic background of supermassive black-hole sources. Gravitational-wave astronomy is a data-rich field, and is only becoming more so with upgraded detectors, additional detectors, and longer observing time; we need robust, fast, and unbiased techniques to analyze that data.

PUBLISHED CONTENT AND CONTRIBUTIONS

[1] Gabriella Agazie et al. The NANOGrav 15 yr Data Set: Evidence for a Gravitational-wave Background. *Astrophys. J. Lett.*, 951(1):L8, 2023. doi: 10.3847/2041-8213/acdac6. **SH** performed analysis comparing Hellings-Downs to Hellings-Downs + sinusoid and computed detection Bayes Factors.

[2] Gabriella Agazie et al. The NANOGrav 15 yr Data Set: Bayesian Limits on Gravitational Waves from Individual Supermassive Black Hole Binaries. *Astrophys. J. Lett.*, 951(2):L50, 2023. doi: 10.3847/2041-8213/ace18a. **SH** performed analysis reweighting likelihoods from those excluding to those including Hellings-Downs correlations (both containing individual SMBHBs).

[3] Bence Bécsy et al. How to Detect an Astrophysical Nanohertz Gravitational Wave Background. *Astrophys. J.*, 959(1):9, 2023. doi: 10.3847/1538-4357/ad09e4. **SH** computed detection statistics for simulations.

[4] Sudarshan Ghonge, Joshua Brandt, J. M. Sullivan, Margaret Millhouse, Katerina Chatziioannou, James A. Clark, Tyson Littenberg, Neil Cornish, **Sophie Hourihane**, and Laura Cadonati. Characterizing the efficacy of methods to subtract terrestrial transient noise near gravitational wave events and the effects on parameter estimation. *Accepted: Phys. Rev. D*, 11 2023. **SH** suggested analyses and advised on BayesWave settings.

[5] Aaron D. Johnson et al. NANOGrav 15-year gravitational-wave background methods. *Phys. Rev. D*, 109(10):103012, 2024. doi: 10.1103/PhysRevD.109.103012. **SH** authored section describing reweighting methods.

[6] Ethan Payne, **Sophie Hourihane**, Jacob Golomb, Rhiannon Udall, Richard Udall, Derek Davis, and Katerina Chatziioannou. Curious case of GW200129: Interplay between spin-precession inference and data-quality issues. *Phys. Rev. D*, 106(10):104017, 2022. doi: 10.1103/PhysRevD.106.104017. Reprinted here as Chapter 6. **SH** helped conceptualize the project, led all BayesWave analyses, and created related figures, and authored related text.

[7] Cailin Plunkett, **Sophie Hourihane**, and Katerina Chatziioannou. Concurrent estimation of noise and compact-binary signal parameters in gravitational-wave data. *Phys. Rev. D*, 106(10):104021, 2022. doi: 10.1103/PhysRevD.106.104021. **SH** mentored first author (undergraduate student).

[8] S. Soni et al. LIGO Detector Characterization in the first half of the fourth Observing run. *arXiv e-prints*, 9 2024. **SH** led LIGO O4 glitch-subtraction analyses.

[9] **Sophie Hourihane** and Katerina Chatziioannou. Glitches far from transient gravitational-wave events do not bias inference. *In Prep*, 2025. Currently in preparation, presented here in 8. **SH** conceptualized the project, led all analyses, created all figures, and authored the text.

[10] **Sophie Hourihane**, Katerina Chatziioannou, Marcella Wijngaarden, Derek Davis, Tyson Littenberg, and Neil Cornish. Accurate modeling and mitigation of overlapping signals and glitches in gravitational-wave data. *Phys. Rev. D*, 106(4):042006, 2022. doi: 10.1103/PhysRevD.106.042006. Reprinted here as Chapter 5. **SH** performed all analyses, authored the text, helped conceptualize and scope the project, and created all figures.

[11] **Sophie Hourihane**, Patrick Meyers, Aaron Johnson, Katerina Chatziioannou, and Michele Vallisneri. Accurate characterization of the stochastic gravitational-wave background with pulsar timing arrays by likelihood reweighting. *Phys. Rev. D*, 107(8):084045, 2023. doi: 10.1103/PhysRevD.107.084045. Printed here as Chapter 10. **SH** conceptualized the project, authored the text, performed all analyses except for P-P tests, and all figures.

[12] Rhiannon Udall, **Sophie Hourihane**, Simona Miller, Derek Davis, Katerina Chatziioannou, Max Isi, and Howard Deshong. Antialigned spin of GW191109: Glitch mitigation and its implications. *Phys. Rev. D*, 111(2):024046, 2025. doi: 10.1103/PhysRevD.111.024046. Reprinted as Chapter 7. **SH** helped conceptualize the project, led all BayesWave analyses, created related figures, and authored related text.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

ACRONYMS

KAGRA Kamioka Gravitational Wave Detector 2, 10

LHO LIGO Hanford observatory xi, 12

LIGO Laser Interferometer Gravitational-Wave Observatory [1] 2, 5, 10, 169

LLO LIGO Livingston observatory xi, 12

LVK LIGO Virgo KAGRA 2, 4, 7, 9, 15, 16, 22, 24, 168, 169

NANOGrav North American Nanohertz Observatory for Gravitational Waves xxx, 4, 173

O2 LVK's second observing run (November 30 2016 - August 25th 2017) 24

O3 LVK's third observing run (April 1, 2019 - March 27, 2020) 22–24

O4a The first half of the LVK's fourth observing run lasting from May 24th 2023 - January 16th, 2024 22, 24

PSD power spectral density 172, 176

SGWB stochastic gravitational-wave background xxx, 5, 20, 168, 169, 171–176

SMBH supermassive black-hole 168

Virgo Virgo observatory [43] 2, 10

Prologue

Chapter 1

OVERVIEW OF GRAVITATIONAL-WAVE ASTRONOMY

When massive stars reach the end of their lives they collapse, then explode ("supernova"), leaving behind remnants, either black holes or neutron stars [188]. Neutron stars represent the densest state of matter that can exist before collapsing into black holes—regions of space where matter becomes so dense that not even light can escape. The term "black hole" is somewhat misleading, as it suggests emptiness, when in fact black holes are regions of such extreme compactness that not even light can escape. When these compact objects orbit each other, their immense gravitational pull causes them to emit gravitational waves—ripples in the fabric of space-time itself. This gravitational radiation carries energy away from the system, causing the objects to spiral inward and eventually combine, or "merge".

Since the first direct detection of gravitational waves in 2015 [20], the global detector network operated by the LIGO Virgo KAGRA (LVK) collaboration has identified more than 200 "mergers" of compact objects [38, 116]. This network includes the Laser Interferometer Gravitational-Wave Observatory [1] (LIGO) with detectors in Hanford, Washington and Livingston, Louisiana, the Virgo observatory [43] (Virgo) in Italy, the GEO600 detector in Germany [48, 149], and the Kamioka Gravitational Wave Detector (KAGRA) in Japan [55, 70, 321].

This ever-growing set of detections allows us to study not only the properties of individual black holes and neutron stars (e.g. their masses and angular momenta (spins)) in binaries, but also the distributions of these properties across the astrophysical population. From these observations, we have learned that compact binaries observable by the LVK can have masses ranging from $\sim 1 - 100\,\mathrm{M}_{\odot}$ and that, on average, they have small spins that are generally aligned with the orbital angular momentum [15, 34, 35].

Current analyses have yet to conclusively determine the origin of the binary black holes observable by the LVK, although there are two leading theories. The first is *isolated binary evolution*, in which two stars orbit together and, after collapsing *individually* into black holes, remain gravitationally bound.

This scenario tends to predict lower total masses and spins that are primarily aligned, though supernova "kicks" may introduce some isotropy [e.g. 74, 75, 168, 205, 239, 241, 277, 286, 350]. The second theory is *dynamical formation* in dense stellar environments, like nuclear or stellar clusters, which predicts higher masses, isotropic spin distributions, and potentially even second-generation mergers[1] [e.g 247, 276, 294, 319]. Understanding the distribution of spin, in particular, is one of the most promising ways to understand binary black hole formation and evolution: spin provides unique insight into angular momentum transfer, mass transfer, and tidal interactions within stellar binary systems [e.g. 75, 165, 286, 326, 390].

Gravitational waves also can contain novel information about matter at nuclear densities. Neutron stars are objects the size of cities with masses $\sim 1.4\ M_{\odot}$, meaning that they are *more dense nucleus of an atom* [258]. When neutron stars are in orbit, they pull and stretch each other "tidally"[2], which modifies their orbit, and thus the gravitational-wave signal, in an observable way. In this way, gravitational-waves from binary neutron-star signals probe how much this nuclear-density matter, which is impossible to create in a laboratory, can bend and stretch [9]. The first detection of gravitational-waves from a binary neutron star merger[3] was actually detected not only as a gravitational-wave signal but also from its optical "kilonova" signal, detected with light by traditional telescopes [18]. This signal, because it was measured both as gravitational-radiation *and* as light, allowed us to constrain the Hubble constant H_0, a measurement of the expansion-rate of the universe, in a completely independent way from historic methods [8].

Gravitational waves are also used to "test" Einstein's theory of general relativity. We use general relativity to create models or "waveforms" that predict what we expect gravitational-waves to look like for systems with some specified parameters: mass, spin, etc. By measuring the deviations between what general relativity predicts and what we measure, we can use our observations

[1]that is, mergers of binary black holes that were themselves created from *other* binary black hole mergers.

[2]To picture this, recall how the gravitational pull from the moon cause the ocean to rise and fall. Now imagine that the moon had the mass of the sun, was the size of a city, and was orbiting the earth 100 times per second. These tidal forces are *strong*.

[3]The first evidence of *gravitational-radiation* however was in the form of a the shrinking orbit of a neutron star (specifically a pulsar) in orbit with another neutron star [196] which was consistent with the predictions of general-relativity [331].

to test the accuracy of the theory. So far general relativity has proved itself to be a remarkably accurate theory [6, 11, 30, 32][4].

Transient gravitational waves from solar-mass compact objects are not the only sources of gravitational waves. Supermassive black-hole binaries (SMBHBs) are galaxy-mass black holes, with masses more than a billion times the mass of our sun, $10^9 - 10^{10}\ M_\odot$ formed by the inspirals of galaxies themselves. Unlike the stellar-mass black holes that merge in seconds[5], these black holes are far from their final merger, and orbit one other anywhere between a few times a year to once a decade [51, 65]. However, unlike the observations of individual black holes as measured by the LVK collaboration, NANOGrav instead is looking for a "stochastic background" of gravitational-waves from a combination of all SMBHBs in the Universe. The background is "stochastic" or random because, unlike LVK-like signals, the individual black hole systems are not resolved. The background is created not by one but by thousands to tens of thousands of these SMBHBs all inspiraling between us on Earth (at redshift 0) to galaxies at redshift 2, roughly 10 billion years ago [304]. This means that NANOGrav is sensitive to gravitational waves that were emitted *6 billion* years before the earth even existed) [6]. Though this background has not yet been "detected" per se, there is strong evidence that the signal measured by Pulsar Timing Arrays is indeed gravitational waves from SMBHBs[8] [50]. If the stochastic gravitational-wave background signal is detected, it will be the first proof that SMBHBs do form, evolve to sub-parsec distances, and eventually coalesce [51]. If instead the gravitational-wave background has a different origin, it would prove an exciting challenge to the standard model and cosmology.

[4]Despite what the bi-weekly emails I receive might have you believe.

[5]Merge in seconds in the LVK frequency band; they inspiral for much longer.

[6]Redshift, z, is a fundamental concept in astronomy and better known in other fields as a Doppler-shift, the result of relative motion between the emitter and the observer[7]. Since the universe is expanding at an accelerating rate, light (and gravitational-waves) get stretched out to lower frequencies by a factor of $1/(1+z)$, the ratio of the size of the Universe at creation and the size at observation. This effect dominates any other Doppler shifts in all but the most local universe. Because of this, z is directly related to other cosmological quantities associated with the expansion of the universe like distance, the age of the universe at emission, and the volume of the universe, among others. A redshift of $z = 1$ is, for instance, the distance away that light (or gravitational-waves) has to come from to have its wavelength stretched by a factor of two (or identically, its frequency halved). Earth is located at redshift $z = 0$.

[7]We witness, or more, hear Doppler shifts every day; for instance an ambulance traveling away from you has a lower pitch than an ambulance heading towards you.

[8]The method by which they are detected I discuss further in Chapter 9

Chapter 2

OVERVIEW OF THESIS

In this thesis I explore analysis methods across the gravitational-wave spectrum. I will focus on two topics—in Part I (Chapters 4–8) I analyze transient gravitational wave data in the presence of non-Gaussian detector noise (glitches) and in Part II (Chapters 9 and 10) I present work that helped researchers studying the Nano Hz SGWB speed up some analyses by orders of magnitude, which assisted in finding evidence of the SGWB [50].

In Chapter 4, I provide a brief introduction to gravitational waves, focusing on transient signals (since I cover the SGWB in Part II). I discuss what gravitational waves are, the astrophysical phenomena that produce them, and how they are detected. The chapter then transitions to an overview of parameter estimation methods for gravitational waves, the techniques employed, and the challenges involved—particularly the presence of non-Gaussian noise in detectors, commonly referred to as "glitches."

Chapter 5 introduces a method for mitigating glitches while accounting for the uncertainty in the mitigation process—a critical component that has been absent in previous LIGO collaboration analyses. Following this, Chapter 6 and Chapter 7 explore the impact of glitches on real gravitational-wave events GW200129 and GW191109, respectively, illustrating how different glitch mitigation strategies can profoundly alter the astrophysical conclusions drawn. Finally, Chapter 8 examines when and where within the gravitational-wave analysis window glitches influence inference, providing a more nuanced understanding of their effects.

In the second part of the thesis I begin with an introduction to the SGWB in Chapter 9, what it is, the SMBHBs that likely produce it, and how it is analyzed. I then discuss my contributions to the detection of the SGWB in Chapter 10.

Finally, I summarize the thesis and future work in Chapter 10.7.

Part I

Gravitational wave transients: Analysis and challenges

Chapter 3

OVERVIEW OF PART I

In Chapter 4 I give a brief introduction to gravitational waves, their detection, how we measure them, and difficulties thereof, primarily, the difficulty being non-Gaussian noise, or glitches. I also introduce Bayesian parameter estimation.

In Chapter 5 I present a project that I helped conceptualize, ran all the analyses for and was the lead paper-writer. There I expanded upon a method originated in [110] that proposed a method to include the uncertainty from glitch mitigation.

In Chapter 6 and Chapter 7 I present applications of the analysis introduced in Chapter 5 on real LVK events. Here we show that that astrophysical conclusions depend quite heavily on the glitch model used in the analysis meaning that for events coincident with glitches, the noise model is incredibly important to astrophyical conclusions.

In Chapter 8 I analyze when and where glitches in the gravitational-wave detectors can actually bias parameter estimation, and crucially where they cannot.

Chapter 4

BACKGROUND ON TRANSIENT ANALYSES

In Part I I will cover the analysis and challenges associated with the recovery of gravitational-wave parameters from transient events. For the non-expert, in this section I will cover first what are gravitational waves, what produce them, and how can we measure their properties.

4.1 What are gravitational-waves? Explanation by analogy

For a mathematical description of gravitational waves as metric perturbations in general relativity, please see e.g. Maggiore [238]; here I only attempt to build intuition.

Gravitational waves are ripples in spacetime. In Einstein's theory of general relativity, gravity is explained as masses curving spacetime. That is the reason, for instance, that we, as massive objects[1], feel gravity. Imagine a pool ball (the earth) on a thin, stretchy surface held above the ground. An ant on that surface would fall towards the pool ball, similar to us falling towards the earth. The pool ball curves that stretchy surface around it, but the further away from that pool ball, the less curved the surface is. The moon, for instance, is both less massive and less dense than the Earth, meaning that it curves spacetime less, and the pull of gravity is less strong on the moon.

If we now send a marble rolling towards that pool ball at an angle such that it rolls around the ball, it will fall into something looking like an orbit because the fabric is warped. Unlike the Newtonian theory of gravity where two masses can orbit each other indefinitely, in general relativity orbits radiate energy; in our fabric system, it takes energy to move around on that stretchy surface. The marble and pool ball are moving and accelerating relative to each other and warping that stretchy fabric as they do so. The energy to send ripples through the fabric then must come from somewhere, on our fabric system it will likely just come from the kinetic energy of the marble and likely a bit of gravitational energy as it falls down the potential well made by the pool ball; in orbital systems the energy comes from the gravitational-attraction between

[1]Here meaning objects with mass.

them.

An observer nearby these two masses might have a difficult time describing the rapidly changing ripples nearby[2] but as one gets further away from the source of the ripples the description of them gets simpler. As the marble falls inwards it will orbit faster and faster around the pool ball meaning that any ripples (or "waves") observed far away will come at smaller and smaller time increments. Finally they will collide and the orbit will stop, meaning that the ripples will as well. If you can picture this simplified example, then you can understand the basic physics of gravitational waves from binary black holes.

Creating observable ripples in spacetime requires much more energy than can be stored in a pool ball. For these ripples in spacetime we need incredibly massive objects. But, recall that as the marble moved closer it started to radiate more energy. So having objects be massive and close to each other will make the signals as loud as possible. Recall also that as the marble and pool ball got too close, they hit each other and stopped moving. So the objects need to not only be massive, but also dense enough that they can get extremely close without touching. While not at all relevant in the marble case, we also need objects that are bound together tightly enough such that they do not rip each other apart as they get closer and closer together, a process known as tidal disruption.

4.2 Detection of gravitational-waves

While any accelerating masses can technically create gravitational waves, the processes that generate measurable gravitational waves on Earth are some of the most energetic in the universe. However, just as a light bulb looks dimmer as you get further from it, gravitational waves get weaker (or lower-amplitude) as they travel away from the source. The sources the LVK measures are billions of light-years away[3]; by the time they reach the LVK interferometers, the gravitational waves change the size of the detector by a distance that is 10,000 times smaller than the nucleus of an atom [20].

The LVK detects gravitational waves as the change of spacetime (measured as an *actual* change of length) of the arms of a Michelson interferometer. Michel-

[2] This area is known as the "near field" or "strong field".

[3] The binary black hole merger associated with the first gravitational wave observation, GW150914, merged 1.4 billion years ago. For a sense of time, on Earth $\sim$ 1.4 billion years ago, the first eukaryotic cells evolved.

son interferometers function as follows: a laser beam, which can be thought of as a simple sinusoidal wave, is split into two beams of equal intensity. Each beam travels along one of two arms of the interferometer which form a 90° angle. These arms extend 4 km, each ending with a mirror that reflects the laser back the source where the light from each arm recombines at a photodiode. If the mirrors are equidistant from the beam splitter, the two beams traverse identical path lengths. Upon recombination at the beam splitter, the waves remain in phase, constructively interfering to produce a measurable signal. If the path lengths differ, for instance, by half the wavelength of the beam, they would interfere destructively. In this way by measuring the intensity of the recombined light, Michelson interferometers can measure distance to incredible precision[4].

A gravitational wave passing over such an interferometer would change the length of the arms, stretching and compressing them as it propagates. Depending on the direction and orientation of the wave, each arm will change different amounts. The waves coming over the detector will cause the recombined light at the photodetector to pulse, the frequency of that pulsing in turn gives us a measurement of the frequency of the gravitational wave.

Gravitational waves were first directly detected on September 14, 2015, by the LIGO interferometers. This landmark event, designated GW150914, follows the convention of naming transient gravitational-wave events based on their detection date [20]. Since this initial discovery, the global network of gravitational-wave observatories has expanded with the inclusion of detectors Virgo and KAGRA. Together, these instruments have enabled the detection of, as of this writing, around 300 gravitational-wave events [116], with many more expected in upcoming observing runs.

4.3 Gravitational Waveforms

Given the physical properties of a compact binary system (like masses, spins, distance, etc), general relativity predicts the orbits of the compact objects and the associated gravitational waves radiated. This waveform is not made of light or matter, but spacetime itself, and as such each is given as "strain"

[4]The sensitivity of the strain ($\frac{\Delta L}{L}$) measured is in part a function of the length L of the interferometer arms. While I offered a simple picture, the actual interferometers contain Fabry Perot cavities which bounce the light back and forth, effectively increasing the path length from 4 km to ~ 1200 km.

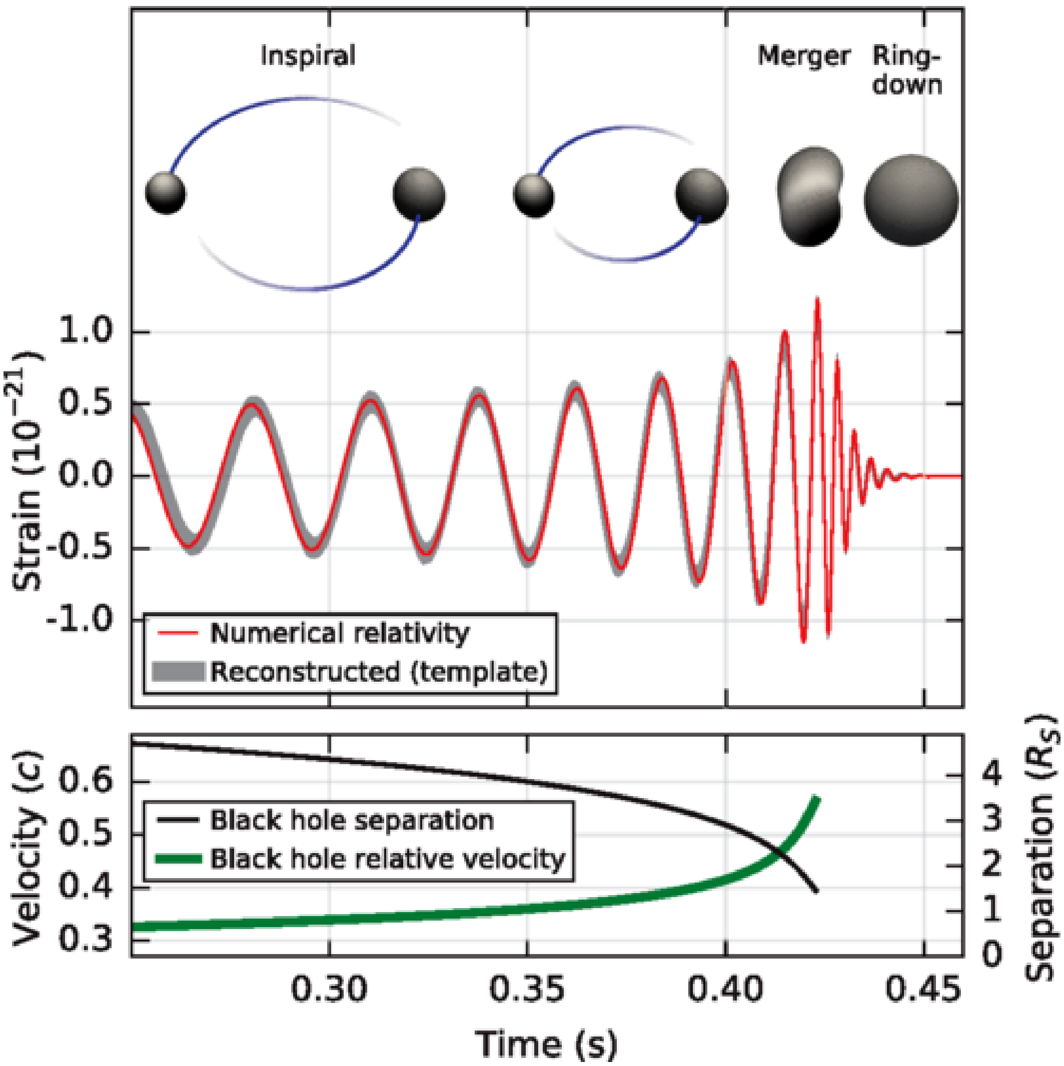

Figure 4.1: Figure from Abbott et al. [20]. Top: Visualization of the inspiral, merger, ringdown of the first detected gravitational-wave event, GW150914. We can see the black holes get closer together as their orbit radiates away energy (inspiral), the frequency of the wave increases until the black holes collide (merger), and spacetime settles down to that outside of a single black hole (ringdown)

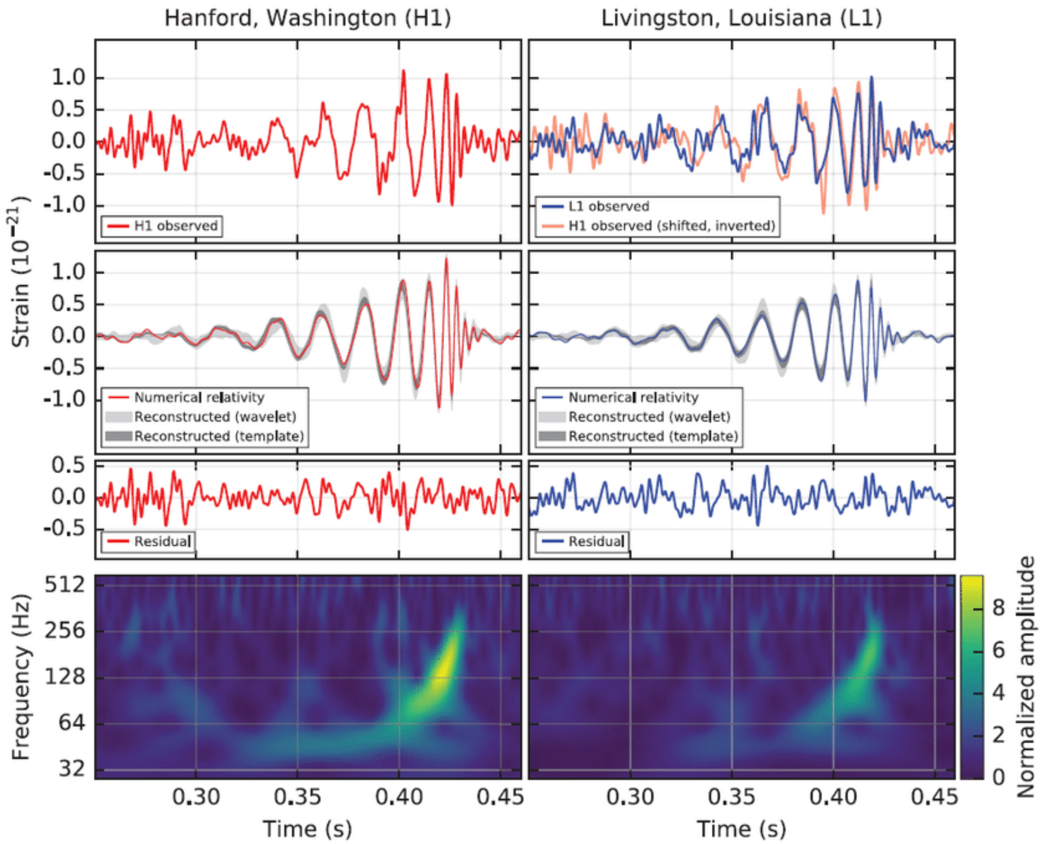

Figure 4.2: Figure from Abbott et al. [20]. Gravitational-wave event GW150914 split between the LHO detector (left) and the LLO detector (right). Top row: Band passed strain. Second row down: Predictions from numerical relativity compared to wavelet and physical waveform (template) predictions. Third row down: Residual once the numerical-relativity waveform had been subtracted from the data. Bottom row: Q-scans or spectrograms of the whitened gravitational-wave data plotted as a function of time vs frequency, colored by power in that time-frequency bin.

as a function of time, a dimensionless change of distance over some reference distance. These gravitational radiation patterns, or "waveforms", usually written as $h(t)$ are solutions to the far-field, general relativistic two-body problem. These waveforms can be used to understand what signal was observed in the gravitational-wave detectors in a process described in Section 4.4.

These waveforms are typically described in three phases, "inspiral" where the two bodies are in orbit, "merger" when the two bodies collide, and "ring-down" where the once-rapidly changing spacetime settles down to a stationary state [371].

Understanding gravitational waveforms

Below I provide some equations that approximate the orbital evolution of a simple binary system to provide a basic understanding of how energy is radiated in gravitational-wave inspirals and the approximate form of that radiation[5]. We take the simplifying assumptions (which are valid only in the inspiral regime) that 1) the energy radiated is small in comparison to the gravitational energy of the orbit, 2) the speeds are not relativistic, 3) the orbits are approximately circular[6], 4) the orbits are approximately Keplerian, and 5) the individual objects have negligible spins.

Under these assumptions, we use Einstein's equations [17, 153] to derive that in this regime, the energy lost to gravitational radiation is given as

$$\frac{\mathrm{d}}{\mathrm{dt}}E_{\mathrm{GW}} = \frac{32}{5}\frac{G}{c^5}\mu^2 r^4 \omega^6, \tag{4.1}$$

where G is the Gravitational constant, c is the speed of light, r is the separation of the objects, ω is the orbital angular velocity, and $\mu = \frac{m_1 m_2}{M}$ is the "reduced mass", where m_i represent the masses of the individual objects and $M = m_1 + m_2$ is the total mass of the system. Since the energy is radiated from the potential energy of the orbit, we have $-\frac{\mathrm{d}}{\mathrm{dt}}E_{\mathrm{GW}} = \frac{\mathrm{d}}{\mathrm{dt}}E_{\mathrm{orb}}$, where $E_{\mathrm{orb}} = \frac{-GM\mu}{2r}$. This leaves $\frac{\mathrm{d}}{\mathrm{dt}}E_{\mathrm{GW}} = -\frac{GM\mu}{2r^2}\dot{r}$, where we use $\dot{x}$ to mean the derivative of a quantity with respect to time. Then, since the orbits are approximately Keplerian, we use Kepler's third law $r^3 = GM/\omega^2$ and its derivative $\dot{r} = -\frac{2}{3}r\dot{\omega}/\omega$ to obtain both

$$\dot{r} = -\frac{64}{5}\frac{G^3 M^2 \mu}{c^5}\frac{1}{r^3}, \tag{4.2}$$

$$\dot{\omega}^3 = \left(\frac{96}{5}\right)^3 \frac{G^5 M^2 \mu^3}{c^{15}}\omega^{11}. \tag{4.3}$$

How do these equations govern the orbits of the system and therefore the observed gravitational wave? First: the change in distance between the objects ($\dot{r}$) is always negative, meaning the objects get closer each orbit. Second: the change in frequency is always positive; the objects orbit around each

[5] A detailed derivation of the basic physics is provided in Abbott et al. [17] and a complete derivation is in Maggiore [238].

[6] This is a surprisingly accurate assumption as gravitational waves actually cause orbits to circularize [271]. Eccentric (i.e. non-circular) systems cannot remain eccentric for long. However, since an inspiraling orbit cannot possibly be actually circular (since the radius of a circle does not get smaller...), this is an approximation.

other faster and faster. Both of these are visible in the example gravitational-waveform displayed in Fig 4.1 as well as the characteristic "chirp" in the bottom of fig Fig. 4.2. Key features to note are the increasing amplitude and frequency of the signal. The increasing frequency is immediately apparent in Eq 4.3. As to the increasing power: gravitational waves drain the gravitational potential energy. As the objects fall closer together, they radiate more energy, causing them to fall even closer together and radiate even more energy. This feedback loop creates an increasing energetic signal until the objects finally merge into a single, and quiescent, merged remnant (usually a black hole).

Gravitational-waveforms and source parameters

So far we have seen how at least the mass parameter affects the gravitational waveform, but other parameters also determine their evolution. Gravitational waveforms from binary black holes are somewhat simple in that they can be entirely characterized with 15 parameters[7,8]. Each compact object has a mass, a three-dimensional spin vector, and (for neutron stars) a tidal deformability. These parameters are known as *intrinsic parameters* because they govern the dynamics of the actual physical system. In this section we discuss how some of those different parameters change the dynamics of the system, and others change the observed waveform. The remaining parameters are the *extrinsic* parameters, which govern what an outside observer sees but do not affect the actual dynamics of the binary system. These include the inclination angle ι[9], which is defined as the angle between the orbital angular momentum of the system and the line-of-sight to the observer. Other extrinsic parameters are the distance to the event, the propagation direction of the gravitational wave (given as a sky location, right ascension, and declination), the time the event occurred, the phase of the gravitational wave at coalescence, and the polarization angle between the plus and cross modes of the gravitational wave.

An important feature of gravitational waveforms is that high mass binaries merge more quickly than binaries with lower masses (from a given frequency) [238]. Why? Orbital energy $E_{\rm orb}$ is proportional to the total mass of the system. More massive systems have more energy and as we can see in Eq 4.2, the

[7] 17 for neutron star mergers since they also have parameters determining their tidal deformability (when using a single nuclear equation of state).

[8] Including the effects of eccentricity also adds two more parameters: the eccentricity e and the mean anomaly.

[9] also called θ_{ln}

higher the mass, the more negative $\dot{r}$ becomes, the more the orbit falls inward. Therefore, if we had a high and a low mass system orbiting at the same frequency, in a single orbit, the higher mass system will radiate more energy away. Therefore the high mass systems will spend less time at any given frequency, meaning the objects will merge more quickly, and spend less time in the LIGO frequency band ranging from $\mathcal{O}(10-10^3)$ Hz. This means that there is a relationship between the frequency at which the binary merges and the mass of the system.

Cosmology research has found that our universe is expanding, and that distant points are becoming even more distant [195]. This causes a "redshift", z, which causes visible light emitted from distant galaxies to shift to lower frequencies (i.e. become more red since that is the lowest frequency of visible light) depending on how far away from us the system itself is[10]. The exact same frequency shift happens to gravitational waves; the gravitational waves observed in the LVK detectors will appear lower frequency (and therefore, higher mass) when observed. We then distinguish between detector mass and source-frame mass, which are related as such:

$$m = (1+z)\, m_s, \tag{4.4}$$

where m represents any mass parameter in the "detector frame"[11] and m_s its corresponding value in the source frame. Since the distance to the source itself is getting stretched as the gravitational wave travels incredible distances, the distance is reported as luminosity distance, D_L, where $D_L = (1+z)D_M$, where D_M is the co-moving transverse distance, i.e. the frozen distance to the source when the gravitational wave is observed.

Gravitational waves are transverse meaning that they stretch and squeeze spacetime perpendicular to their propagation direction. General relativity predicts that gravitational waves have two polarization: the plus polarization which stretch and squeeze masses along the horizontal axis, and the cross polarization which acts along the diagonal. These polarizations, in the time

[10]Since the universe is expanding, light (and gravitational waves) get stretched out to lower frequencies. This is similar to the Doppler effect where an ambulance traveling away from you sounds lower in pitch than an ambulance heading towards you. A redshift of $z = 1$ is the distance away that light (or gravitational-waves) has to come from to have its wavelength stretched by a factor of two (or identically, its frequency halved). Earth is located at redshift $z = 0$.

[11]i.e. the mass corresponding to the frequency as measured in the detector

domain, are related like so

$$h_+(t) = \frac{1}{2}\mathcal{A}_{\rm GW}\left(1+\cos^2(\iota)\right)\cos\left(\phi_{\rm GW}(t)\right), \tag{4.5}$$

$$h_\times(t) = \mathcal{A}_{\rm GW}\cos(\iota)\sin\left(\phi_{\rm GW}(t)\right), \tag{4.6}$$

where $\cos\iota = \hat{\mathbf{L}}\cdot\hat{\mathbf{N}}$ is the "inclination angle" between the orbital angular momentum $\mathbf{L}$ and the line of sight from the observer to the source, $\mathbf{N}$, $\mathcal{A}_{\rm GW}$ is the amplitude of the signal and $\phi_{\rm GW}$ is the phase evolution of the wave [161].

The transverse nature of gravitational waves means that a wave traveling in the plane of an LVK-like interferometer along one arm will not stretch or squeeze *that* arm (although it will still affect the other). In contrast, if the wave travels perpendicular to the interferometer, the arms are maximally stretched and squeezed[12]. The detector sensitivity to gravitational-wave direction and polarization is represented with the scalar valued "detector response functions", $F_+(\alpha,\delta,\psi,t_c)$ and $F_\times(\alpha,\delta,\psi,t_c)$, where α and δ specify the sky location of the wave (right ascension and declination), and ψ represents its polarization, and the time t_c at which the event was observed[13]. These functions ultimately allow us to calculate $\tilde{h}(f)$, the strain in the detector as a function of the gravitational wave's frequency,

$$\begin{aligned}\tilde{h}(f) = \Big[&F_+(\alpha,\delta,\psi,t_c)\ \tilde{h}_+(f,\{m_1,\boldsymbol{\chi}_1,\Lambda_1\},\{m_2,\chi_2,\Lambda_2\},\iota,D_L,\phi_c)\\ &+F_\times(\alpha,\delta,\psi,t_c)\ \tilde{h}_\times(f,\{m_1,\boldsymbol{\chi}_1,\Lambda_1\},\{m_2,\chi_2,\Lambda_2\},\iota,D_L,\phi_c)\Big]e^{-2\pi i f t_c},\end{aligned} \tag{4.7}$$

where $\boldsymbol{\chi}_i = \hat{\mathbf{S}}$ is the dimensionless spin vector of object i, Λ_i is the tidal deformability [14], D_L is the luminosity distance, ϕ_c is the phase at coalescence, and t_c is the coalescence time, which just shifts the whole waveform in time[15]. Here I omitted the eccentricity and mean anomaly of the system, the effects of which are expected to be small except, perhaps, in systems formed dynamically [271].

[12]When also the polarization of the wave is aligned with the arms

[13]The event time is necessary because Earth's rotation alters detector sensitivity to different celestial directions. For transient analyses lasting seconds to minutes, we assume this rotation does not significantly change the detector response over the course of the event in the detector, although doing so is possible.

[14]Relevant only for neutron stars.

[15]A time shift $h(t-t_c)$ in the time domain becomes a simple phase shift $e^{-2\pi i t_c}$ in the frequency domain

Spin-effects on gravitational-waveforms

Systems where the component spins are large and aligned with the orbital angular momentum experience a so-called "orbital hangup effect" where the the merger is delayed relative to an otherwise identical system with smaller spin [97]. When instead the spins of the system are mis-aligned to the angular momentum of the system the orbital plane and spins "precess" about the total angular momentum [61, 210]. Whereas aligned spin waveforms evolve with increasing amplitude, spin-orbit precession causes the orbital angular momentum of the system, $\mathbf{L}$, to *precess* around the total angular momentum $\mathbf{J} = \mathbf{L} + \mathbf{S}$. This means that as the orbit evolves and $\mathbf{L}$ precesses, the binary changes its orbital alignment to an outside observer, ι. Because "face-on" ($\iota = 0$) mergers are louder than "edge on" ($\iota = \frac{\pi}{2}$) as in Eq 4.5, characteristic amplitude modulations are created in the inspiral portion of the waveform [62, 211]. Though these modulations are typically weak [318, 373, 375], a strong measurement of the effect is highly sought after as spin-precession could distinguish between formation channels of binary black holes [240, 388, 391].

Actually computing gravitational waveforms

In Sec 4.3 we made many approximations in order to get a heuristic understanding of how orbits evolve when accounting for the emission of gravitational-radiation. The most accurate way to compute gravitational-waveforms is by evolving the system using numerical relativity. Such numerical relativity codes solve the *full* Einstein Field Equations, creating waveforms accurate to within numerical error [176]. However, solving the Einstein Field Equations is computationally expensive; each simulation takes anywhere from *weeks to months to complete*. The $\sim 4,000$ waveforms in Simulating Extreme Spacetimes (SXS) catalog are therefore insufficient for determining the parameters of the system (which require $\mathcal{O}(10^6)$ waveform evaluations for each event: see Sec 4.4); there are simply not enough computational resources to *ever* cover the entire parameter space.

So-called "waveform approximants" have been developed to bridge this computational gap. These waveforms include include inspiral-only post-Newtonian waveform models [63, 82–84, 251, 310, 369], inspiral-merger-ringdown models that are calibrated to numerical relativity [166, 178, 198, 207, 279, 281], effective one-body approximants [86, 87, 90, 131–135, 284, 330], and "surrogate" models, which are templates created by interpolating numerical relativity waveforms [79–81, 202, 362].

4.4 Parameter estimation

Now that we have discussed how the detectors work in Section 4.2 and described how general relativity relates the source properties to what is observed in the detector in Section 4.3, we are ready to determine the parameters of a gravitational-wave source from data, $\mathbf{d}$, containing a signal observed by a detector. In an idealized scenario—free from any noise and with perfect waveform models $\mathbf{h}(\theta)$—this task would reduce to finding parameters θ such that $|\mathbf{d} - \mathbf{h}(\theta)| = 0$. However, three significant challenges complicate this process:

1. **Detector Noise:** Real detectors introduce noise, making the signal extraction less straightforward.

2. **Model Imperfections:** Waveform models are approximations and inherently imperfect.

3. **High-Dimensional Parameter Space:** The parameter space of θ spans 15–19 dimensions, is highly correlated, and often features sharply peaked structures. Additionally, evaluating gravitational-wave models in this complex space is computationally intensive.

Because of the noise and model imperfections, we are instead tasked with finding the probability distribution $p(\theta|\mathbf{d})$, the probability that, given some data, we observed a gravitational wave with parameters θ.

Though the goal is straightforward, sampling a highly correlated 15-19 dimensional space is difficult, and as such, these challenges form the core of my thesis work. In this section, I will detail how we determine the probability distribution of parameters that are most likely to describe the gravitational-wave signal observed in the detectors.

Bayesian Inference

In Sec 4.4 I describe how the essential challenge of gravitational-wave parameter estimation is sampling a probability distribution in many dimensions. Here I lay out what that probability distribution actually is.

We are interested in finding the "posterior" distribution,

$$p(\theta|d), \tag{4.8}$$

the probability p of a gravitational-wave with parameters θ[16] given that we have observed data d. The posterior is normalized such that

$$\int d\theta \; p(\theta|d) = 1. \tag{4.9}$$

Using only conditional probability, we know that $p(d|\theta)p(\theta) = p(\theta|d)p(d)$. We rewrite this in a perhaps more familiar form, Bayes Theorem:

$$p(\theta|d) = \frac{p(d|\theta)p(\theta)}{p(d)}, \tag{4.10}$$

and then again, this time with more suggestive notation

$$p(\theta|d) = \frac{\mathcal{L}(d|\theta)\pi(\theta)}{Z}. \tag{4.11}$$

Here $p(d|\theta)$ became $\mathcal{L}(d|\theta)$, our "likelihood". The likelihood is the probability of observing some data d given that there is an event with some parameters θ. The likelihood function is determined by our noise distribution, and the assumptions required to describe the noise; we discuss how to formulate our likelihood in Sec 4.4.

$\pi(\theta)$ is our "prior", it incorporates the belief about θ before we actually take our measurement and, like the likelihood, it is something that we get to choose. The prior is where the physical assumptions of our models are baked in. Sometimes there is an obvious choice of a prior; we might expect that the gravitational waves we observe to have no preferred sky location, so we might weight every patch on the sky as equally likely. In other cases, the prior is less obvious, and we might instead express our ignorance by choosing an "uninformative" or "flat" prior, usually either uniform or log-uniform.

The final piece is Z, our "evidence", or the probability of observing that data at all (given a noise model and a prior). That is, we can rewrite the evidence

[16] Here the parameters represent the masses, spins, etc of a gravitational wave. We call these the parameters or the model parameters, so there is an implicit assumption that a model is being chosen. Some authors prefer to write $p(\theta|d, M)$ where M represents some model, but unless two models are being compared, this notation is clunky.

as

$$Z = \int d\pi \ \mathcal{L}(d|\theta) \tag{4.12}$$

$$= \int_\pi d\theta \ \pi(\theta) \ \mathcal{L}(d|\theta). \tag{4.13}$$

Essentially, the evidence is a normalization term of equation, e.g. it is defined such that Eq 4.9 is true. The evidence is the expected value of the likelihood over the prior. It is in some ways an absolute measurement of the goodness-of-fit of a model. It is common to directly compare the evidences of two models and take the ratio of their evidences as the "Bayes factor". The one with a higher evidence will therefore be the more statistically favored model. However it is important to note that as in Eq 4.12, changing the prior can have an outsized affect on the evidence [374]. Unless comparing models with the same or similar priors, direct model comparison can become misleading.

Likelihood and noise model

To choose our likelihood model, we have to describe what our data looks like without the presence of gravitational waves[17]. In other words, this is what our background or "noise" model is.

For transient analyses where we analyze data of 2 min or less[18], we assume that (i) the detector noise is uncorrelated between the detectors, (ii) our noise follows a Gaussian distribution with a zero-mean and (iii) the noise is stationary, i.e. the mean and standard deviation do not change over the course of the analysis window. The first assumption could be violated through e.g. distant lightning strikes activating Schumann resonances in the Earth's magnetic field that can cause correlated detector noise [204]. Such noise could cause glitches (see Sec 4.5) in both detectors and affect the interpretation of a SGWB [96, 125, 190, 191, 246, 338, 340]. The second and third assumptions are that the noise is Gaussian and stationary over short timescales [109]. Longer signals, however, might be subject to noise non-stationarity, motivating other studies [118, 152, 253, 354].

Now we lay out how to get the functional form of the likelihood distribution

[17] or other non-Gaussianities in the data

[18] more often $4 - 8\,\mathrm{sec}$, but of course the analysis length depends on the mass of the system(s) we are interested in as in Sec 4.3

given these assumptions. Gravitational-wave data $\mathbf{d}$[19] from detectors is a combination of a gravitational-wave signal $\mathbf{h}$ and noise $\mathbf{n}$.

The assumption that $\mathbf{n}$ is stationary means that the noise is uncorrelated between frequency bins. Gaussianity means that the noise is described by a zero-mean Gaussian distribution at each frequency. This per-frequency Gaussian distribution is entirely described by the variance at each frequency, its power spectrum $\mathbf{S_n}$. That is, the real an imaginary component of the noise at each frequency f_j is distributed according to a Gaussian with variance proportional to $\mathbf{S}_{\mathbf{n}k}$, that is,

$$\Re\left(\tilde{\mathbf{n}}_j\right) \sim N\left(0, \sqrt{\frac{1}{4\Delta f}\mathbf{S}_{\mathbf{n}j}}\right), \tag{4.14}$$

where Δf is the frequency resolution [20], which is likewise true for $\Im\left(\tilde{\mathbf{n}}_j\right)$.

The stationarity assumption means that the behavior of noise in one frequency bin does not affect the behavior of the noise in another. Then, in the absence of non-Gaussian noise or gravitational wave signal, the distribution of the noise in each frequency bin is

$$p(\tilde{n}_j) = \frac{\Delta f}{2\pi S_{nj}} \exp\left(-2\Delta f \frac{|\tilde{n}_j|^2}{S_{nj}}\right). \tag{4.15}$$

Note that compared to a one-dimensional Gaussian, the normalization factor is missing a square root in the prefactor. This is because the frequency domain data are complex, and so the distribution is the product of two Gaussian distributions (one for each of the real and imaginary components) [122, 339, 378]. Now again because we assumed that our frequencies are independent, then the log probability of observing the whole realization of our noise $\mathbf{n}$ is

$$\begin{aligned}
\ln p(\tilde{\mathbf{n}}) &= \ln \prod_i p(\tilde{n}_i) && (4.16)\\
&= \ln \prod_i \frac{\Delta f}{2\pi S_{ni}} \exp\left(-2\Delta f \frac{|\tilde{n}_i|^2}{S_{ni}}\right) && (4.17)\\
&= -\frac{1}{2}\langle \tilde{\mathbf{n}} | \tilde{\mathbf{n}} \rangle - \sum_i \ln 2\pi S_{ni}\,, && (4.18)
\end{aligned}$$

[19] We use $\mathbf{d}$ to represent the time-domain data, which corresponds to the strain $\frac{\Delta L}{L}$ measured by the detectors. This strain is the relative change in the interferometer arm length, ΔL, divided by the arm length L. The strain is sampled at discrete times t_i, forming a time series $\mathbf{d}$. In the frequency domain, the Fourier-transformed data is represented as $\tilde{\mathbf{d}}$. The values of the strain at a specific time or frequency are denoted by d_i and $\tilde{d}_j$, respectively.

[20] The exact normalization will be a function of your choice of Fourier transform normalization

where we have defined our noise-weighted inner product

$$\langle \mathbf{a}, \mathbf{b} \rangle = 4\Delta f \sum_j \Re \frac{a_j^* b_j}{S_{nj}} \,. \tag{4.19}$$

Now that we know what the noise distribution is at each frequency, let's take another look at the data. The data are just going to be whatever model $\mathbf{m}$ we use to describe the signal $\mathbf{h}$ and the noise $\mathbf{n}$

$$\mathbf{d} = \mathbf{m}(\theta) + \mathbf{n}. \tag{4.20}$$

So therefore, our data, once you subtract away whatever model you have, should be distributed the same way that $\mathbf{n}$ is distributed. That is, our likelihood function will become

$$\ln \mathcal{L}\left(\tilde{\mathbf{d}}|\theta\right) = -\frac{1}{2}\left\langle \mathbf{d} - \mathbf{m}(\theta)|\mathbf{d} - \mathbf{m}(\theta)\right\rangle - \sum_i \ln 2\pi S_{ni}. \tag{4.21}$$

4.5 Glitches

In Section 4.4 we laid out all the assumptions that go into the likelihood model for transient gravitational-wave parameter estimation. What I have yet to mention are all the ways that the real detector data breaks these assumptions, specifically, with the presence of glitches.

Glitches are transient, non-Gaussian noise that appear in gravitational-wave detectors as bursts of excess power. They originate from a variety of environmental and/or instrumental couplings. For instance, laser light scattering inside the detectors [322], thunder [142], ravens pecking at ice nearby the detector [126], camera shutters [38], and a litany of other sources that are both identified and unidentified. Examples of the time-frequency morphology of a subset of these glitches are shown as a spectrogram in Fig. 5.3.

LVK detectors have gotten increasingly sensitive between observing runs allowing us, in The first half of the LVK's fourth observing run lasting from May 24th 2023 - January 16th, 2024 (O4a), to observe gravitational-wave events approximately once every two days [116] in comparison to LVK's third observing run (April 1, 2019 - March 27, 2020) (O3) which observed gravitational-wave events approximately once a week. In contrast, both in O4a and O3, the glitch rates were $\sim 1/\text{min}$, meaning $\mathcal{O}(1500)$ per day *in each detector* [8, 33, 38].

4.6 Glitch impact on parameter estimation

Glitches break the noise assumptions laid out in Section 4.4, primarily the assumption that the noise is stationary and Gaussian[21]. When left unmitigated, glitches can bias gravitational-wave parameter estimation, as shown in [150, 265, 278]. This issue is discussed in depth in Chapters 5 (regarding what can be done about glitches), 6 and 7 (in the context of glitch mitigation on individual events), and 8 (on determining where in an analysis segment glitches impact parameter estimation).

4.7 Glitch mitigation strategies

Just as there are a litany of glitch types, there are an array of different strategies to get rid of them. These strategies fall into a few broad categories: (i) subtract the glitch from the data [115, 137, 150, 243, 252, 259, 345, 354, 368, 384], (ii) zero-out the data containing the glitch (which throws away signal information) [18, 307, 327, 353, 385, 393], (iii) fit for both the signal and the glitch simultaneously (thus marginalizing over the glitch parameters and their associated uncertainty; see Chapters 5–7 for details and references), and (iv) avoid the data containing the glitch (if possible; see Sec. 8 for details).

Scientists are often able to determine what the glitch was and stop them from occurring in the first place. This was done to mitigate e.g. scattering glitches in O3 [142], but here I focus on the methods to remove the glitch once it has already occurred.

Glitch subtraction

By and large, the most popular method for glitch mitigation is glitch subtraction. Glitch subtraction is the process of removing a glitch from the gravitational-wave data by subtracting a *glitch* model from the data, and analyzing the more Gaussian ($\mathbf{d}$ − glitch model) instead of the original data, $\mathbf{d}$. To perform this, a model of the glitch is needed.

While there are a few glitch classes that can be described with physical models, such as slow and fast scattering [351], and in some cases a model can be constructed using a transfer function of a "witness" channel [137], for the majority of glitches, there is neither physical models to parameterize them nor a well characterized witness channel. Instead, a phenomenological model that

[21] Generally, glitches do not violate the assumption that the noise is uncorrelated between detectors, though there are possible exceptions [204].

fits for glitches as sums of sine-Gaussian wavelets is used [124, 222][22].

To perform the glitch mitigation for LVK's second observing run (November 30 2016 - August 25th 2017) (O2) and O3, sine-Gaussian wavelets were used to model both the gravitational waves *and* glitches in the data. Then a single draw from the glitch posterior was subtracted to create the glitch subtracted data used by all downstream parameter estimation analyses (see [138]). Starting in O4a, now instead glitch subtraction is performed using the analysis I study in Chapter 5 where the glitch is sampled alongside a physical waveform model, although a single realization is still subtracted.

Removing contaminated data

The simplest approach to handling glitches in the data is to not analyze data containing glitches. Specifically, if the glitch has frequency components that are either exclusively higher or lower than those of the gravitational-wave signal, one can restrict the frequency window of the segment without altering other aspects of the analysis. However, if one tried to do this on the first binary neutron star event GW170818, there *would be no data left to analyze* since the glitch contaminated every frequency bin [18]. In this way, this method carries the risk of discarding potentially valuable data. In O3, this approach was not employed [33, 140].

Some more brute-force methods zero the contaminated data in the time domain (known as "gating"; see e.g. [393]) whereas others "inpaint" or fill the zeroed data back with simulated Gaussian noise [385]. These methods are fast, but in areas where you destroy part of the gravitational-wave signal, they can remove important data or bias results. While gating is used in gravitational-wave searches, because it "throws away" detector data, it has not been used in official LVK transient parameter estimation results.

Glitch marginalization

So-called glitch marginalization is the process by which we model the glitch alongside the gravitational-wave parameters. This process is outlined in Chapter 5 and Chapter 6 for an aligned-spin waveform modeled alongside sine-Gaussian wavelets, and in Chapter 7 for a spin-precessing waveform model.

[22] Sine-Gaussian wavelets are chosen because they form an over-complete basis in smooth functional space and can thus be used model essentially any function, but particularly for transient (short-duration) functions.

This chapter first discusses a physical model for scattering glitches [351] and later examines the use of sine-Gaussian wavelets.

The upside of this glitch marginalization method is that the uncertainty in the glitch model is propagated to the uncertainty of the gravitational-wave parameters (unlike the subtraction method described above). The downside of glitch marginalization is that it is slow and either requires sampling a large parameter space (see Chapter 5) or having a model for the glitch class which exists for only a small subset of all glitches (see Chapter 7).

4.8 BayesWave and BayesWave++

Given the frequent references to `BayesWave` throughout this work, it merits its own dedicated section. A significant portion of this thesis focuses on the development of both the `BayesWave` software and its upgraded version, `BayesWave` ++. While the earlier iteration of `BayesWave` is thoroughly discussed in Cornish and Littenberg [121], Littenberg and Cornish [226] and in Chapter 5, this section provides an overview and outlines the primary distinctions between `BayesWave` and `BayesWave` ++.

`BayesWave` is a parallel tempered reversible jump Markov chain Monte Carlo (MCMC) sampler that has been developed and widely utilized for gravitational-wave data analysis. Specifically, `BayesWave` plays a central role in estimating the power spectral density (PSD) around gravitational-wave events[23] [13, 33, 38], performing consistency checks for general relativity e.g. [32], potentially following up burst events [249], and facilitating glitch subtraction [8, 18, 38, 140].

What distinguishes `BayesWave` from other parameter estimation codes used within the LIGO-Virgo-KAGRA (LVK) collaboration, such as `bilby` [69, 223], is its implementation of a "trans-dimensional" or "reversible jump" algorithm. This functionality enables `BayesWave` to sample multi-dimensional posteriors, a feature that sets it apart from other samplers. The subsequent sections will explain how this method is applied across the different models implemented in `BayesWave`.

`BayesWave` models

The different `BayesWave` models are described below.

[23] More precisely, power spectral density estimation is done by the subroutine `BayesLine`

- The *noise model*, or PSD model, represents the noise power spectral density as a combination of Akima splines and Lorentzian functions. The Akima splines provide a smooth fit to the broadband background noise, while the Lorentzians account for narrow, prominent spectral features.
- The *glitch model* models transient noise artifacts in the detector as a sum of sine-Gaussian wavelets. Each detector has its own independent set of wavelets, allowing the model to adapt to specific noise features. The model's complexity is dynamic, as wavelets can be added or removed to optimize the fit.
- The *CBC model* captures compact binary coalescence (CBC) signals by utilizing waveform templates like those described in Sec 4.3. In `BayesWave` itself, the signal model is restricted to aligned spin waveforms. However, in `BayesWave` ++, there is no such restriction and any waveform model available in LAL can be sampled over[24].
- The *coherent wavelet model* (formerly called the signal model) is designed to capture unmodeled astrophysical signals e.g. [169]. Like the glitch model, it uses sine-Gaussian wavelets, but with the added constraint that the wavelets must exhibit coherence across the network, as expected for genuine astrophysical signals. If this signal model were run on data with a single detector, it should return an identical posterior as the glitch model run on the same data (with the main difference being (likely uninformative) extrinsic parameters returned as well for the coherent wavelet model).

In `BayesWave` ++, any combination of these models can be used at once and I have extensively tested the glitch model and the CBC model individually. See Figure 5.2 for how the sampler actually works in action.

[24]Eccentric models will need to be implemented.

Chapter 5

ACCURATE MODELING AND MITIGATION OF OVERLAPPING SIGNALS AND GLITCHES IN GRAVITATIONAL-WAVE DATA

This chapter contains work from

Sophie Hourihane, Katerina Chatziioannou, Marcella Wijngaarden, Derek Davis, Tyson Littenberg, and Neil Cornish. Accurate modeling and mitigation of overlapping signals and glitches in gravitational-wave data. *Phys. Rev. D*, 106(4):042006, 2022. doi: 10.1103/PhysRevD.106.042006. Reprinted here as Chapter 5. **SH** performed all analyses, authored the text, helped conceptualize and scope the project, and created all figures.

5.1 Abstract

The increasing sensitivity of gravitational-wave detectors has brought about an increase in the rate of astrophysical signal detections as well as the rate of "glitches"; transient and non-Gaussian detector noise. Temporal overlap of signals and glitches in the detector presents a challenge for inference analyses that typically assume the presence of only Gaussian detector noise. In this study we perform an extensive exploration of the efficacy of a recently proposed method that models the glitch with sine-Gaussian wavelets while simultaneously modeling the signal with compact-binary waveform templates. We explore a wide range of glitch families and signal morphologies and demonstrate that the joint modeling of glitches and signals (with wavelets and templates respectively) can reliably separate the two. We find that the glitches that most affect parameter estimation are also the glitches that are well modeled by such wavelets due to their compact time-frequency signature. As a further test, we investigate the robustness of this analysis against waveform systematics like those arising from the exclusion of higher-order modes and spin-precession effects. Our analysis provides an estimate of the signal parameters; the glitch waveform to be subtracted from the data; and an assessment of whether some detected excess power consists of a glitch, signal, or both. We analyze the low-significance triggers (191225_215715 and 200114_020818) and find that they are both consistent with glitches overlapping high-mass signals.

5.2 Introduction

Gravitational-wave (GW) analyses require accurate models for both the astrophysical signals and the detector noise [19]. The majority of source properties inference for transient signals such as compact binary coalescences (CBCs) is based on three assumptions about the detector noise that inform the functional form of the likelihood function: (i) the detector noise is uncorrelated between the detectors, (ii) it follows a Gaussian distribution, and (iii) it is stationary, i.e. its mean and covariance do not change with time. Violation of these assumptions could impact detection and inference efforts. For example, Schumann resonances in the Earth's large-scale magnetic field could cause correlated detector noise and affect detection and interpretation of a stochastic GW background [96, 125, 190, 191, 246, 338, 340]. Additionally, the detector Gaussian noise is stationary over short timescales [109], but longer signals might be subject to noise nonstationarity which has motivated relevant studies [118, 152, 253, 354].

Transient noise artifacts, i.e., glitches, in a detector violate the assumption of Gaussianity and could bias parameter inference when they overlap with signals [150, 265, 278]. In the recent third observing run (O3), LIGO [1] and Virgo [43] have detected an astrophysical event approximately once every five days [33, 38]. Glitches, however, appear in the detectors far more frequently. The average rates for glitch transients with signal-to-noise ratio (SNR) > 6.5 in the first and second half of the third observing run were $0.3\,\mathrm{min}^{-1}$ in LIGO Hanford (LHO), $1.13\,\mathrm{min}^{-1}$ in LIGO Livingston (LLO), and $0.75\,\mathrm{min}^{-1}$ in the Virgo detector [33, 38]. Overall, in O3 a total of 18 events required some form of glitch mitigation [33, 38]. Glitches are most likely to intersect lower-mass, long-duration events such as binary neutron star (BNS) mergers; indeed, both such detected events have overlapped with a glitch and required mitigation [4, 18]. As detector sensitivity improves not only will the event rate increase, but also the glitch rate might increase as weaker glitches that are currently below the noise floor could emerge above it. Additional detectors such as KAGRA [54] in the next observing run (O4) also increase the likelihood that a glitch will appear in at least one detector. In order to have an accurate and unbiased catalog of GW events, effective and generic methods for separating signal and glitch power are necessary. Proposed approaches for glitch subtraction include removing the contaminated data [18, 307, 327, 353, 385, 393] or subtracting detector noise based on data from auxiliary channels [115, 137, 150, 243, 252, 259, 345, 354, 368, 384]. The glitches discussed in this paper are those that remain after the noise mitigation described in [137, 368].

A complementary analysis was proposed in Ref. [110] based on the `BayesWave` algorithm [121, 124, 226]. This analysis expands glitch-mitigation techniques already applied to LIGO and Virgo data [18, 33, 38], where it was used to subtract glitches in 15 out of 18 O3 events that required glitch mitigation [33, 38]. The analysis of [110] simultaneously models the signal and glitch using waveform templates and wavelets respectively. The waveform templates are models for CBC signals that are obtained as solutions to the two-body problem in General Relativity. The glitch model is based on a sum of sine-Gaussian wavelets and it is flexible enough to be able to reliably describe a wide range of potential glitch morphologies. This first study presented a number of examples of data containing binary black hole (BBH) signals and common glitch types and demonstrated the ability of the analysis to reliably separate the two [110]. In this study we expand upon this analysis by considering a wider rage of glitch classes, more instances of each class, and CBC injections of varying masses.

Our analysis results in a posterior distribution for the parameters of the glitch and the CBC signal. Depending on the exact placement of the signal in relation to the glitch, correlations between the two might exist and the resulting CBC posteriors might not be identical to those from data with no glitches. This is unsurprisingly most prominent for signals and glitches with similar time-frequency morphology. Despite this, the CBC parameters are correctly estimated within the extent of the posterior. As a point of comparison for each case, we also examine the bias incurred on CBC parameters by performing a standard analysis that ignores the presence of the glitch in the data entirely.

Our process allows us to obtain both a model for the glitch that can be subtracted from the data and parameter estimation results for the CBC signal, though the latter are restricted by the assumption of aligned-spins in the current algorithm implementation. The glitch-subtracted data are ready for downstream analyses with more sophisticated waveform or detector calibration models. We demonstrate that the lack of spin-precession and higher-order modes in our CBC waveform models does not hinder accurate CBC-glitch separation. We also test whether we can do glitch subtraction on single-detector data. The additional examples and checks presented here demonstrate that the analysis is ready to tackle incoming data in O4 [16].

Finally, we consider some low-significance triggers and attempt to distinguish between CBC signals, Gaussian noise, glitches, or a combination of all three present in the data. Standard detection efforts consider only the possibility that either a CBC signal or a glitch is present in some data. It is therefore possible that the significance of a CBC signal could be impacted if it overlaps with a glitch as this could make the data inconsistent with our CBC model. We find that trigger 191225_215715 (hereon S191225) [38] and trigger 200114_020818 (hereon S200114) [40] are consistent with the presence of both a glitch and a CBC signal. The above results are subject to the caveat that we use a CBC waveform model without spin-precession or higher-order modes.

The paper continues as follows. In Sec. 5.3, we describe the `BayesWave` algorithm, focusing on the CBC (templated) and the glitch models. In Sec. 5.4 we present our injection and recovery scheme. In Sec. 5.5 we present our results on an array of different glitch classes. We test the systematics and limitations of our CBC sampler in Sec. 5.6 by including injections containing higher order modes, spin-precession, and events in a single detector. Finally, in Sec. 5.7, we apply our analysis on triggers S191225 and S200114 in order to assess the presence of signals and/or glitches in the data. In Sec. 5.8 we conclude.

5.3 General Methodology Description

In this section we describe the main features of our analysis that estimates parameters of GW events using templates while jointly modeling detector glitches using wavelets. We briefly introduce our inference scheme, then discuss the features of `BayesWave` relevant for this study including the models, priors, and the joint sampler.

Brief Introduction to Inference Scheme

GW parameter estimation aims to compute $p(\theta_h|d)$, the posterior probability distribution that a model h with parameters θ_h describes the given data, d. The quantity h can contain any component of the data we attempt to model, for example a CBC signal, a glitch, or Gaussian noise. In the case of CBC signals, h is typically a waveform template and θ_h are parameters that describe the binary. The posterior $p(\theta_h|d)$ is proportional to the prior $p(\theta_h)$ times the likelihood of observing data d given the model $h(\theta_h)$, $p(d|\theta_h)$. The likelihood function encodes our assumption about the detector noise. For LIGO analyses

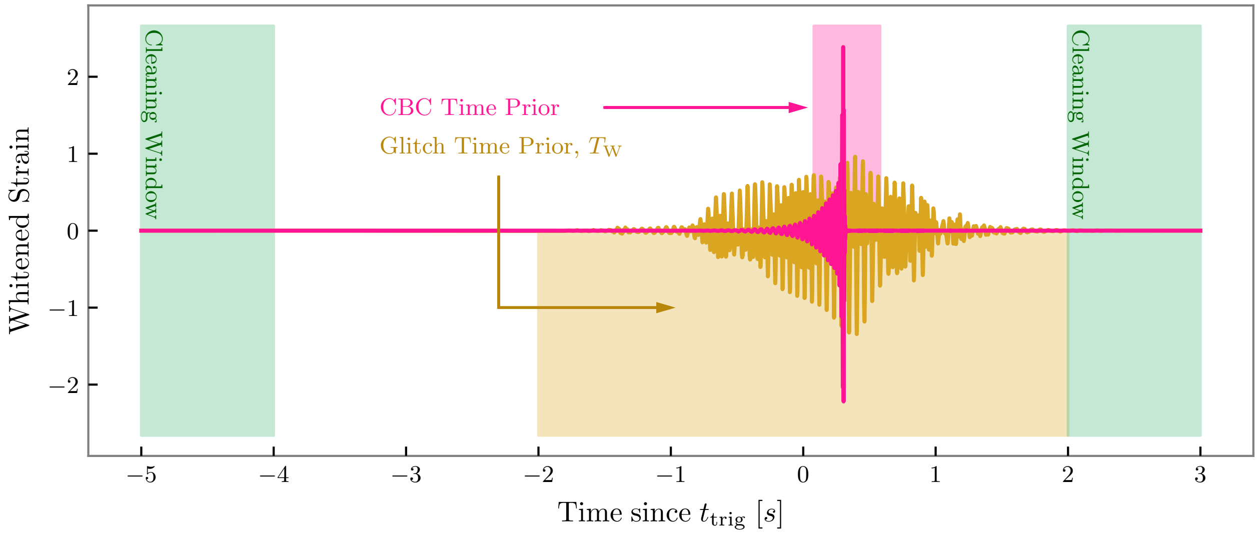

Figure 5.1: Analysis setup with the different time intervals and priors for the analysis of a CBC signal (magenta) and a glitch (gold). The CBC (glitch) time prior is depicted by the magenta (gold) shaded region, although the two regions need not share a common center. The preprocessing cleaning phase is used to permanently subtract wavelets from the data in the green region.

and stationary, Gaussian noise with power spectral density (PSD) $S_n(f)$ in each detector, the likelihood is given by Eq. 3 in [124]. Under this formulation, the multidimensional parameter posterior $p(\theta_h|d)$ is typically explored with stochastic sampling methods such as Markov Chain Monte Carlo (MCMC) and nested sampling [299, 367].

`BayesWave` Analysis

`BayesWave` [121, 124, 222, 226] is a flexible data analysis algorithm that models various components in GW data including signals, glitches, and noise. `BayesWave` explores the joint posterior distribution of its models using a combination of MCMC and reversible jump (RJ) MCMC [174] samplers. The algorithm has a wide range of applications including an unmodeled search pipeline [14, 31], on-source PSD generation [33, 38, 109], tests of General Relativity [21, 23, 30, 111] and consistency tests [169], and various studies of poorly modeled or unmodeled sources including binary neutron star (BNS) postmergers [22, 105, 349], eccentric BBH mergers [130], and supernova [289]. `BayesWave` has also been used to perform glitch subtraction [18, 33, 38, 265]. With this flexibility to model a wide range of signal and glitch morphologies in hand, in Ref. [110] we extended the algorithm capabilities to include a model for CBC signals based on CBC waveform templates. The algorithm implementation is described in detail in [121, 124, 226, 379] and we briefly summarize the most relevant features here.

Whereas most GW parameter estimation analyses [299, 367] use single, point estimates of the PSD and assume that the data contains no glitches, `BayesWave` can relax both assumptions by modeling the CBC signal, the noise PSD, and any potential transient noise all at once.[1] The full `BayesWave` model for this study is the union of a CBC (waveform template) model, a glitch model, and a noise PSD model. The different `BayesWave` models are described below.

- The *noise model*, or PSD model, expresses the noise PSD as a sum of splines and Lorentzians. The splines capture the smooth underlying broadband noise whereas the Lorentzians capture the sharp spectral peaks. Within this paper, we use the color grey to represent the noise model.

- The *glitch model* expresses excess detector transient noise as the sum of sine-Gaussian wavelets: accordingly, each detector has its own, independent, set of glitch wavelets. The set of all sine-Gaussian wavelets form an overcomplete basis over smooth function space, and are thus able to describe a glitch of any morphology provided that it is sufficiently loud. The number of wavelets and hence the model dimensionality is not fixed and wavelets can be added or subtracted as needed. Each wavelet is described by five parameters (θ^{glitch} :) central time, frequency, quality factor, amplitude, and phase. The quality factor is related to the damping time of the sine-Gaussian, and together with the frequency determines the duration of each wavelet. The functional form of the wavelet is given by Eq. 4 in [124]. Within this paper, we use gold to represent the glitch model.

- The *CBC model* uses waveform templates to capture the CBC signal in a manner similar to traditional GW parameter estimation [299, 367]. Details of the implementation of the CBC model in `BayesWave` are given in [379]. For this analysis we restrict to quasicircular CBC signals whose

[1] The effect of PSD uncertainty in CBC analyses has also been explored in [67, 73, 77, 151, 227, 305, 306, 329, 355].

spins are aligned with the orbital angular momentum. Such signals are characterized by up to 13 parameters, namely four intrinsic CBC parameters (the two masses m_1, m_2 and two dimensionless spins χ_1, χ_2 projected onto the Newtonian orbital angular momentum) and seven extrinsic parameters (a time and phase, the right ascension and declination, the luminosity distance D_L, the inclination ι, and the polarization angle). For binary neutron stars (BNSs), we also have two tidal parameters Λ_1, Λ_2. In what follows, we express the masses through the chirp mass $\mathcal{M}_c \equiv (m_1 m_2)^{3/5} (m_1 + m_2)^{-1/5}$ which determines the GW phase evolution to leading order and the mass ratio $q \equiv m_2/m_2 < 1$. We also express the spin through $\chi_{\text{eff}} \equiv (m_1\chi_1 + m_2\chi_2)/(m_1 + m_2)$, which is conserved approximately throughout the binary inspiral [287]. Within this paper, we use pink to represent the CBC model.

- Though not used here, for completeness we also mention the *signal model* that fits for coherent, excess power ("unmodeled" astrophysical signals) using again sine-Gaussian wavelets. Unlike the glitch model, the signal model enforces that the wavelets must be coherent across the detector network as a genuine astrophysical signal would be. Both the signal and the CBC model have the potential to capture a CBC source though the former is more flexible, and thus less sensitive, particularly to weak or long-duration signals.

Priors

The priors for the glitch and CBC model parameters remain mostly unchanged compared to [110, 379]. However, for some combinations of glitch and CBC signals, we find that additional flexibility is required in the time placement of glitch wavelets and the CBC template. By construction, `BayesWave` analyzes data of duration T_{obs} around some trigger time t_{trig}. The prior on the central time of the glitch wavelets then has support within a "window" of length $T_{\text{w}} < T_{\text{obs}}$ around t_{trig}, while the CBC time prior is by default $(t_{\text{trig}} - 0.5\text{s}, t_{\text{trig}} + 1.5\text{s})$. Here we relax the requirement that glitches and CBC signals have a time prior around a common time t_{trig} and allow for them to be placed in different time intervals, though still within T_{obs}. The priors for the noise PSD model are all unchanged compared to [110, 226]. Figure 5.1 shows the relation between the different time intervals.

Though the wavelets that model the glitch can have central times only within T_{w}, the current `BayesWave` implementation employs a preprocessing "cleaning phase". During this phase, the algorithm is run with only the glitch model activated and $T_{\mathrm{w}} = T_{\mathrm{obs}}$, i.e., wavelets can be placed anywhere in the analyzed data segment. At the end of the cleaning phase and before proceeding to the main analysis, wavelets with time centered within a specified interval are permanently subtracted from the data. We typically use a 1s cleaning window at the beginning and end of the data segment. This procedure removes any glitch power that might be present in the analysis segment but not necessarily close to the CBC signal itself as well as data artifacts caused by the finite segment duration and that could bias the PSD estimation.

Sampler

To characterize the joint glitch, CBC, and noise posterior, `BayesWave` uses a combination of MCMC and RJMCMC samplers stringed together within a blocked Gibbs sampler. The blocked samplers give us the flexibility to trivially turn models on and off during an analysis. Here we sketch the workflow for the "CBC+Glitch" analysis, but other `BayesWave` running modes with different model combinations vary only in which samplers are active.

Instead of sampling all parameters concurrently, we separate them into groups of related parameters that are sampled together while other parameters are held fixed. The order of the corresponding samplers and a breakdown of the fixed or varying parameters within each sampler is displayed in Fig. 5.2. Each component sampler runs for a predetermined number of iterations, typically $\mathcal{O}(100)$ and returns its last iteration to be used as fixed parameters by the other samplers. We iterate through the Gibbs sampler loop $\mathcal{O}(10^4)$ times before completing.

Before the Gibbs sampler begins, each model (i.e. PSD, CBC, and glitch) needs to be initialized. An initial estimate for the PSD is generated by the methods described in [124]. To initialize the CBC parameters we follow [120]. An optional `GlitchBuster` step finds initial parameters for the glitch wavelets by iteratively estimating the PSD, wavelet-transforming the data, and removing excess power wavelets [348]. The procedure is described in more detail in [124]. Without `GlitchBuster`, the glitch model begins with no wavelets.

The sampling procedure, and specifically the integration of the CBC sampler is

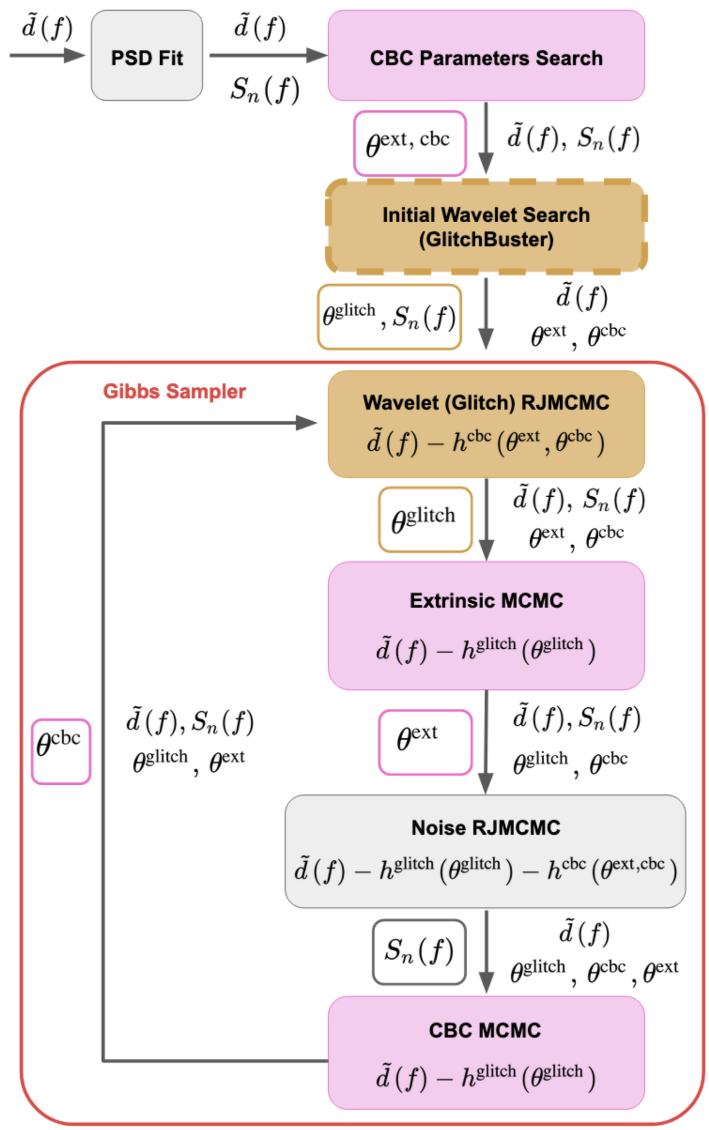

Figure 5.2: Visual depiction of the BayesWave workflow described in Sec. 5.3. Each colored box represents a component sampler and displays its name (bolded) and its input data (in LaTeX). Pink (gold, grey) boxes indicate the sampler that searches over the CBC (glitch, PSD) parameter space. Underneath each colored box, the sampled (fixed) parameters are boxed (unboxed). The Gibbs sampler (i.e., the cycle of component samplers) is boxed in red. Before entering the Gibbs sampler, we use the Fourier domain strain data, $\tilde{d}(f)$, to obtain initial points for the noise and CBC samplers through the procedure described in [120]. We also optionally (dotted box) obtain an initial point for the glitch model [124], otherwise the glitch model begins with 0 wavelets. Within the Gibbs sampler, each component (RJ)MCMC goes through $\mathcal{O}(10^2)$ iterations before passing its values to the next sampler. After each Gibbs sampler loop, a single sample for all parameters is returned. The Gibbs sampler loops $\mathcal{O}(10^4)$ times.

Signal	$m_1(M_\odot)$	$m_2(M_\odot)$	$\min(T_{\rm obs})$ (s)	Λ_1	Λ_2
HM BBH	36	29	4	-	-
LM BBH	12	7	16	-	-
BNS	1.5	1.4	128	115	320

Table 5.1: Parameters of the injected signals we consider. For all the injections we set $\chi_{\rm eff} = 0$, $\cos\iota = 0.88$, $\phi_0 = 1.23$, and $\psi = 0.3$. The sky location is fixed overhead LLO, while the distance is varied to keep the network SNR fixed at ~ 15. For each injection type, we display the minimum segment length $T_{\rm obs}$ necessary to contain the signal, but certain long-duration glitches required an increased segment length.

described in detail in [379]. Below we briefly describe each individual sampler.

- The "wavelet (glitch) RJMCMC" updates the parameters of one of the current wavelets or adds/subtracts a wavelet. Details of the RJMCMC implementation are presented in [121, 124].
- The "extrinsic MCMC" updates the extrinsic parameters of the signal, namely the distance, sky-location, inclination and polarization angles, and time[2]. Details are provided in [121, 124, 379].
- The "CBC MCMC" updates the intrinsic CBC parameters (masses, spins, tides) as well as parameters that are correlated with them, namely the distance, time, and phase. We use waveforms implemented in the `LALSimulation` suite of models [242]. The sampler can operate with any non-precessing model available in `LALSimulation` and in this study we rely on `IMRPhenomD` [198, 207]. We currently do not account for nonaligned spin degrees of freedom but plan to extend our analysis to include the effect of spin-precession in the future. The sampling proposals as well as the heterodyne procedure [119, 120] used to speed up the likelihood calculation are described in [379].
- The "noise RJMCMC" updates the number and parameters of the splines and Lorentzians that describe the noise PSD. Details are provided in [226].

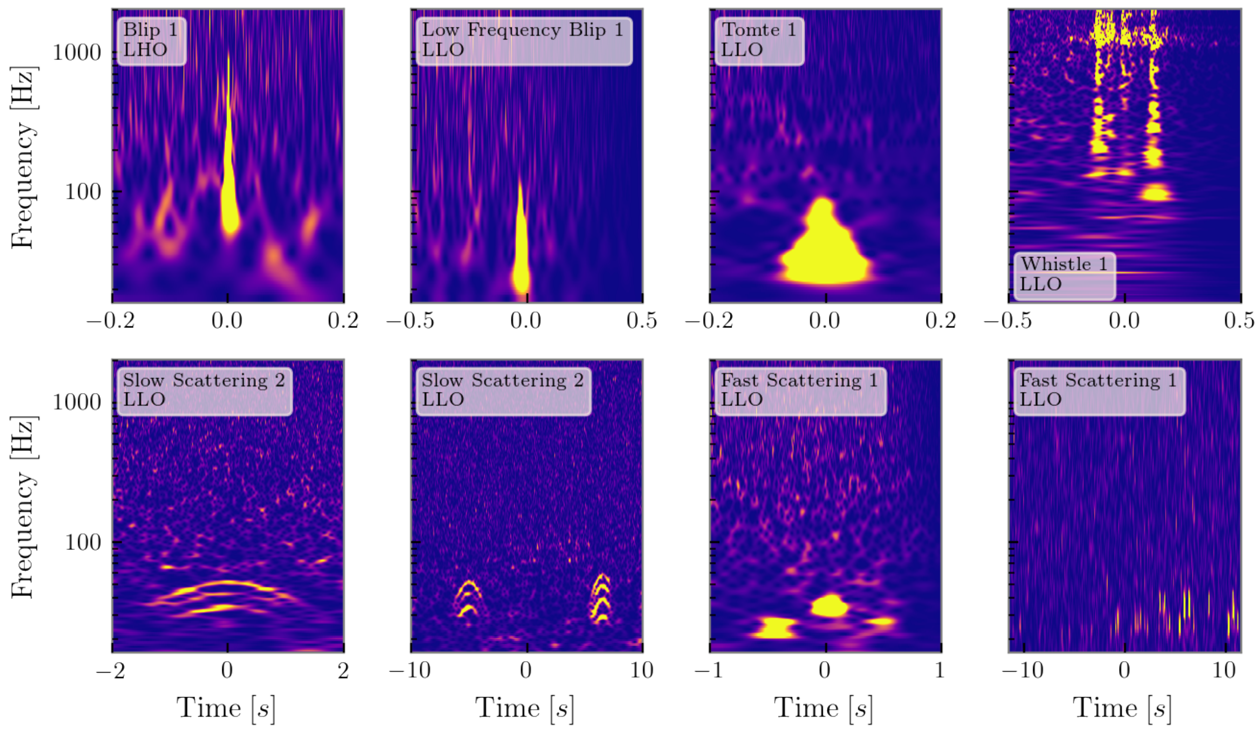

Figure 5.3: Spectrograms of representative glitches for each glitch family we consider in our study. We display two spectrograms for the same fast and slow scattering glitches to demonstrate the long- and short-term behavior of such glitch types. See Tab. 5.4 for glitch GPS times.

Quantity	Data	Recovery	Description
$g_{\mathrm{CBC+G}}$	glitch and CBC	CBC+Glitch	recovered glitch
g_{G}	glitch	GlitchOnly	recovered glitch
$h^{\mathrm{rec}}_{\mathrm{CBC+G}}$	glitch and CBC	CBC+Glitch	recovered CBC
$h^{\mathrm{rec}}_{\mathrm{CBC}}$	glitch and CBC	CBCOnly	recovered CBC
h^{inj}	N/A	N/A	injected CBC

Table 5.2: Summary of quantities used in the overlaps. From left to right columns provide the symbol, the relevant data, the models active during recovery, and a description of what each quantity is.

5.4 Data and Analysis Setup

To test the efficacy of our analysis, we use LIGO data from O2 and O3, available through the GW Open Science Center [39]. Since we do not have exact first-principles models for glitches[3], we identify data containing genuine detector glitches, which were classified through Gravity Spy [386]. We then inject a known CBC signal on top of the glitch, and analyze the data by simultaneously modeling the CBC, glitch, and noise. We label such analyses as "CBC+Glitch". In each case we also analyze the same data using only the CBC and noise models, i.e., ignoring the presence of the glitch. We label such analyses as "CBCOnly". To create a point of reference for the glitch, we also consider the original data with no CBC injection with a "GlitchOnly" run using only the glitch and noise models.

Going beyond the study of [110], here we analyze a more extensive set of glitch types (as classified by GravitySpy [323, 390]), glitch instances, and a set of both BBH and BNS injections. The injected signals include high-mass (HM) BBH (GW150914-like [20]), low-mass (LM) BBH (GW170608-like [7]), and BNS (GW170817-like [18]) injections; their parameters are provided in Table 5.1. Though initially we targeted specific glitches, in many cases secondary glitches (occurring in either or both detectors) were present in the data, speaking to the high occurrence rate of glitches. In those cases we do not discard the data; we analyze them nonetheless and attempt to model all glitches. These analyses also show that `BayesWave` can differentiate between signals and glitches even when they occur in multiple detectors. A spectrogram of one glitch per glitch family is shown in Fig. 5.3. Although the "worst case scenario" for CBC analyses is a glitch with SNR similar to or greater than that of the signal that has a similar time-frequency morphology, in our validation studies we include a wide range of combinations of signals and glitches.

We fix the intrinsic parameters and the sky location within injections of the same CBC type. For each combination of glitch and CBC type, we inject the signal at different times with respect to the glitch in order to vary the amount of overlap between the two. Each signal has a network signal-to-noise ratio (SNR) of 15 and we use the `IMRPhenomD` [198, 207]

---

[2]Though not used in the study, the `BayesWave` *signal model* that describes a GW signal with coherent sine-Gaussian wavelets also makes use of the extrinsic sampler.

[3]Phenomonological models for some glitch types have been constructed, for example [244].

(IMRPhenomD_NRTidalv2 [146–148]) model for the BBH (BNS) signals for injection and recovery unless otherwise indicated. For computational efficiency and since the BNSs overlap with all glitches at low frequencies (below 40Hz), we use a low sampling rate that does not allow us to extract tidal parameters. The possibility of separating BNSs and glitches when they overlap in the high frequencies relevant for tidal effects was shown in [103] using the same analysis as here.

The recovered CBC parameters can be compared with the injected ones directly to assess the recovery reliability. However, no such comparison is possible for glitches as they are obtained from real data. We therefore use various overlaps ($\mathcal{O}$, defined in Eq. 8 of [169]) and mismatches (one minus the overlap, $\mathcal{M} = 1 - \mathcal{O}$) between the recovered CBC and glitch to quantify the quality of the CBC-glitch separation. Definitions are given in Table 5.2. Specifically, for each glitch we also analyze the original data with no CBC injection using only the glitch and noise models. This gives us access to a baseline estimate for the glitch that can be compared to estimates of the same glitch recovered in data coincident with a signal. We then compute the mismatch between the median glitch recovery from analyses with ($g_{\rm CBC+G}$) and without ($g_{\rm G}$) injected CBC signals as a way to test whether the recovered glitch subsumes part of the CBC signal power. A high mismatch could suggest that either the glitch model captures part of the CBC (and leads to it being inadvertently subtracted from the data) or the glitch model misses part of the glitch and fails to subtract all glitch power. Additionally, we compute the posterior for $\mathcal{M}(h^{\rm rec}_{\rm CBC}|h^{\rm inj})$, the mismatch between the injected and the recovered CBC signal whose expected value is a function of SNR [107].

In each case we also carry out a "CBCOnly" analysis that is akin to traditional parameter estimation (with the difference that we also marginalize over the PSD). Though such an analysis is not well motivated as the model cannot exactly match the data (and we encounter increased convergence issues), we find it instructive to compare its results to the full "CBC+Glitch" analysis. We do so by plotting both posteriors when discussing each glitch, and also via the Jensen-Shannon divergence in Sec. 5.5. We denote the recovered CBC signal from such an analysis as $h^{\rm rec}_{\rm CBC}$ and compute $\mathcal{M}(g_{\rm G}|h^{\rm rec}_{\rm CBC})$ as an estimate of how much of the glitch the CBC model can recover. We also compare $\mathcal{M}(h^{\rm rec}_{\rm CBC+G}|h^{\rm inj})$ against $\mathcal{M}(h^{\rm rec}_{\rm CBC}|h^{\rm inj})$ to test the effect of assuming pure

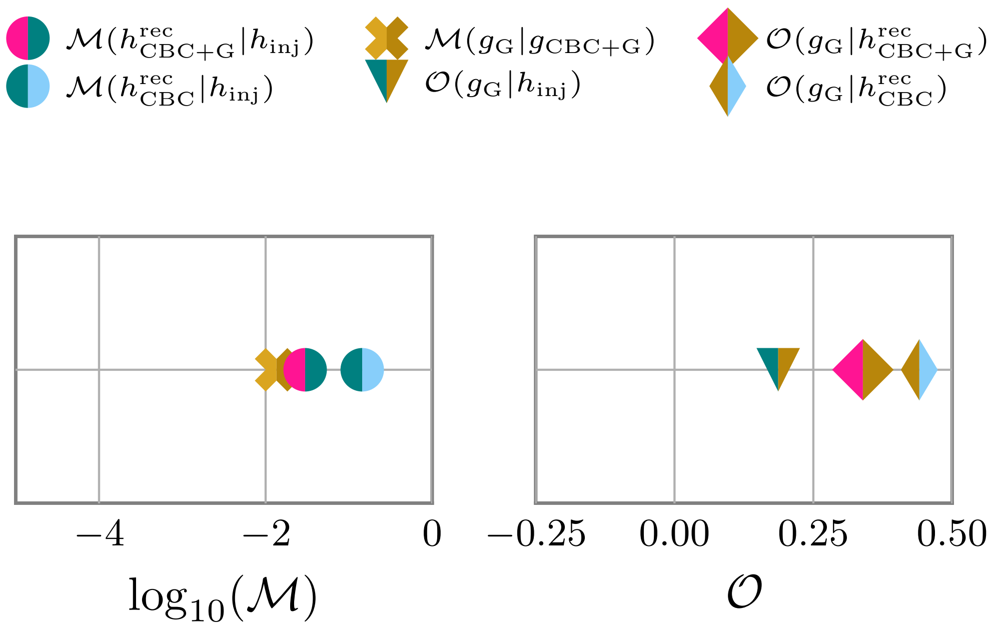

Figure 5.4: Example result for the mismatches defined in Table 5.3. In this case the glitch recovery is not impacted by the presence of the CBC as $\mathcal{M}(g_{\rm G}|g_{\rm CBC+G})$ (gold cross; left panel) is low. Additionally $\mathcal{M}(h^{\rm rec}_{\rm CBC+G}|h^{\rm inj}) < \mathcal{M}(h^{\rm rec}_{\rm CBC}|h^{\rm inj})$ (magenta/teal and blue/teal; left panel) so the CBC model is recovering part of the glitch if one uses an analysis that assumes pure Gaussian noise. We see $\mathcal{O}(g_{\rm G}|h^{\rm inj}) < \mathcal{O}(g_{\rm G}|h^{\rm rec}_{\rm CBC+G}) < \mathcal{O}(h^{\rm rec}_{\rm CBC}|g_{\rm G})$ (teal/gold, magenta/gold, and blue/gold; right panel) meaning that the CBC model is absorbing part of the glitch power, but doing less so when the glitch is part of the model. However, such overlaps are quite low, so the effect is small.

Gaussian noise on CBC recovery in the presence of a glitch. An example result for the various overlaps is presented in Fig. 5.4 and described in detail in Table 5.3. Generally speaking, good CBC-glitch separation is achieved when the quantities on each panel of Fig. 5.4 and equivalent figures in later sections are small.

Another potential test that was explored over the course of this analysis is the Anderson-Darling statistic [58]. This test can be used to assess the degree of Gaussianity in the residual and has been proposed in the context of PSDs [109]. Specifically, we explored the option of subtracting some point estimate (such as the median, or a fair posterior draw) for the CBC and the glitch models from the data and then computing the Anderson-Darling statistic. However,

Mismatch	Symbol	Interpretation
$\mathcal{M}(g_{\mathrm{G}}\|g_{\mathrm{CBC+G}})$		Mismatch between the median glitch reconstruction from data with and without a CBC injection. A high value means that the recovered glitch has been impacted by the presence of the CBC.
$\mathcal{M}(h^{\mathrm{rec}}_{\mathrm{CBC+G}}\|h^{\mathrm{inj}})$		Median mismatch between injected and recovered CBC signal when accounting for the glitch. The lower its value, the better the CBC recovery, though the expected value depends on the signal SNR.
$\mathcal{M}(h^{\mathrm{rec}}_{\mathrm{CBC}}\|h^{\mathrm{inj}})$		Median mismatch between injected and recovered CBC signal without accounting for the glitch. If $\mathcal{M}(h^{\mathrm{rec}}_{\mathrm{CBC+G}}\|h^{\mathrm{inj}}) < \mathcal{M}(h^{\mathrm{rec}}_{\mathrm{CBC}}\|h^{\mathrm{inj}})$, then the CBC model is subsuming glitch power when the latter is left unaccounted for.
$\mathcal{O}(g_{\mathrm{G}}\|h^{\mathrm{inj}})$		Overlap between injected CBC signal and the median glitch recovered without a CBC injection. The absolute value of $\mathcal{O}(g_{\mathrm{G}}\|h^{\mathrm{inj}})$ indicates how similar the CBC and the glitch are, and thus how difficult the separation will be.
$\mathcal{O}(g_{\mathrm{G}}\|h^{\mathrm{rec}}_{\mathrm{CBC+G}})$		Overlap between the median CBC recovered when accounting for glitch power and the median glitch recovered without a CBC injection. If $\mathcal{O}(g_{\mathrm{G}}\|h^{\mathrm{rec}}_{\mathrm{CBC+G}}) > \mathcal{O}(g_{\mathrm{G}}\|h^{\mathrm{inj}})$, then the CBC model might be absorbing undue glitch power.
$\mathcal{O}(h^{\mathrm{rec}}_{\mathrm{CBC}}\|g_{\mathrm{G}})$		Overlap between the median CBC recovered without accounting for glitch power and the median glitch recovered without a CBC injection. If $\mathcal{O}(h^{\mathrm{rec}}_{\mathrm{CBC}}\|g_{\mathrm{G}}) > \mathcal{O}(g_{\mathrm{G}}\|h^{\mathrm{inj}})$, then the CBC model might be absorbing undue glitch power.

Table 5.3: Detailed description of the various mismatches we compute using the reconstructions defined in Table 5.2. An example result for these mismatches is plotted in Fig. 5.4. Throughout, pink and light blue are used for CBC reconstructions with and without the glitch model respectively, gold is used for glitch reconstructions, and teal is used for CBC injections. The split colors indicate the two models used for the overlap or mismatch.

Run Label	GPS Time	Injected signal	Glitches	$T_{\rm obs}$[s]	$T_{\rm w}$[s]	$Q_{\rm max}$	$D_{\rm max}$	$f_{\rm low}$[Hz]
B1_HM_BBH	1168989748	HM BBH	Blip LHO	4	1	40	100	16
B1_LM_BBH	1168989748	LM BBH	Blip LHO	16	1	40	100	16
B1_BNS	1168989748	BNS	Blip LLO	128	1	60	100	16
B2_HM_BBH	1165578732	HM BBH	2×Blip LHO	4	1	40	100	16
B2_LM_BBH	1165578732	LM BBH	2×Blip LHO	16	1	40	100	16
B3_HM_BBH	1171588982	HM BBH	Blip LLO / Unclassified LHO	4	1	40	100	16
B3_LM_BBH	1171588982	LM BBH	Blip LLO / Unclassified LHO	16	1	40	100	16
S1_HM_BBH	1172917780	HM BBH	Slow Scattering LLO	8	4	160	100	8
S1_LM_BBH	1172917780	LM BBH	Slow Scattering LLO	16	4	160	100	8
S2_HM_BBH	1166358283	HM BBH	Slow Scattering / Blip / High Frequency Lines LLO	8	4	250	100	16
S2_LM_BBH	1166358283	LM BBH	Slow Scattering / Blip / High Frequency Lines LLO	16	4	250	100	16
S2_BNS	1166358288	BNS	2×Slow Scattering / Blip / High Frequency Lines LLO	128	30	250	100	16
S3_HM_BBH	1177523957	HM BBH	Slow Scattering LLO	8	4.5	250	100	8
S3_LM_BBH	1177523957	LM BBH	Slow Scattering LLO	16	4.5	250	100	8
FS1_HM_BBH	1238326223	HM BBH	Fast Scattering LLO	4	2	60	100	16
FS1_LM_BBH	1238326212	LM BBH	Fast Scattering LLO	32	23	60	200	16
FS1_BNS	1238326221	BNS	Fast Scattering LLO	128	84.7	60	100	16
FS2_HM_BBH	1265656683	HM BBH	Fast Scattering LLO	4	3	60	100	16
FS2_LM_BBH	1265656673	LM BBH	Fast Scattering LLO	64	32	60	100	16
FS3_HM_BBH	1266384078	HM BBH	Fast Scattering LLO/ Blip LHO	4	3	60	100	16
FS3_LM_BBH	1266384070	LM BBH	Fast Scattering LLO/ Blip LHO	32	30	60	100	16
BL1_HM_BBH	1256909978	HM BBH	Low-Frequency Blip LLO	4	1	40	100	16
T1_HM_BBH	1243679046	HM BBH	Tomte LLO/Blip LHO	4	1	40	100	8
T1_LM_BBH	1243679046	LM BBH	Tomte LLO/Blip LHO	16	3	100	100	16
W1_HM_BBH	1253426470	HM BBH	Whistle LLO	4	1	100	400	16

Table 5.4: Settings for the analyses of Sec. 5.5. From left to right, columns correspond to the run label used in subsequent plots, the approximate GPS time of the given glitch (center of $T_{\rm w}$ as in Fig. 5.1), the type of injected signal, the glitches encountered and the affected detector(s), the segment length $T_{\rm obs}$, the duration of the time window where glitch wavelets can be placed $T_{\rm w}$, the maximum quality factor of the glitch wavelets $Q_{\rm max}$, the maximum number of wavelets allowed $D_{\rm max}$, and the low-frequency cutoff $f_{\rm low}$. All BBH runs have a sampling rate of 2048Hz whereas the BNS runs used a sampling rate of 1024 Hz (which precludes any possibility of recovering tidal parameters). Horizontal lines group analyses that target overlapping data, though the exact center of the glitch window (GPS time) might be shifted according to long- and short-term glitch behavior.

we found that the test is very forgiving and even fails to identify large amounts of residual glitch power. We attribute this to the fact that the form of the Anderson-Darling test we employ is better suited for identifying large non-Gaussian tails in distributions (and thus it is well suited for PSDs [109, 329]) than for coherent non-Gaussian residual that affects only a few data points.

5.5 Results on overlapping signals and glitches

In this section we present results from our injection study. Each subsection corresponds to a different glitch family and injections using three different CBC sources, see Table 5.1. In many cases, especially for the longer signal analyses, the data contain additional glitches beyond what was intended. We point these cases out, but our inability to easily find glitch-free data of duration 1–2 minutes speaks to their prevalence in LIGO data. Details about the times analyzed and the glitches that we encountered either intentionally or accidentally are provided in Table 5.4, together with analysis settings.

Blip Glitches

Blip glitches are short-duration glitches that occur in the most sensitive frequency band of the detectors with a frequency range from ~ 32Hz up through 1024Hz. Because of their short duration and prevalence, they challenge analyses of high-mass events [92], especially for sources with unequal masses and high spins [68]. Their origin is largely unknown; $< 8\%$ of LHO glitches and $< 2\%$ of LLO glitches were identified with auxiliary channels during the first and second observing runs [92]. Details about the blip glitches used for this study are provided in Table 5.4 while a spectrogram is provided in Fig. 5.3. By chance, the data around GPS time 1165578732 contain a second glitch in LHO, 1s later. The presence of the additional glitch did not require any modification of the priors since it occurs entirely after the signal and has a low SNR. Figure 5.5 presents our results, with runs labeled according to Table 5.4. Each row corresponds to the same data with a given glitch and the same CBC signal injected at various times relative to the glitch. The merger time of the CBC relative to the glitch center is given on the y axis.

The first and second column show various mismatches following the format of Fig. 5.4. The set in the first column measure how well the models were reconstructed: the lower the mismatch, the more faithful each model recovery is. Specifically in all cases we find $\mathcal{M}(g_{\rm G}|g_{\rm CBC+G}) \leq 0.01$, suggesting that the recovered glitch model does not consume any significant amount of the injected CBC signal, nor does it miss part of the glitch due to the presence of the signal. The first column also shows that typically $\mathcal{M}(h^{\rm rec}_{\rm CBC+G}|h^{\rm inj}) \sim \mathcal{M}(h^{\rm rec}_{\rm CBC}|h^{\rm inj})$, though we also occasionally find $\mathcal{M}(h^{\rm rec}_{\rm CBC+G}|h^{\rm inj}) < \mathcal{M}(h^{\rm rec}_{\rm CBC}|h^{\rm inj})$; in these cases the CBC model captures part of the glitch if we ignore the presence of the latter by analyzing the data assuming pure Gaussian noise.

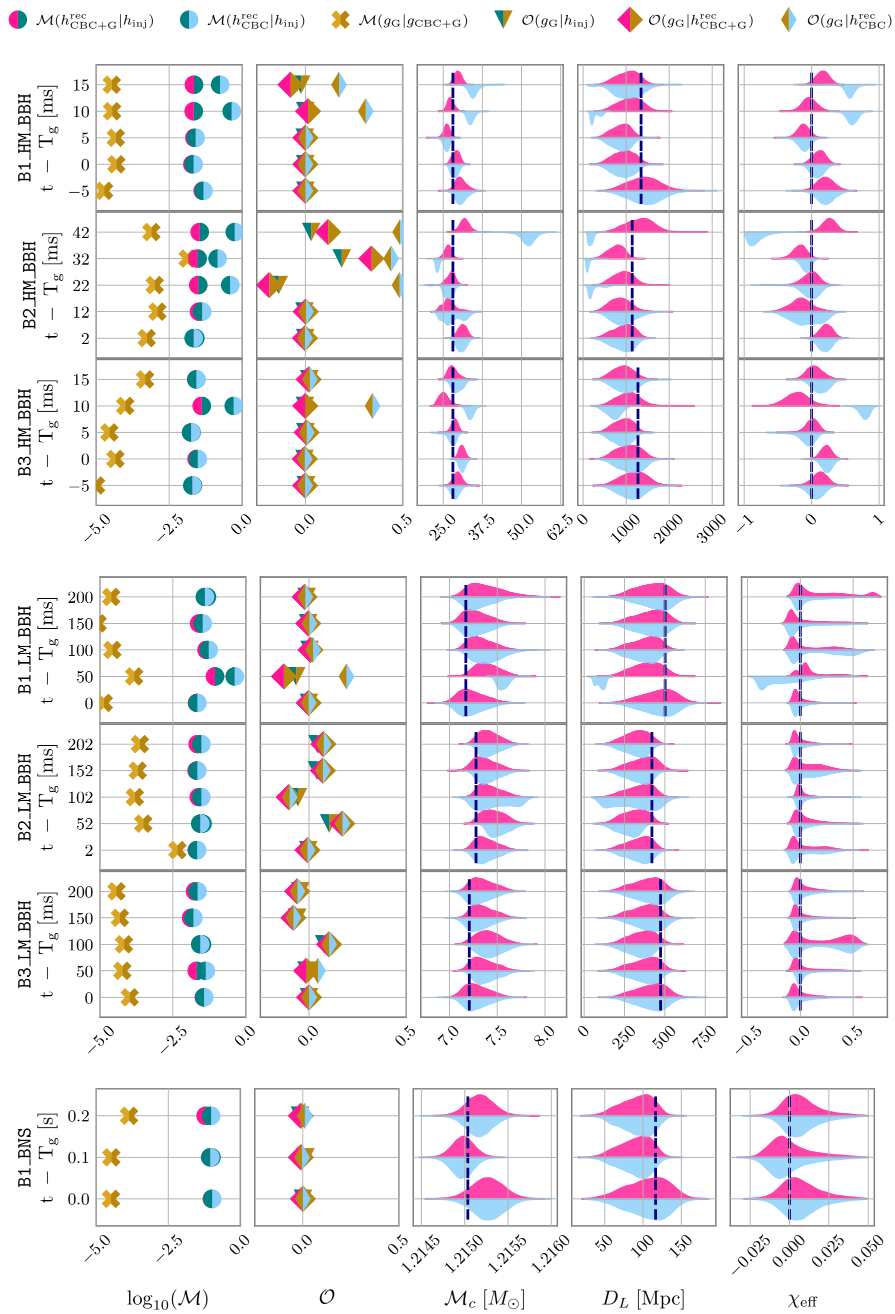

Figure 5.5: Results for CBC signals injected on top of blip glitches for high-mass BBH (top), low-mass BBH (middle), and BNS (bottom). Each row represents an instance of a glitch and a CBC signal injected at different times (y axis) with respect to the glitch; compare run labels to Table 5.4. The first two columns follow Fig. 5.4 and Table 5.3. The violin plots show marginalized posteriors for select recovered parameters, specifically from left to right: detector-frame chirp mass $\mathcal{M}_c$, luminosity distance D_L, and effective spin $\chi_{\rm eff}$. The pink (light blue) violin plots show the posteriors recovered with a "CBC+Glitch" ("CBCOnly") analysis. The correct, injected value is plotted with a dashed, navy blue line. Pink and light blue are used for $h^{\rm rec}_{\rm CBC}$ and $h^{\rm rec}_{\rm CBC+G}$ respectively, gold is used for glitch reconstructions, teal is used for CBC injections. The split colors indicate the two models used to calculate an overlap or mismatch.

In the second column we present information quantifying how similar the glitch the CBC are and, by extension, how difficult it is to separate them. Again the format is similar to Fig. 5.4. Since $\mathcal{O}(g_{\rm G}|h^{\rm inj})$ is evaluated on the injected CBC parameters and not maximized over CBC parameters, it does not directly correspond to how well a CBC template can recover the glitch, but it is a conservative estimate of the similarity between the glitch and injected signal. All overlaps are small, however these is some clear variation. Specifically, we find $\mathcal{O}(g_{\rm G}|h^{\rm inj}) \sim \mathcal{O}(g_{\rm G}|h^{\rm rec}_{\rm CBC+G}) \leq \mathcal{O}(h^{\rm rec}_{\rm CBC}|g_{\rm G})$ which means that in the "CBCOnly" analysis the CBC model absorbs part of the glitch power. This is not the case with the full "CBC+Glitch" model; indeed in all cases $\mathcal{O}(g_{\rm G}|h^{\rm rec}_{\rm CBC+G})$ is closer to the original value of $\mathcal{O}(g_{\rm G}|h^{\rm inj})$.

The remaining columns show the marginalized posterior distributions for the (detector frame) chirp mass $\mathcal{M}_c$, luminosity distance D_L, and effective spin $\chi_{\rm eff}$ respectively. The light blue posterior correspond to $h^{\rm rec}_{\rm CBC}$ ("CBCOnly") whereas the magenta posteriors correspond to $h^{\rm rec}_{\rm CBC+G}$ ("CBC+Glitch"). In all cases the parameters recovered with the full "CBC+Glitch" model are consistent with the injected value. We find differences in the posteriors for the same CBC signal injected at different times with respect to the glitch. This is expected for two reasons. Firstly, the glitch and CBC posteriors are not completely uncorrelated, and hence the marginal CBC posterior will not be exactly the same as if there was no glitch. Secondly, each CBC is injected at slightly different times, and hence is subject to a different realization of the detector Gaussian noise. This distinction becomes more important for shorter signals, and indeed we find that the posteriors become more similar as the CBC signal duration increases from top to bottom. Additionally, we find numerous instances where the "CBCOnly" posteriors are significantly shifted and even inconsistent with the injected value. These cases are typically accompanied by $\mathcal{M}(h^{\rm rec}_{\rm CBC+G}|h^{\rm inj}) < \mathcal{M}(h^{\rm rec}_{\rm CBC}|h^{\rm inj})$ (first column) and/or $\mathcal{O}(h^{\rm rec}_{\rm CBC}|g_{\rm G}) < \mathcal{O}(g_{\rm G}|h^{\rm rec}_{\rm CBC+G})$ (second column). Similar biases were reported in [232] for extrinsic signal parameters computed in low latency.

Blip glitches are one of the most common glitch types and are very similar to high-mass BBHs. However, they are also one of the most straightforward glitch types to deal with due to their compactness in time and similarity to wavelets. The analyses presented here typically used the default `BayesWave` glitch settings (apart from cases where there were additional glitches in the data beyond blips), and could be easily automated.

Slow-Scattering Glitches

Glitches from slow-scattered light appear in the detectors as long duration, $\mathcal{O}(4s)$, arches in the time-frequency domain, evenly stacked in frequency, usually in the range $8-64$ Hz. Each set of arches often recurs multiple times in the detector as shown in Fig. 5.3. Unlike blip glitches, they are not morphologically similar to CBC signals, yet they create long periods of non-Gaussianity and nonstationarity in the data, thus posing a challenge for noise PSD estimation. Their rate of occurrence increased during O3, when slow scattered light glitches overlapped with nine events throughout the observing run [33, 38]. Due to their morphology, slow scattered light glitches required longer analysis segments and wavelets of higher quality factors compared to short duration glitches such as blip glitches. A few of the glitches also extended to lower frequencies than other glitch types, so we used $f_{\mathrm{low}} = 8$ Hz for come cases. See Table 5.4 for run settings.

Figure 5.6 presents our results. All recovered CBC posteriors from the full "CBC+Glitch" analysis are consistent with the injected values. Unlike the blip glitch case discussed in Sec. 5.5, the "CBCOnly" analysis that ignores the presence of the glitch returns largely unbiased posteriors as well, exhibiting mostly small shifts. This is likely due to the fact that fast scattering glitches are morphologically very different than the types of CBC signals we consider.

The left column shows some variation between the median recovered glitch reconstructions. Though $\mathcal{M}(g_{\mathrm{G}}|g_{\mathrm{CBC+G}})$ remains mostly low and around 0.01, it can reach ~ 0.1 in some cases mostly for the second scatter glitch (runs whose label starts with `S2`). We explore this further in Fig. 5.7 where we plot the spectrum of the data, signal, and glitch as well as spectrogram of the data for the `S2`_HM_BBH injection at -500ms compared to the glitch. In each row, we plot the spectrum (left panel) and subtract from the data (right panel) the median glitch reconstruction (top panel) or a fair draw from the posterior (middle panel). In the middle right panel, more of the glitch has been subtracted compared to the top right panel. This is due to the low SNR of the glitch (11.9 for `S2` compared to 15.1 for `S1`; computed by the Omicron pipeline [293]), which results in some of the scattering arches residing in the threshold for reconstruction by the glitch model and thus not

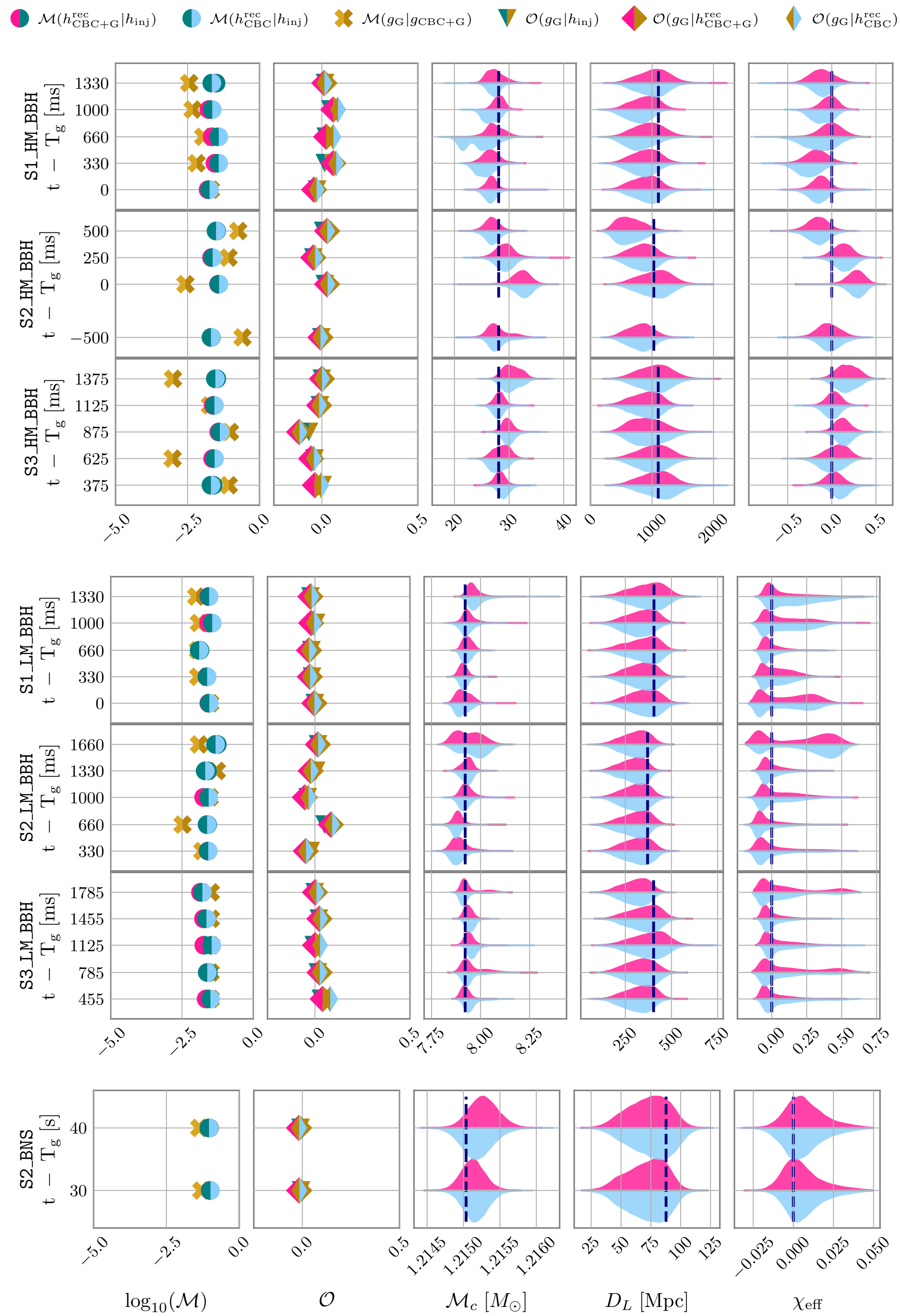

Figure 5.6: Same as Fig. 5.5 but for the analyses anchored around slow scattering glitches. See Table 5.4 for run settings and labels.

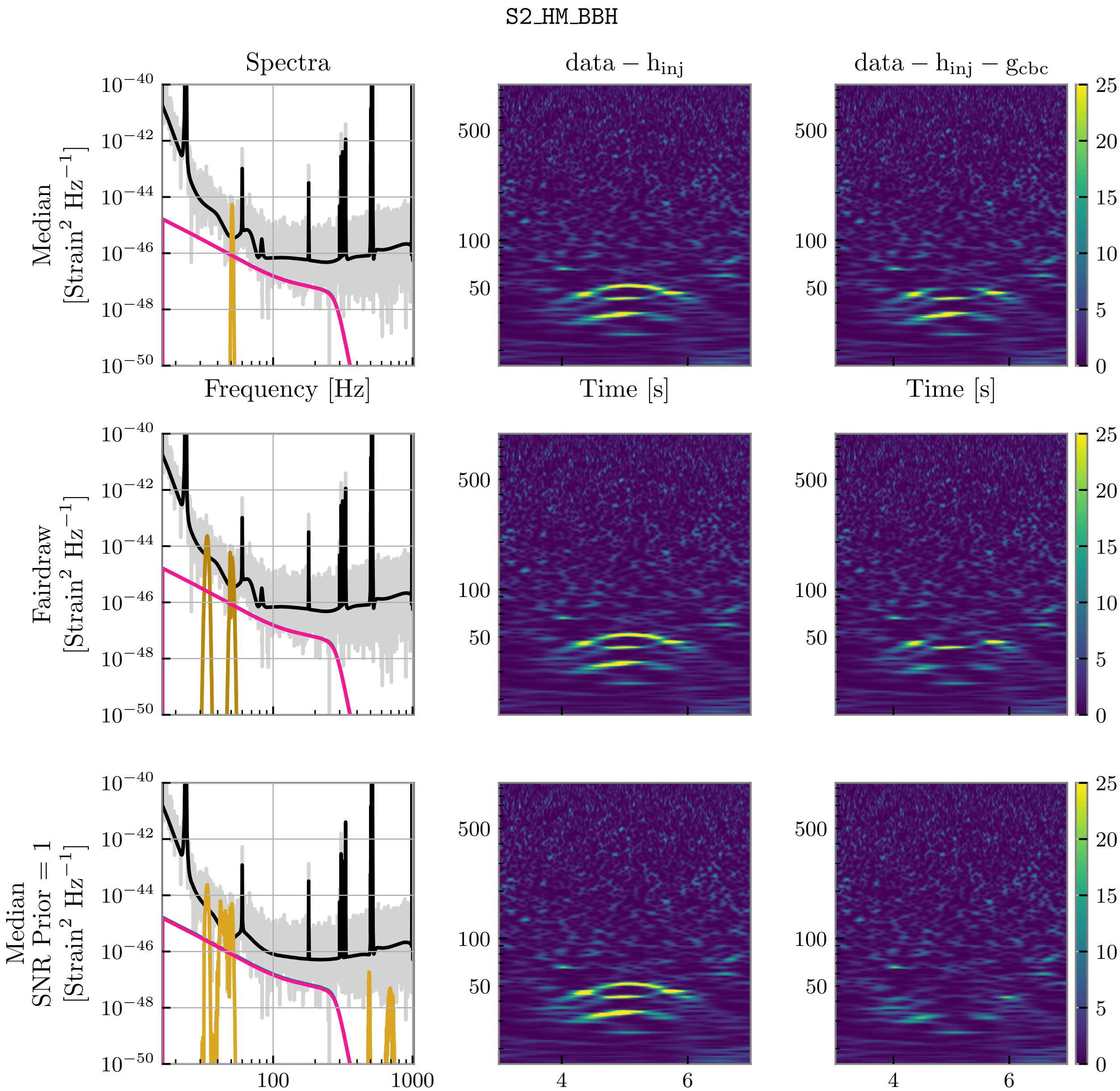

Figure 5.7: Spectra and residual plots for the S2_HM_BBH injection at −500ms from the glitch, described in Sec. 5.5. The first column shows the power spectra of h^{inj} (teal), the raw data (grey) and a point estimate for the PSD (black), $h^{\mathrm{rec}}_{\mathrm{CBC+G}}$ (pink), and $g_{\mathrm{CBC+G}}$ (gold). The second column shows spectrograms of the original data without the injection, while the third column shows spectrograms of the data where an estimate of the glitch has also been subtracted. The first and second row show results with the median glitch reconstruction and a fair draw from the posterior respectively. The third row shows results where the glitch amplitude prior set to peak at SNR=1 that aids identification of lower-SNR glitches. Due to the low SNR of the glitch, we find that all these choices impact the quality of glitch reconstruction and the subtraction.

consistently included in the median. As a result, the median reconstruction has a large variation between different analyses, resulting in higher values for $\mathcal{M}(g_{\mathrm{G}}|g_{\mathrm{CBC+G}})$. In such cases, the glitch-subtracted data are sensitive to the choice of which glitch reconstruction to subtract (median or some fair draw) and additional case-by-case attention is needed.

Motivated by the low SNR of the S2 glitch, we also considered the effect of the default priors in BayesWave. The prior on the amplitude of the wavelets is broad, but peaks at SNR=5 per wavelet by default, see Fig. 5 of [124]. The bottom panel of Fig. 5.7 shows results with a wavelet amplitude prior that peaks at SNR=1 per wavelet. Clearly more of the low-SNR glitch is subtracted. We leave further tests of the prior tunings such as this to future work. We conclude that although analysis of high-SNR slow scattering glitches can be potentially automated, lower-SNR instances will require some user attention.

Fast Scattering Glitches

Fast-scattering glitches (also referred to as "crowns" [323]) are long duration glitches composed of many short bliplike bursts in frequencies from $10-60$ Hz. Fast-scattering glitches have been linked to light scattered off the LIGO optical systems, particularly during ground motion [323]. They were the most common glitch type in LLO in O3, comprising 27% of all glitches [323]. Two spectrograms of fast-scattering glitches are given in Fig. 5.3 and display the long- and short-term glitch behavior; relevant run settings are presented in Table 5.4.

Since a single fast-scattering glitch contains multiple, time-separated bursts, some adjusted settings are required. Such glitches create long-term nonstationarity, particularly at low frequencies so we increase the duration of the analyzed segments from 16 to 32 seconds for the low-mass BBHs. Despite the overall long duration of the glitch, we found that an increase in $Q_{\max}$ is not necessary, as each glitch consists of individual shorter burst that are each reconstructed by a few low-Q wavelets.

Figure 5.8 presents our results. Overall, we find similar results to the slow-scattering case of Fig. 5.6, as the full "CBC+Glitch" analysis is able to separate the signals and glitches while reliably estimating the parameters of the former. The "CBCOnly" analysis returns mostly unbiased results, with the exception of isolated cases. Additionally, we find $\mathcal{M}(g_{\mathrm{G}}|g_{\mathrm{CBC+G}}) < 0.01$ in the first

Figure 5.8: Same as Fig. 5.5 but for the analyses anchored around fast-scattering glitches. See Table 5.4 for run settings and labels.

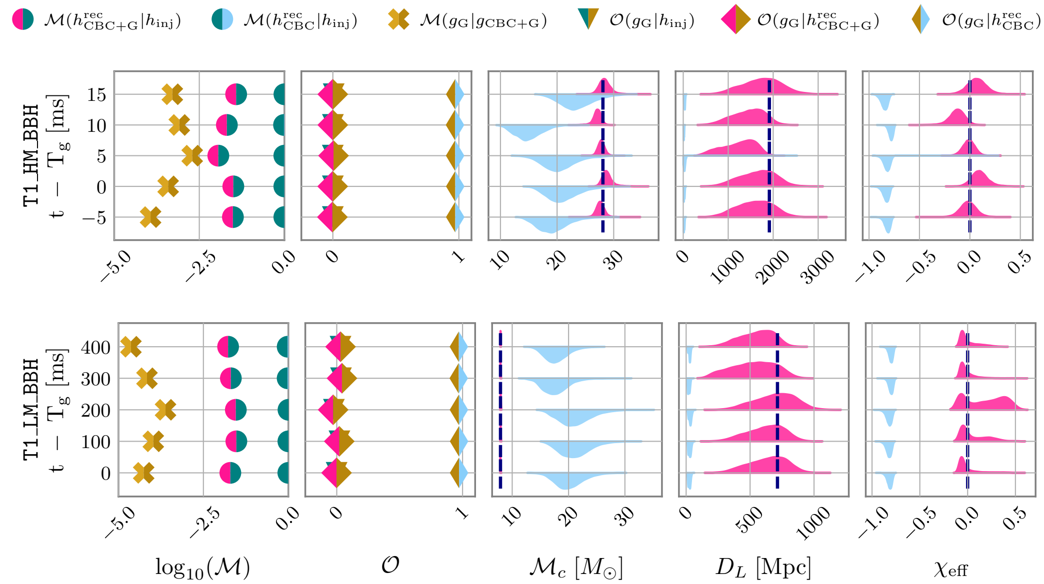

Figure 5.9: Same as Fig. 5.5 but for the analyses anchored around a tomte glitch. See Table 5.4 for run settings and labels. The recovered CBC posteriors without simultaneous glitch modeling (light-blue violin plots) are heavily biased.

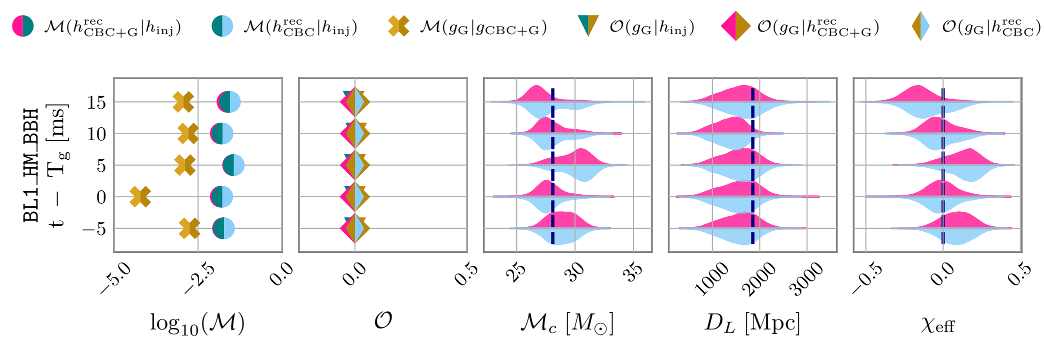

Figure 5.10: Same as Fig. 5.5 but for the analyses anchored around a low-frequency blip glitch. See Table 5.4 for run settings and labels.

column which indicates that the glitch reconstruction is reliable even in the presence of a CBC signal. Due to the adjusted settings required for this analysis, automating analyses of fast-scattering glitches will be challenging.

Tomte Glitches

Tomte glitches are similar to blip glitches in that they can resemble CBCs with high, unequal masses and high spins [68]. We again initially considered various instances of tomte glitches in LIGO data. However, we found the

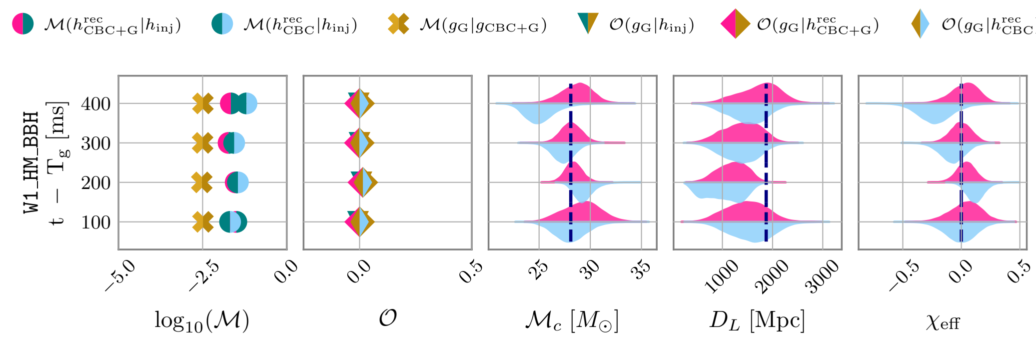

Figure 5.11: Same as Fig. 5.5 but for the analyses anchored around the whistle glitch. See Table 5.4 for run settings and labels.

various tomte glitches to be morphologically very similar to each other, and therefore here restrict to a single instance. The glitch spectrogram is again given in Fig. 5.3 and results are presented in Fig. 5.9. We find broadly similar results to the blip glitch case.

Similar to blip glitches, the "CBCOnly" analysis leads to large biases for all CBC parameters, both intrinsic and extrinsic. This suggests that of all glitches analyzed so far, tomtes are the ones most morphologically similar to CBCs. This is also evident in the second column of Fig. 5.9, where $\mathcal{O}(h^{\mathrm{rec}}_{\mathrm{CBC}}|g_{\mathrm{G}}) \sim 1$ and $\mathcal{O}(g_{\mathrm{G}}|h^{\mathrm{inj}}) \sim 0$, which means that in the "CBCOnly" analysis the CBC model ignored the signal entirely in favor of the glitch. However, even in this challenging case, the joint "CBC+Glitch" analysis is able to separate the signal and the glitch and result in reliable CBC parameter estimates and glitch reconstruction.

Low-Frequency Blip Glitches

Low-frequency blips, as the name suggests, are similar in morphology to blip glitches except that they infect lower frequency bands, see the spectrogram in Fig. 5.3. Given their similarity to blips, we consider a single instance of a low-frequency blip glitch and show results in Fig. 5.10. We obtain similar results to the blip glitch case, Fig. 5.5, with small mismatches $\mathcal{M}(g_{\mathrm{G}}|g_{\mathrm{CBC+G}}) \leq 0.01$ and recovered posteriors that are consistent with the injected values. However, low-frequency blips do not cause as significant a bias on the recovered CBC parameters when the glitch is not included in the model. This might be due to their low-frequency nature, which means that they do not significantly overlap with the CBCs in the most sensitive detector frequency band.

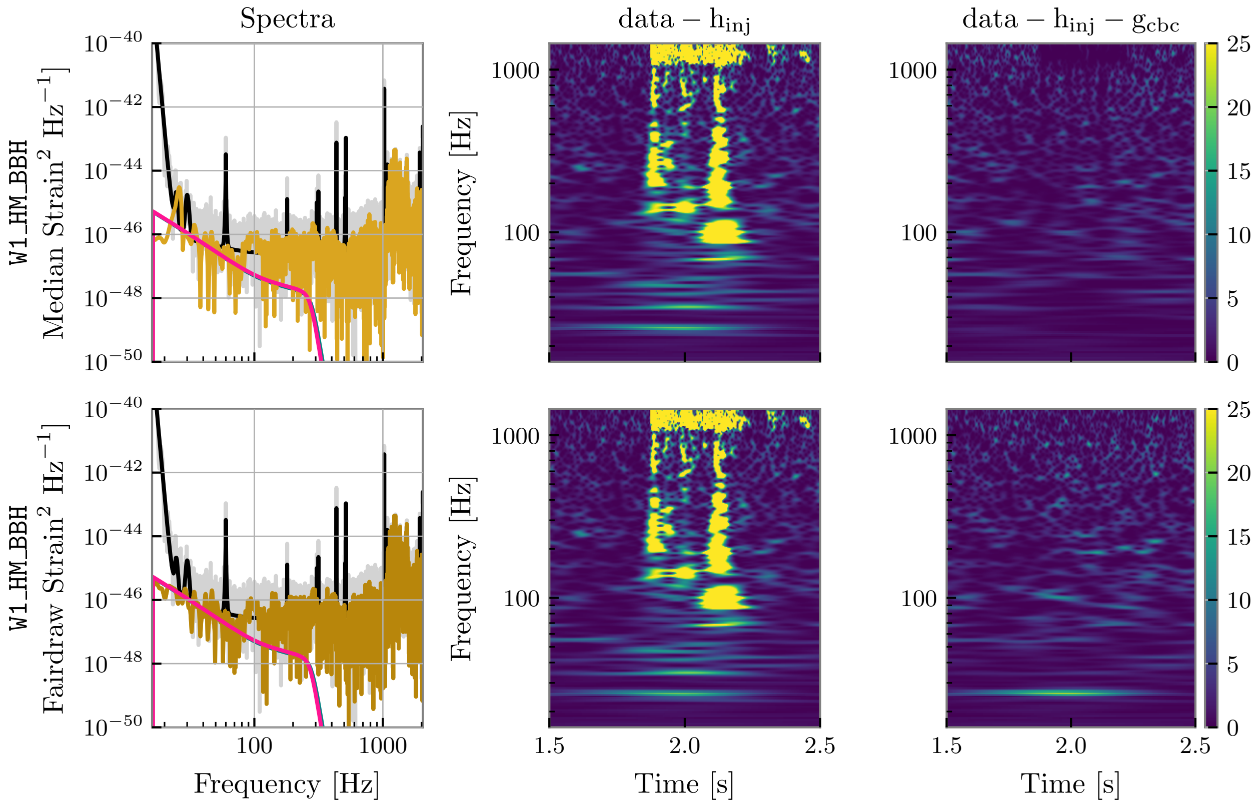

Figure 5.12: Spectra and residual plots for the `W1_HM_BBH` injection at 100ms from the glitch, described in Sec. 5.5 in similar format as Fig 5.7. The right plot shows the final data where we have subtracted the median glitch reconstruction (top panel) or a fair draw from the glitch posterior (bottom panel), leaving behind Gaussian noise. The median reconstruction leads to oversubtraction of the glitch; we therefore favor the fair draw.

Whistle

Whistle glitches are fairly loud glitches with a characteristic morphology depicted in the spectrogram of Fig. 5.3. Our chosen instance of this glitch has an SNR of ~ 275. Given their strength, more than 200 wavelets are required to model them accurately, which poses a considerable challenge for sampler convergence. Given this fact, we only attempted injections on short duration segments.

To aid convergence, we use GlitchBuster to initialize the glitch model, see Sec. 5.3. Despite the short duration, the high frequency of the glitch results in a lot of waveform cycles, we therefore also increase the maximum quality factor of the wavelets. Finally, we also increase the number of iterations within the wavelet RJMCMC (see Fig. 5.2) from 10^2 to 10^3 as this glitch reconstruction requires upwards of 230 wavelets at every posterior sample. By default, we retain one out of 100 samples, and only update (i.e., add/remove/change) a single glitch wavelet at each sampler step. These default settings would therefore not lead to independent samples as not all wavelets have a chance to be updated before a posterior sample is retained. Details about the run settings are provided in Table 5.4.

Results are presented in Fig. 5.11 where we find that we are able to subtract the glitch consistently as well as estimate the CBC parameters. Despite the strength of the glitch, the "CBCOnly" analysis returns mostly unbiased parameter posteriors, possibly due to the fact that whistle glitches are not morphologically similar to high-mass BBHs. The quality of glitch modeling and subtraction is further explored in Fig. 5.12 for the analysis of Fig. 5.11, specifically the injection at 100 ms relative to the glitch. Comparison between the middle and right panel shows that we can efficiently subtract the glitch power. In this case, we also find that the median glitch reconstruction (top panel) results in an *oversubtraction* of the glitch. The fair draw (bottom panel) leads to data that look more consistent with Gaussian noise. For this and the reasons discussed in Sec. 5.5 we generally prefer working with fair posterior draws rather than median glitch reconstructions when making glitch-subtracted data.

Jensen-Shannon Divergence

As a final test, we compute a simple summary statistic for the differences between the "CBCOnly" and the "CBC+Glitch" posteriors: the Jensen-Shannon (JS) divergence. The JS divergence describes the similarity of two distributions with JS $= 0$ for identical distributions and JS $= 1$ for disparate distributions. We plot the median and maximum/minimum JS for D_L, χ_{eff}, and $\mathcal{M}_c$ across CBC injections on the same glitch at different times in Fig. 5.13. With the exception of the tomte glitch (discussed in Sec. 5.5) where the posteriors are completely different, we recover the general trend that the JS divergence is smaller for low masses, which implies that the "CBCOnly" and "CBC+Glitch" posteriors are more similar for lower-mass events. This again supports the previous conclusion that though glitches are more likely to overlap long duration events, glitch subtraction is more important for high-mass than low-mass signals. Among the different parameters, the chirp mass is the one with the lowest

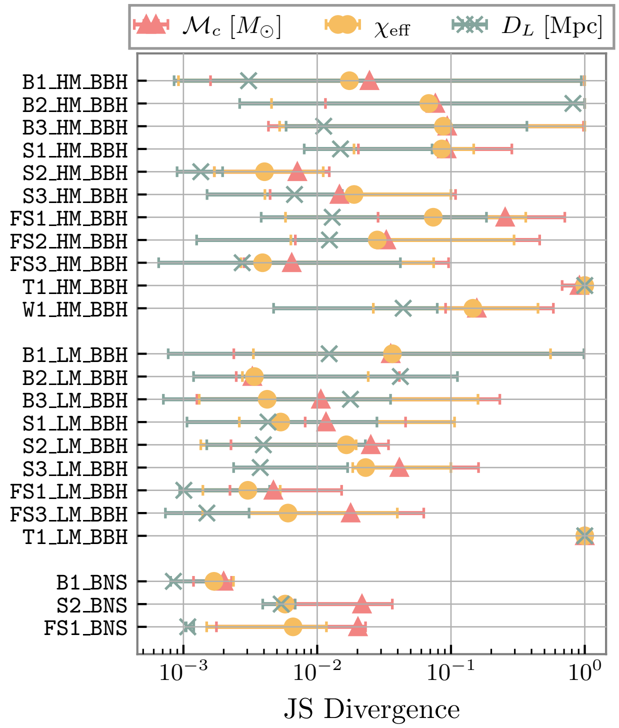

Figure 5.13: Jensen-Shannon (JS) divergence between the "CBC only" and the "CBC+Glitch" marginalized one-dimensional posteriors for D_L, $\chi_{\rm eff}$, and $\mathcal{M}_c$. We then plot the median (marker) and minimum to maximum values (error bars) over CBC injections at different times with respect to the same glitch. The breaks in the y axis of the plot indicate different mass ranges, increasing upwards from the origin. The general trend is that JS increases with the signal mass, again suggesting that glitches affect high-mass systems more.

JS on average, which is expected given the fact that it is the best measured intrinsic source parameter.

5.6 Further validation studies

The results of Sec. 5.5 show that the full "CBC+Glitch" analysis can separate signals and glitches. In this section we provide some further validation tests regarding robustness against waveform systematics in the case of injections including higher-order modes or spin precession. We also assess the performance for signals that are observed by a single detector. Run settings and injection parameters for this section are presented in Tables 5.5 and 5.6 respectively.

Run Label	GPS Time	Injected signal	Glitches	$T_{\rm obs}$ [s]	$T_{\rm w}$ [s]	$Q_{\rm max}$	$D_{\rm max}$	$f_{\rm low}$ [Hz]
`B1_HM_BBH_HOM`	1168989748	HM BBH w/ HOM	Blip LHO	4	1	40	100	16
`T1_HM_BBH_SPIN`	1243679046	HM BBH w/ SPIN	Tomte LLO	4	1	40	100	8
`B1_HM_BBH_SING`	1168989748	HM BBH w/ SING	Blip LHO	4	3	40	100	16

Table 5.5: Settings for the runs of Sec. 5.6 that test the effect of omitting higher-order modes (`HOM`), spin precession (`SPIN`), or considering only a single detector (`SING`). Columns give the same information as those of Table 5.4.

Signal	Injected Waveform	Varied param.	min.	max.
`HOM`	`IMRPhenomHM`	$\cos(\iota)$	-1	1
`SPIN`	`IMRPhenomPv2`	χ_p	0.23	0.60

Table 5.6: Parameters of the injected signals for the tests of higher-order modes (`HOM`) and spin precession (`SPIN`). For high-order modes (spin precession) we vary $\cos(\iota)$ (χ_p) between a minimum and a maximum value (third and fourth columns) to modify the strength of the deviation between the `IMRPhenomD` recovery waveform and the injected waveform.

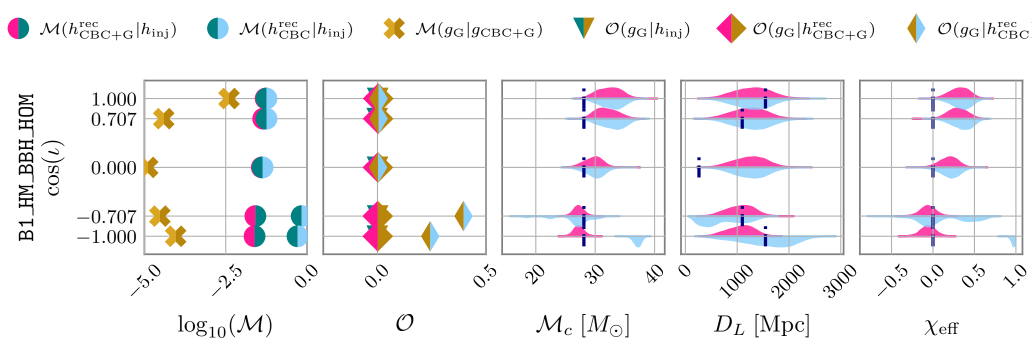

Figure 5.14: Similar to Fig. 5.5 for a blip glitch and different high-mass BBH signals injected with higher-order modes but recovered without. The y axis now shows the binary inclination. See Table 5.5 for run settings and labels.

Waveform Systematics: Higher-Order Modes

The results of Sec. 5.5 are based on IMRPhenomD [198, 207], a waveform model that does not include higher-order modes, i.e., power from spherical harmonics beyond the dominant $l = |m| = 2$ mode. Such modes change the waveform morphology and thus neglecting them will lead to biases, especially for high SNR, unequal-mass systems, observed "edge-on" ($\cos\iota = 0$) [93, 94, 112, 270, 360, 361][4]. Our current CBC sampler can work with waveform models that include higher-order modes such as IMRPhenomHM [228], however, we

[4]The inclination $\iota \in [0, \pi]$ is defined as the angle between our line-of-sight and the binary's Newtonian orbital angular momentum.

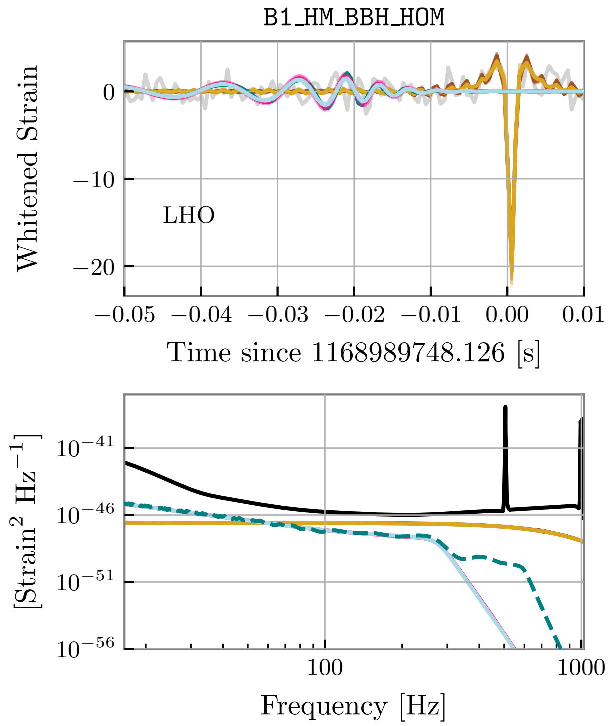

Figure 5.15: Whitened time-domain reconstruction (top) and spectrum (bottom) for a high-mass BBH injected with higher-order modes and edge-on. We show medians and 90% credible intervals for the CBC signal (magenta), the glitch (gold), and the noise PSD (grey/black) from the "CBC+glitch" analysis. The injected signal is given with a dashed teal line. The higher-order modes result in oscillations in the inspiral spectral amplitude as well as additional power at high frequencies. The CBC waveform used for the reconstruction does not include higher-order modes, however the reconstructed glitch model does not recover the excess power from higher-order modes. The CBC model from a "CBCOnly" analysis (light blue) is similar to the one from the "CBC+glitch" analysis.

do not perform such runs here because we lack an implementation of the heterodyne procedure [119, 120] that speeds up the likelihood calculation for such waveforms. Such an extension was described in [221] and we plan to implement it in the future.

Because a real signal will inevitably contain some amount of higher-order modes, recovery with a waveform that neglects them could induce a systematic error in parameter extraction. Perhaps even more worrisome would the possibility that the glitch model subsumes some of the higher-order mode power which is then inadvertently subtracted from the data together with the glitch. We check for both effects by injecting the high-mass BBH signal from Table 5.1 using IMRPhenomHM [228] with varying inclination $\cos\iota \in [-1, 1]$ onto one of our blip glitches and recover them again with IMRPhenomD. Figure 5.14 shows recovered parameters for different system inclinations and Fig. 5.15 shows the recovered CBC and glitch reconstruction for the case with $\cos\iota = 0$.

Compared to the top row of Fig. 5.5, i.e., the same injected CBC and glitch but without higher-order modes and different inclinations, we find that $\mathcal{M}(g_{\rm G}|g_{\rm CBC+G})$ has increased, but still remains ≤ 0.01. The other mismatches are comparable between Figs. 5.14 and 5.5. Despite the increase in $\mathcal{M}(g_{\rm G}|g_{\rm CBC+G})$, its value remains small and comparable to results from other blip glitches without higher-order modes, for example the second row of Fig. 5.5. We attribute this to the fact that the higher-order mode power is still too low to overcome the parsimony requirement of the glitch model to be picked up. Figure 5.15 further reinforces this picture, by showing that the main effect of higher-order modes is additional high-frequency power that cannot be captured by the template (compare the magenta reconstruction to the teal injection). However, this residual power is still two orders of magnitude below the glitch power in the same frequency range.

Figure 5.14 also confirms that the amount of bias expected on the CBC parameters (notably D_L) is a function of the binary inclination. Furthermore, we find that not including the glitch in the model now leads to more pronounced parameter biases (blue violin plots). Regardless, even in the edge-on case the glitch and CBC signal can be separated sufficiently well as demonstrated by the mismatches in the first column. Higher SNR signals or, in general, a signal with more than SNR$\sim 6-7$ in the higher-order modes could

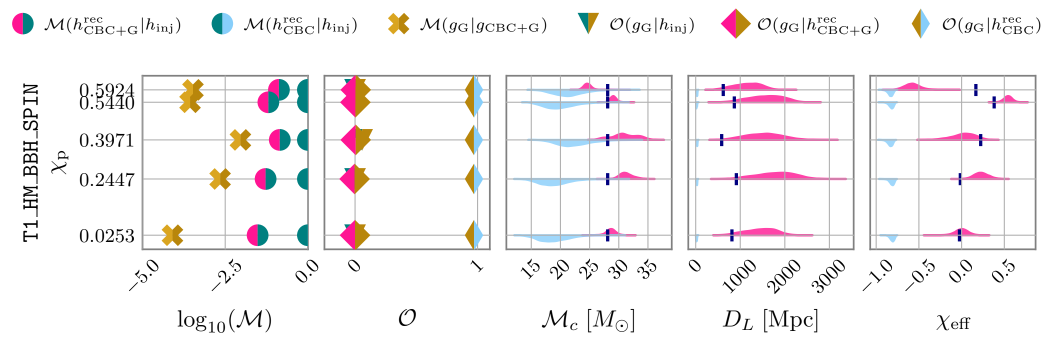

Figure 5.16: Similar to Fig. 5.5 but for a tomte glitch and different precessing high-mass BBH signals as a function of the injected binary precession parameter $\chi_{\rm p}$. See Table 5.5 for run settings and labels.

have more noticeable deviations from waveforms without higher-order modes where their power could then be picked up by the glitch model. However, current events with detectable higher-order mode content are below this SNR threshold [24, 25, 193, 250, 387].

Waveform Systematics: Spin Precession

In physical scenarios where the component spins are misaligned with the orbital angular momentum, the binary system experiences spin-precession which modulates the observed waveform [61]. The current implementation of the CBC sampler used here only accounts for spins aligned with the orbital angular momentum, so we assumed that the injected waveforms in our main analysis were also nonprecessing, motivated also by the lack of strong precession effects in event catalogues [37, 38]. However, signals with large in-plane spins, unequal masses, and/or observed edge-on could exhibit strong precessional effects [62, 104, 106, 206, 208, 209, 261, 280, 281, 313, 314]. The CBC sampler will be extended to include misaligned spin degrees of freedom in the future.

Similar to our analysis of the impact of higher-order modes, we study the impact of using nonprecessing templates by performing injections of high-mass BBHs with misaligned spins on the tomte glitch. The tomte glitch family was selected for this study as it is similar morphologically to highly-spinning, massive BBHs [68], and also because it consistently leads to the largest biases in CBC parameters when mismodeled (see Fig. 5.9). For the injections we use IMRPhenomPv2 [178] and we recover the signals with the

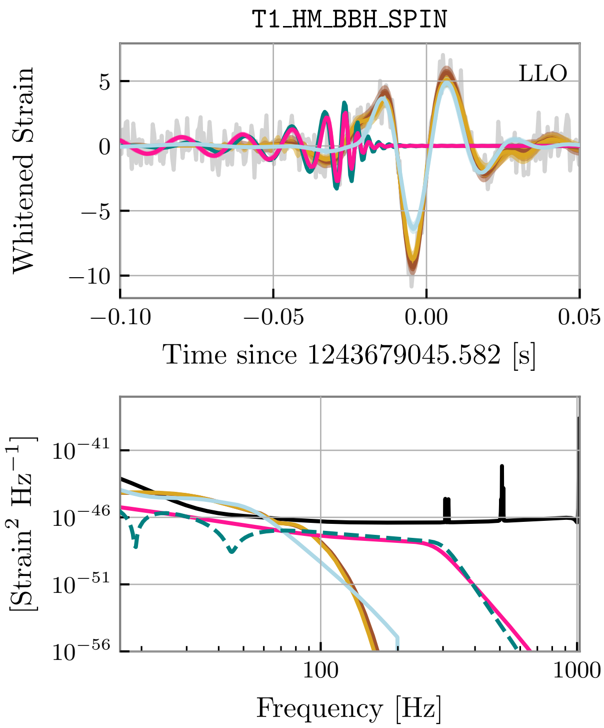

Figure 5.17: Similar to Fig. 5.15 for a spin-precessing high-mass BBH signal with $\chi_\mathrm{p} = 0.592$. Spin precession induces oscillations in the inspiral spectral amplitude, most visible at around 40 Hz. The CBC waveform template used for the reconstruction assumes the spins are aligned with the orbital angular momentum, however the reconstructed glitch model appears unaffected. The CBC model from a "CBCOnly" analysis instead attempts to recover the glitch.

same IMRPhenomD waveform as before. The degree of spin-precession in a signal is commonly characterized by the parameter χ_p which is proportional to a mass-weighted maximum (over the two compact objects) of the in-plane spin magnitude [315]. Its range is $\chi_\mathrm{p} \in [0, 1]$, where $\chi_\mathrm{p} = 0$ describes a system with aligned spins (no precession) and $\chi_\mathrm{p} = 1$ is a maximally precessing system. Figure 5.16 shows results for different values of χ_p and Fig. 5.17 shows the recovered CBC and glitch reconstructions for the case the largest χ_p.

Compared to Fig. 5.9 that shows results with the same signal and glitch but with a nonprecessing injection, we find that parameter recovery is increasingly biased as the injected χ_p increases. Again the "CBCOnly" analysis (blue violin plots) leads to overwhelming parameter biases, which seems to be a generic feature of tomte glitches. This is again reflected in the fact that $\mathcal{O}(h^\mathrm{rec}_\mathrm{CBC}|g_\mathrm{G}) \sim 1$ whereas $\mathcal{O}(g_\mathrm{G}|h^\mathrm{inj}) \sim 0$. Importantly, however, $\mathcal{M}(g_\mathrm{G}|g_\mathrm{CBC+G}) \leq 0.01$ and

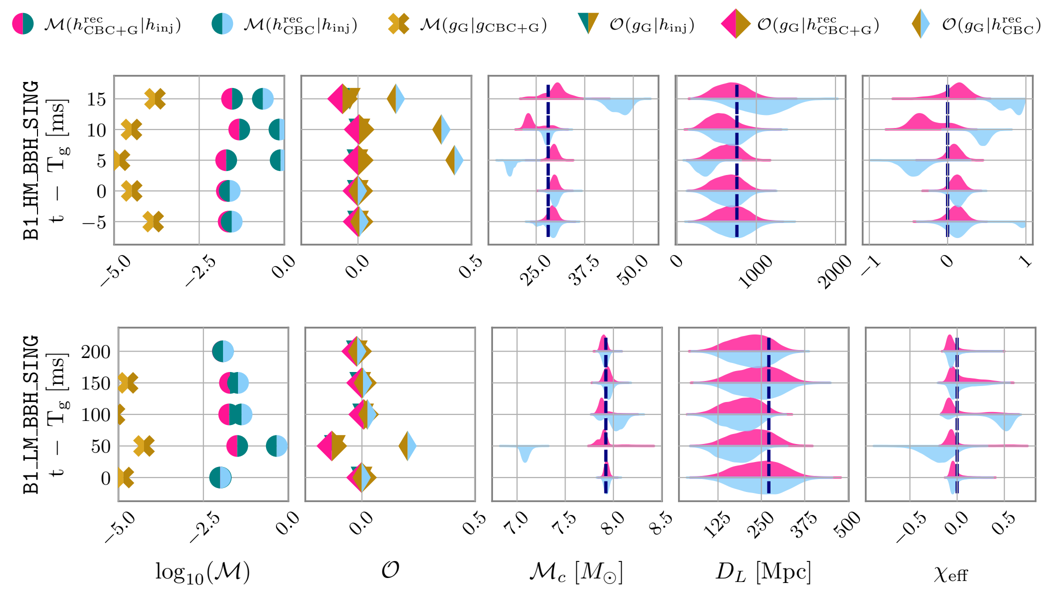

Figure 5.18: Similar to Fig. 5.5 but for a blip glitch and high-mass BBH (top) and low-mass BBH (bottom) signals observed in a single LHO detector. See Table 5.5 for run settings and labels.

is similar to the corresponding mismatches of Fig. 5.9; this means that the excess power due to spin-precession is not recovered by the glitch model, as its power is too low to be significant to the glitch model.

A similar conclusion is drawn from Fig. 5.17. The recovered signal is noticeably different from the injected signal, most prominently shown in the bottom panel where the precession-induced amplitude oscillations are absent from the posterior. However, again the glitch reconstruction appears to be unaffected, most likely because the relevant residual power is 2–3 orders of magnitude below the glitch power in the relevant frequency range. Higher binary inclinations and higher signal SNRs might make the difference between precessing and nonprecessing waveforms more stark, causing the glitch model to capture any residual power due to spin precession. However, with current signal strengths and inferred amounts of spin precession, we find this to be unlikely. Though the CBC parameters recovered are clearly biased, the glitch modeling appears to be robust. Glitch-subtracted data can therefore be constructed and further analyzed with more complete waveform models. Nonetheless, we plan to extend the analysis to include spin-precession effects in the CBC sampler.

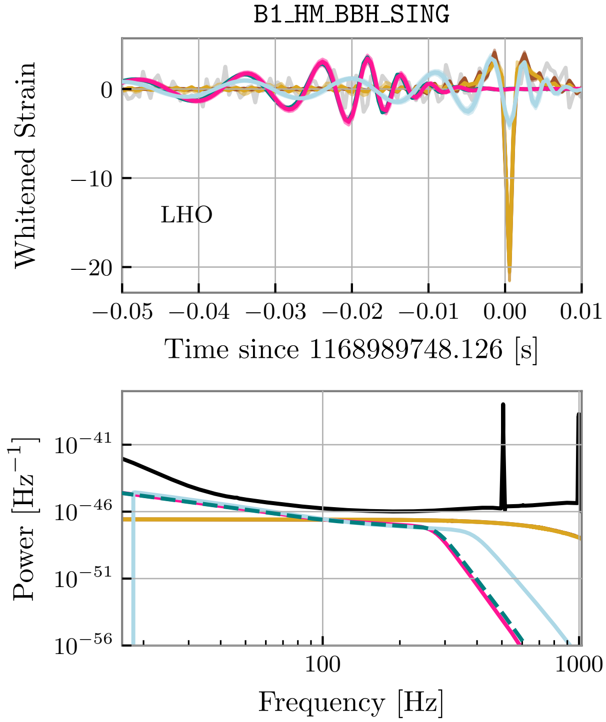

Figure 5.19: Similar to Fig. 5.15 for the B1_HM_BBH_SING injection at 100ms from the glitch. When using the full "CBC+Glitch" model, $h^{\rm rec}_{\rm CBC+G}$ and $g_{\rm CBC+G}$ recover the CBC signal and glitch respectively even when data from a single detector only are available. The CBC model from a "CBCOnly" analysis instead attempts to recover the glitch.

Single-detector signals

The joint analysis of CBCs and glitches assumes that the astrophysical signal is coherent across the detector network, while the glitch is not. Hence, the results presented so far are based on data from both LIGO detectors. However, single-detector candidates have been reported [38], and we therefore test here if our analysis could separate them from glitches. The CBC waveform template we employ is a fairly restrictive model and we indeed find that this allows us in some cases to separate them from glitches even in single-detector data. Such a separation would be inherently impossible for the previous `BayesWave` analyses that distinguished between signal and glitch solely via coherence within a detector network [124].

We revisit runs `B1_BBH_HM` and `B1_BBH_LM` from Table 5.4 and analyze now only the LHO data that contain the glitch. We decrease the distance so that

Trigger	GPS Time	$T_{\rm obs}$ [s]	$T_{\rm w}$ [s]	$Q_{\rm max}$	$D_{\rm max}$	$f_{\rm low}$ [Hz]
S191225	1261346253	4	1	40	100	16
S200114	1263002916	4	2	40	100	16
Tomte 1	1243679046	4	1	40	100	16

Table 5.7: Settings for the analyses of Sec. 5.7. The first two columns provide the trigger name and GPS time. The remaining columns are the same as Table 5.4. Additionally, S191225 used a sampling rate of 1024 Hz whereas S200114 used a sampling rate of 2048 Hz.

the single-detector SNR is 15 for consistency with all other analyses. We present parameter results in Fig. 5.18 and find that we are largely able to seperate the signal and the glitch. Even in this single-detector case the full "CBC+Glitch" model outperforms the simpler "CBCOnly" analysis and the glitch reconstruction is consistent with the one from data with no CBC injections. An example reconstruction plot is shown in Fig. 5.19 where again the CBC and glitch components of the full "CBC+Glitch" model recover their corresponding data component. The "CBCOnly" analysis, on the other hand, largely mistakes the glitch for a signal. Although these preliminary results are promising, we remain cautious of such cases. If presented with a similar scenario during an actual observing run, our analysis would require additional case-by-case attention and testing.

5.7 Classifying triggers

The joint "CBC+Glitch" analysis simultaneously models CBC signals and glitches, however, the priors for both the CBC and the glitch model allow for the possibility of no CBC and/or no glitch in the data. In the glitch case, this is straightforward, as the model allows for 0 wavelets in all detectors, as was for example recovered in the case of GW150914 in [110]. While the CBC priors ensure the presence of a CBC in the data, the luminosity distance prior extends to 10Gpc, which effectively corresponds to a signal with negligibly small SNR. This suggests that our analysis can be used to assess whether certain detected excess power consists of a CBC signal, a glitch, both, or neither. We revisit two low-significance candidates from [33, 40] and analyze them with the joint "CBC+Glitch" and the "CBCOnly" analysis. Analysis settings are provided in Table 5.7.

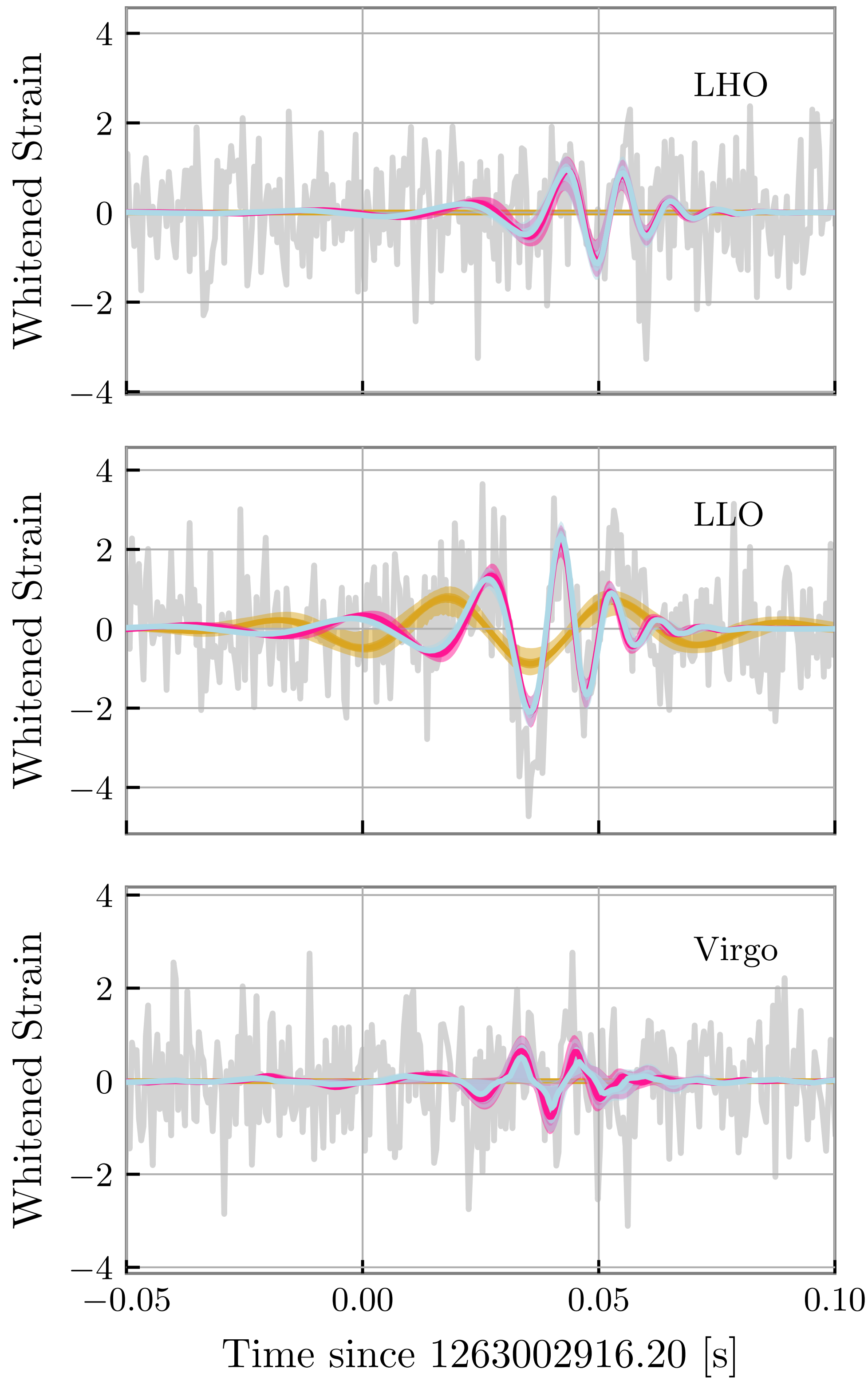

Figure 5.20: Whitened time-domain reconstruction for S200114 in each detector. We show medians and 90% credible intervals for $h^{\mathrm{rec}}_{\mathrm{CBC+G}}$ (magenta), $g_{\mathrm{CBC+G}}$ (gold), and $h^{\mathrm{rec}}_{\mathrm{CBC}}$ (blue). The data are consistent with the presence of both a CBC signal (i.e., coherent power that is morphologically similar to a CBC), and a glitch (i.e., additional incoherent power in LLO).

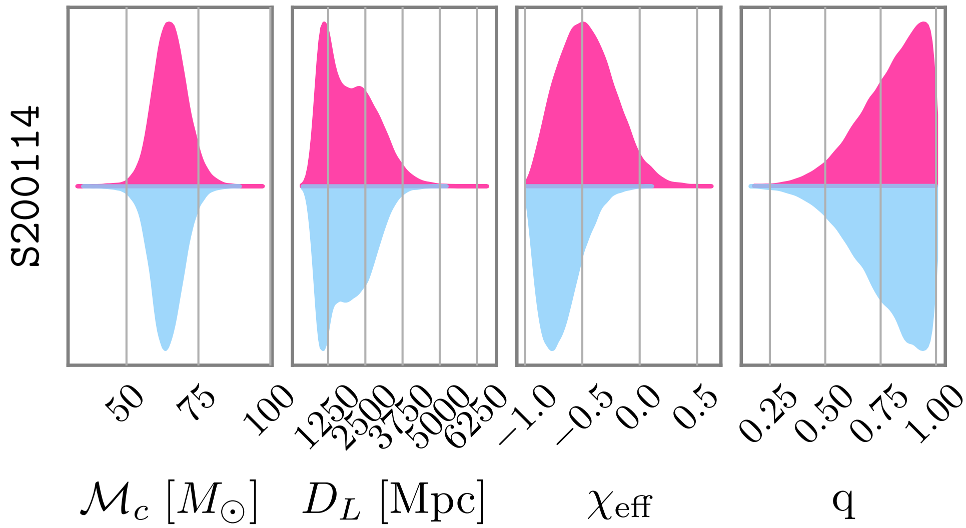

Figure 5.21: Posteriors for select parameters for S200114 from the "CBC+Glitch" analysis (pink) to the "CBCOnly" analysis (blue). Differences between these posteriors are smaller than those between different waveform models reported in [40].

S200114

We begin with 200114_020818 (referred to as S200114 from now on), which was identified by Coherent Wave Burst [212] with a false-alarm rate of 0.058 yr^{-1} [40]. Despite the low false alarm rate, the conclusion was that although an astrophysical origin could not be excluded, the trigger is of likely glitch origin since the estimated CBC parameters depended heavily on the choice of waveform model.

Figure 5.20 shows the time-domain reconstruction in each detector and Fig. 5.21 shows a few marginalized parameter posteriors. The joint "CBC+Glitch" analysis is consistent with the presence of both a CBC signal (at the 90% credible level) and a low-frequency glitch in LLO. The morphology of the LLO glitch is consistent with a low-frequency blip or a tomte. This suggests that while there is excess power that is morphologically similar to a CBC and is coherent between LHO and LLO, there is also additional incoherent power in LLO that is captured by the glitch model. The CBC reconstruction is consistent with zero in Virgo at the 90% credible level. Additionally, the "CBCOnly" analysis finds a broadly consistent CBC reconstruction, with small differences at low frequencies. Differences between the "CBCOnly" and "CBC+glitch" CBC parameter posteriors are much smaller than the waveform systematics reported in [40].

The presence of some amount of coherent power is consistent with the low false-alarm rate reported by a coherent detection pipeline in [40]. The additional incoherent power in LLO could explain the inconsistent parameter estimation results between different waveform models presented in [40], especially since the `NRSur7dq4` [362], `SEOBNRv4PHM` [261], and `IMRPhenomXPHM` [281] waveform models used in [40] include the effects of spin-precession and higher-order modes. Thus they can account for more complicated morphologies [95, 264] in the CBC signal than the `PhenomD` model used here.

The glitch in LLO is similar to a tomte which we have found to be recovered as a CBC in "CBCOnly" analyses. To check whether a single tomte glitch could trick the "CBC+glitch" analysis into concluding that both a CBC and a glitch are present in the data, we revisit the tomte glitch from Table 5.4, perform no CBC injection, and carry out a "CBC+Glitch" analysis. Reassuringly, the sampler indeed converges to the correct answer as shown in Fig. 5.22, namely that no CBC is present in the data, rather only a glitch. This suggests that the results of Fig. 5.20 cannot be the outcome of a single tomte glitch and the "CBC+glitch" analysis does not recover coherent power when there is none.

However, we do not attempt to obtain a background estimate and thus cannot assess the probability that such a combination of coherent/incoherent power has a terrestrial origin. Such a full background estimate could result in the calculation of a false alarm rate similarly to the `BayesWave` analysis in [14, 27] based on the signal and glitch models only. In our case, we would use the "CBC+glitch" analysis to estimate Bayes Factors for the various models of interest and compare them to similar results obtained from data that have been shifted in time between LHO and LLO.

S191225

We then consider 191225_215715 (labeled S191225 from now on), a low-significance LLO-Virgo trigger found by the PyCBC Live [257] and the PyCBC-IMBH [101] searches with false-alarm rates of 0.4 yr^{-1} and 0.47 yr^{-1} respectively in O3 [38, 40]. This candidate was ultimately deemed a glitch due to similar detector behavior surrounding the event. The reconstructed time-

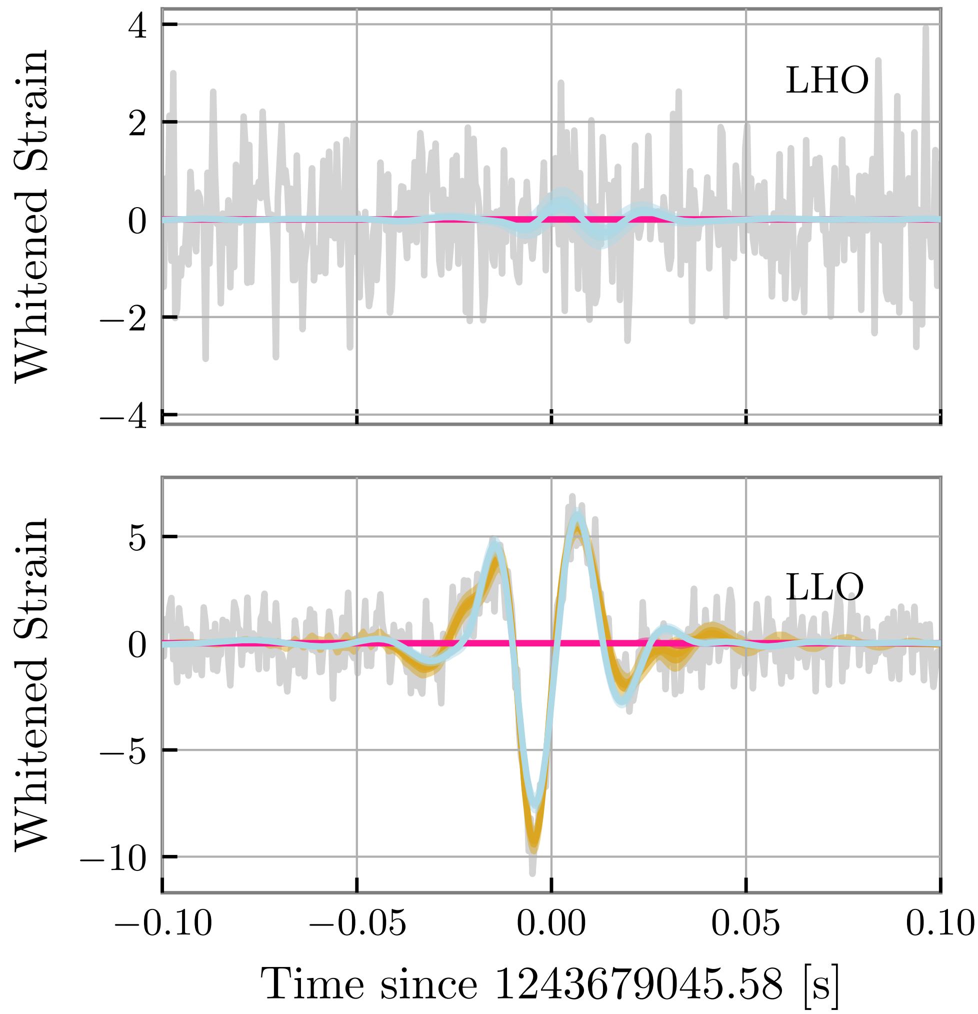

Figure 5.22: Whitened time-domain reconstruction for an analysis of data that contain only a single tomte glitch. We show medians and 90% credible intervals for $h^{\mathrm{rec}}_{\mathrm{CBC+G}}$ (magenta), $g_{\mathrm{CBC+G}}$ (gold), and $h^{\mathrm{rec}}_{\mathrm{CBC}}$ (blue). The data are consistent with the presence of solely a tomte glitch in LLO and do not recover any coherent power (the magenta CBC reconstruction is consistent with zero) when there is none.

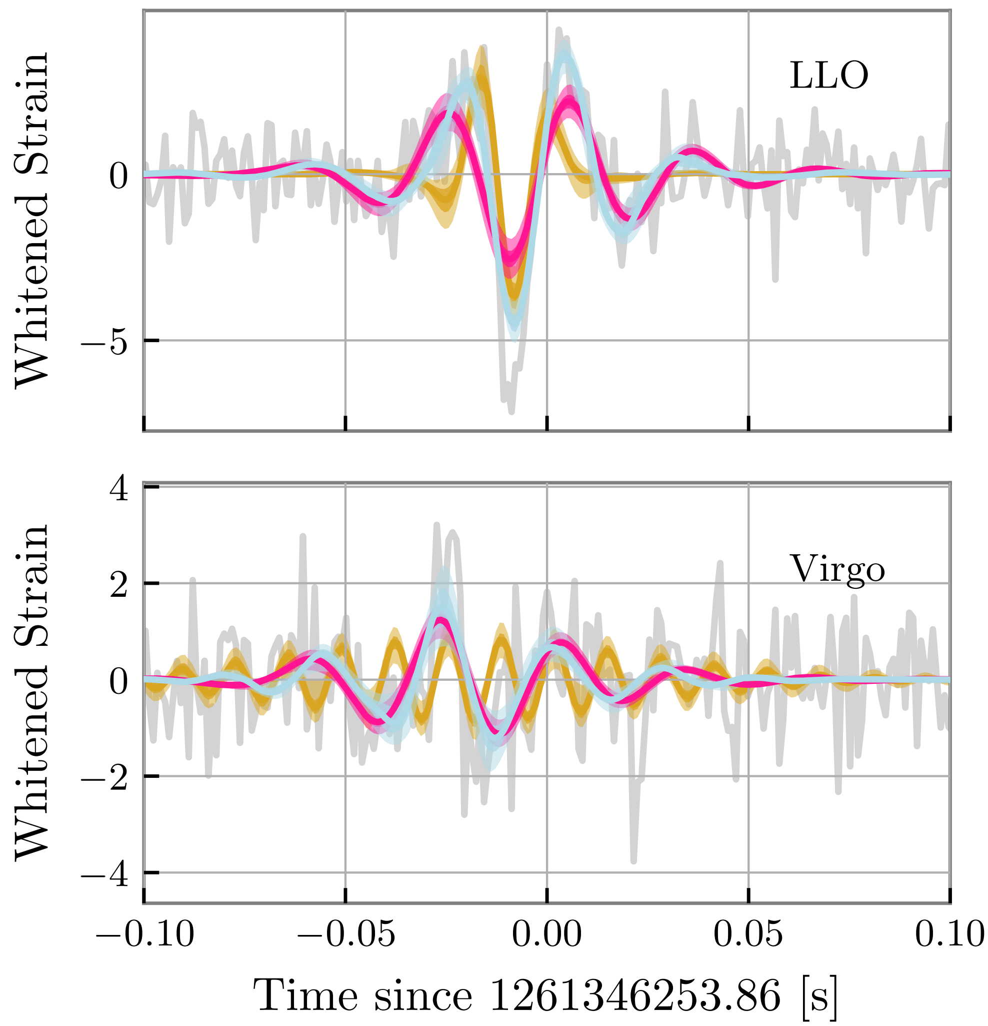

Figure 5.23: Whitened time-domain reconstruction for S191225 in LLO (top) and Virgo (bottom). We show the median and 90% credible intervals for $h^{\mathrm{rec}}_{\mathrm{CBC+G}}$ (magenta), $g_{\mathrm{CBC+G}}$ (gold), and $h^{\mathrm{rec}}_{\mathrm{CBC}}$ (blue). Some low-frequency coherent power is recovered by $h^{\mathrm{rec}}_{\mathrm{CBC+G}}$ whereas the high-frequency power is largely incoherent and is recovered by the glitch model.

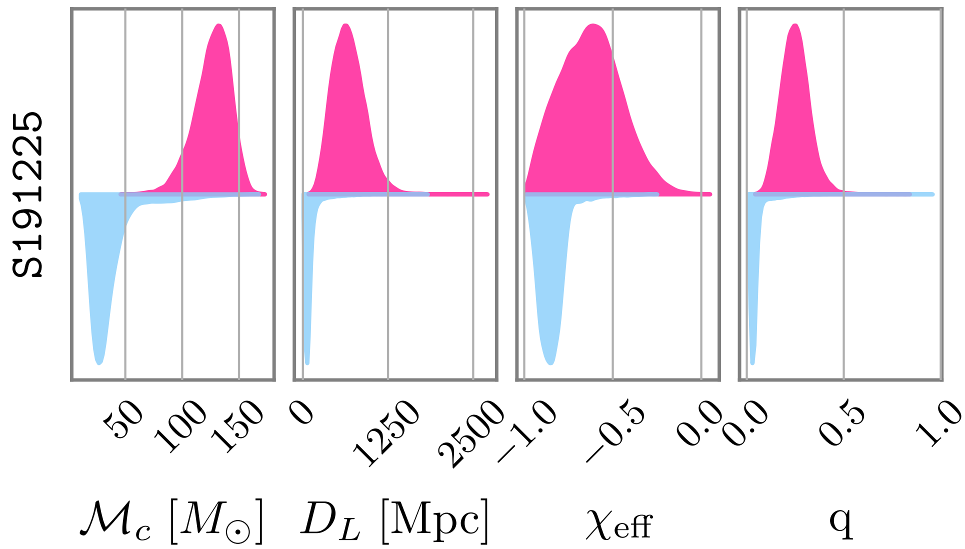

Figure 5.24: Posteriors for select parameters for S191225 from the "CBC+Glitch" analysis (pink) to the "CBCOnly" analysis (blue).

domain signal is shown in Fig. 5.23 for each detector, where we find that the data are consistent with a very high mass CBC and a glitch.

When contrasting parameters from the "CBCOnly" analysis to the "CBC+Glitch" analysis. Fig. 5.24, we find that the former displays the telltale signs of a glitch; negative $\chi_{\rm eff}$ and unequal masses [68]. The latter still recovers some coherent power with recovered parameters instead pointing to a much higher mass binary (total mass $> 300 M_\odot$). Since the new recovered mass is much larger than the original one, we might expect the false-alarm rate for this event to increase, although a full background estimate is outside the scope of this study.

Overall, we find that S191225 is consistent with a glitch in LLO (possibly a tomte) atop of some coherent power, which agrees and expands upon with the conclusions of [38, 40].

5.8 Conclusions

The various models that form the `BayesWave` algorithm allow us to analyze GW data that include multiple components, specifically noise PSD, glitches, and a CBC signal. We present multiple tests with injected signals that overlap with real LIGO glitches and show that we can reliably separate signals and glitches, estimate the CBC parameters, and provide estimates for the glitch to

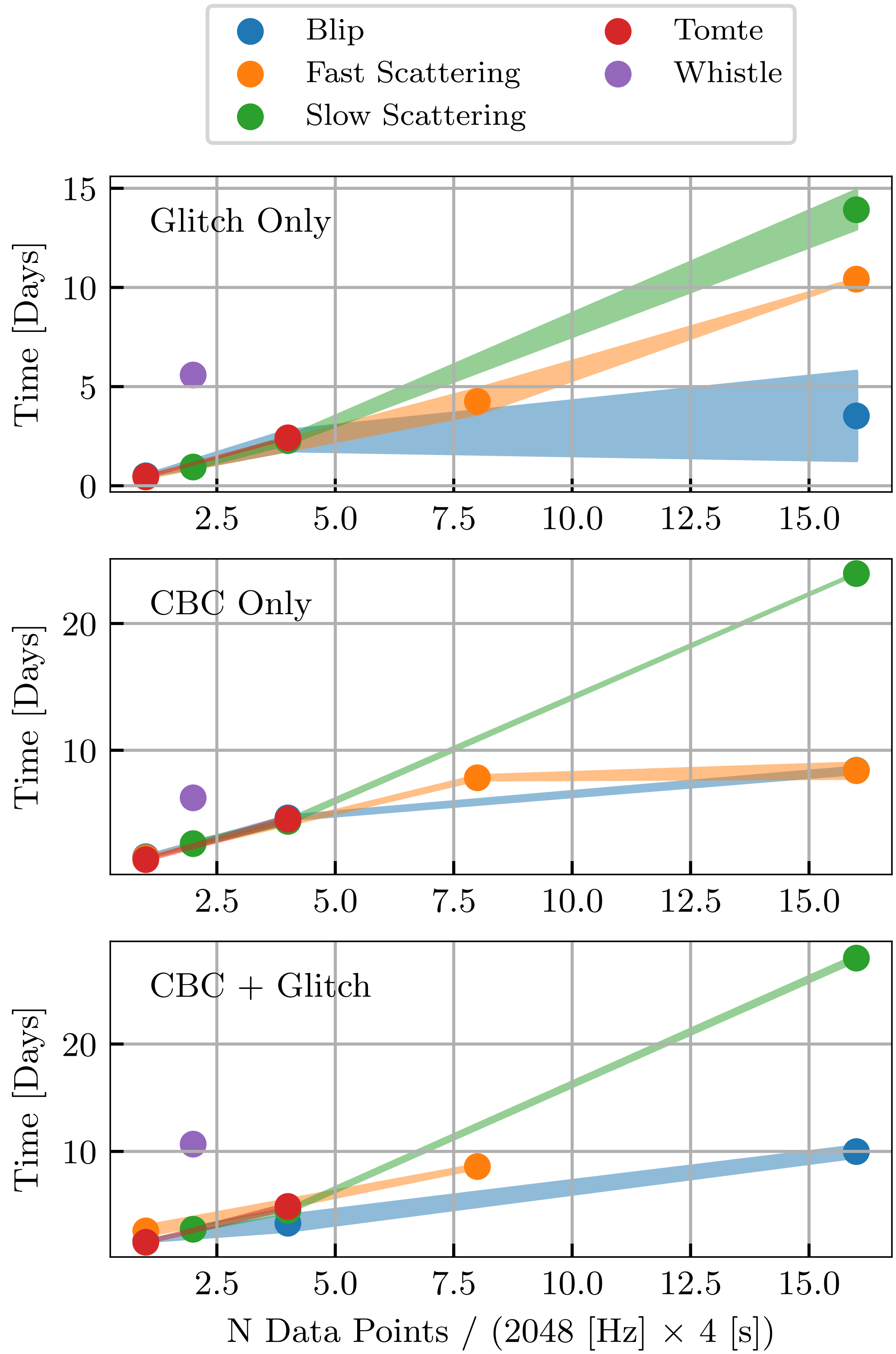

Figure 5.25: Run time estimates (90% intervals) for each glitch type and run setting we employ as a function of data points N (segment length $\times$ sample rate). The x-axis is normalized by the shortest runs performed. Since runtime is (approximately) linear with the number of MCMC samples and number of chains, we rescale estimates to Number of Chains = 20 and Number of Iterations = 4×10^6, which are the default settings. Lighter settings can be used to expedite certain analyses.

be subtracted from the data. Runtime estimates for the various analyses are presented in App. 5.10.

Our analysis is able to identify all glitches analyzed. It is particularly reliable for short-duration glitches such as blips, tomtes, and low-frequency blips which could be tackled in an automated way with default analysis settings. These are also the glitch types that cause the largest biases for CBC parameters when left unaccounted for. Long-duration glitches such as fast and slow scattering glitches are more challenging and in some cases need specialized settings. Crucially, the necessary settings (such as maximum wavelet quality factor or analysis segment length) differ even between glitches of the same family, suggesting that automation is not yet feasible. This effect is particularly prominent for fast-scattering glitches that create long periods of nonstationarity that challenge PSD estimation, especially at low frequencies (below 40 Hz). Luckily, these glitches incur smaller biases in CBC parameter, most likely for the same reason they are difficult to model: they have a large time-frequency footprint that does not resemble a CBC chirp.

In this study, (with one exception) changes on glitch wavelet priors concerned their ranges, while their shapes were unaltered compared to default settings. However, dedicated glitch priors that target particularly problematic glitch families would improve glitch modeling. For example, a prior that favors wavelets with lower amplitude can lead to improved results for the low-SNR slow-scattering glitch (`S2`). Further examples of dedicated priors include a trained model based on a principal component analysis for tomte glitches [244] or prior information about the frequency spacing of the scattering arches [41, 322]. Depending on the characteristics of the most prevalent and problematic glitch types in O4, we plan to explore such dedicated priors in the future.

5.9 Acknowledgements

This research has made use of data, software and/or web tools obtained from the gravitational-wave Open Science Center (https://www.gw-openscience.org), a service of LIGO Laboratory, the LIGO Scientific Collaboration and the Virgo Collaboration. Virgo is funded by the French Centre National de Recherche Scientifique (CNRS), the Italian Istituto Nazionale della Fisica Nucleare (INFN) and the Dutch Nikhef, with contributions by Polish and Hungarian institutes. This material is based upon work supported by NSF's LIGO Laboratory which is a major facility fully funded by the National Science Foundation. The authors are grateful for computational resources provided by the LIGO Laboratory and supported by NSF Grants PHY-0757058 and PHY-0823459. S.H. and K.C. were supported by NSF Grant No. PHY-2110111. D.D. was supported by the National Science Foundation as part of the LIGO laboratory T.L. was supported by funding from the NASA LISA study office. N.J.C. was supported by NSF Grant No. PHY-1912053. Software: gwpy [235], matplotlib [197].

5.10 Appendix

Runtimes

Including a variable-dimensional glitch and PSD model comes at an additional computational cost compared to standard CBC analyses. We display the runtimes for the different glitch types and analyses presented in this paper in Fig. 5.25. There is an essentially linear dependence between the number of data points (the segment length times the sampling rate) and the run time. The whistle glitch runs are outliers as additional settings were required to subtract such high-SNR glitches (see Sec. 5.5). Since runtime is (approximately) linear with the number of parallel chains and the number of iterations, we rescale time by the default settings of 20 chains and 4×10^6 iterations, though lighter settings can be used for expedited results. For example, the BNS runs for Fast-Scattering and Blip glitches used 10 chains for speed, and the whistle glitch was run with 25 chains for convergence. The rest were run with 20 chains. These estimates are further based on single-core runs and will be sped up accordingly by ongoing work to parallelize the chains.

Chapter 6

IMPACT OF GLITCH MITIGATION ON GW200129

This chapter contains work from

Ethan Payne, **Sophie Hourihane**, Jacob Golomb, Rhiannon Udall, Richard Udall, Derek Davis, and Katerina Chatziioannou. Curious case of GW200129: Interplay between spin-precession inference and data-quality issues. *Phys. Rev. D*, 106(10):104017, 2022. doi: 10.1103/PhysRevD.106.104017. Reprinted here as Chapter 6. **SH** helped conceptualize the project, led all BayesWave analyses, and created related figures, and authored related text.

6.1 Abstract

Measurement of spin-precession in black hole binary mergers observed with gravitational waves is an exciting milestone as it relates to both general relativistic dynamics and astrophysical binary formation scenarios. In this study, we revisit the evidence for spin-precession in GW200129 and localize its origin to data in LIGO Livingston in the 20–50 Hz frequency range where the signal amplitude is lower than expected from a non-precessing binary given all the other data. These data are subject to known data quality issues as a glitch was subtracted from the detector's strain data. The lack of evidence for spin-precession in LIGO Hanford leads to a noticeable inconsistency between the inferred binary mass ratio and precessing spin in the two LIGO detectors, something not expected from solely different Gaussian noise realizations. We revisit the LIGO Livingston glitch mitigation and show that the difference between a spin-precessing and a non-precessing interpretation for GW200129 is smaller than the statistical and systematic uncertainty of the glitch subtraction, finding that the support for spin-precession depends sensitively on the glitch modeling. We also investigate the signal-to-noise ratio ~ 7 trigger in the less sensitive Virgo detector. Though not influencing the spin-precession studies, the Virgo trigger is grossly inconsistent with the ones in LIGO Hanford and LIGO Livingston as it points to a much heavier system. We interpret the Virgo data in the context of further data quality issues. While our results do not disprove the presence of spin-precession in GW200129, we argue that any such inference is contingent upon the statistical and systematic uncertainty of the glitch mitigation. Our study highlights the role of data quality investi-

gations when inferring subtle effects such as spin-precession for short signals such as the ones produced by high-mass systems.

6.2 Introduction

GW200129_065458 (henceforth GW200129) is a gravitational wave (GW) candidate reported in GWTC-3 [38]. The signal was observed by all three LIGO-Virgo detectors [1, 43] operational during the third observing run (O3) and it is consistent with the coalescence of two black holes (BHs) with source-frame masses $34.5^{+9.9}_{-3.2}\,M_\odot$ and $28.9^{+3.4}_{-9.3}\,M_\odot$ at the 90% credible level. Though the masses are typical within the population of observed events [37], the event's signal-to-noise-ratio (SNR) of $26.8^{+0.2}_{-0.2}$ makes it the loudest binary BH (BBH) observed to date. Additionally, it is one of the loudest triggers in the Virgo detector with a detected SNR of 6–7 depending on the detection pipeline [38]. The signal temporally overlapped with a glitch in the LIGO Livingston detector, which was subtracted using information from auxiliary channels [138]. The detection and glitch mitigation procedures for this event are recapped in App. 6.9.

The interpretation of some events in GWTC-3 was impacted by waveform systematics, with GW200129 being one of the most extreme examples. As part of the catalog, results were obtained with the IMRPhenomXPHM [281] and SEOBNRv4PHM [261] waveform models using the parameter inference algorithms Bilby [69, 299] and RIFT [381] respectively. Both waveforms correspond to quasicircular binary inspirals and include high-order radiation modes and the effect of relativistic spin-precession arising from interactions between the component spins and the orbital angular momentum. All analyses used the glitch-subtracted LIGO Livingston data. The IMRPhenomXPHM result was characterized by large spins and a bimodal structure with peaks at ~ 0.45 and ~ 0.9 for the binary mass ratio. The SEOBNRv4PHM results, on the other hand, pointed to more moderate spins and near equal binary masses. Both waveforms, however, reported a mass-weighted spin aligned with the Newtonian orbital angular momentum of $\chi_{\rm eff} \sim 0.1$, and thus the inferred large spins with IMRPhenomXPHM corresponded to spin components in the binary orbital plane and spin-precession. Such differences between the waveform models are not unexpected for high SNR signals [285]. Waveform systematics are also likely more prominent when it comes to spin-precession, as modeling prescriptions vary and are not calibrated to numerical relativity simulations featuring spin-precession [261, 280, 281]. Data quality issues could further lead to evidence for spin-precession [68]. Due to differences in the inference algorithms and waveform systematics, GWTC-3 argued that definitive conclusions could not be drawn regarding the possibility of spin-precession in this event [38].

Stronger conclusions in favor of spin-precession [179] and a merger remnant that experienced a large recoil velocity [364] were put forward by means of a third waveform model. `NRSur7dq4` [362] is a surrogate to numerical relativity simulations of merging BHs that is also restricted to quasicircular orbits and models the effect of high-order modes and spin-precession. This model exhibits the smallest mismatch against numerical relativity waveforms, sometimes comparable to the numerical error in the simulations. It is thus expected to generally yield the smallest errors due to waveform systematics [362]. This fact was exploited in Hannam *et al.* [179] to break the waveform systematics tie and argue that the source of GW200129 exhibited relativistic spin-precession with a primary component spin magnitude of $\chi_1 = 0.9^{+0.1}_{-0.5}$ at the 90% credible level.

During a binary inspiral, spin-precession is described through post-Newtonian theory [62, 210]. Spin components that are not aligned with the orbital angular momentum give rise to spin-orbit and spin-spin interactions that cause the orbit to change direction in space as the binary inspirals, e.g., [88, 89, 106, 107, 167, 178, 206, 288, 313, 314]. The emitted GW signal is modulated in amplitude and phase, and morphologically resembles the beating between two spin-aligned waveforms [159] or a spin-aligned waveform that has been "twisted-up" [313, 314]. As the binary reaches merger, numerical simulations suggest that the direction of peak emission continues precessing [260]. Parameter estimation analyses using `NRSur7dq4` find that spins and spin-precession can be measured from merger-dominated signals for certain spin configurations [78], however the lack of analytic understanding of the phenomenon means that it is not clear how such a measurement is achieved.

The main motivation for this study is to revisit GW200129 and attempt to understand how spins and spin-precession can be measured from a heavy BBH with a merger-dominated observed signal. In Sec. 6.3 we use `NRSur7dq4` to conclude that the evidence for spin-precession originates exclusively from the LIGO Livingston data in the 20–50 Hz frequency range, where the inferred signal amplitude is lower than what a spin-aligned binary would imply given the rest of the data. This range coincides with the known data quality issues described in App. 6.9 and first identified in GWTC-3 [38]. LIGO Hanford is consistent with a spin-aligned signal, causing an inconsistency between the inferred mass ratio q and precession parameter χ_p inferred from each LIGO detector separately. By means of simulated signals, we argue that such $q - \chi_\mathrm{p}$ inconsistency is unlikely to be caused solely by the different Gaussian noise realizations in each detector at the time of the signal, rather pointing to remaining data quality issues beyond the original glitch-subtraction [38]. We also re-analyze the LIGO Livingston data above 50 Hz, (while keeping the original frequency range of the LIGO Hanford data) and confirm that all evidence for spin-precession disappears.

In the process, we find that the Virgo trigger, though consistent with a spin-aligned BBH, is *inconsistent* with the signal seen in the LIGO Hanford and LIGO Livingston detectors. Specifically, the Virgo data are pointing to a much heavier BBH that merges ∼20 ms earlier than the one observed by the LIGO detectors. We discuss Virgo data quality considerations in Sec. 6.4 within the context of a potential glitch that affects the inferred binary parameters if unmitigated. As a consequence, we do not include Virgo data in the sections examining spin-precession unless otherwise stated. The Virgo-LIGO inconsistency can be resolved if we use BayesWave [121, 124, 226] to simultaneously model a CBC signal and glitches with CBC waveform models and sine-Gaussian wavelets respectively [110, 192]. The Virgo data are now consistent with the presence of both a signal that is consistent with the one in the LIGO detectors and an overlaping glitch with SNR ~ 4.6.

In Sec. 6.5 we revisit the LIGO Livingston data quality issues and compare the original glitch-subtraction based on gwsubtract [137, 138] that uses information from auxiliary channels and the glitch estimate from BayesWave that uses only strain data. Though the CBC model used in BayesWave does not include the effect of spin-precession, we show that differences between the reconstructed waveforms from a non-precessing and spin-precessing analysis for GW200129 are *smaller* than the statistical uncertainty in the glitch inference. Such differences can therefore not be reliably resolved in the presence of the glitch and its subtraction procedure. The two glitch estimation methods give similar results within their statistical errors, however gwsubtract yields typically a lower glitch amplitude. We conclude that any evidence for spin-precession from GW200129 is contingent upon the systematic and statistical uncertainties of the LIGO Livingston glitch subtraction. Given the low SNR of the LIGO Livingston glitch and the glitch modeling uncertainties, we can at present not conclude whether the source of GW200129 exhibited spin-precession or not.

In Sec. 6.6 we summarize our arguments that remaining data quality issues in LIGO Livingston cast doubt on the evidence for spin-precession. Besides data quality studies (i.e., spectrograms, glitch modeling, auxiliary channels), our investigations are based on comparisons between different detectors as well as different frequency bands of the same detector. We propose that similar investigations in further events of interest with exceptional inferred properties could help alleviate potential contamination due to data quality issues.

6.3 The origin of the evidence for spin-precession

Our main goal is to pinpoint the parts of the GW200129 data that are inconsistent with a non-precessing binary and understand the relevant signal morphology. Due to different orientations, sensitivities, and noise realizations, different detectors in the network do not observe an identical signal. The detector orientation, especially, affects the signal polarization content and thus the degree to which spin-precession might be measurable in each detector. Motivated by this, we begin by examining data using different detector combinations.

We perform parameter estimation using the `NRSur7dq4` waveform and examine data from each detector separately (left panel) as well as the relation between the LIGO and the Virgo data (right panel) and show posteriors for select intrinsic parameters in Fig. 6.1. Analysis settings and details are provided in App. 6.9 and in all cases we use the same LIGO Livingston data as GWTC-3 [38] where the glitch has been subtracted. Though we do not expect the posterior distributions for the various signal parameters inferred with different detector combinations to be identical, they should have broadly overlapping regions of support. If the triggers recorded by the different detectors are indeed consistent, any shift between the posteriors should be at the level of Gaussian noise fluctuations.

The left panel shows that the evidence for spin-precession arises primarily from the LIGO Livingston data, whereas the precession parameter χ_{p} poste-

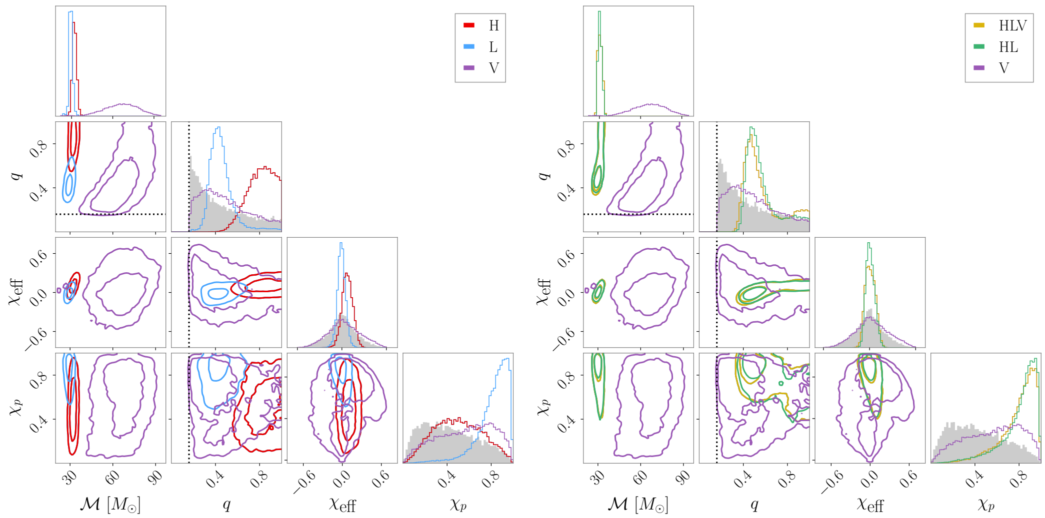

Figure 6.1: One- and two-dimensional marginalized posteriors for select intrinsic binary parameters: detector frame chirp-mass $\mathcal{M}$, mass ratio q, effective spin $\chi_{\rm eff}$, and precessing spin $\chi_{\rm p}$. See Table 6.1 for analysis settings and App. 6.9 for detailed parameter definitions. Two-dimensional panels show 50% and 90% contours. The black dashed line marks the minimum bound of q=1/6 in NRSur7dq4's region of validity. Shaded regions shows the prior for q, $\chi_{\rm eff}$, $\chi_{\rm p}$. The $\mathcal{M}$ prior increases monotonically to the maximum allowed value (see App. 6.9 for details on choices of priors). Left panel: comparison between analyses that use solely LIGO Hanford (red; H), LIGO Livingston (blue; L), and Virgo (purple; V) data. Right panel: comparison between analyses of all three detectors (yellow; HLV), only LIGO data (green; HL) and only Virgo data (purple; V). The evidence for spin-precession originates solely from the LIGO Livingston data as the other detectors give uninformative $\chi_{\rm p}$ posteriors. Additionally, the binary masses inferred based on Virgo only are inconsistent with those from the LIGO data.

rior is much closer to its prior when only LIGO Hanford or Virgo data are considered. A similar conclusion was reached in Hannam *et al.* [179]. There is reasonable overlap between the two-dimensional distributions that involve the chirp mass $\mathcal{M}$, the mass ratio q, and the effective spin $\chi_{\rm eff}$ inferred by the two LIGO detectors, as expected from detectors that observe the same signal under different Gaussian noise realizations. The discrepancy between the spin-precession inference in the two LIGO detectors, however, is evident in the $q-\chi_{\rm p}$ panel. The two detectors lead to non overlapping distributions that point to either unequal masses and spin-precession (LIGO Livingston), or equal masses and no information for spin-precession (LIGO Hanford).

Besides an uninformative posterior on $\chi_{\rm p}$, the left panel points to a bigger issue

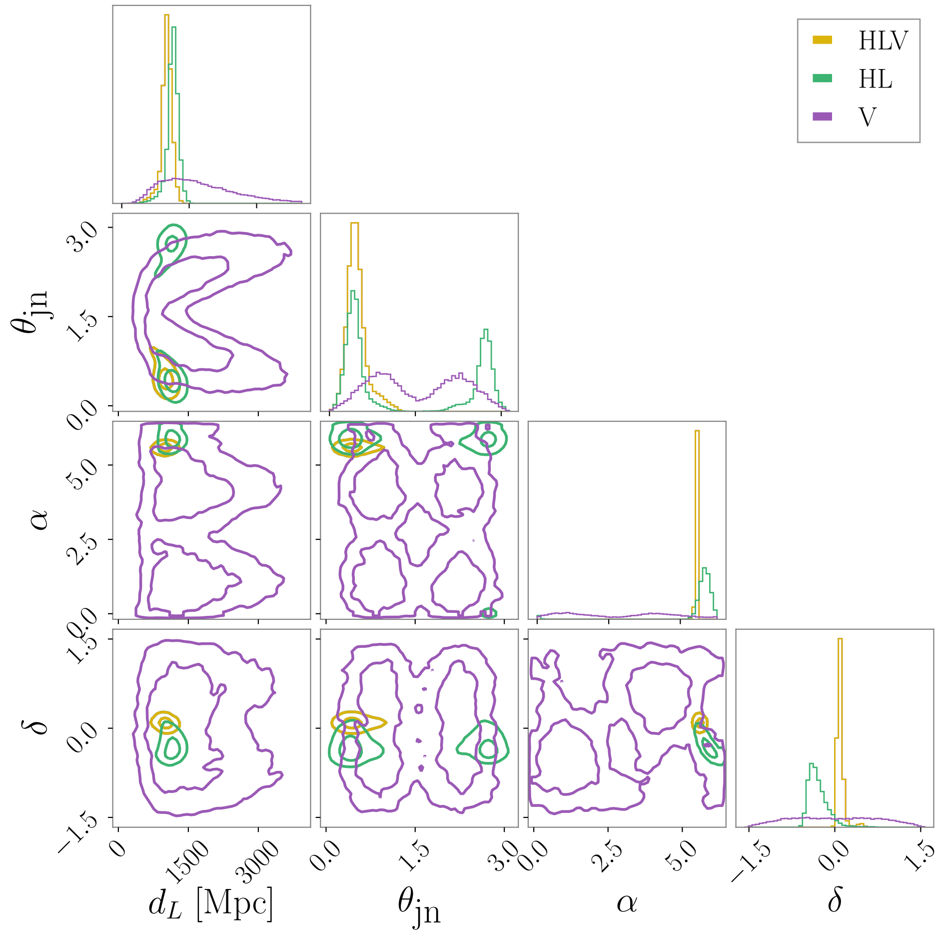

Figure 6.2: Similar to the right panel of Fig. 6.1 but for select extrinsic parameters: luminosity distance d_L, angle between total angular momentum and line of sight $\theta_{\rm jn}$, right ascension α, and declination δ. For reference, the median optimal SNR for each run is HLV: 27.6, HL: 26.9, V: 6.7.

with the Virgo data: inconsistent inferred masses. The right panel examines the role of Virgo in more detail in comparison to the LIGO data. Due to the lower SNR in Virgo, the intrinsic parameter posteriors are essentially identical between the HL and the HLV analyses. The lower total SNR means that the Virgo-only posteriors will be wider, but they are still expected to overlap with the ones inferred from the two LIGO detectors. However, this is not the case for the mass parameters as is most evident from the two dimensional panels involving the chirp mass. While the LIGO data are consistent with a typical binary with (detector-frame) chirp mass $30.3^{+2.5}_{-1.6}\,M_\odot$ at the 90% credible level, the Virgo data point to a much heavier binary with $66.7^{+19.7}_{-22.6}\,M_\odot$ at the same credible level.

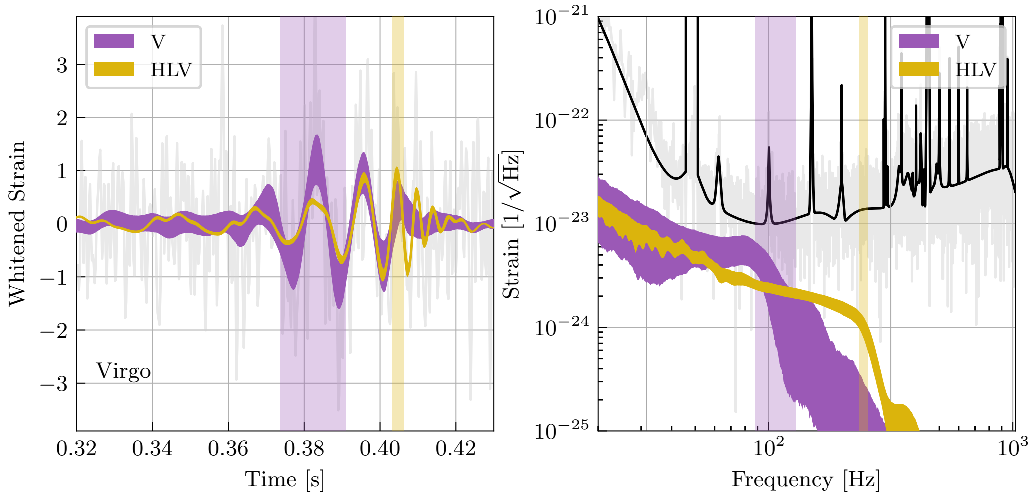

Figure 6.3: 90% credible intervals for the whitened time-domain reconstruction (left) and spectrum (right) of the signal in Virgo from a Virgo-only (purple; V) and a full 3-detector (yellow; HLV) analysis, see Table 6.1 for analysis settings. The data are shown in gray and the noise PSD in black. The time on the left plot is relative to GPS 1264316116. The high value of the PSD at $\sim 50\,\mathrm{Hz}$ was imposed due to miscalibration of the relevant data [38]. Vertical shaded regions at each panel correspond to the 90% credible intervals of the merger time (left; defined as the time of peak strain amplitude) and merger frequency (right; approximated via the dominant ringdown mode frequency as computed with `qnm` [325], merger remnant properties were computed with `surfinBH` [363]). The Virgo data point to a heavier binary that merges ~ 20ms earlier than the full 3-detector results that are dominated by the LIGO detectors.

The role of Virgo data on the inferred binary extrinsic parameters is explored in Fig. 6.2. In general, Virgo data have a larger influence on the extrinsic than the intrinsic parameters as the measured time and amplitude helps break existing degeneracies. The extrinsic parameter posteriors show a large degree of overlap. The Virgo distance posterior does not rail against the upper prior cut off, suggesting that this detector does observe some excess power. The HL sky localization also overlaps with the Virgo-only one, though the latter is merely the antenna pattern of the detector that excludes the four Virgo "blind spots". We use the HL results to calculate the projected waveform in Virgo and calculate the 90% lower limit on the signal SNR to be 4.2. This suggests that given the LIGO data, Virgo should be observing a signal with at least that SNR at the 90% level.

In order to track down the cause of the discrepancy in the inferred mass parameters, we examine the Virgo strain data directly. Figure 6.3 shows the whitened time-domain reconstruction (left panel) and the spectrum (right panel) of the signal in Virgo from a Virgo-only and a full 3-detector analysis. Compared to Figs. 6.1 and 6.2, here we only consider a 3-detector analysis as the reconstructed signal in Virgo inferred from solely LIGO data would not be phase-coherent with the data, and thus would be uninformative. Given the higher signal SNR in the two LIGO detectors, the signal reconstruction morphology in Virgo is driven by them, as evident from the intrinsic parameter posteriors from the right panel of Fig. 6.1.

The two reconstructions in Fig. 6.3 are morphologically distinct. The 3-detector inferred signal is dominated by the LIGO data and resembles a typical "chirp" with increasing amplitude and frequency. This signal is, however, inconsistent with the Virgo data as it underpredicts the strain at $t\sim 0.382\,\mathrm{s}$ in the left panel. The Virgo-only inferred signal matches the data better by instead placing the merger at earlier times to capture the increased strain at $t\sim 0.382\,\mathrm{s}$ as shown by the shaded vertical region denoting the merger time. Rather than a "chirp," the signal is dominated by the subsequent ringdown phase with an amplitude that decreases slowly over $\sim$2 cycles. As also concluded from the inferred masses in Fig. 6.1, the Virgo data point to a heavier binary with lower ringdown frequency (vertical regions in the right panel).

Despite these large inconsistencies, the issues with the Virgo data do not affect our main goal, which is identifying the origin of the evidence for spin-precession. In order to avoid further ambiguities for the remainder of this section we restrict to data from the two LIGO detectors unless otherwise noted. In Fig. 6.1 we concluded that LIGO Livingston alone drives this measurement and here we attempt to further zero in on the data that support spin-precession by comparing results from a spin-precessing and a spin-aligned analysis with `NRSur7dq4` (see App. 6.9 for details). Figure 6.4 shows the whitened time-domain reconstruction (left panel) and the spectrum (right panel) in LIGO Hanford (top) and LIGO Livingston (bottom). The two reconstructions remain phase-coherent, however there are some differences in the inferred amplitudes, with the spin-aligned amplitude being slightly larger at $\sim$30–50 Hz and slightly smaller for $\gtrsim 100\,\mathrm{Hz}$. Comparison to the estimate for the glitch that was subtracted from the data based on information from auxiliary channels with gwsubtract shows that the glitch overlaps with the part of the signal

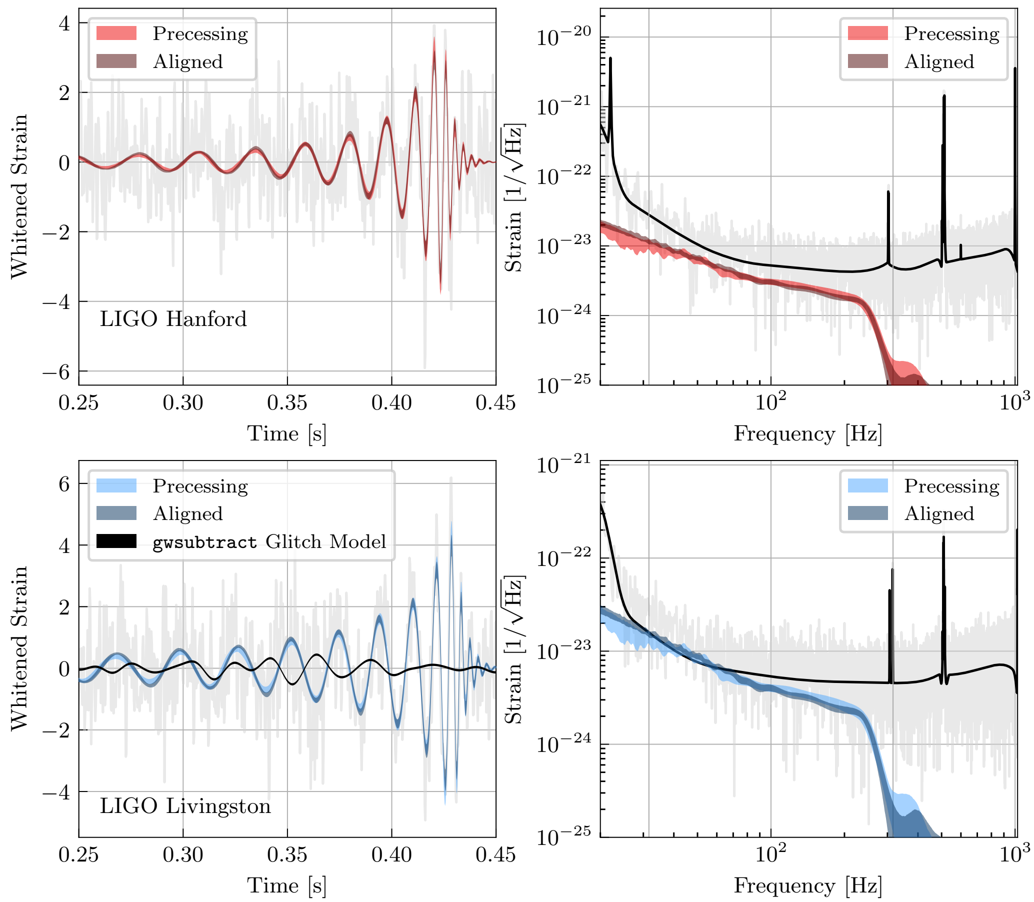

Figure 6.4: Whitened time-domain reconstruction (left) and spectrum (right) of GW200129 in LIGO Hanford (top) and LIGO Livingston (bottom). Shaded regions show the 90% credible intervals for the signal using a spin-precessing (light blue and red) and a spin-aligned (dark blue and red) analysis based on NRSur7dq4; see Table 6.1 for run settings. In gray we show the analyzed data where the gwsubtract estimate for the glitch (black line) has already been subtracted. The black line in the right panels is the noise PSD. The glitch overlaps with the part of the inferred signal where the spin-aligned amplitude is on average larger than the spin-precessing one.

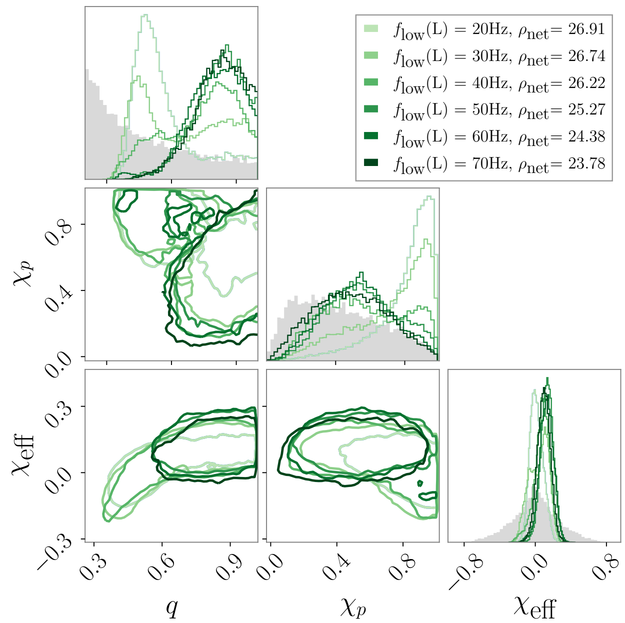

Figure 6.5: One- and two-dimensional marginalized posterior for the mass ratio q, the precession parameter $\chi_{\rm p}$, and the effective spin parameter $\chi_{\rm eff}$ for analyses using a progressively increasing low frequency cutoff in LIGO Livingston but all the LIGO Hanford data, see Table 6.1 for details. The median network SNR for each value of the frequency cutoff is given in the legend. Contours represent 90% credible regions and the prior is shaded in gray. As the glitch-affected data are removed from the analysis, the posterior approaches that of an equal-mass binary and becomes uninformative about $\chi_{\rm p}$. This behavior does not immediately indicate data quality issues and we only use this increasing-$f_{\rm low}(L)$ analysis to isolate the data which contribute the evidence of spin-precession when compared to the rest of the data to within 20–50 Hz.

where the spin-precessing amplitude is smaller than the spin-aligned one. The glitch subtraction and data quality issues are therefore related to the evidence for spin-precession.

We confirm that the low-frequency data in LIGO Livingston (in relation to the rest of the data) are the sole source of the evidence for spin-precession, by carrying out analyses with a progressively increasing low frequency cutoff in LIGO Livingston only, while leaving the LIGO Hanford data intact. Figure 6.5 shows the effect on the posterior for $\chi_{\rm p}$, q, and $\chi_{\rm eff}$. When we use the full data bandwidth, $f_{\rm low}(L) = 20\,{\rm Hz}$, we find that q and $\chi_{\rm p}$ are correlated and their two-dimensional posterior appears similar to the combination of the individual-detector posteriors from Fig. 6.1. However, as the low frequency cutoff in LIGO Livingston is increased and the data affected by the glitch are removed, the posterior progressively becomes more consistent with an equal-mass binary and χ_p approaches its prior. By $f_\mathrm{low}(L) = 50\,\mathrm{Hz}$, χ_p is similar to its prior and further increasing $f_\mathrm{low}(L)$ has a marginal effect. This confirms that *given all the other data*, the LIGO Livingston data in 20–50 Hz drive the inference for spin-precession.

The signal network SNR (i.e., the SNR in both detectors added in quadrature) is given in the legend for each value of the low frequency cutoff. By $f_\mathrm{low}(L) = 50\,\mathrm{Hz}$ where all evidence for spin-precession has been eliminated, the SNR reduction is only 1.5 units, suggesting that the large majority of the signal is consistent with a non-precessing origin. This might also suggest that χ_p inference is not degraded solely due to loss of SNR, as the latter is very small. The χ_eff posterior is generally only minimally affected, with a small shift to higher values driven by the $q - \chi_\mathrm{eff}$ correlation [128]. We have verified that these conclusions are robust against re-including the Virgo data (using their full bandwidth).

The above analysis is *not* on its own an indication of data quality issues in LIGO Livingston, but we now turn to an observation that might be more problematic: the $q - \chi_\mathrm{p}$ inconsistency between LIGO Hanford and LIGO Livingston identified in Fig. 6.1. In order to examine whether such an effect could arise from the different Gaussian noise realizations in each detector, we consider simulated signals. We use 100 posterior samples obtained from analyzing solely the LIGO Livingston data, make simulated data that include a noise realization with the same noise PSDs as GW200129, and analyze data from the two LIGO detectors separately. To quantify the degree to which the LIGO Hanford and LIGO Livingston posteriors overlap, we compute the Bayes factor for overlapping posterior distributions relative to if the two distributions do not overlap [180, 181],

$$\mathcal{B}^\mathrm{overlapping}_\mathrm{not\ overlapping} = \iint \mathrm{d}\chi_\mathrm{p}\mathrm{d}q\, \frac{p_L(\chi_\mathrm{p}, q|d)p_H(\chi_\mathrm{p}, q|d)}{\pi(\chi_\mathrm{p}, q)}, \tag{6.1}$$

where we compute the overlap within the $q-\chi_\mathrm{p}$ plane, $p_L(\chi_\mathrm{p}, q|d)$ and $p_H(\chi_\mathrm{p}, q|d)$ are the LIGO Livingston and LIGO Hanford posteriors, and $\pi(\chi_\mathrm{p}, q)$ is the prior. While evaluating this quantity is subject to sizeable sampling uncertainty for events where the two distributions are more distinct (i.e., the case of

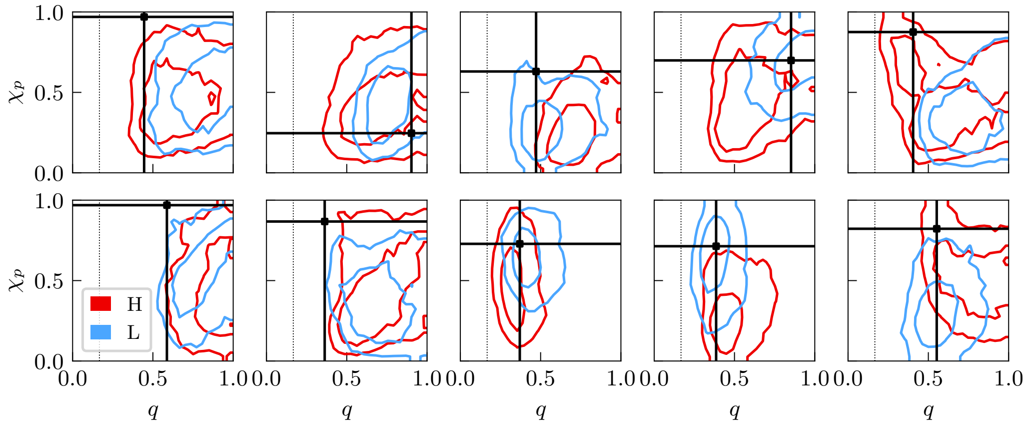

Figure 6.6: 90% contours for the two-dimensional marginalized posteriors for the mass ratio q and the precessing parameter χ_p obtained from analyzing data from each LIGO detector separately for 10 simulated signals. The signal parameters are drawn from the posterior for GW200129 when using LIGO Livingston data only and true values are indicated by black lines. Due to the spin priors disfavoring large χ_p, the injected value is outside the two-dimensional 90% contour in some cases. We only encounter an inconsistency between LIGO Hanford (red; H) and LIGO Livingston (blue; L) as observed for GW200129 in Fig. 6.1 in $\mathcal{O}(5/100)$ injections.

GW200129), we find $\mathcal{O}(5/100)$ injections have a similar overlap as GW200129 (Fig. 6.1). Figure 6.6 shows a selection of $q - \chi_\mathrm{p}$ posteriors for 10 injections as inferred from each detector separately. The posteriors typically overlap, though they are shifted with respect to each other as expected from the different noise realizations.

We conclude that the evidence for spin-precession originates exclusively from the LIGO Livingston data that overlapped with a glitch. This causes an inconsistency between the LIGO Hanford and LIGO Livingston that we typically do not encounter in simulated signals in pure Gaussian noise. This inconsistency suggests that there might be residual data quality issues in LIGO Livingston that were not fully resolved by the original glitch subtraction. Though inconsequential for the spin-precession investigation, we also identify severe data quality issues in Virgo. Before returning to the investigation of spin-precession, we first examine the Virgo data in detail in Sec. 6.4 and argue that they should be removed from subsequent analyses. We reprise the spin-precession investigations in Sec. 6.5.

6.4 Data quality issues: Virgo

Having established that the Virgo trigger is coincident but not fully coherent with the triggers in the two LIGO detectors, we explore potential reasons for this discrepancy. Figure 6.7 shows a spectrogram of the data in each detector centered around the time of the event. A clear chirp morphology is visible in the LIGO detectors but not in Virgo, though this might also be due to the low SNR of the Virgo trigger. Within a few seconds of the trigger, however, a number of other glitches are also present in Virgo, mostly assigned to scattered light. We estimate the SNR of the Virgo trigger without assuming it is a CBC signal (i.e., without using a CBC model) through Omicron [293] and BayesWave using its glitch model that fits the data with sine-Gaussian wavelets; see Table 6.2 for run settings[1]. The former finds a matched-filter Omicron SNR[2] of 7.0, while the latter finds an optimal SNR of 7.3 for the median glitch reconstruction.

Given the prevalence of glitches, the first option is that the Virgo trigger is actually a detector glitch that happened to coincide with a signal in the LIGO detectors. To estimate the probability that such a coincidence could happen by chance, we consider the glitch rate in Virgo. In O3, the median rate of glitches in Virgo was 1.11/min, with significant variation versus time [38]. When we consider the hour of data around the event, the rate of glitches with Omicron SNR > 6.5 is 10.2/min. Most of the glitches in Virgo at this time are due to scattered light [41, 44, 229, 230, 322]. While Fig. 6.7 shows that there are scattered light glitches in the Virgo data near the time of GW200129, the excess power from these glitches are concentrated at frequencies $< 30\,\mathrm{Hz}$. To account for the excess power corresponding to GW200129 in Virgo, there must be a different type of glitch present in the data. The rate of glitches at frequencies similar to the signal is much lower; using data from 4 days around the event, the rate of glitches with frequency 60-120 Hz is only 0.06/hr. Given this rate, we calculate the probability that a glitch occurred in Virgo within a 0.06 s window (roughly corresponding to twice the light-travel time between the LIGO detectors and Virgo) around a trigger in the LIGO detectors. We find that if glitches at any frequency are considered, the probability of coin-

[1]=The BayesWave analyses described here does not concurrently marginalize over the PSD uncertainty.

[2]The SNR reported by Omicron is normalized so that the expectation value of the SNR is 0, rather than $\sqrt{2}$ [293]. To highlight this difference, we use the phrase "Omicron SNR" whenever a reported result uses this normalization.

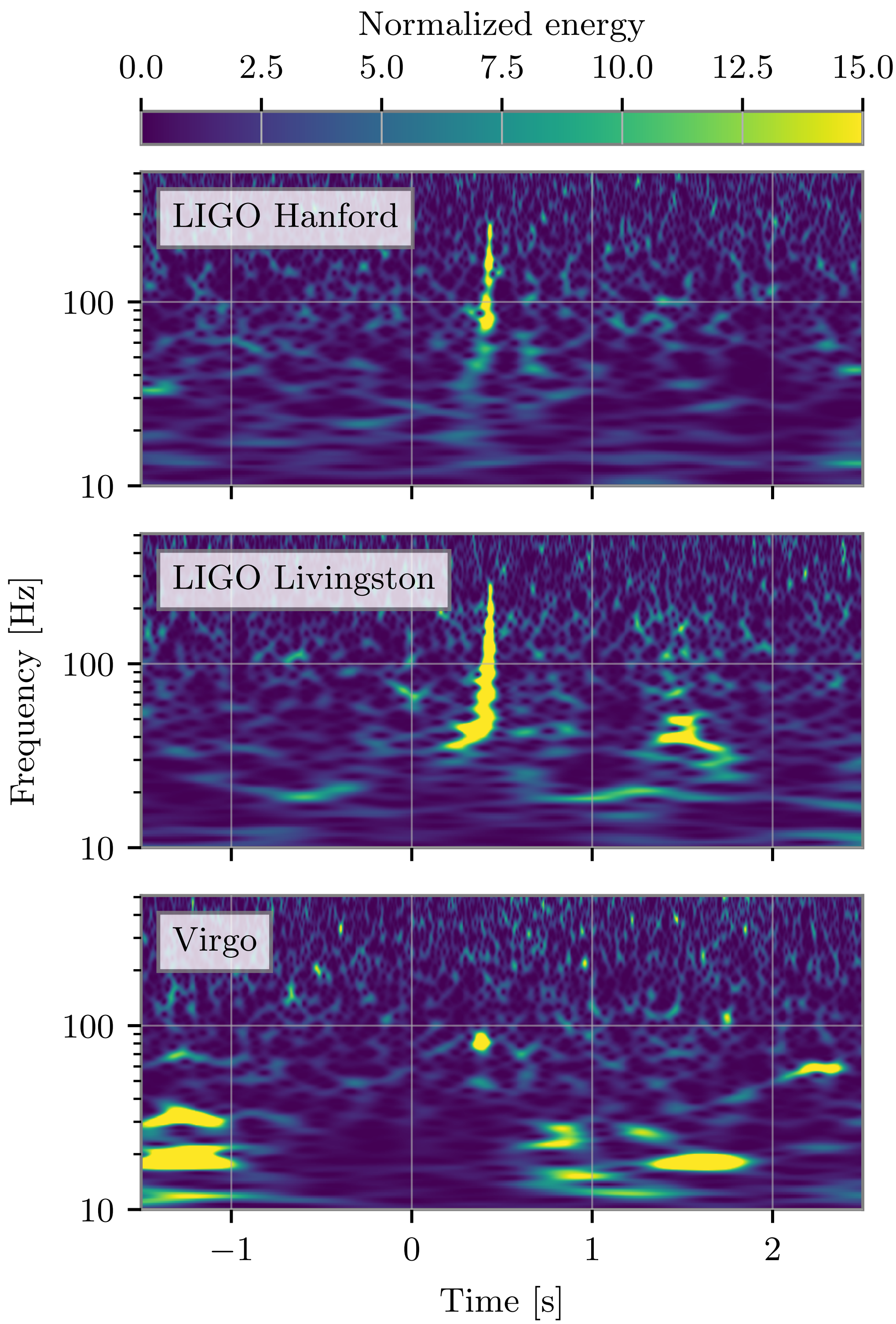

Figure 6.7: Spectrogram of the data in each detector, plotted using the Q-transform [102, 237]. Listed times are with respect to GPS 1264316116. Besides the clear chirp morphology in LIGO, there is visible excess power $\sim 1\,\mathrm{s}$ after the signal in LIGO Livingston. Virgo demonstrates a high amount of excess power, though most is due to scattered light and concentrated at frequencies $< 30\,\mathrm{Hz}$. The excess power in Virgo that is coincident with GW200129 does not have a chirp morphology.

cidence per event is $\mathcal{O}(0.01)$, and if only glitches with similar frequencies are considered, the same probability is $\mathcal{O}(10^{-5})$.

Another option is that the Virgo trigger is a combination of a genuine signal and a detector glitch. We explore this possibility using BayesWave [121, 124, 226] to simultaneously model a potential CBC signal that is coherent across the detector network and overlapping glitches that are incoherent [110, 192]. In this "CBC+glitch" analysis, BayesWave models the CBC signal with the IMRPhenomD waveform [198, 207] and glitches with sine-Gaussian wavelets. Details about the models and run settings are provided in App. 6.9. An important caveat here is that IMRPhenomD does not include the effects of higher-order modes and spin-precession. A concern is, therefore, that the CBC model could fail to model precession-induced modulations in the signal amplitude and instead assign them to the glitch model. This precise scenario is tested in Hourihane *et al.* [192] where the analysis was shown to be robust against such systematics. Below we argue that the same is true here for the Virgo data, especially since they are consistent with a spin-aligned binary as shown in Fig. 6.1.

Figure 6.8 compares BayesWave's reconstruction in Virgo with the one obtained with the NRSur7dq4 analysis from Fig. 6.3 that ignores a potential glitch but models spin-precession and higher order modes. All results are obtained using data from all three detectors. The CBC reconstruction from BayesWave with IMRPhenomD is consistent with the one from NRSur7dq4 to within the 90% credible level at all times. This is unsurprising given Fig. 6.1 that shows that Virgo data are consistent with a spin-aligned BBH. Crucially, there is no noticeable difference between the two CBC reconstructions for times when the inferred glitch is the loudest. This suggests that the lack of higher-order modes and spin-precession in IMRPhenomD does not lead to a noticeable difference in the signal reconstruction and could thus not account for the inferred glitch. The differences between the inferred signals using IMRPhenomD and NRSur7dq4 are much smaller than the amount of incoherent power present in Virgo. In fact, the glitch reconstruction is larger than the signal at the 50% credible level, though not at the 90% level. This result suggests that a potential explanation for the trigger in Virgo is a combination of a signal consistent with the one in the LIGO detectors and a glitch.

Figure 6.9 summarizes the various SNR estimates for the excess power in

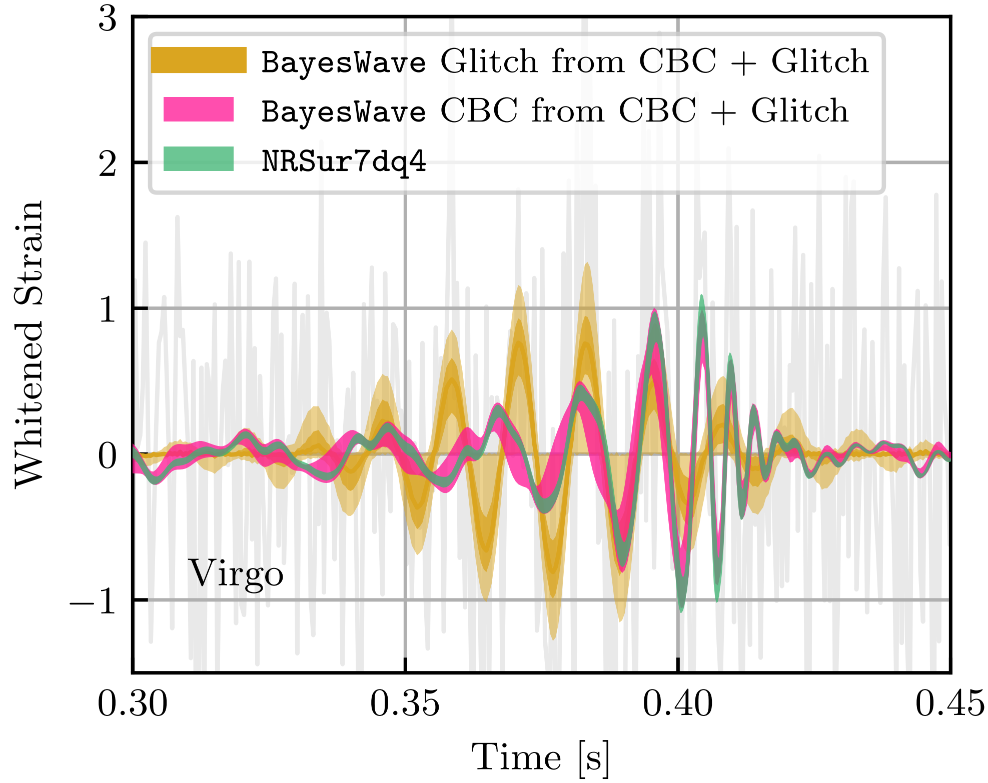

Figure 6.8: Whitened time-domain reconstruction of the signal in Virgo obtained after analysis of data from all three detectors relative to GPS 1264316116. Shaded regions correspond to 90% and 50% (where applicable) credible intervals. Green corresponds to the same 3-detector result obtained with NRSur7dq4 as Fig. 6.3, while pink and gold correspond to the CBC and glitch part of the "CBC+glitch" analysis with BayesWave. See Tables 6.1 and 6.2 for run settings. The two CBC reconstructions largely overlap, suggesting that the lack of spin-precession in BayesWave's analysis does not affect the reconstruction considerably. A glitch overlapping with the signal is, however, recovered.

Virgo. We plot an estimate of the SNR in Virgo suggested by LIGO data; in other words it is the SNR that is consistent with GW200129 as observed by LIGO. In comparison, we also show the SNR from a Virgo-only analysis and the SNR from BayesWave's "glitchOnly" analysis that models the excess power with sine-Gaussian wavelets without the requirement that it is consistent with a CBC. The fact that the SNR inferred from HL data is smaller than the other two again suggests that the Virgo trigger is not consistent with the one seen by LIGO and contains additional power. BayesWave's "CBC+glitch" analysis is able to separate the part of the trigger that is consistent with a CBC and recovers a CBC SNR that is consistent to the one inferred from LIGO only. The "remaining" power is assigned to a glitch with SNR ~ 4.6 (computed through the median BayesWave glitch reconstruction).

Based on the glitch SNR calculated by the BayesWave "CBC+glitch" model, we revisit the probability of overlap with a signal based on the SNR distribution of Omicron triggers. Since the lowest SNR recorded in Omicron analyses is 5.0, we fit the SNR distribution of glitches with Omicron SNR > 5.0 with a power-law and extrapolate to SNR 4.6. We find that the rate of glitches with frequencies similar to the one in Fig. 6.8 with SNR > 4.6 is 0.31/min and the probability of overlap with a signal in Virgo is $\mathcal{O}(10^{-3})$. Given the 60 events from GWTC-3 that were identified in Virgo during O3, the overall chance of at least one glitch of this SNR overlapping a signal is $\mathcal{O}(0.1)$.

The above studies suggest that the most likely scenario is that the Virgo trigger consists of a signal and a glitch. However, due to the low SNR of both, this interpretation is subject to sizeable statistical uncertainties and we therefore do not attempt to make glitch-subtracted Virgo data. Such data would be extremely dependent on which glitch reconstruction we chose to subtract, for example the median or a fair draw from the BayesWave glitch posterior. For these reasons and due to its low sensitivity, we do not include Virgo data in what follows.

6.5 Data quality issues: LIGO Livingston

The data quality issues in LIGO Livingston were identified and mitigated in GWTC-3 [38] through use of information from auxiliary channels [137, 138] and the gwsubtract pipeline as also described in App. 6.9. The comparison of Figs. 6.1 and 6.6, however, suggest that residual data quality issues might

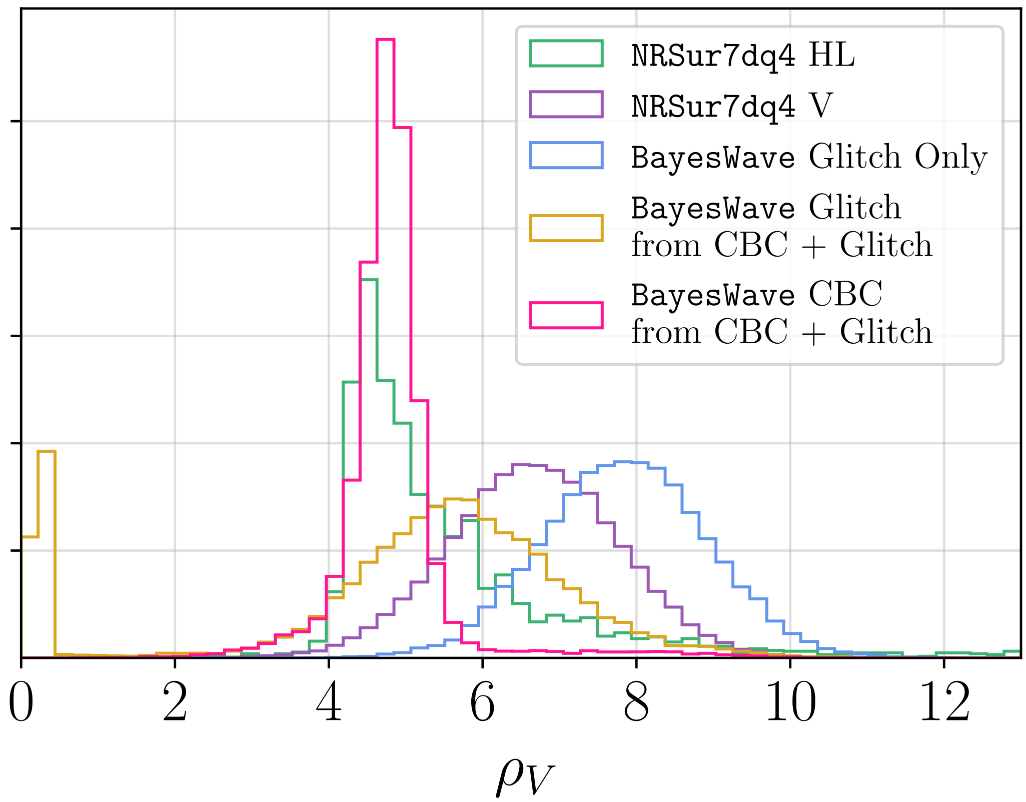

Figure 6.9: Comparison of optimal SNR estimates for Virgo from different analyses. In green is the posterior for the expected SNR in Virgo from just the LIGO data using the `NRSur7dq4` waveform (HL analysis of Fig. 6.1), while purple corresponds to the SNR from an analysis of the Virgo data only (V analysis of Fig. 6.1). The CBC and glitch SNR posterior from BayesWave's full "CBC+glitch" model (Fig. 6.8) are shown in pink and orange respectively. Part of the latter is consistent with zero, which corresponds to no glitch (as also seen from the 90% credible interval in Fig. 6.8). The SNR posterior from a "glitchOnly" BayesWave is shown in blue.

remain, as the two LIGO detectors result in inconsistent inferred $q-\chi_{\rm p}$ parameters beyond what is expected from typical Gaussian noise fluctuations. Here we revisit the LIGO Livingston glitch with `BayesWave` and again model both the CBC and potential glitches. This analysis offers a point of comparison to `gwsubtract` as it uses solely strain data to infer the glitch instead of auxiliary channels. Additionally, `BayesWave` computes a posterior for the glitch, rather than a single point estimate, and thus allows us to explore the statistical uncertainty of the glitch mitigation. In all analyses involving `BayesWave` we use the original LIGO Livingston data without any of the data mitigation described in App. 6.9.

Figure 6.10 shows `BayesWave`'s CBC and glitch reconstructions in LIGO Livingston compared to the one based on the `NRSur7dq4` (from glitch-mitigated data) and the glitch model computed with `gwsubstract`. All analyses use data from the two LIGO detectors only. Unsurprisingly, now, the CBC reconstructions based on `IMRPhenomD` and `NRSur7dq4` do not fully overlap around t=0.3 s, though they are consistent during the signal merger phase. This is expected from the fact that LIGO Livingston supports spin-precession as well as Fig. 6.4. However, this difference is *smaller* than the statistical uncertainty in the inferred glitch from `BayesWave` (yellow) and well as differences between the `BayesWave` and the `gwsubtract` glitch estimates. This suggests that even though the `BayesWave` glitch estimate might be affected by the lack of spin-precession in its CBC model, this effect is smaller than the glitch uncertainty.

We also model the signal as a superposition of coherent wavelets in addition to the incoherent glitch wavelets using `BayesWave` [121, 124, 226]. This approach has been previously utilized for glitch subtraction [38]. However, we do not recover strong evidence for a glitch overlapping the signal in LIGO Livingston when running with this "signal+glitch" analysis. The "signal+glitch" analysis attempts to describe both the signal and the glitch with wavelets and hence it is significantly less sensitive than the "CBC+glitch" model. In the data of interest, both the signal and the glitch whitened amplitudes are $\sim 1\sigma$ and as such they are difficult to separate using coherent and incoherent wavelets. Given that we know (based on the auxiliary channel data) that there is some non-Gaussian noise in LIGO Livingston, we find that the "signal+glitch" analysis is not sensitive enough for our data.

The large statistical uncertainty in the glitch reconstruction (yellow bands

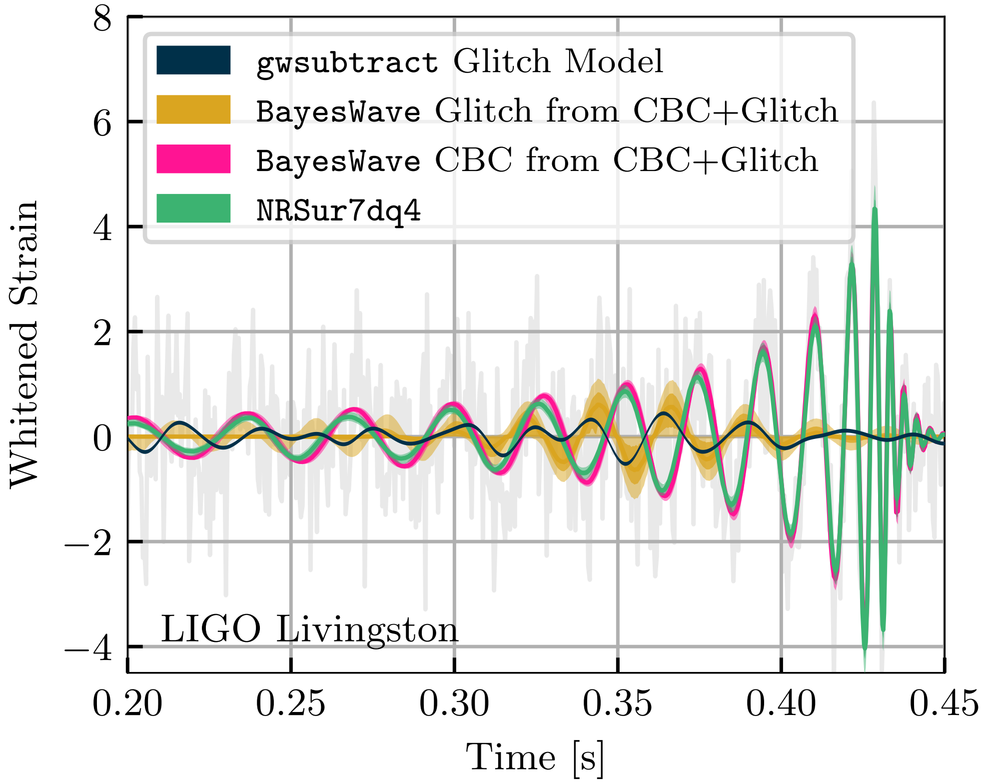

Figure 6.10: Whitened time-domain reconstruction of the data in LIGO Livingston obtained after analysis of data from the two LIGO detectors. Shaded regions correspond to 90% and 50% (where applicable) credible intervals and gray gives the original data without any glitch mitigation. Green corresponds to the same 2-detector result obtained with NRSur7dq4 as Fig. 6.4, while pink and gold correspond to the CBC and glitch part of the joint "CBC+glitch" analysis with BayesWave. The black line shows an estimate for the glitch obtained through auxiliary channels. All analyses use only LIGO data.

in Fig. 6.10) implies that the difference between the spin-precessing and non-precessing interpretation of GW200129 cannot be reliably resolved. To confirm this, we select three random samples from the glitch posterior of Fig. 6.10, subtract them from the unmitigated LIGO Livingston data, and repeat the parameter estimation analysis with NRSur7dq4. The BayesWave glitch-subtracted frames and associated NRSur7dq4 parameter estimation results are available in [269]. For reference, we also analyze the original unmitigated data (no glitch subtraction whatsoever). Figure 6.11 confirms that the spin-precession evidence depends sensitively on the glitch subtraction. The original unmitigated data and the gwsubtract subtraction yield the largest evidence for spin-precession, but this is reduced -or completely eliminated- with different realizations of the BayesWave glitch model. In general, larger glitch amplitudes lead to less support for spin-precession, suggesting that the evidence for spin-precession is increased when the glitch is *undersubtracted*.

Figure 6.12 compares the corresponding $q-\chi_{\rm p}$ posterior inferred from LIGO Hanford and LIGO Livingston separately under each different estimate for the glitch. Each of the three BayesWave glitch draws results in single-detector posteriors that fully overlap, thus resolving the inconsistency seen in $q-\chi_{\rm p}$ when using the gwsubtract glitch estimate. Due to the lack of spin-precession modeling in the "CBC+glitch" analysis of Fig. 6.10, however, we cannot definitively conclude that any one of the new glitch-subtracted results is preferable. The 3 BayeWave glitch draws result in different levels of support for spin-precession, it is therefore possible that GW200129 is still consistent with a spin-precessing system. We do conclude, though, that the evidence for spin-precession is contingent upon the large statistical uncertainty of the glitch subtraction.

As a further check of whether the lack of spin-precession in BayesWave's CBC model could severely bias a potential glitch recovery, we revisit the 10 simulated signals from Fig. 6.6 and analyze them with the "CBC+glitch" model. These signals are consistent with GW200129 as inferred from LIGO Livingston data only, and thus exhibit the largest amount of spin-precession consistent with the signal. In all cases we find that the glitch part of the "CBC+glitch" model has median and 50% credible intervals that are consistent with zero at all times. This again confirms that the differences between the spin-precessing and the spin-aligned inferred signals in Fig. 6.10 is smaller than the uncertainty in the glitch. This tests suggests that the glitch model is not strongly

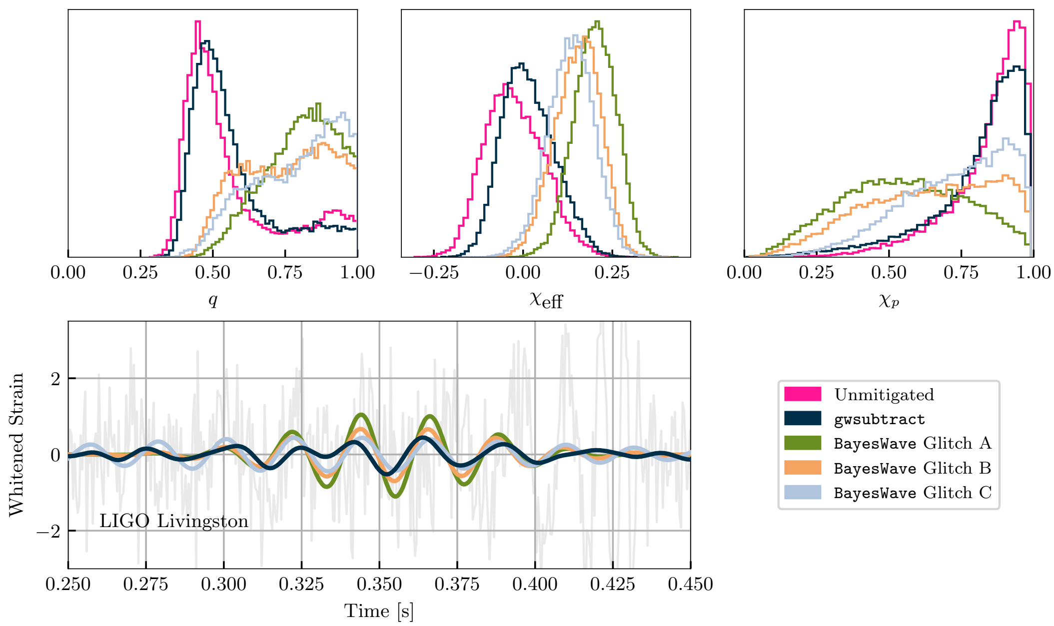

Figure 6.11: Bottom: Whitened, time domain reconstructions of various glitch reconstructions subtracted from LIGO Livingston data. The green line corresponds to the glitch reconstruction obtained from auxiliary data using `gwsubtract`. The rest are glitch posterior draws from the `BayesWave` "CBC+Glitch" analysis on HL unmitigated data. Top: Marginalized posterior distributions corresponding to parameter estimation performed with the `NRSur7dq4` waveform model on HL data where each respective glitch realization was subtracted from LIGO Livingston (same colors). Pink corresponds to the original data without any glitch subtraction. Larger glitch reconstruction amplitudes roughly lead to less informative $\chi_{\rm p}$ posteriors and eliminate the $q - \chi_{\rm p}$ inconsistency between LIGO Hanford and LIGO Livingston.

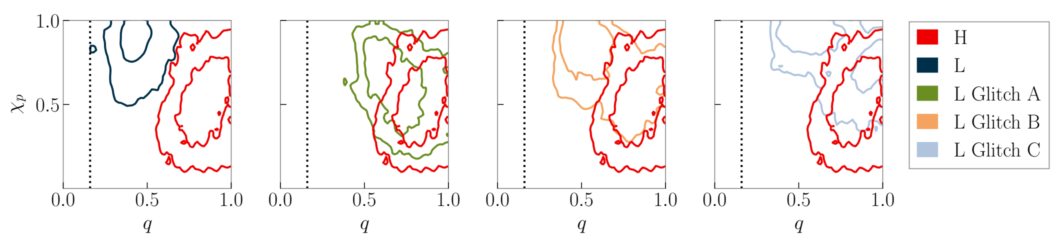

Figure 6.12: Two-dimensional posterior distributions for $\chi_{\rm p}$ and q (50% and 90% contours) from single-detector parameter estimation runs. The far left panel shows the same tension as the LIGO Hanford and LIGO Livingston data plotted in Fig. 6.1 when using the `gwsubtract` estimate for the glitch. Subsequent figures show inferred posterior distributions using data where the same three different `BayesWave` glitch models as Fig. 6.11 have been subtracted. These results show less tension between the two posterior distributions.

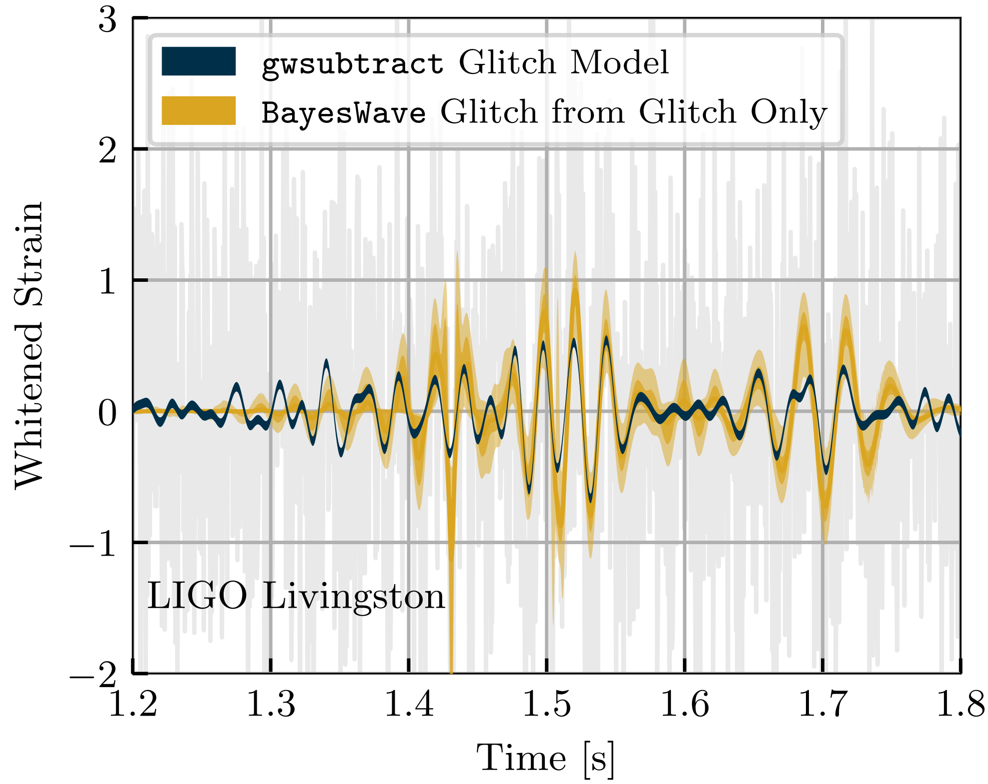

Figure 6.13: Comparison between the two glitch reconstruction and subtraction methods for a glitch in LIGO Livingston $\sim 1\,$s after GW200129, see the middle panel of Fig. 6.7. We plot the original data with no glitch mitigation (grey), the glitch reconstruction obtained from auxiliary channels with 90% confidence intervals (black), and the 50% and 90% credible intervals for the glitch obtained with BayesWave that uses only the strain data (gold).

biased by the lack of spin-precession, however it does not preclude small biases (within the glitch statistical uncertainty); it is therefore necessary but not sufficient.

As a final point of comparison between BayesWave's glitch reconstruction that is based on strain data and the gwsubtract glitch reconstruction based on auxiliary channels, we consider a *different* glitch in LIGO Livingston approximately 1s after the signal, see Fig. 6.7. Studying this glitch offers the advantage of direct comparison of the two glitch reconstruction methods without contamination from the CBC signal and uncertainties about its modeling. We analyze the original data with no previous glitch mitigation around that glitch using BayesWave's glitch model and plot the results in Fig. 6.13. For the gwsubtract reconstruction we also include 90% confidence intervals, as described in App. 6.9.

The two estimates of the glitch are broadly similar but they do not always overlap within their uncertainties. The main disagreement comes from the sharp data "spike" at $t = 1.43\,\mathrm{s}$ that is missed by gwsubtract, but recovered by BayesWave. The reason is that the the maximum frequency considered by gwsubtract was 128 Hz and thus cannot capture such a sharp noise feature [138]. Away from the "spike," the two glitch estimates are approximately phase-coherent. On average BayesWave recovers a larger glitch amplitude as the gwsubtract result typically falls on BayesWave's lower 90% credible level.

Figures 6.10 and 6.13 broadly suggest that BayesWave recovers a higher-amplitude glitch. Figure 6.11 shows that the evidence for spin-precession is indeed reduced, the LIGO Hanford-LIGO Livingston inconsistency is alleviated (Fig. 6.12), and the LIGO Livingston data become more consistent across low and high frequencies (Fig. 6.5) if the glitch was originally undersubtracted. However, due to the low SNR of the glitch and other systematic uncertainties it is not straightforward to select a "preferred" set of glitch-subtracted data. All studies, however, indicate that the statistical uncertainty of the glitch amplitude is larger than the difference between the inferred spin-precessing and spin-aligned signals.

6.6 Conclusions

Though it might be possible to infer the presence of spin-precession and large spins in heavy BBHs, our investigations suggest that in the case of GW200129 any such evidence is contaminated by data quality issues in the LIGO Livingston detector. In agreement with [179] we find that the evidence for spin-precession originates exclusively from data from that detector. However, we go beyond this and also demonstrate the following.

1. The evidence for spin-precession in LIGO Livingston is localized in the 20–50 Hz band in comparison to the rest of the data, precisely where the glitch overlapped the signal. Excluding this frequency range from the analysis, we find that GW200129 is consistent with an equal-mass BBH with an uninformative χ_{p} posterior; it is thus similar to the majority of BBH detections [12, 29, 37]. However, the fact that there is no evidence for spin-precession if $f_{\mathrm{low}}(L) > 50\,\mathrm{Hz}$ is not on its own cause for

concern as it might be due to Gaussian noise fluctuations or the precise precessional dynamics of the system.

2. LIGO Hanford is not only uninformative about spin-precession (which again could be due to Gaussian noise fluctuations or the lower signal SNR in that detector), but it also yields an *inconsistent* $q - \chi_{\rm p}$ posterior compared to LIGO Livingston. Using simulated signals, we find that the latter, i.e., the $q - \chi_{\rm p}$ inconsistency, is larger than $\mathcal{O}(95\%)$ of results expected from Gaussian noise fluctuations.

3. Given the LIGO Livingston glitch's low SNR, the statistical uncertainty in modeling it is *larger* than the difference between a spin-precessing and a non-precessing analysis for GW200129. Inferring the presence of spin-precession requires reliably resolving this difference, something challenging as we found by using different realizations of the glitch model from the BayesWave glitch posterior. Crucially, any evidence for spin-precession in GW200129 depends sensitively on the glitch model and priors employed.

4. Given the large statistical uncertainty in modeling the glitch, evidence for systematic differences between BayesWave and gwsubtract that use strain and auxiliary data respectively is tentative. However, the BayesWave estimate typically predicts a larger glitch amplitude, which would reduce the evidence for spin-precession and alleviate the tension between LIGO Hanford and LIGO Livingston. Additionally, we do not recover any support for a glitch when injecting spin-precessing signals from the LIGO Livingston-only posterior distribution into Gaussian noise. This indicates that BayesWave is unlikely to be strongly biasing the glitch recovery due to its lack of spin-precession.

Overall, given the uncertainty surrounding the LIGO Livingston glitch mitigation, we cannot conclude that the source of GW200129 was spin-precessing. We do not conclude the opposite either, however. Though we obtain tentative evidence that the glitch was undersubtracted, we can at present not estimate how much it was undersubtracted by due to large statistical and potential systematic uncertainties. It is possible that some evidence for spin-precession remains, albeit reduced given the glitch statistical uncertainty.

In addition, we verify that this uncertainty in the glitch modeling is larger than uncertainty induced by detector calibration. We repeat select analyses in Appendix 6.9 and confirm that the inclusion of uncertainty in the calibration of the gravitational-wave detectors negligibly impacts the spin-precession inference, as expected. Indeed, the glitch impacts the data at a level comparable to the signal strain, c.f., Fig. 6.10, whereas the calibration uncertainty within 20 to 70 Hz is only $\sim 5\%$ in amplitude and 5° in phase [328]. Therefore, the glitch in LIGO Livingston's data dominates over uncertainties about the data calibration.

Though not critical to the discussion and evidence for spin-precession, we also identified data quality issues in Virgo. The inconsistency between Virgo and the LIGO detectors is in fact more severe than the one between the two LIGO detectors, however the Virgo data do not influence the overall signal interpretation due to the low signal SNR in Virgo. Nonetheless, we argue that the most likely explanation is that the Virgo data contain both the GW200129 signal and a glitch.

These conclusions are obtained with `NRSur7dq4`, which is expected to be the more reliable waveform model including spin-precession and higher-order modes in this region of the parameter space [179, 362]. We repeated select analyses with `IMRPhenomXPHM` which also favored a spin-precessing interpretation for GW200129 [38]. We found largely consistent but not identical results between `NRSur7dq4` and `IMRPhenomXPHM`, suggesting that there are additional systematic differences between the two waveform models. Appendix 6.10 shows some example results. Nonetheless, our results are directly comparable to the ones of [179, 364] as they were obtained with the same waveform model.

Our analysis suggests that extra caution is needed when attempting to infer the role of subdominant physical effects in the detected GW signals, for example spin-precession or eccentricity. Low-mass signals are dominated by a long inspiral phase that in principle allows for the detection of multiple spin-precession cycles or eccentricity-induced modulations. However, the majority of detected events, such as GW200129, have high masses and are dominated by the merger phase. The subtlety of the effect of interest and the lack of analytical understanding might make inference susceptible not only to waveform systematics, but also (as argued in this study) potential small data quality issues.

Indeed, Fig. 6.11 shows that a difference in the glitch amplitude of $< 0.5\sigma$ can make the difference between an uninformative $\chi_{\rm p}$ posterior and one that strongly favors spin-precession. This also demonstrates that low-SNR glitches are capable of biasing inference of these subtle physical effects. Low-SNR departures from Gaussian noise have been commonly observed by statistical tests of the residual power present in the strain data after subtracting the best-fit waveform of events [23, 30, 32]. If indeed such low-SNR glitches are prevalent, they might be individually indistinguishable from Gaussian noise fluctuations. Potential ways to safeguard our analyses and conclusions against them are (i) the detector and frequency band consistency checks performed here, (ii) extending the BayesWave "CBC+glitch" analysis to account for spin-precession and eccentricity while carefully accounting for the impact of glitch modeling and priors especially for low SNR glitches, and (iii) and modeling insight on the morphology of subtle physical effects of interest such as spin-precession and eccentricity in relation to common detector glitch types.

6.7 Acknowledgements

We thank Aaron Zimmerman, Eric Thrane, Paul Lasky, Hui Tong, Geraint Pratten, Mark Hannam, Charlie Hoy, Jonathan Thompson, Steven Fairhurst, Vivien Raymond, Max Isi, and Colm Talbot for useful discussions. We also thank Vijay Varma for providing a version of LALSuite optimized for running NRSur7dq4, as well as suggestions for some of our configuration settings. This research has made use of data, software and/or web tools obtained from the Gravitational Wave Open Science Center (https://www.gw-openscience.org), a service of LIGO Laboratory, the LIGO Scientific Collaboration and the Virgo Collaboration. Virgo is funded by the French Centre National de Recherche Scientifique (CNRS), the Italian Istituto Nazionale della Fisica Nucleare (INFN) and the Dutch Nikhef, with contributions by Polish and Hungarian institutes. This material is based upon work supported by NSF's LIGO Laboratory which is a major facility fully funded by the National Science Foundation. The authors are grateful for computational resources provided by the LIGO Laboratory and supported by NSF Grants PHY-0757058 and PHY-0823459. This research utilized the OzStar Supercomputing Facility at Swinburne Unersity of Technology. The OzStar facility is partially funded by the Astronomy National Collaborative Research Infrastructure Strategy (NCRIS) allocation provided by the Australian Government. This research was enabled in part by computing resources provided by Simon Fraser University and the Digital Research Alliance of Canada (alliancecan.ca). SH and KC were supported by NSF Grant PHY-2110111. Software: gwpy [237], matplotlib [197], numpy [182], pandas [263], scipy [370], qnm [325], surfinBH [363], Bilby [223], LALSuite [242], BayesWave [222].

6.8 Appendix

6.9 Analysis details

In this appendix we provide details and settings for the analyses presented in the main text. All data are obtained via the GW Open Science Center [39]. Throughout we use geometric units, $G = c = 1$.

Detection and Glitch-subtracted data

GW200129 was identified in low latency [224] by GstLAL [177, 245], cWB [212], PyCBC Live [129, 257], MBTAOnline [46], and SPIIR [114]. The quoted false alarm rate of the signal in low latency was approximately 1 in 10^{23} years, making this an unambiguous detection. Below we recap the detection and glitch mitigation process from [38].

Multiple data quality issues were identified in the data surrounding GW200129. As a part of the rapid response procedures, scattered light noise [41, 44] was identified in the Virgo data, as seen in Fig. 6.7 in the frequency range 10–60 Hz. These glitches did not overlap the signal, and no mitigation steps were taken with the Virgo data. During offline investigations of the LIGO Livingston data quality, a malfunction of the 45 MHz electro-optic modulator system [5] was found to have caused numerous glitches in the days surrounding GW200129. To help search pipelines differentiate these types from glitches, a data quality flag was generated for this noise source [139]. These data quality vetoes are used by some pipelines to veto any candidates identified during the data quality flag time segments [142]. The glitches from the electro-optic modulator system directly overlapped GW200129, meaning that the time of the signal overlapped the time of the data quality flag.

Although clearly an astrophysical signal, the data quality issues present in LIGO Livingston introduced additional complexities into the estimation of the significance of this signal [38]. Due to the data quality veto, the signal was not identified in LIGO Livingston by the PyCBC [136, 256] MBTA [71], and cWB [212] pipelines. PyCBC was still able to identify GW200129 as a LIGO Hanford – Virgo detection, but the signal was not identified by MBTA due to the high SNR in LIGO Hanford and cWB due to post-production cuts. The GstLAL [98, 307] analysis did not incorporate data quality vetoes in its O3 analyses and was therefore able to identify the signal in all three detectors.

The excess power from the glitch directly overlapping GW200129 in LIGO Livingston was subtracted before estimation of the signal's source properties [38, 138] using the `gwsubtract` algorithm [137]. This method relies on an auxiliary sensor at LIGO Livingston that also witnesses glitches present in the strain data. The transfer function between the sensor and the strain data channel is measured using a long stretch of data by calculating the inner product of the two time series with a high frequency resolution and then averaging the measured value at nearby frequencies to produce a transfer function with lower frequency resolution [57]. This transfer function is convolved with the auxiliary channel time series to estimate the contribution of this particular noise source to the strain data. Therefore, the effectiveness of this subtraction method is limited by the accuracy of the auxiliary sensor and the transfer function estimate. This tool was previously used for broadband noise subtraction with the O2 LIGO dataset [137], but this was the first time it was used for targeted glitch subtraction. Additional details about the use of `gwsubtract` for the GW200129 glitch subtraction can be found in Davis *et al.* [138].

The `gwsubtract` glitch model does not include a corresponding interval that accounts for all sources of statistical errors as is done by `BayesWave`. However, a confidence interval based on only uncertainties due to random correlations between the auxiliary channel and the strain data can be computed. For the GW200129 glitch model, this interval is ± 0.022 in the whitened strain data [138]. Additional systematic uncertainties due to time variation in the measured transfer function and effectiveness of the chosen auxiliary channel are expected to be present but are not quantified. The relative size of these uncertainties is dependent on the specific noise source that is being modeled and chosen auxiliary channel.

`Bilby` parameter estimation analyses

Quasicircular BBHs are characterized by 15 parameters, divided into 8 intrinsic and 7 extrinsic parameters. Each component BH has source frame mass m_i^s,

Figure(s)	Waveform Model	Detector Network	Glitch mitigation	$f_{\rm low}$ (Hz)
6.1, 6.12	NRSur7dq4	H	gwsubtract	20
6.1, 6.12	NRSur7dq4	L	gwsubtract	20
6.1, 6.2, 6.3	NRSur7dq4	V	gwsubtract	20
6.1, 6.2, 6.3, 6.8	NRSur7dq4	HLV	gwsubtract	20
6.1, 6.2, 6.4, 6.10, 6.11, 6.14	NRSur7dq4	HL	gwsubtract	20
6.4	NRSur7dq4 spin-aligned	HL	gwsubtract	20
6.5	NRSur7dq4	HL	gwsubtract	{20,30,40,50,60,70} in L, 20 in H
6.11	NRSur7dq4	HL	No mitigation	20
6.11	NRSur7dq4	HL	BayesWave fair draws	20
6.12	NRSur7dq4	L	BayesWave fair draws	20
6.14	IMRPhenomXPHM	HL	gwsubtract	20

Table 6.1: Table of Bilby runs and settings. All analyses use 4 s of data, and a sampling rate of 4096 Hz. Columns correspond to the main text figures each analysis appears in, the waveform model, the detector network used (H: LIGO Hanford, L: LIGO Livingston, V: Virgo), the type of glitch mitigation in LIGO Livingston, and the low frequency cutoff of the analysis. Figure 6.6 also presents results for a set of 10 injections drawn from the LIGO Livingston only posterior distribution with $f_{\rm low}(L) = 20\,\mathrm{Hz}$. These analyses use the same settings as above with $f_{\rm low}(L) = 20\,\mathrm{Hz}$.

$i \in \{1, 2\}$. In the main text we mainly use the corresponding detector frame (redshifted) masses $m_i = (1+z)m_i^s$, where z is the redshift, as we are interested in investigating data quality issues and detector frame quantities better relate to the signal as observed. Each component BH also has dimensionless spin vector $\vec{\chi}_i$, and χ_i is the magnitude of this vector. We also use parameter combinations that are useful in various contexts: total mass $M = m_1 + m_2$, mass ratio $q = m_2/m_1 < 1$, chirp mass $\mathcal{M} = (m_1 m_2)^{3/5}(m_1 + m_2)^{-1/5}$ [82, 161, 273], effective orbit-aligned spin parameter [53, 287, 309]

$$\chi_{\rm eff} = \frac{\vec{\chi}_1 \cdot \vec{L} + q\vec{\chi}_2 \cdot \vec{L}}{1+q}\,, \tag{6.2}$$

where $\vec{L}$ is the Newtonian orbital angular momentum, and effective precession spin parameter [178, 315]

$$\chi_{\rm p} = \max\left(\chi_{1\perp}, q\chi_{2\perp}\frac{3q+4}{4q+3}\right)\,, \tag{6.3}$$

where $\chi_{1\perp}$ is the $\vec{\chi}_i$ component that is perpendicular to $\vec{L}$. The remaining parameters are observer dependent, and hence referred to as extrinsic. The right ascension α and declination δ designate the location of the source in the sky, while the luminosity distance to the source is d_L. The angle between total angular momentum and the observer's line of sight is θ_{jn}; for systems without perpendicular spins it reduces to the inclination ι, the angle between the orbital angular momentum and observer's line of sight. The time of coalescence t_c is the geocenter coalescence time of the binary. The phase of the signal ϕ is defined at a given reference frequency, and the polarization angle ψ completes the geometric description of the sources position and orientation relative to us; neither of these are used directly in this work.

Parameter estimation results are obtained with parallel Bilby [69, 299, 320] using the nested sampler, Dynesty [324]. The numerical relativity surrogate, NRSur7dq4 [362], is used for all main results due to its accuracy over the regime of highly precessing signals. Its space of validity is limited by the availability of numerical simulations [85] to $q > 1/4$ and component spin magnitudes $\chi < 0.8$, though it maintains reasonable accuracy when extrapolated to $q > 1/6$ and $\chi < 1$ [362].

The majority of our analyses use the publicly released strain data, including the aforementioned glitch subtraction in LIGO Livingston [138], and noise power spectral densities (PSDs) [38]. The exception to the publicly released data was the construction of glitch-subtracted strain data using BayesWave for LIGO Livingston, as discussed in Sec. 6.5. We do not incorporate the impact of uncertainty about the detector calibration as the SNR of the signal is far below the anticipated regime where calibration uncertainty is non-negligible [157, 267, 372, 376]. Furthermore, we confirm that including marginalization of calibration uncertainty does not qualitatively change the recovered posterior distributions or our main conclusions by also directly repeating select runs.

As is done in GWTC-3 [38], we choose a prior that is uniform in detector frame component masses, while sampling in chirp mass and mass ratio. The mass ratio prior bounds are 1/6 and 1, where we utilize the extrapolation region of NRSur7dq4. Since NRSur7dq4 is trained against numerical relativity simulations which typically have a short duration, only a limited number of cycles are captured before coalescence. With a reduced signal model duration, our analysis is restricted to heavier systems so that the model has content spanning the frequencies analyzed (20 Hz and above). We therefore enforce an additional constraint on the total detector-frame mass to be greater than $60\,M_\odot$. We verify that our posteriors reside comfortably above this lower bound. The luminosity distance prior is chosen to be uniform in comoving volume. The prior distribution on the sky location is isotropic with a uniform distribution

Figure(s)	Models	Detector Network
6.8, 6.9	CBC+glitch	HLV
6.10, 6.11	CBC+glitch	HL
6.9	glitch	V
6.13	glitch	L

Table 6.2: Table of BayesWave runs and settings. All analyses use 4 s of data, a low frequency cut-off of $f_{\rm low} = 20$ Hz, a sampling rate of 2048 Hz, and the IMRPhenomD waveform when the CBC model is used. Furthermore, all analyses use the original strain data without the glitch mitigation described in Sec. 6.9. Columns correspond to the main text figures each analysis appears in, the BayesWave models that are used, and the detector network (H: LIGO Hanford, L: LIGO Livingston, V: Virgo). While not plotted in any figure, we also performed "CBC+Glitch" analyses on injections into the HL detector network as a glitch background study on GW200129-like sources, see Sec. 6.5.

on the polarization angle. Finally, for most analyses, the prior on the spin distributions is isotropic in orientation and uniform in spin magnitude up to $\chi = 0.99$. For the spin-aligned analyses, a prior is chosen on the aligned spin to mimic an isotropic and uniform spin magnitude prior. These settings and data are utilized in conjunction with differing GW detector network configurations and minimum frequencies in LIGO Livingston. The differences between runs and their corresponding figures are presented in Tab. 6.1.

BayesWave CBC and glitch analyses

BayesWave [121, 124, 226] is a flexible data analysis algorithm that models combinations of coherent generic signals, glitches, Gaussian noise, and most recently, CBC signals that appear in the data [110, 192, 379]. To sample from the multi-dimensional posterior for all the different models, BayesWave uses a "Gibbs sampler" which cycles between sampling different models while holding the parameters of the non-sampling model(s) fixed.

For this analysis, we mainly use the CBC and glitch models (a setting we refer to as "CBC+Glitch"). The CBC model parameters (see App. 6.9) are sampled via a fixed-dimension Markov Chain Monte Carlo sampler (MCMC) using the priors described in Wijngaarden *et al.* [379]. The glitch model is based on sine-Gaussian wavelets and samples over both the parameters of each wavelet (central time, central frequency, quality factor, amplitude, phase [121]) and the number of wavelets via a trans-dimensional or Reverse-jump MCMC. In some cases, we also make use of solely the glitch model (termed "glitchOnly" analyses) that assumes no CBC signal and the excess power is described only with wavelets. The differences between runs and the figures in which they appear are presented in Tab. 6.2.

Though BayesWave typically marginalizes over uncertainty in the noise PSD [226], in this work we use the same fixed PSD as the Bilby runs for more direct comparisons. Additionally, we use identical data as App. 6.9 for the LIGO Hanford and Virgo detectors. However, when it comes to LIGO Livingston we use the original (i.e., "unmitigated," without any glitch subtraction) data in order to independently infer the glitch. We do not marginalize over uncertainty in the detector calibration.

6.10 Select results with IMRPhenomXPHM

In this Appendix, we present select results obtained with the IMRPhenomXPHM [281] waveform model that also resulted in evidence for spin-precession in GWTC-3 [38]. Even though IMRPhenomXPHM and NRSur7dq4 both support spin-precesion, in contrast to SEOBNRv4PHM, there are still noticeable systematic differences between them. Figure 6.14 shows that while NRSur7dq4 and IMRPhenomXPHM generally have overlapping regions of posterior support, IMRPhenomXPHM shows slightly more preference for higher q and less support for extreme precession when compared to NRSur7dq4. Waveform systematics are expected to play a significant role in GW200129's inference (e.g. Refs. [38, 179, 194]), which motivates utilizing NRSur7dq4 for all of our main text results.

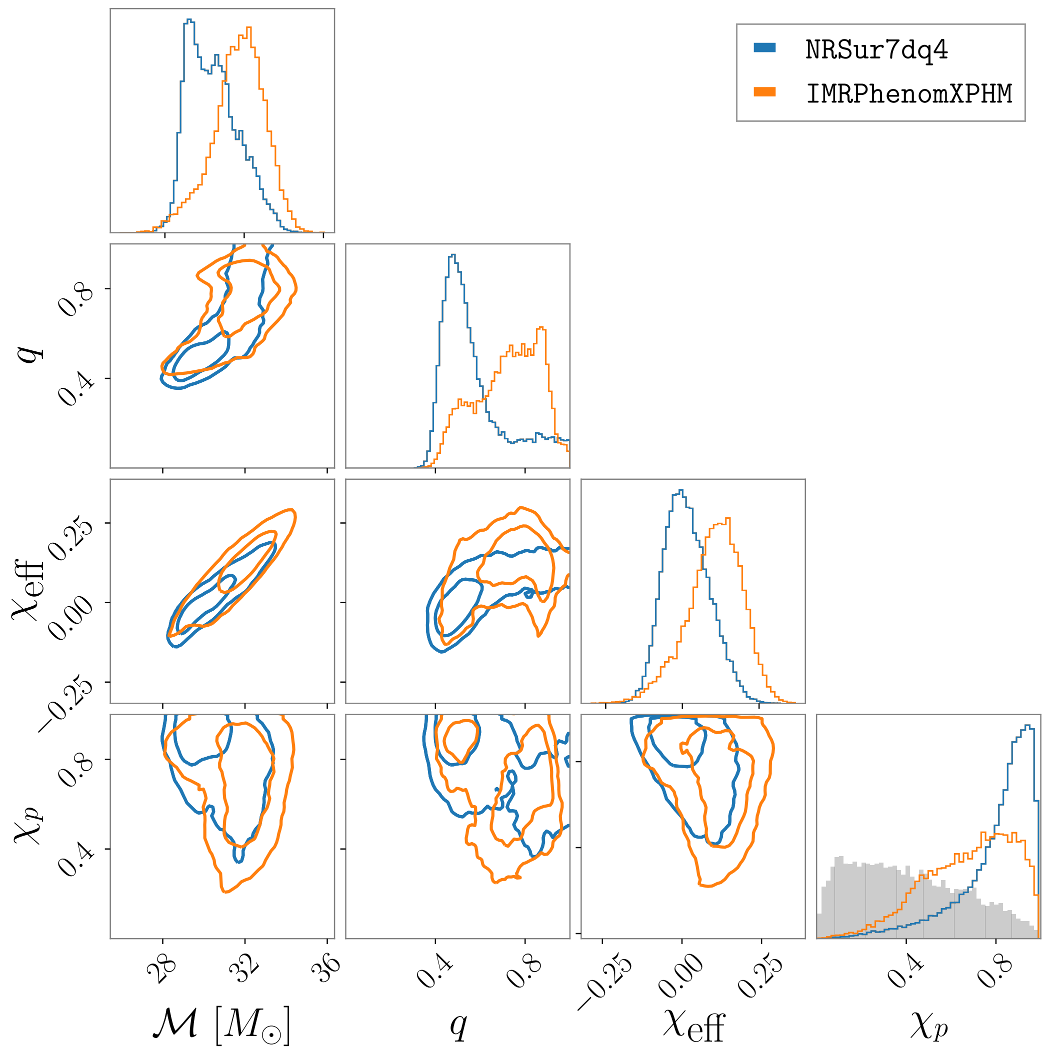

Figure 6.14: Similar to Fig. 6.1, using data from LIGO Livingston and LIGO Hanford. The comparison shows slight tension between results when using `NRSur7dq4` and `IMRPhenomXPHM`, though qualitatively `IMRPhenomXPHM` also seems to support the evidence for spin-precession.

Chapter 7

IMPACT OF GLITCH MITIGATION ON GW191109

This chapter contains work from

Rhiannon Udall, **Sophie Hourihane**, Simona Miller, Derek Davis, Katerina Chatziioannou, Max Isi, and Howard Deshong. Antialigned spin of GW191109: Glitch mitigation and its implications. *Phys. Rev. D*, 111(2): 024046, 2025. doi: 10.1103/PhysRevD.111.024046. Reprinted as Chapter 7. **SH** helped conceptualize the project, led all BayesWave analyses, created related figures, and authored related text.

7.1 Abstract

With a high total mass and an inferred effective spin anti-aligned with the orbital axis at the 99.9% level, GW191109 is one of the most promising candidates for a dynamical formation origin among gravitational wave events observed so far. However, the data containing GW191109 are afflicted with terrestrial noise transients, i.e., detector glitches, generated by the scattering of laser light in both LIGO detectors. We study the implications of the glitch(es) on the inferred properties and astrophysical interpretation of GW191109. Using time- and frequency-domain analysis methods, we isolate the critical data for spin inference to $35-40\,\mathrm{Hz}$ and $0.1-0.04\,\mathrm{s}$ before the merger in LIGO Livingston, directly coincident with the glitch. Using two models of glitch behavior, one tailored to slow scattered light and one more generic, we perform joint inference of the glitch and binary parameters. When the glitch is modeled as slow scattered light, the binary parameters favor anti-aligned spins, in agreement with existing interpretations. When more flexible glitch modeling based on sine-Gaussian wavelets is used instead, a bimodal aligned/anti-aligned solution emerges. The anti-aligned spin mode is correlated with a weaker inferred glitch and preferred by $\sim 70:30$ compared to the aligned spin mode and a stronger inferred glitch. We conclude that if we assume that the data are only impacted by slow scattering noise, then the anti-aligned spin inference is robust. However, the data alone cannot validate this assumption and resolve the anti-aligned spin and potentially dynamical formation history of GW191109.

7.2 Introduction

Reported in the third gravitational wave (GW) transient catalog (GWTC-3) [38], GW191109_010717 (more concisely GW191109) stands out among existing binary black hole (BBH) signals. With source-frame primary and secondary masses of $m_1 = 65^{+11}_{-11}\,M_\odot$ and $m_2 = 47^{+15}_{-13}\,M_\odot$ (90% symmetric credible intervals), it is among the most massive events. Furthermore, there is significant support for black hole (BH) spins anti-aligned with the orbital angular momentum: the mass-weighted effective spin [53, 287, 309] is $\chi_{\rm eff} = -0.29^{+0.42}_{-0.31}$. For these reasons, as well as support for unequal masses, $q = m_2/m_1 = 0.73^{+0.21}_{-0.24}$, spin-precession, and hints of eccentricity [175, 303], the binary is potentially of dynamical and/or hierarchical origin [59, 391] and impacts population inference [45, 347].

Multiple GW191109 properties hint toward a dynamical origin. High masses, above the pair-instability supernova (PISN) limit of $45 - 70\,M_\odot$ (depending on modeling assumptions) [160, 380], may require a hierarchical mechanism in order to form and merge. Asymmetric masses, in particular, might imply the merger of a second- and a first-generation BH [59]. Furthermore, population synthesis simulations of isolated formation scenarios typically find little support for spins anti-aligned with the orbital angular momentum, unless supernova kicks are exceptionally high [168, 205, 391]. Finally, eccentricity would also be challenging to explain except by dynamical processes [302, 303, 389], due to the rapid orbit circularization by GW emission [271].

Given their astrophysical implications, the inferred properties of GW191109 are worth scrutinizing. The first potential source of systematics is the waveform used to model the signal. GWTC-3 employed the IMRPhenomXPHM [281] and SEOBNRv4PHM approximants [261], with inference performed by Bilby [69, 299] and RIFT [218] respectively. Both models include the physical effects of higher-order modes and spin-precession, and headline results (as quoted above) are their average. However, GW191109 is flagged for systematic differences between approximants [38], especially for the binary inclination (edge-on versus face-on/off respectively) and the longer $\chi_{\rm eff} > 0$ tail with IMRPhenomXPHM. A third waveform, NRSur7dq4 [362], was employed in Ref. [203]. A direct surrogate of numerical relativity simulations, NRSur7dq4 is expected to be the most accurate available model for systems with high masses and spins [179, 203, 362]. These results bolster the evidence for dynamical origin, with a more negative spin, $\chi_{\rm eff} = -0.38^{+0.21}_{-0.20}$, asymmetric masses, $q = 0.65^{+0.20}_{-0.19}$, and a precessing spin parameter [315] of $\chi_p = 0.59^{+0.26}_{-0.27}$. While waveform systematics remain relevant, the broad agreement between three waveforms (including a direct surrogate to numerical relativity) that $\chi_{\rm eff} \lesssim 0$ to varying credibility, suggests that subsequent interpretations of its formation history remain valid.

A second potential source of systematics concerns modeling the detector noise. Around GW191109's arrival, both LIGO [1] detectors experienced a terrestrial noise transient known as a scattered light glitch [38, 138, 351]. The Virgo detector [42] was offline at this time, and so only the LIGO detectors contributed to the observation. In LIGO Hanford (LHO), the glitch power was at a nadir while the event was in the detection band, making its impact on the inferred parameters negligible, see App. 7.8. As such, we ignore the LHO glitch going forward. By contrast, glitch power in the Livingston detector (LLO) was directly coincident in time and frequency with the signal, a circumstance which could bias astrophysical inference [4, 110, 192, 265, 268]. Specifically, glitch power extends up to $\sim 40\,\mathrm{Hz}$, coincident with the signal, see Fig. 7.1. Spin parameters might be particularly susceptible to such data quality issues due to the relatively smaller imprint they leave on signals compared to, e.g., the BH masses. For example, GW200129 shows evidence of spin-precession [38, 179], but its significance depends on how the glitch that overlapped that signal is modeled [233, 268].

The headline GWTC-3 results were obtained after an estimate for the glitch had been subtracted from the data. The two-step process involved first modeling the signal and the glitch with a flexible sum of coherent and incoherent wavelets respectively with BayesWave [121, 124, 226]. Second, a fair draw from the glitch posterior was subtracted and the system parameters were inferred as quoted above. This procedure has been shown to generally lead to unbiased mass and (aligned) spin inference [192, 265]. However, uncertainties remain related to BayesWave's glitch model and in the fair draw chosen to be subtracted. These effects were investigated in Ref. [138], albeit with a simpler waveform model with single-spin precession and no higher-order modes, IMRPhenomPv2 [178]. Glitch mitigation was found to affect the $\chi_{\rm eff}$ inference by a similar amount as waveform systematics. Completely removing the glitch-affected data, i.e. all LLO data below $40\,\mathrm{Hz}$, instead resulted in a dramatic shift of $\chi_{\rm eff}$ to positive values $\chi_{\rm eff} = 0.27^{+0.24}_{-0.48}$.

The stark impact of glitch-affected data on astrophysically-impactful spin inference motivates our study. In Sec. 7.4 we extend Ref. [138] to explore the manner in which the data inform the system parameters. Using NRSur7dq4 and a frequency-domain analysis, we find that the LLO data between 30 and 40 Hz are crucial for spin inference: excluding $30-40$ Hz data shifts the probability of $\chi_{\rm eff} < 0$ from 99.4% to 32.2%, effectively wiping out any preference for for anti-aligned spins. A similar time-domain analysis [248] highlights the role of the data $0.1-0.04$ s prior to merger. These data, which inform the $\chi_{\rm eff} < 0$ measurement, coincide in time and frequency with excess power in LLO, see Fig. 7.2 and in particular the excess power at ~ 36 Hz. To check whether such dramatic shifts in support for $\chi_{\rm eff} < 0$ are possible from Gaussian noise alone, we analyze 100 simulated signals consistent with GW191109. We find that shifts of this magnitude are unlikely but not impossible as 6% of the simulations experience a larger shift than GW191109.

In Sec. 7.5, we focus on the 36 Hz excess power and address the key question: is the excess power part of the signal (and hence $\chi_{\rm eff} < 0$) or is it part of the glitch (and hence inference has been affected by systematics)? Rather than the two-step process of glitch fitting and subtraction, we perform a full analysis where we *simultaneously* model both the signal and the glitch. Using a physically motivated model for scattered light glitches [351] we find $\chi_{\rm eff} < 0$ at the 99.9% level using NRSur7dq4. We attribute this to the fact that the 36 Hz power is more contained in time than expected for scattered light glitches that are characterized by extended arches in time-frequency. This analysis, therefore, attributes the 36 Hz power to the signal and thus prefers $\chi_{\rm eff} < 0$. It is, however, possible that not all terrestrial power is due to scattered light or that the physical model of Ref. [351] does not capture all scattered light power. Instead, using a more flexible model for the glitch based on wavelets and BayesWave and IMRPhenomXPHM we obtain a bimodal solution for the spin. One mode, preferred at the $70:30$ level, attributes most of the 36 Hz power to the signal and results in $\chi_{\rm eff} < 0$. The second mode attributes this power to the glitch and results in $\chi_{\rm eff} > 0$. Given the low signal-to-noise ratio (SNR) of the 36 Hz power, these results are impacted by the priors of the glitch model parameters at the few percent level.

In Sec. 7.6 we summarize our conclusions. Physically grounded assumptions about the behavior of scattered light glitches lend support to $\chi_{\mathrm{eff}} < 0$ for GW191109, and thus a dynamical origin. However, both systematic limitations on scattered light models and statistical uncertainty due to low SNR of the excess power and the impact of glitch priors prevent us from making that determination confidently. While the crucial 36 Hz power is not part of the scattered light glitch as modeled in Ref. [351], we cannot rule out glitch mismodeling or other types of terrestrial noise.

7.3 Modeling signals and glitches

The relevant data contain the GW191109 signal, glitch power, and Gaussian noise. In this section, we describe how we model the signal (Sec. 7.3), the glitch (Sec. 7.3), and methods for glitch mitigation (Sec. 7.3). We focus on the respective strengths and weaknesses of each approach and what unique information each supplies. All analyses model the Gaussian noise component with the power spectral densities (PSDs) from the GWTC-3 data release [3, 36]. Detailed settings and identification numbers for all analyses are given in Table 7.1 in App. 7.9.

Modeling the Compact Binary Signal

We use both time- and frequency-domain techniques to model the signal with either waveform approximants for compact binary signals or, more generically, with sine-Gaussian wavelets. All analyses consider data surrounding the nominal trigger time of GW191109, GPS time 1257296855.22, and employ a sampling rate of 1024 Hz, with the maximum analysis frequency set to 7/8 of the Nyquist frequency. Unless otherwise noted, analyses that model only the compact binary (and not the glitch) use a minimum frequency of 20 Hz in both detectors. We use standard compact-binary priors [299], notably uniform in detector-frame component masses and spin magnitude and orientation.

Frequency-domain inference

Frequency domain analyses with waveform approximants are based on Bilby [69, 299] with its implementation of the dynesty sampler [324] and BayesWave [110], both analyzing 4 s of data. The former models the signal with NRSur7dq4 [362] and the latter with IMRPhenomXPHM [281] (though for consistency we also perform checks with the former using IMRPhenomXPHM in App. 7.10). NRSur7dq4 supports a minimum mass ratio of 0.25 and minimum detector-frame chirp mass of $35\,M_\odot$; neither restriction affects the analysis. We extend into the extrapolation region in spins, setting a maximum spin magnitude of 0.99. For comparison, we also perform analyses with BayesWave where the signal is modeled as a flexible sum of coherent sine-Gaussian wavelets [121, 124]. Settings are similar to the glitch wavelet analysis described in Sec. 7.3, only here, the wavelets are coherently projected across the two detectors rather than being independent.

Time-domain inference

While GW inference is typically conducted in the frequency domain for computational efficiency, it can equivalently be conducted in the time domain [100, 199–201]. Frequency domain analyses are non-local in time; to avoid non-trivial likelihood modifications [99], time-domain inference is necessary in order to isolate purely temporal features of the data. Below, we truncate the GW191109 data at different times around the 36 Hz excess power, and independently conduct inference on the pre- or post-cutoff-time data. For this, we use the time-domain inference code employed in Ref. [248] to study the GW190521 properties and which was based on time-domain implementations targeting post-merger data [200, 201]. All time-domain results are based on regions of 1 s of data around GW191109's trigger time and employ NRSur7dq4 [362]. The same PSDs are used in the time domain analyses are the same as those in the frequency domain analyses, i.e. from the GWTC-3 data release [36].

Modeling the Glitch

Both LHO and LLO experienced slow scattering noise around the time of GW191109. We use two models for the glitch power: a physically motivated model tailored to slow scattering, implemented in Bilby, and a more flexible wavelet model, implemented in BayesWave.

Physically-parameterized scattering

As the name implies, scattered light glitches arise due to laser light that scatters off the main beam path, bounces off a surface, and recombines with the main beam [41, 142, 322, 346, 351]. During periods of significant ground motion when the scattering surface moves, this light acquires a phase offset, resulting in excess noise. Figure 7.1 shows a spectrogram of the LLO data, along with the frequency tracks of the scattering excess noise as predicted by

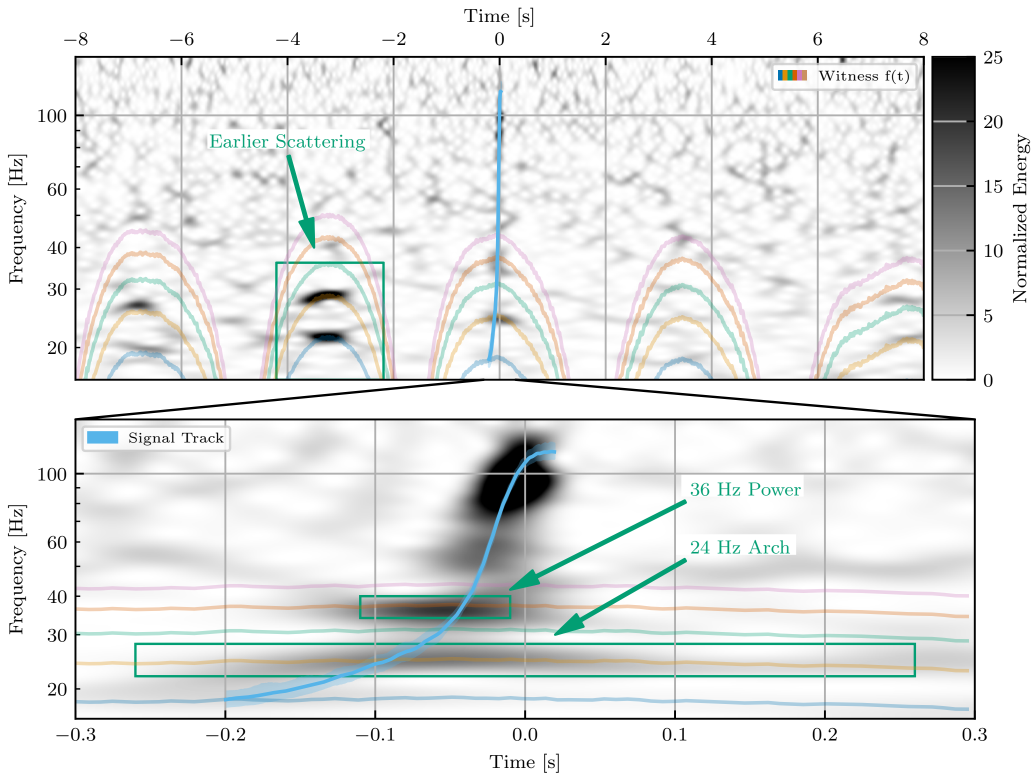

Figure 7.1: Spectrograms of the original (before glitch mitigation) data in LLO centered around the time of GW191109. The top panel shows $\pm 8\,\mathrm{s}$ of data, while the bottom panel zooms in around the event. Onto this, we plot the time-frequency tracks of the scattered light glitch, as predicted by the motion observed in the witness channel `L1:SUS-ETMX_L2_WIT_L_DQ`. This is the witness to the penultimate stage of the reaction chain pendulum for the X-arm end test mass. The scattering surface is the final stage of the reaction chain, and so this witness does not perfectly capture the motion of the scattering surface; to compensate, we apply a static coefficient of 1.38 to the predicted frequency, such that it is calibrated to the prominent scattering arches $\sim 3\,\mathrm{s}$ before the event. We also plot the inferred signal from a NRSur7dq4 analysis of full-bandwidth data after glitch subtraction (Run 1 in Table 7.1). We annotate three regions of interest: the prominent scattering before the event (top panel), the long-duration excess power at 24 Hz (bottom panel), and the short-duration excess power at 36 Hz (bottom panel). Both the 24 Hz and the 36 Hz excess power coincide with expected glitch arches, however only the former has an arch-like shape.

a witness data stream that captured the motion of the suspected scattering surface. The effect of scattering is most easily discernible 3 s before the signal, taking the form of a "stack" of arches, which is characteristic of slow scattering. Slow scattering results from low-frequency ground motion, ∼0.05−0.3 Hz, driving slow movement of the scattering surface [322]. This induces phase noise with frequency [41]

$$f(t) = \left|\frac{2v_{sc}(t)}{\lambda}\right|, \tag{7.1}$$

with $v_{sc}(t)$ being the velocity of the scattering surface and $\lambda = 1064\,\mathrm{nm}$ is the wavelength of the laser. In order for the glitch frequency to reach the analysis band, the scattered light must bounce multiple times, yielding a fixed frequency ratio between arches as the same amount of phase offset is accumulated with each successive bounce.

This picture forms the basis for a parametrized model for slow scattering that treats the scattering surface as a simple harmonic oscillator. We use the physically parameterized scattering model proposed in Ref. [351]. The model is a sum of frequency-modulated sinusoids with $2N + 4$ parameters, where N is the number of arches:

$$g(t) = \sum_{k=0}^{N} A_k \sin\left[\frac{f_{h,0} + k\delta f_h}{f_{\mathrm{mod}}} \sin\left(2\pi f_{\mathrm{mod}}(t - t_c)\right) + \phi_k\right]. \tag{7.2}$$

The peak frequency of the lowest arch is $f_{h,0}$ and the spacing in peak frequencies between adjacent arches is δf_h, such that the peak frequency of the kth arch is $f_{h,0} + k\delta f_{h,0}$.[1] The modulation frequency f_{mod} corresponds to the motion of the scattering surface (and hence the driving ground motion) and sets the width of the arch, while t_c is the time of peak frequency. Each arch k further has an independent amplitude A_k and phase ϕ_k.

Priors on these parameters reflect the physical slow scattering picture. For δf_h and $f_{h,0}$, we place uniform priors around the approximate values read from Fig. 7.1, $\delta f_h \sim \mathcal{U}(5, 8)\,\mathrm{Hz}$ and $f_{h,0} \sim \mathcal{U}(18, 20)\,\mathrm{Hz}$, while for ϕ_k we set a uniform periodic prior, $\phi_k \sim \mathcal{U}(0, 2\pi)$. We employ two sets of priors on f_{mod}. "Physical" priors limit the modulation frequency to the microseism band $f_{\mathrm{mod}} \sim \mathcal{U}(0.05 - 0.3)\,\mathrm{Hz}$ [322]. "Targeted" priors further restrict the modulation based on the witness motion $f_{\mathrm{mod}} \sim \mathcal{U}(0.05 - 0.15)\,\mathrm{Hz}$. While the

[1] Unlike Ref. [351], we fix the frequency ratio between arches to $\delta f_{h,0}$, thus eliminating $N - 1$ parameters.

former choice is more agnostic, the latter maximizes information from witness channels. Since the detector sensitivity varies by orders of magnitude in the frequency region spanned by the arches, we explore both a uniform and log-uniform amplitude for the amplitude A_k. We do not impose a relation between the arch amplitudes; while amplitudes might be expected to decrease with each arch, this is not universally the case [351].

The number of arches N is fixed and not a parameter of the model that is varied, unlike the flexible glitch model with BayesWave discussed in Sec. 7.3. The choice of the number of arches, therefore, impacts the results, especially for the uniform amplitude prior. Motivated by Fig. 7.1 we set $N = 5$, a choice which we investigate in App. 7.11. All analyses that model the glitch with the slow scattering model further employ a reduced minimum frequency of 16 Hz in LLO. Though the signal SNR, ρ, is negligible between 16 and 20 Hz (0.16% of ρ^2 in LLO), this setting accommodates the $\sim$ 18 Hz arch, which in turn informs the upper arches.

If the glitch overlapping GW191109 is consistent with the physical picture that motivates the slow scattering model, corresponding analyses provide the most sensitive results on the system properties. However, the model is also restricted to an interpretation of slow scattering and does not provide a means to test this assumption. If other non-Gaussian transient noise is present or if the physical picture does not fully capture the glitch morphology, biases might arise.

Wavelet glitch model

To mitigate against glitch modeling systematics, we also employ a more flexible approach with BayesWave which models transient, non-Gaussian noise independently in each detector as sums of sine-Gaussian, Morlet-Gabor wavelets [121, 124]. Such wavelets are an overcomplete basis and any smooth function can be described with some linear combination of wavelets. Thus, this glitch model is flexible enough to fit a wide range of non-Gaussian transients without fine-tuning, including slow scattering [110, 192]. Unlike the parameterized scattering model, the BayesWave glitch model is purely phenomenological, though motivated by the generic morphology of the LIGO glitches. Each wavelet is described by five parameters: central time t and frequency f, quality factor Q describing how quickly it is damped, amplitude A, and phase ϕ. We employ uniform priors over all parameters other than the amplitude, which is set through a prior on the wavelet SNR that peaks at 5 [121]. In addition to these parameters, the number of wavelets in each detector is also a variable and sampled over with a uniform prior. Uniform prior bounds are wide enough so as to not affect the posterior.

Glitch Mitigation Approaches

We employ three approaches to mitigate and study the impact of the glitch on inference: (1) discarding the affected data, (2) subtracting an estimate for the glitch from the data, and (3) simultaneously modeling the signal and glitch and obtaining source parameters for the former by marginalizing over the latter.

Discarding Affected Data

The most straightforward way to mitigate the impact of a glitch is to discard the affected data, either by band-passing in the frequency domain or by analyzing limited segments in the time domain [138, 265, 268]. While straightforward to implement, such methods forego all information in the discarded data, making them suboptimal. We instead follow Refs. [138, 268] and discard glitch-affected data only as a consistency check and to study the impact of the glitch, or its residual, on inference. Such analyses confirm that mitigation is necessary and provide insights into the detailed behavior of the data.

Subtraction of a Glitch Estimate

GWTC-3 results on GW191109 were obtained after an estimate of the glitch was subtracted from the data [38]. In most cases, the estimate for the glitch is a fair draw from a previous analysis with BayesWave [10, 13, 33, 38] but estimates generated from witness channels such as in GWsubtract are also possible [138]. Glitch-subtracted data are then used for downstream source inference. This method retains all the data and information available and is, therefore, more suitable for production analyses. However, its efficacy hinges on the subtracted glitch estimate since the true morphology of the glitch cannot be perfectly known. In the fair draw case, the expected glitch residual SNR is non-zero due to statistical uncertainty [127]. In the witness channel case, the relevant transfer functions induce further systematic and/or statistical uncertainty [268]. Residual glitch power that could bias inference is therefore expected.

Marginalization Over Glitch Realizations

Since selecting a single glitch estimate to subtract results in residual glitch SNR, the final method is to marginalize over the glitch. This approach is the most robust, but it is also typically more difficult to implement. Given some parameterized glitch model $g(\phi)$, we can model the data as

$$d = n + h(\theta) + g(\phi). \tag{7.3}$$

From this, we may extend the typical likelihood in a single detector to include the glitch:

$$\ln \mathcal{L}(d|\theta, \phi) = -\frac{1}{2} \sum_k \left\{ \frac{[d_k - h_k(\theta) - \phi_k(\phi)]^2}{S_n(f_k)} + \ln(2\pi S_n(f_k)) \right\}, \tag{7.4}$$

where k indexes the frequency bins being summed over, and $S_n(f_k)$ is the power spectral density in the k'th frequency bin. In detectors without glitches this reduces to the standard CBC likelihood, and they combine in the usual way. Using this formulation, one may then sample over both $h(\theta)$ and $g(\phi)$ simultaneously. From these samples, one may then marginalize over ϕ to produce CBC posteriors which reflect uncertainties in the modeling of the glitch.

We perform three glitch-marginalized analyses on GW191109. First, using BayesWave, we combine the signal model with IMRPhenomXPHM described in Sec. 7.3 and the sine-Gaussian glitch model described in Sec. 7.3. Compared to previous relevant analyses [110, 192, 268] we have extended the signal model to support waveforms with spin-precession and higher-order modes. Second, again using BayesWave, we combine the coherent wavelet signal model described in Sec. 7.3 and the incoherent wavelet glitch model described in Sec. 7.3 [4]. This analysis uses a more flexible—and thus less sensitive—model for the GW signal; it is thus used as an additional check. Even though BayesWave has the capability to also marginalize over the Gaussian noise PSD [109, 226], we fix it for consistency with other analyses and since its effect on source inference is generally minimal [7]. Third, we implemented the physically-motivated scattered light glitch model of Sec. 7.3 in BILBY. This allows us to jointly use the slow scattering model and the NRSUR7DQ4 approximant for the signal.

7.4 Understanding the GW191109 Inference

In this section, we explore the relation between the GW191109 inference, especially the $\chi_{\text{eff}} < 0$ measurement, and the glitch-affected data. In Fig. 7.1 we show spectrograms of the original data (without any glitch mitigation) in LLO at the time of the event.[2] Arch-like traces (multiple colors) show the glitch time-frequency tracks as predicted by a witness channel. The light blue track corresponds to GW191109 as inferred with NRSUR7DQ4 from data after the glitch was subtracted (Run 1 in Table 7.1). The upper panel presents 16 s of data; scattering arches are visible leading up to the event. In the bottom panel, we focus on the vicinity of the signal and highlight the intersection of the signal track with visible excess power along the projected scattering tracks. The first is at ∼24 Hz and has the expected duration and morphology of a scattering arch. The second is at ∼36 Hz and while it coincides with the glitch track predicted by the witness, the excess power duration is short and does not match the expected behavior of slow scattering. As noted in Ref. [138], this 36 Hz excess power is not included in the original BAYESWAVE glitch reconstruction and thus not subtracted in the GWTC-3 data.

We begin by confirming and extending the results of Ref. [138] with NRSUR7DQ4. Analyzing data from each detector separately (Runs 8 and 9 in Table 7.1) we confirm that the measurement is driven solely by LLO, which prefers $\chi_{\text{eff}} < 0$ at 99.6%, compared to 20.0% in LHO. Coherent analysis of both detectors (Run 1 in Table 7.1) tends to the LLO conclusion due to LLO's higher sensitivity in the relevant frequency range, shown below to be 20−40 Hz. Indeed, the maximum likelihood waveform from the coherent analysis accumulates 20% (8%) of its SNR squared in LLO (LHO) for frequencies below 40 Hz. This estimate further suggests that LHO data cannot aid in determining whether the critical ∼36 Hz excess power is part of the signal or the glitch.

Similar differences in parameter inference per detector are present for other parameters as well, notably the detector-frame total mass M and luminosity dis-

[2] A similar plot for the LHO data showing that the scattered light glitch does not overlap with the signal is given in App. 7.8.

tance D_L; see footnote 3 for a discussion of the correlation between $\chi_{\rm eff}$ and D_L. For example, in individual detector analyses (Runs 8 and 9 in Table 7.1) the detector-frame total mass is $M = 133^{+14}_{-14}\,{\rm M}_\odot$ ($M = 162^{+21}_{-20}\,{\rm M}_\odot$) in LLO (LHO), while the luminosity distance is $D_L = 1630^{+1360}_{-850}$ Mpc ($D_L = 2760^{+2300}_{-1570}$ Mpc) in LLO (LHO). The corresponding source-frame total mass remains the same as the increases in detector-frame mass and distance effectively "cancel out". Though different, these estimates are still consistent with each other within statistical uncertainties so there is no indication of a discrepancy across detectors as was the case for GW200129 [268]. Moreover, these differences do not lead to diverging astrophysical interpretations like the $\chi_{\rm eff}$ inference; we therefore focus on the latter in what follows.

Tracing inference across frequencies

To more precisely track the origin of the $\chi_{\rm eff} < 0$ measurement across LLO data, we perform a series of coherent 2-detector analyses where we successively restrict the LLO frequencies, incrementing the minimum frequency f_L by 5 Hz from 20 to 45 Hz (Runs 1–5 in Table 7.1). We use the glitch-subtracted data where the 24 Hz arch from Fig. 7.1 has been subtracted, but the 36 Hz excess power has not [38]. A subset of these results are shown in Fig. 7.2 (pink shading). The top panel shows a spectrogram of the glitch-subtracted data; compared to Fig. 7.1, there is no excess noise at ∼24 Hz.

Marginalized posteriors for $\chi_{\rm eff}$ are shown in the bottom panel. The legend denotes the percentage of the total SNR squared ρ^2 (computed based on the maximum-likelihood full-band signal) that remains in the analysis window after each restriction. Removing data between 20−30 Hz (solid vs dashed horizontal lines in the top panel and histograms in the bottom panel) or 30−35 Hz (dashed vs dotted) removes 6% of ρ^2 but does not dramatically alter inference: $\chi_{\rm eff} < 0$ is still preferred at 96.3% for $f_L = 35$ Hz. Such small shifts are likely consistent with the SNR reduction and regression to the prior (gray). Removing data 35 − 40 Hz (dotted vs dot-dashed) removes an additional 6% of ρ^2 and instead results in an abrupt shift in $\chi_{\rm eff}$, with $\chi_{\rm eff} < 0$ now only at 32.2%, a moderate preference for positive values.[3] Further bandwidth reduc-

[3] Since the $\chi_{\rm eff}$ prior is centered at zero, this shift to mildly positive values goes beyond regression to the prior. We attribute this to a mild $\chi_{\rm eff} - D_L$ degeneracy that arises for merger-only signals. The uniform-in-volume prior favors larger D_L and results in larger $\chi_{\rm eff}$ to compensate for the amplitude reduction. This degeneracy is less pronounced when the signal inspiral is visible, as then $\chi_{\rm eff}$ is constrained by the inspiral phase evolution beyond

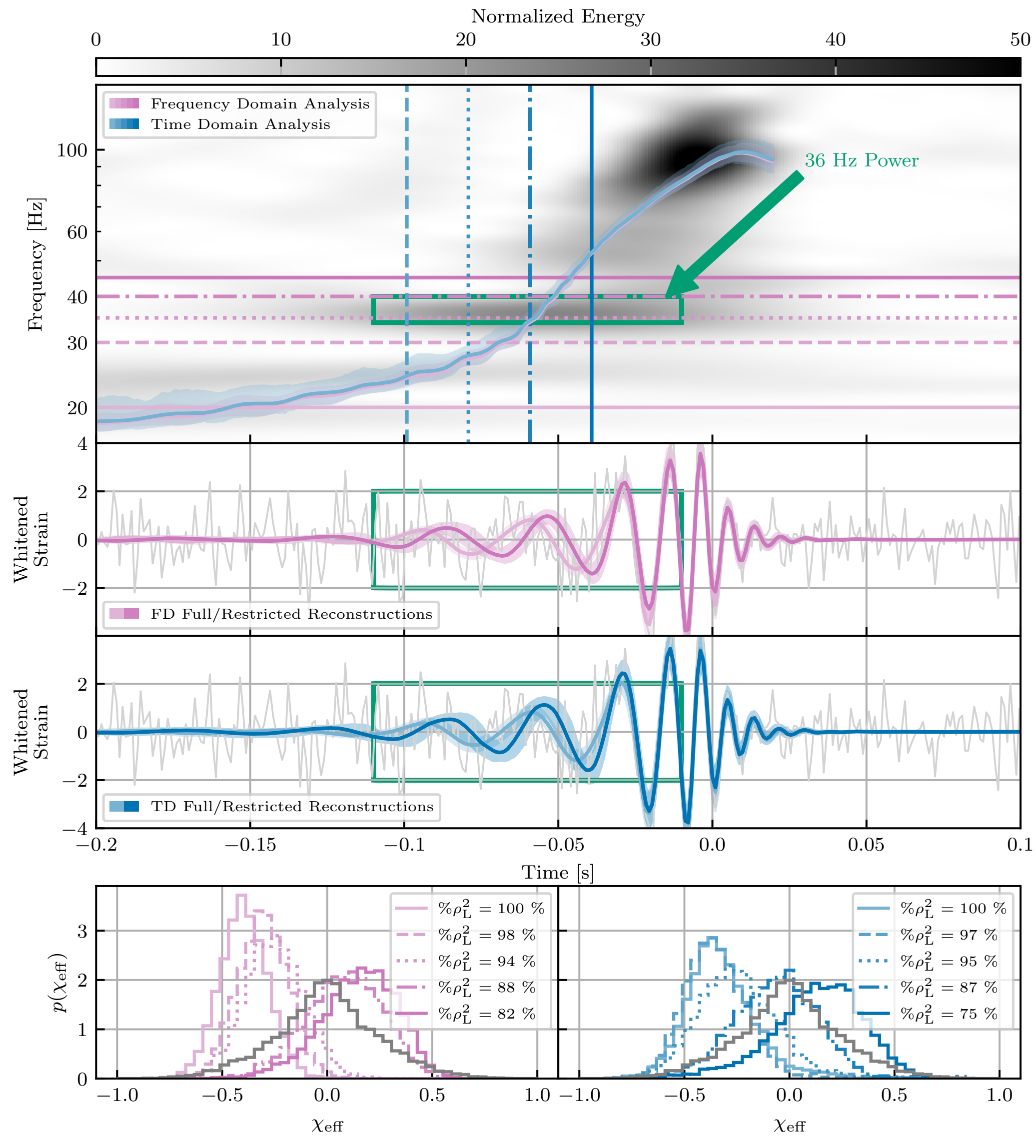

Figure 7.2: Tracing the $\chi_{\rm eff}$ inference across frequencies and times. The top panel shows the spectrogram of the glitch-subtracted data around GW191109, with residual excess power at 36 Hz highlighted along with the signal track. We progressively remove data in the frequency domain (pink) and the time domain (blue) and reanalyze the restricted data. Vertical and horizontal lines in the top panel denote the time and frequency cuts, respectively; only data to the left or above these lines are analyzed. The two middle panels show the whitened time-domain data (grey) and signal reconstruction (pink and blue). Lighter colors correspond to the analyses of the full data, while darker colors correspond to the most restricted data (frequencies above 40 Hz and times from -0.04 s before merger onwards). The bottom row shows the $\chi_{\rm eff}$ prior (gray) and $\chi_{\rm eff}$ marginal posteriors from analyses with varying levels of data restriction, each corresponding to the lines on the top panel. The legend notes the SNR squared ρ^2 fraction in Livingston that remains in the analysis band after each data restriction.

tion does not modify the $\chi_{\rm eff}$ posterior substantially (solid dark pink). These results indicate that it is the LLO data between 35 and 40 Hz that are crucial for measuring $\chi_{\rm eff}$, coinciding with the 36 Hz excess power visible in both the original data, Fig. 7.1, and the glitch subtracted data, Fig. 7.2.

The second panel from the top of Fig. 7.2 shows the whitened time-domain reconstructions. We compare signal reconstructions from two analyses with dramatically different $\chi_{\rm eff}$ posteriors: the full bandwidth analysis that prefers $\chi_{\rm eff} < 0$ against the $f_L = 40\,$Hz analysis with a mildly positive $\chi_{\rm eff}$. While the two analyses are conducted on different data subsets, we can still evaluate the waveforms across the same times and plot them together. The reconstructions are consistent during the merger (corresponding to high frequencies included in both analyses), but start diverging $2-3$ cycles before merger. By eye, the full-band reconstruction better matches the data for $t \approx -0.6\,$s, corresponding to the 36 Hz excess power. When that power is included in the analysis, the signal model absorbs it by setting $\chi_{\rm eff} < 0$ and pushing the GW cycle to earlier times. If that power is not part of the analysis, $\chi_{\rm eff}$ is no longer required to be negative and the 36 Hz excess power is left unaccounted for.

This conclusion raises the question of whether the 36 Hz excess power is part of the signal or part of a glitch that remained unsubtracted. Though the shift in the $\chi_{\rm eff}$ posterior is suggestive of anomalous noise, it is possible that it is at least partly due to loss of information as 6% of ρ^2 in LLO is contained in the 35$-$40 Hz frequency band. In Sec. 7.4 we contextualize this $\chi_{\rm eff}$ shift with simulated signals.

Tracing inference across times

Having identified the crucial frequencies for $\chi_{\rm eff}$ inference, here we do the same across time with the time-domain analysis described in Sec. 7.3. When used on the full dataset, frequency- and time-domain analyses should yield equivalent results. Indeed, we find consistent posteriors for $\chi_{\rm eff}$ when analyzing GW191109 in the frequency and time domains, as seen by the solid histograms in Fig. 7.2.[4]

just the merger amplitude.

[4] The time- and frequency-domain analyses employ different priors on masses, luminosity distance, and time. The time-domain inference uses priors which are uniform in detector-frame total mass, mass ratio, and luminosity distance; and are normally distributed in geocenter time, centered at 1257296855.2114642 with a width of 0.005 seconds. We confirm that the differences in time and the mass priors effect the posteriors minimally. Reweight-

However, as Fig. 7.1 shows, there is no 1-to-1 mapping between time and frequency for the glitch. Though not as apparent, the same is true for the signal beyond the inspiral regime or due to spin-precession and higher-order modes. Truncating the data in the time domain is, therefore, not equivalent to truncating in the frequency domain, as the former allows us to probe the effect of individual cycles (or parts of cycles) of the signal or the glitch.

Results from progressively excluding the earlier portion of the signal in the time-domain (Run 10 in Table 7.1) are shown in Fig. 7.2 (blue shading). We find broadly similar results as the frequency-domain analysis: the full data yield preference for $\chi_{\rm eff} < 0$. As the segment that contains the 36 Hz excess power is progressively removed (blue vertical lines in the top panel), the $\chi_{\rm eff}$ posterior shifts to being principally positive (equivalent blue histograms in the bottom panel). Overall, the data $0.1-0.04$ s before merger are crucial for $\chi_{\rm eff} < 0$ inference. Compared to the frequency-domain results, the shift in the $\chi_{\rm eff}$ posterior is more gradual, likely due to the fact that the 36 Hz power is more concentrated in frequency, hence no time "cut" abruptly completely excludes it. Waveform reconstructions from the time-domain analysis (third panel from the top in Fig. 7.2) yield consistent conclusions.

Simulated signals

We investigate the degree to which the abrupt shift in the $\chi_{\rm eff}$ posterior in Fig. 7.2 is consistent with SNR loss from removing data with simulated signals. We simulate 100 signals drawn from the GW191109 full-band posterior (Run 1 in Table 7.1), add them to Gaussian noise drawn from the GW191109 PSDs in LLO and LHO, and analyze the full data versus the > 40 Hz data in LLO independently (Runs 22–221 in Table 7.1). Signals have true values $\chi_{\rm eff} < 0$ but as data and signal SNR are removed when $f_L = 40$ Hz, we expect the posterior to become more prior-like and shift toward $\chi_{\rm eff} = 0$. For each simulated signal, Fig. 7.3 shows the probability of $\chi_{\rm eff} \leq 0$ from the full-data, $f_L = 20$ Hz, and the restricted-data, $f_L = 40$ Hz, analysis.

For almost all signals removing low-frequency LLO data results in a $\chi_{\rm eff}$ posterior that shifts closer to the prior and positive values (lying below the diagonal)

ing between the two luminosity distance priors proves difficult due to finite sampling and upweighting portions of parameter space with no support in the posterior. However, the luminosity distance posteriors from the time- and frequency-domain analyses are in high agreement when the full data is analyzed, despite using different priors.

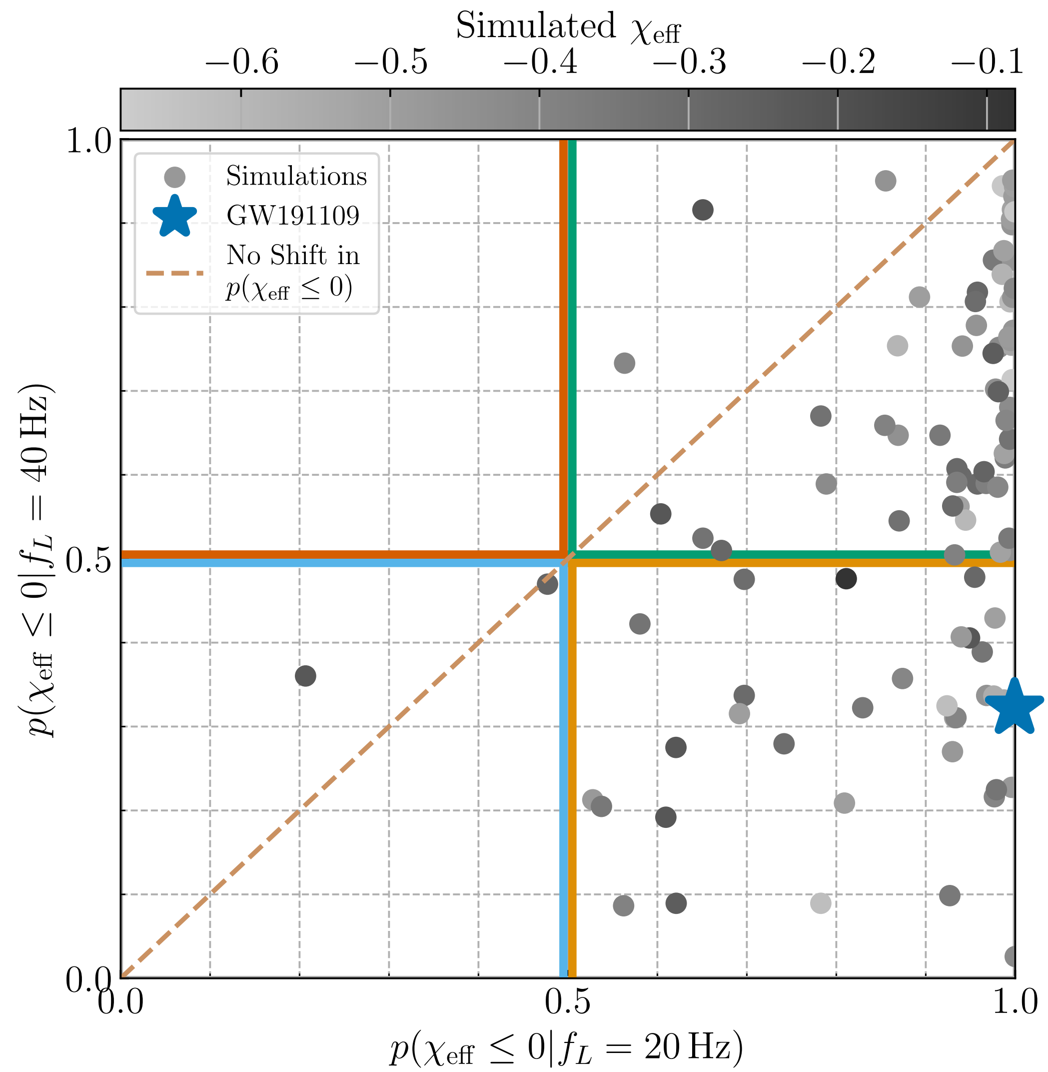

Figure 7.3: Shifts in the probability of $\chi_{\rm eff} \leq 0$ for 100 simulated signals consistent with GW191109 in Gaussian noise (dots) and the real signal (cross). The x-axis corresponds to a full-band analysis, while the y-axis corresponds to a restricted-band analysis with $f_L = 40\,{\rm Hz}$. Going clockwise, the top left quadrant (red-orange axes) would contain cases where the posterior shifted from majority positive to negative (of which there were none), the top right quadrant (green axes) contains cases which were majority negative in both full- and restricted-band analyses, the bottom right quadrant (orange axes) contains cases which started majority negative and became majority positive (including GW191109), and the bottom left quadrant (blue axes) contains cases which were consistently majority positive. The $x = y$ line (dashed brown) corresponds to no shift in the probability for $\chi_{\rm eff} \leq 0$.

as expected for signals with true values of $\chi_{\text{eff}} < 0$. In most cases, this shift is marginal, and posteriors stay majority-negative, as evidenced by the high density (64% of all signals) in the top right quadrant (green axes). The next most likely outcome is the bottom right quadrant (orange axes), which contains 34% of the signals, including GW191109: here the χ_{eff} posterior shifts from favoring negative to positive values. Among these, GW191109 is one of the more extreme cases, exhibiting a shift more significant than 94% of the simulations. Therefore, we conclude that the χ_{eff} shift presented in Fig. 7.2 is *unlikely*, but not impossible, to be explained by a random Gaussian noise instantiation, i.e. without needing to invoke residual glitch power. In App. 7.12 we present further results based on a χ^2 test used in search algorithms that tracks how SNR is accumulated along the signal [56, 141, 353]. Consistent with Fig. 7.3, the test is inconclusive: the full-band analysis (Run 1 in Table 7.1) has behavior more extreme than most simulations, but it is not strongly inconsistent with them.

7.5 Glitch-Marginalized Inference

Having established that the 35–40 Hz data drive the negative χ_{eff} inference, we turn to the question of whether these data are meaningfully impacted by residual glitch power. We go beyond subtracting a single estimate for the glitch and simultaneously model both the signal and the glitch as described in Sec. 7.3. All analyses in this section use the original data in both detectors with no prior glitch mitigation. While this approach is robust against residual glitch power from subtracting a single glitch estimate, it is still impacted by modeling choices, specifically both the parametrized model (physical scattering model or wavelets) and the corresponding glitch parameter priors.

Since the 36 Hz excess power coincides in frequency with an arch predicted by the witness channel, Fig. 7.1, it is reasonable to expect it to be part of the scattering event and thus a prime target for the slow scattering model [351]. However, the time-frequency morphology of the 36 Hz excess power does not resemble scattering arches, which motivates the alternative wavelet-based glitch model. In principle, BayesWave can fit any excess power by adding enough wavelets. Such a many-wavelet fit might be statistically disfavored, though, as it relies on a large number of parameters and a reduced posterior-to-prior volume. The exact quantitative impact of this Occam penalty is controlled by the wavelet parameter priors, which influence whether it is statistically favor-able to add a wavelet to capture the excess power or instead attribute it to the signal. The most influential prior is likely the one for the wavelet amplitude, which—although broad—favors wavelets with SNR ∼5. The situation is further complicated by the the low LHO sensitivity in the relevant frequencies, which weakens its contribution to the likelihood, making the discrimination between glitch and signal even more dependent on the prior shape.

Slow scattering glitch model

We begin with the scattered light model in Fig. 7.4 (Run 13 in Table 7.1), which models five arches with a uniform amplitude prior and the "Targeted" modulation prior that is informed by the witness motion. The signal is modeled with NRSur7dq4. The top panel shows a spectrogram of the data and the signal and glitch posteriors. The inferred glitch arches (multiple colors) match the witness prediction for the arch peak frequency spacing (∼6 Hz) in the region of maximum glitch power. The optimal SNR ρ_{opt} posterior for each arch is shown in the bottom right panel, which reveals that three non-consecutive arches are confidently recovered with $\rho_{opt} > 0$: the first one at 18 Hz (blue), the second at 24 Hz (yellow), and the fourth at 36 Hz (orange). The third arch at 30 Hz has negligible SNR, $\rho_{opt} < 2$ at 88% credibility. Though seemingly surprising given the physical interpretation of scattered light based on bounces off of moving surfaces, a varying arch amplitude is commonly observed and the SNR further depends on the noise PSD that decreases with frequency in this range. The full glitch reconstruction in the time domain is plotted in the middle panel (blue) along with the signal (pink). As expected from the presence of multiple arches, the glitch does not have a constant frequency.

The $\chi_{\rm eff}$ inference is presented in the bottom row. The bottom left panel shows the marginalized $\chi_{\rm eff}$ posterior from this analysis (pink). For comparison, we also plot the posterior from the standard two-step analysis where the glitch has been pre-subtracted and only the signal is analyzed (Run 1 in Table 7.1; blue). Under glitch marginalization, $\chi_{\rm eff}$ remains definitively negative at ∼100% credibility, though the median increases from −0.40 to −0.33. The glitch and $\chi_{\rm eff}$ inference are uncorrelated, as shown in the bottom middle panel through a scatter plot for $\chi_{\rm eff}$ and the optimal SNR of each arch.[5] This suggests that

[5]Rather than the 36 Hz power, we attribute the small shift in the $\chi_{\rm eff}$ median in Fig. 7.4 to the particular glitch realization that was subtracted for the GWTC-3 analysis. Indeed when we analyze the original data (no glitch mitigation) with only a signal (Run 12 in

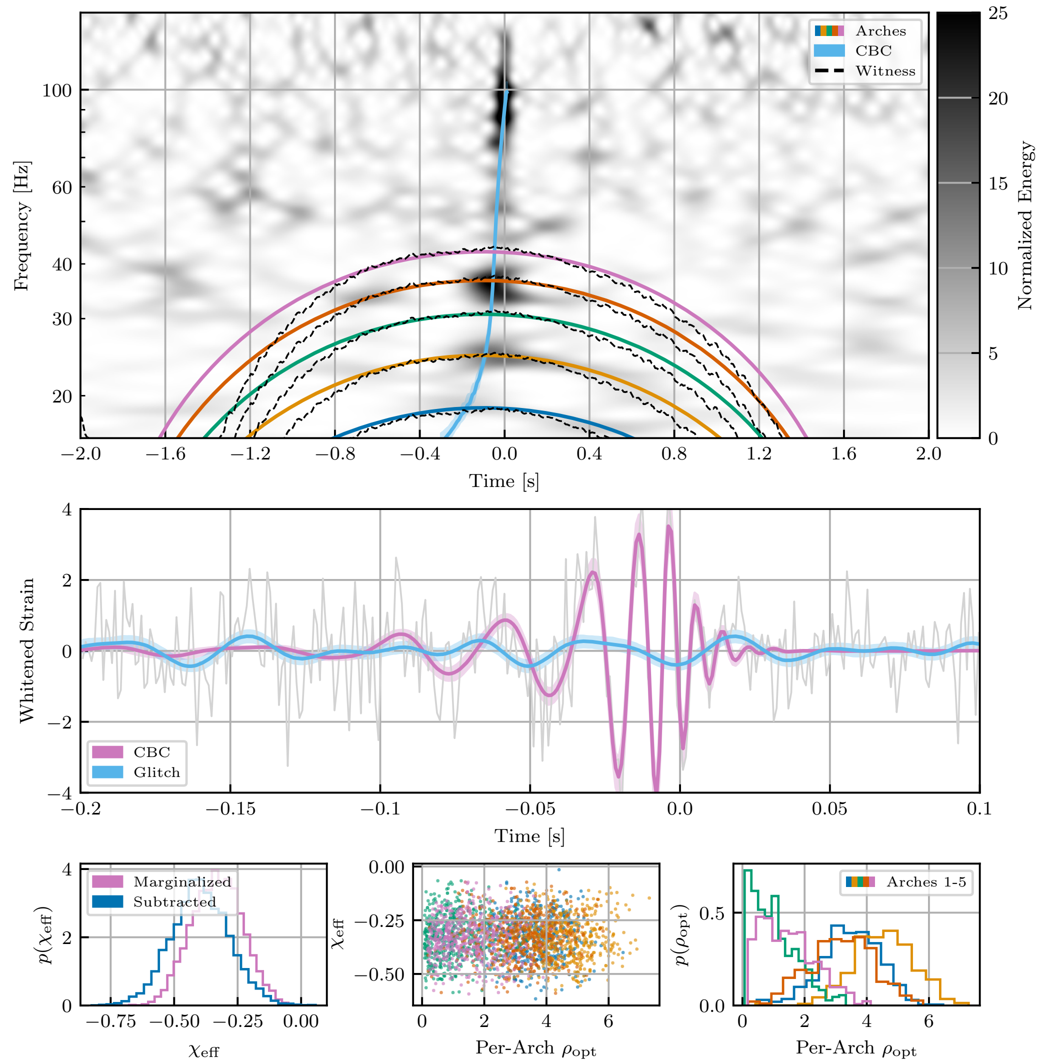

Figure 7.4: Jointly modeling the glitch with the physical slow scattering model and the signal with NRSur7dq4 (Run 13 in Table 7.1). In the top panel, we show a spectrogram of the data, along with the posterior for the glitch arches (median and 90% credible intervals; multiple colors), the signal track (blue), and the prediction of the witness channel (black dashed). In the middle panel, we show the whitened time-domain posterior reconstruction for the glitch (blue) and the signal (CBC; pink). In the bottom left panel, we show the marginalized $\chi_{\rm eff}$ posterior from this analysis (pink), along with the equivalent result from glitch-subtracted data (Run 1 in Table 7.1; blue). In the bottom right panel, we show the marginalized posterior for the optimal SNR of each individual arch. Finally, in the bottom middle panel, we show a scatter plot of individual posterior samples in the $\rho_{\rm opt} - \chi_{\rm eff}$ plane for each arch, showing that no correlation exists.

even though there is a glitch arch at 36 Hz its time-frequency morphology does not match the 36 Hz excess power. Even when the signal and glitch are simultaneously modeled, most of the 36 Hz excess power is attributed to the signal and results in $\chi_{\rm eff} < 0$. The time-domain reconstructions in the middle panel confirm this interpretation, with the signal reconstruction closely resembling those in Fig. 7.2, while a lower-amplitude glitch oscillation accounts for the remainder. We have verified that these $\chi_{\rm eff}$ results are robust under alternative, yet reasonable, priors for the glitch: log-uniform in amplitude and the "Physical" modulation prior discussed in Sec. 7.3 (Runs 14, 15, and 16 in Table 7.1 for uniform amplitude with physical modulation, log-uniform amplitude with targeted modulation, and log-uniform amplitude with physical modulation respectively). We have also verified that other parameters, such as the binary total mass and mass ratio, remain consistent between glitch-subtracted and glitch-marginalized analyses.

To summarize, we conclude that the 36 Hz power is not exclusively due to the signal. Not only does the witness channel predict some glitch power, but also the slow scattering model places an $\rho_{opt} \sim 3$ arch, notably louder than its adjacent arches. However, the excess power is not entirely attributed to scattered light as it is morphologically inconsistent with a slow scattering arch.[6] The $\chi_{\rm eff} < 0$ inference, therefore, persists under the physical slow scattering interpretation of this glitch.

Wavelet glitch model

The physically-motivated slow scattering model finds some glitch power at 36 Hz but cannot account for the entire 36 Hz excess power. This might be because of modeling systematics, the presence of other (beyond slow-scattering) non-Gaussian noise, or simply because the 36 Hz excess power is indeed part of the signal. We explore these possibilities with BayesWave and its more flexible wavelet-based glitch model as described in Sec. 7.3. We present two analyses: both marginalize over the glitch with wavelets but the GW signal is modeled with either the compact binary model IMRPhenomXPHM or with

Table 7.1), we obtain a $\chi_{\rm eff}$ posterior more similar to that of the marginalized analysis with a median $\chi_{\rm eff}$ of -0.36.

[6] In App. 7.11 we show that unphysical priors on the slow scattering parameters can indeed twist the model into fully absorbing the 36 Hz power and eliminating the $\chi_{\rm eff} < 0$ inference. Such priors are, however inconsistent with slow scattering, which forms the basis of the glitch model to being with.

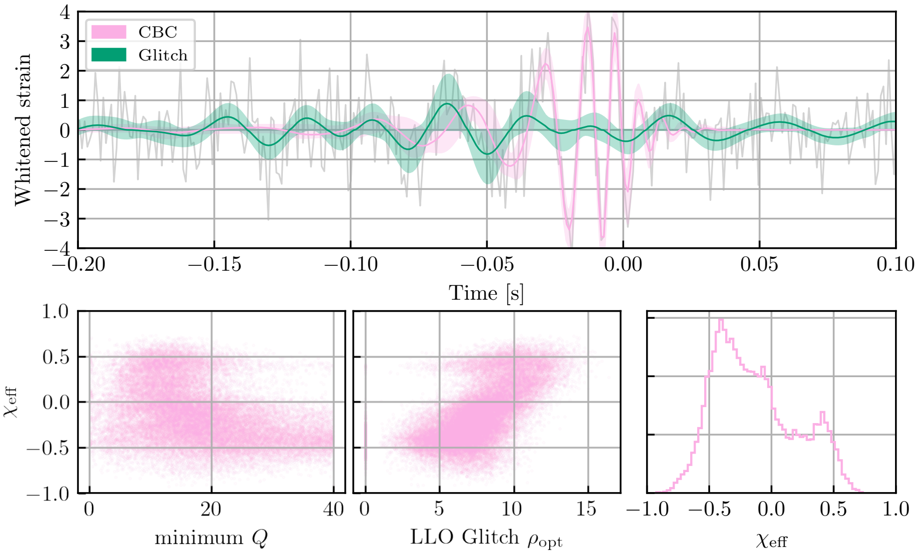

Figure 7.5: Jointly modeling the glitch with sine-Gaussian wavelets and the signal with IMRPhenomXPHM (Run 21 in Table 7.1). The top panel shows the whitened time-domain data (grey) and median and 90% credible intervals for the glitch (green) and signal (CBC; pink). The bottom row displays marginalized posteriors. The right panel shows the glitch-marginalized $\chi_{\rm eff}$ posterior, which displays a much larger spread than the results of Fig. 7.4, now being consistent with $\chi_{\rm eff} = 0$. The left panel shows the scatter plot between $\chi_{\rm eff}$ and the minimum quality factor Q among all wavelets of each posterior sample. Positive $\chi_{\rm eff}$ is correlated with low Q. Scattered light is characterized by larger Q-values [192], confirming that $\chi_{\rm eff} > 0$ only if the glitch does not match the expected scattered light morphology. The middle panel shows a scatter plot between $\chi_{\rm eff}$ and the glitch SNR which are again correlated: higher glitch power leads to a more positive $\chi_{\rm eff}$.

coherent wavelets.

IMRPhenomXPHM

In Fig. 7.5 we show results from the joint analysis with IMRPhenomXPHM for the signal and wavelets for the glitch (Run 21 in Table 7.1). The top panel shows the whitened time-domain reconstructions. Compared to the reconstructions in Fig. 7.4 there is now increased uncertainty around the 36 Hz excess power, i.e. between times −0.09 and −0.04 s. This is due to the larger flexibility of the glitch model, which can now compete with the signal for the data around −0.06 s, leading to larger uncertainties for both models. The

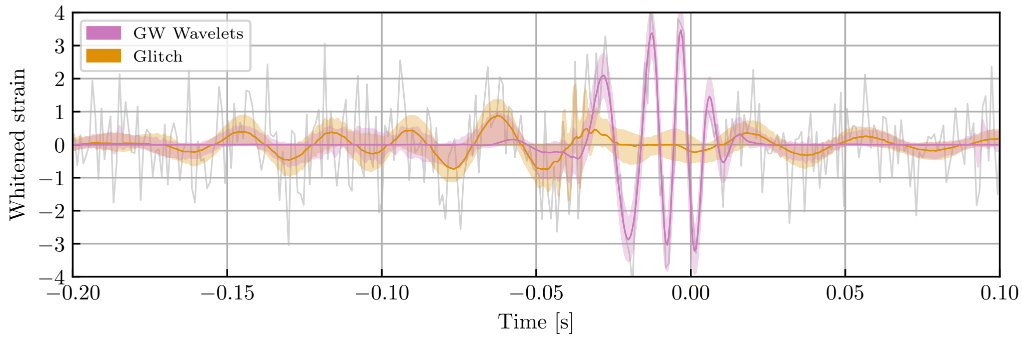

Figure 7.6: Jointly modeling both the glitch and the signal with sine-Gaussian wavelets (Run 20 in Table 7.1). We plot the whitened time-domain data (grey) and median and 90% credible intervals for the glitch (orange) and signal (GW Wavelets; purple). The 36 Hz excess power is consistent with originating from either the glitch or the signal at the 90% credible level.

larger uncertainty is also reflected in the glitch-marginalized $\chi_{\rm eff}$ posterior shown in the bottom right panel. Compared to Fig. 7.4, the $\chi_{\rm eff}$ posterior is now much wider and entirely consistent with zero. It displayes a broadly bimodal structure with one mode favoring $\chi_{\rm eff} < 0$ and peaking at ~ -0.4 and the other favoring $\chi_{\rm eff} > 0$ and peaking at at $\sim$0.4. The antialigned mode is weakly favored at 70% of the posterior samples have $\chi_{\rm eff} < 0$.

The increased $\chi_{\rm eff}$ uncertainty is entirely due to the glitch and the competition between the signal and the glitch models. The bottom middle panel shows a posterior scatter plot for $\chi_{\rm eff}$ and the SNR of the glitch in LIGO Livingston.[7] The glitch SNR is strongly correlated with $\chi_{\rm eff}$: a higher glitch power results in a more positive $\chi_{\rm eff}$. A small fraction of posterior samples, $\sim$ 6%, have vanishing glitch SNR (zero wavelets) and a strongly negative $\chi_{\rm eff}$, consistent with results from Fig. 7.4. Besides the glitch power, we examine the recovered glitch morphology in the bottom left panel, where we plot $\chi_{\rm eff}$ against the minimum quality factor among wavelets in a particular posterior sample. The quality factor corresponds to the number of cycles in a wavelet, therefore scattering arches are characterized by larger values of Q [192]. This plot confirms the conclusions of Fig. 7.4: if the glitch is scattering-like (large Q), the model cannot capture the 36 Hz power, and $\chi_{\rm eff}$ tends to be negative. Support for $\chi_{\rm eff} > 0$ requires low values of Q which morphologically do not resemble

[7] This analysis allows for glitches in both detectors, but the Hanford data are consistent with no glitch power in the analysis window.

scattering arches.

These results are qualitatively robust against different glitch priors. When using a prior for the amplitude of each wavelet that peaks at an SNR of 3 (instead of the default value of 5), we recover the same bimodal solution for χ_{eff} and the glitch SNR. However, the preference for the antialigned mode shifts from 70% to 60% suggesting that our quantitative results are impacted by the glitch prior at the few percent level. This shift is attributed to the fact that the updated prior makes it easier to low-SNR wavelets to be added to the posterior and thus capture the 36 Hz excess power away from the signal model. The impact of glitch priors is akin to the impact of compact-binary parameters on inference [374] and is expected to be more prominent for low-SNR glitches.

We perform a final sanity check by comparing the total (signal plus glitch) reconstructions of posterior samples with $\chi_{\text{eff}} > 0$ to those with $\chi_{\text{eff}} < 0$. Although the two posterior modes result in different interpretations of which parts of the data are signal and which are glitch, their sums are consistent with each other. This is expected as it is the *total* strain of signal-plus-glitch that is compared to the data to calculate the likelihood. So any solution must result in the same total strain. While we view this as a sanity check on the analysis convergence, it also suggests that there are two distinct ways to model the data, and this analysis does not strongly prefer one over the other.

Coherent wavelet model

For completeness, we present a final analysis where both the glitch and the GW are modeled with sums of wavelets [4] (Run 20 in Table 7.1). Since the signal model is now also phenomenological, we do not extract any binary parameters such as χ_{eff} which has thus far been guiding our conclusions. Instead, we directly interpret the time-domain reconstructions in Fig. 7.6. As expected, using more flexible models results in increased uncertainties. The 36 Hz (-0.06 s in the plot) power is still traded between the two models, and neither can rule out that it belongs to them at the 90% credible level. In contrast to the signal reconstructions thus far, Figs. 7.4 and 7.5, the coherent wavelet model is not able to confidently recover the signal inspiral between times -0.1 and -0.04 s. This is again due to the large flexibility of the wavelet signal model, which needs to extract each portion of the signal independently of the others [169] as opposed to the waveform model that coherently models

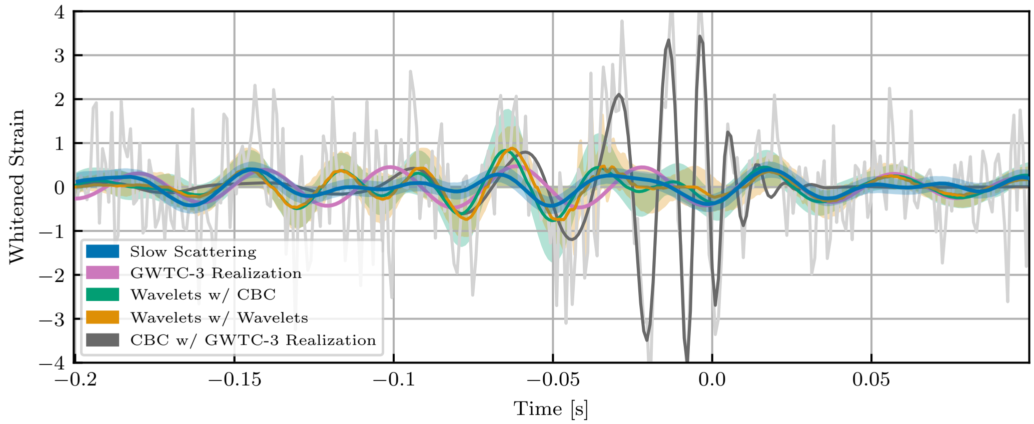

Figure 7.7: Comparison of reconstructions for the LIGO Livingston glitch that overlapped with GW191109 obtained by various analyses. The data are shown in grey, and for reference, we also show the maximum-likelihood GW reconstruction from the full-band analysis on the glitch-subtracted data in black (Run 1 in Table 7.1). The single realization subtracted for the GWTC-3 analysis is shown in pink [38]. The glitch inferred from the joint slow scattering and NRSur7dq4 analysis (Run 13 in Table 7.1) is shown in blue. The glitch inferred with wavelets is shown in green when the signal is modeled with IMRPhenomXPHM (Run 21 in Table 7.1) and orange when the signal is also modeled with wavelets (Run 20 in Table 7.1).

the whole signal across inspiral and merger.

Comparing glitch reconstructions

Finally, we compare glitch reconstructions from the various glitch inferences considered in Fig. 7.7. The comparison includes the single glitch realization considered in GWTC-3 [38] and the three glitch-marginalized analyses presented in this study, Figs. 7.4, 7.5, and 7.6. The glitch reconstructions are largely consistent with each other, with the largest differences encountered in the crucial -0.06 s region. As expected, the wavelet-based reconstructions have a larger statistical uncertainty due to the larger model flexibility. This allows them to reach a larger amplitude at -0.06 s which is necessary in order to capture the 36 Hz excess power.

7.6 Conclusions

When seeking to interpret GW data in the presence of glitches, absolute confidence in all aspects of the analysis is impossible. Unlike compact-binary signals for which we have exact numerical relativity simulations to compare models against, glitch modeling does not have the luxury of a "ground truth" solution. Nonetheless, we have sought an understanding of GW191109, its astrophysically-influential $\chi_{\text{eff}} < 0$ inference, and the overlapping glitch within the limitations of imperfect glitch models and large statistical uncertainties.

We showed that the $\chi_{\text{eff}} < 0$ measurement is attributed to a segment of LIGO Livingston data occurring between 0.1 and 0.04 s before the merger, and between 30 and 40 Hz. These data are impacted by excess scattered light non-Gaussian noise, consistent with Ref. [138]. Simultaneously modeling the GW signal with compact-binary waveforms and the glitch yields results that depend on the glitch model. A physical glitch model tailored to slow scattering glitches cannot morphologically match the excess power observed in the 36 Hz range. Therefore the $\chi_{\text{eff}} < 0$ measurement still stands. A more flexible wavelet-based glitch model is instead able to fully account for the 36 Hz excess power and wipe out all support for $\chi_{\text{eff}} < 0$. Though witness channel information suggests that slow scattering was indeed what occurred during GW191109, we cannot rule out shortcomings of the slow scattering parametrized model or additional non-Gaussian noise.

Given this, we cannot make absolute statements about the properties of GW191109. If, as expected from witness channel information, the data contain Gaussian noise, a well-modeled slow scattering glitch, and a GW signal, then GW191109 likely had asymmetric masses and $\chi_{\text{eff}} < 0$, strongly implying a dynamical origin [391]. However, if other non-Gaussian noise was present in the data, or the glitch morphology varied from classical slow scattering, spin inference becomes uninformative—though in any situation, GW191109 remains one of the heaviest observations to date. Distinguishing between these interpretations is challenging. Firstly, LIGO Hanford's sensitivity in the relevant frequency range is diminished, it can therefore not contribute to the question of whether the crucial 36 Hz power is coherent (and thus part of the signal) or incoherent (and thus part of the glitch). Secondly, the overall low SNR of the glitch makes results depend on the glitch model priors, e.g. the BayesWave glitch prior explored in Sec. 7.5.

Our analysis builds upon Refs. [138, 268] to propose a framework for in-depth analyses of glitch-afflicted data. The framework includes cross-detector comparisons, band- and time-limited analyses, simulated signals, marginalizing over the glitch, and exploring different glitch models (tailored to a specific glitch family or flexible) and prior assumptions.

As GW astronomy collects more data and seeks to constrain increasingly more subtle effects, mitigating systematics related to data quality presents a complementary challenge to waveform systematics. Similar to waveform systematics, data quality systematics can be particularly troublesome for spin inference, which typically leaves a subtle imprint on the data and is concentrated on a small (time or frequency) region of data. Studies such as the ones presented here and in Ref. [268] are based on targeted, intensive follow-up of selected events, hand-chosen for the astrophysically important inference. Data quality systematics aggregating over catalogs of detections require additional care to identify and mitigate in an automated way, e.g., [186]. Such efforts will be significantly aided by the work of experts in reducing the absolute rate of glitches, in characterizing the state of the detectors, and in developing efficient and statistically sound analyses in the presence of glitches. In this work we present techniques to help address these challenges moving forward.

7.7 Acknowledgments

We thank Jess McIver, Niko Lecoeuche, Eric Thrane, Paul Lasky, Hui Tong, Maya Fishbach, Ethan Payne, and Jacob Golomb for helpful discussions about this analysis. We thank Lucy Thomas and Aaron Zimmerman for insightful discussions on the behavior of merger dominated waveforms and the $\chi_{\rm eff} - D_L$ degeneracy. We thank Isobel Romero-Shaw and Tousif Islam for assistance with accessing and interpreting previous analyses. We thank Colm Talbot for assistance in the implementation of glitch inference in BILBY. We thank Carl-Johan Haster for helpful comments on the manuscript.

RPU and DD are supported by NSF Grant PHY-2309200. SM, SH, HD, and KC were supported by NSF Grant PHY-2110111 and NSF Grant PHY-2308770. SH was supported by the National Science Foundation Graduate Research Fellowship under Grant DGE-1745301. This work was partially supported by the generosity of Eric and Wendy Schmidt by recommendation of the Schmidt Futures program. The Flatiron Institute is funded by the Simons Foundation.

This material is based upon work supported by NSF's LIGO Laboratory which is a major facility fully funded by the National Science Foundation. LIGO was constructed by the California Institute of Technology and Massachusetts

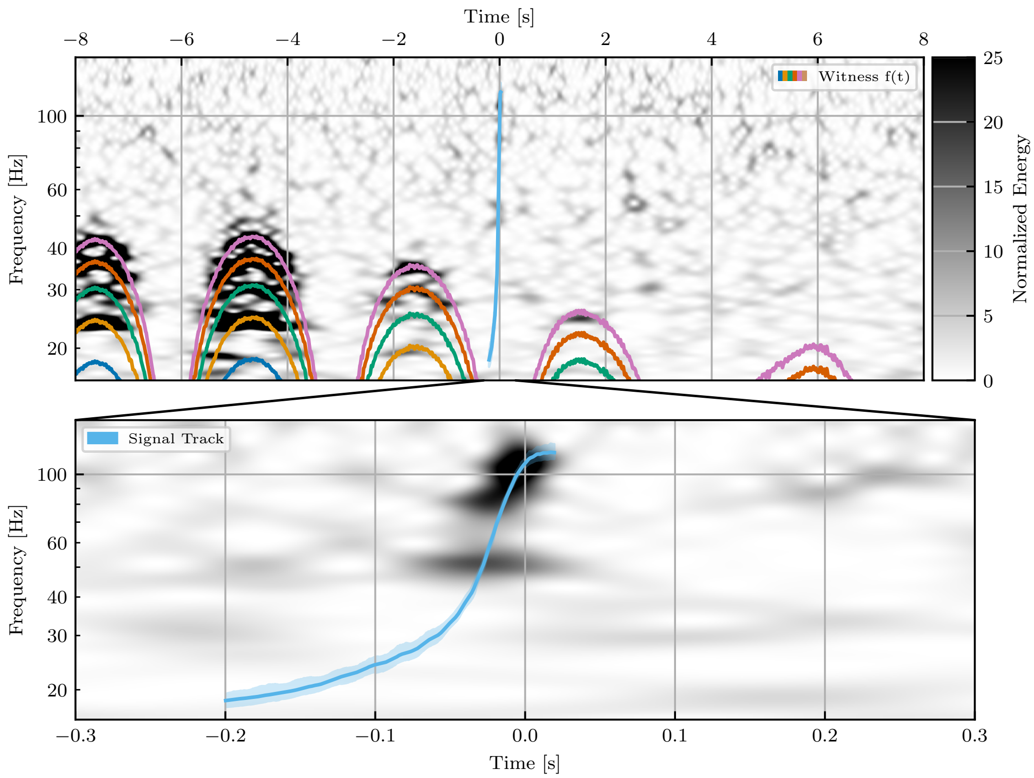

Figure 7.8: Similar to Fig. 7.1 but for LHO data at the time of GW191109, with the scattering tracks predicted by the motion of the witness channel `H1:SUS-ETMX_L2_WIT_L_DQ`. The absolute intensity of the slow scattering was significantly worse than in LLO, but the signal occurred at a minimum in the scattering, such that there is no overlap in time and frequency between the glitch arches and the GW191109 track (blue).

Institute of Technology with funding from the National Science Foundation, and operates under cooperative agreement PHY-2309200. The authors are grateful for computational resources provided by the LIGO Laboratory and supported by National Science Foundation Grants PHY-0757058 and PHY-0823459.

This work made use of NUMPY [182], SCIPY [370], MATPLOTLIB [197], LALSUITE [242], DYNESTY [324], GWPY [234], ASTROPY [282], BILBY [69], BILBY_PIPE [299], PYCBC [255], GWDETCHAR [236], and BAYESWAVE [121, 124, 226].

7.8 Scattered light glitches in LHO

In Fig. 7.8 we show a spectrogram of the data in LHO at the time of GW191109 and the scattering tracks predicted by the witness channel. When GW191109 entered the LHO frequency band, the scattering surface motion was at a minimum, so that the signal and the glitch are disjoint in time and frequency. Accordingly, we expect source inference to be unaffected by the glitch. Reference [138] reached similar conclusions. We confirm this expectation by performing analyses which restrict the frequency band in LHO in a similar fashion to main-text LLO analyses. When restricting to $> 40\,\mathrm{Hz}$ in LHO (Run 6 in Table 7.1) but with no LLO restrictions, the $\chi_{\rm eff}$ posterior remains almost entirely negative ($\chi_{\rm eff} < 0$ at 99.9% credibility, the same as Run 1 in Table 7.1 that uses all data in both detectors). When removing sub-40 Hz data in both detectors (Run 7 in Table 7.1), we obtain a modestly positive result ($\chi_{\rm eff} < 0$ at 33.3% credibility), but no more so than when we only restricting the LLO data (32.2%) (Run 4 in Table 7.1). While it is mildly surprising that removing so much low-frequency data in LHO has so little apparent effect on inference, we attribute this to the significant difference between LLO and LHO low-frequency sensitivity.

7.9 Detailed analysis settings

In this appendix we provide details about the settings of all analyses presented in this study. Table 7.1 identifies all analyses with a unique index, referenced throughout the text. We also list the data analyzed, the relevant glitch and signal models, any restrictions applied to the data being analyzed, and the analysis type (both the software used and the data domain in which it operates). Data for these analyses is made public in the associated zenodo dataset [352].

7.10 IMRPhenomXPHM Analyses with bilby

To assess whether differences between BayesWave results and bilby results are due to waveform systematics, we also perform two analyses using bilby and IMRPhenomXPHM: one on subtracted data (Run 11 in Table 7.1), and one using the slow scattering glitch model (Run 19 in Table 7.1). The analysis on subtracted data found $\chi_{\rm eff} \leq 0$ at 99.3% credibility, while the analysis marginalizing over the slow scattering model found $\chi_{\rm eff} \leq 0$ at 99.9% credibility. From this we conclude that the observed differences between bilby and BayesWave are due to the choice of glitch model, rather than the choice of waveform approximant.

Run ID	Data	Glitch Model	Signal Model	Data Restrictions	Analysis Type
1	Subtracted	—	NRSur7dq4	—	Bilby-FD
2	Subtracted	—	NRSur7dq4	$f_L = 30\,\mathrm{Hz}$	Bilby-FD
3	Subtracted	—	NRSur7dq4	$f_L = 35\,\mathrm{Hz}$	Bilby-FD
4	Subtracted	—	NRSur7dq4	$f_L = 40\,\mathrm{Hz}$	Bilby-FD
5	Subtracted	—	NRSur7dq4	$f_L = 45\,\mathrm{Hz}$	Bilby-FD
6	Subtracted	—	NRSur7dq4	$f_H = 40\,\mathrm{Hz}$	Bilby-FD
7	Subtracted	—	NRSur7dq4	$f_L = f_H = 40\,\mathrm{Hz}$	Bilby-FD
8	Subtracted	—	NRSur7dq4	No LHO	Bilby-FD
9	Subtracted	—	NRSur7dq4	No LLO	Bilby-FD
10	Subtracted	—	NRSur7dq4	Various t_H, t_L	TD
11	Subtracted	—	IMRPhenomXPHM	—	Bilby-FD
12	Original	—	NRSur7dq4	—	Bilby-FD
13	Original	Slow Scattering (Uniform + Targeted)	NRSur7dq4	—	Bilby-FD
14	Original	Slow Scattering (Uniform + Physical)	NRSur7dq4	—	Bilby-FD
15	Original	Slow Scattering (Log-Uniform + Targeted)	NRSur7dq4	—	Bilby-FD
16	Original	Slow Scattering (Log Uniform + Physical)	NRSur7dq4	—	Bilby-FD
17	Original	Slow Scattering (Uniform + Targeted, N=4)	NRSur7dq4	—	Bilby-FD
18	Original	Slow Scattering (Uniform + Unphysical)	NRSur7dq4	—	Bilby-FD
19	Original	Slow Scattering (Uniform + Targeted)	IMRPhenomXPHM	—	Bilby-FD
20	Original	Wavelets	Wavelets	—	BayesWave-FD
21	Original	Wavelets	IMRPhenomXPHM	—	BayesWave-FD
22-121	Simulated	—	NRSur7dq4	—	Bilby-FD
122-221	Simulated	—	NRSur7dq4	$f_L = 40\,\mathrm{Hz}$	Bilby-FD

Table 7.1: Settings and properties for all analyses presented in this work. We list from left to right: a unique run ID hyperlinked in the text, the type of data used (original or glitch-subtracted GWTC-3 data [2, 36]), how the glitch is modeled per Sec. 7.3, how the CBC signal is modeled per Sec. 7.3, frequency or time cuts on the data on top of the default settings, and the analysis type (software and data domain - FD for frequency and TD for time). Analyses based on glitch-subtracted data use the data provided by GWTC-3 [38], while analyses that marginalize over the glitch employ the original unmitigated data. Frequency bands are described by f_H and f_L designating the minimum frequency of analysis in LHO and LLO respectively. For runs which us the parameterized slow scattering model, the parenthetical descriptions correspond to the choice of amplitude prior and modulation frequency prior respectively for each run. All slow scattering analyses model five slow scattering arches, with the exception of Run 17.

7.11 Alternate Slow Scattering Glitch Priors

To test our assumption that it is appropriate to use the slow scattering model with five scattering arches, we also perform a test using four scattering arches with uniform amplitude and targeted modulation frequency priors (Run 17 in Table 7.1). This result finds $\chi_{\rm eff} \leq 0$ with $\sim 100\%$ credibility, indicating that the inclusion of an arch around 42 Hz does not alter the conclusions of this work.

The slow scattering model under the physically expected range of modulation frequencies $f_{\rm mod} \sim \mathcal{U}(0.05 - 0.3)$ Hz results in arches that are too extended in time to match the 36 Hz excess power morphology. We explore what values of f_{mod} are required in order to impact $\chi_{\rm eff}$ inference, with an analysis that employs a uniform amplitude prior and a maximum modulation frequency of 5 Hz (Run 18 in Table 7.1). We recover a tri-modal structure favoring $f_{\rm mod} = 1.5\,{\rm Hz}$ and a less negative $\chi_{\rm eff}$, with $\chi_{\rm eff} < 0$ at 77.1% credibility. However, $f_{\rm mod} = 1.5\,{\rm Hz}$ is 10 times larger than the scattering surface motion witnessed by the channel `L1:SUS-ETMX_L2_WIT_L_DQ`. Such a result would presume the existence of some alternative source of frequency modulated phase noise, either due to another scattering surface driven at a different frequency, or some non-scattering mechanism, coincidentally aligned in time and frequency with the known scatterer. While we cannot rule out the existence of such a source, there is no physical motivation to presuppose its existence. We instead use this analysis to emphasize the conclusion from the BayesWave study, namely that sufficiently flexible glitch models allow for a wider range of possibilities.

7.12 Frequency Bin χ Test

Tests which assess the Gaussianity of data [231, 383] may be applied to residual data after glitch and signal subtraction, but these do not address whether the signal model is capturing any glitch power. In this Appendix we instead consider the frequency bin χ^2 test as employed by search algorithms [56, 141, 353]. Qualitatively, it assesses tension between the signal waveform and the data over the entire frequency band, and hence measures deviations due both to model misspecification and to distribution of power not characteristic of a CBC, e.g., a glitch.

For each posterior waveform, we divide the frequency band into p bins of equal optimal SNR. If the data are consistent with the sum of the waveform in question and Gaussian noise, then the matched-filter SNR will also be evenly distributed over these bins. For the jth bin, the matched-filter SNR $\rho_{{\rm mf},j}$ will deviate from the mean $\rho_{\rm mf}$

$$\Delta\rho_{{\rm mf},j} = \rho_{{\rm mf},j} - \frac{\rho_{\rm mf}}{p}\,. \tag{7.5}$$

The statistic

$$\chi^2 = p\sum_{j=1}^{p} |\Delta\rho_{{\rm mf},j}|^2\,, \tag{7.6}$$

is distributed according to a χ^2 distribution with $2p-2$ degrees of freedom

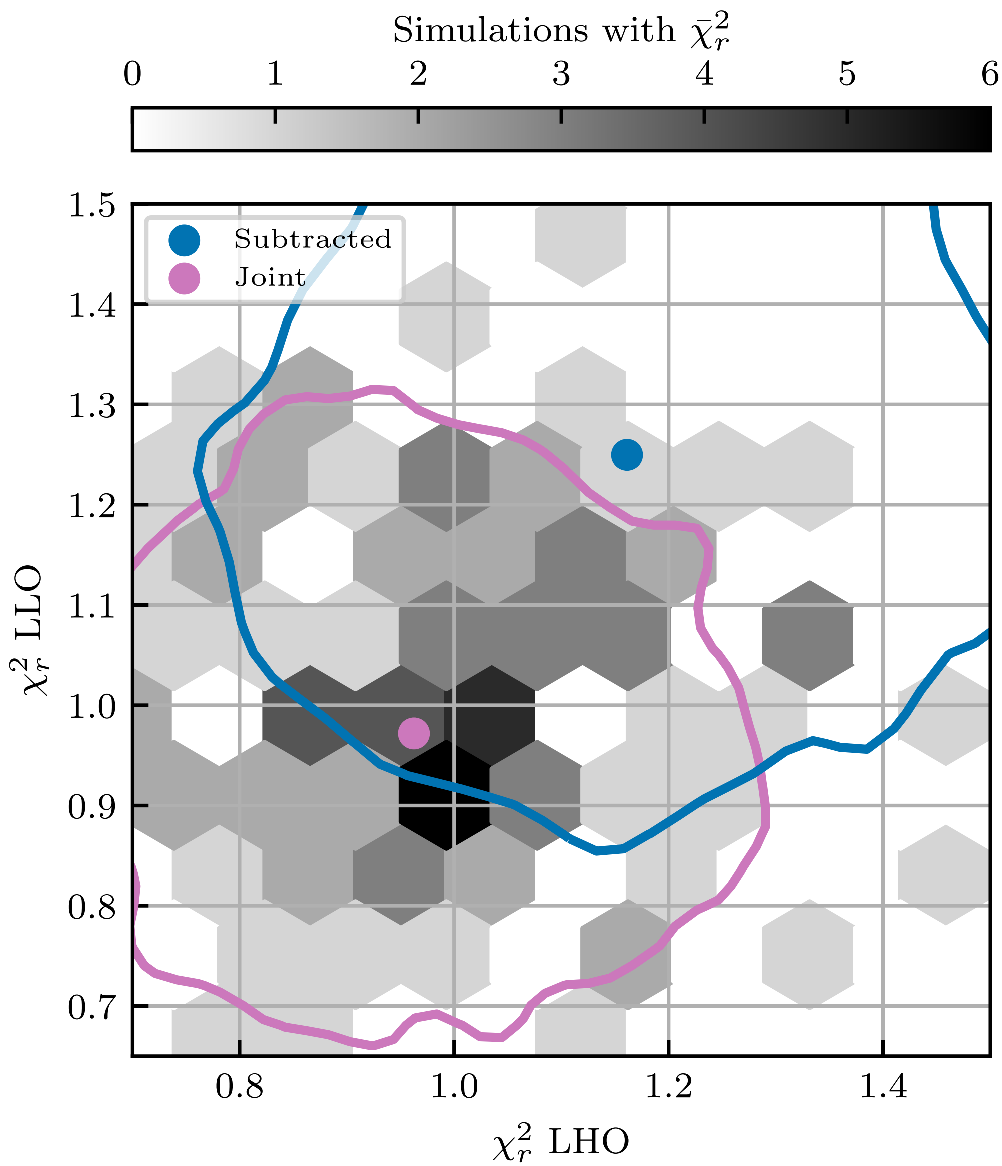

Figure 7.9: Distribution of χ_r^2 in both detectors for the glitch-subtracted analysis (blue; Run 1) and the glitch-marginalized analysis (purple; Run 13). The colormap corresponds to the distribution of $\bar{\chi}_r^2$ from simulated signals consistent with GW191109, Runs 22-121. Dots denote the distribution mean and contours denote the 90% level. For the reference distribution, we histogram the $\bar{\chi}_r^2$ values in LHO and LLO from each simulation.

under Gaussian noise[56].[8] The normalized statistic

$$\chi_r^2 = \frac{\chi^2}{2p-2}\,, \tag{7.7}$$

will then have an expected value of 1. Deviations indicate that the data might not be solely described by the waveform plus Gaussian noise, likely due to a glitch. We compute χ_r^2 for each GW191109 signal posterior sample on data where the corresponding glitch posterior sample has been subtracted. We denote the mean statistic over posterior samples as $\bar{\chi}_r^2$. We then compare against corresponding results from the simulated signals of Sec. 7.4. The reason we compare against simulations rather than directly the frequentist expectation for Eq. (7.7) is that the distribution over the posterior samples is not equivalent to a distribution over many Gaussian noise realizations.

In Fig. 7.9 we plot the statistic distribution over posterior samples in both detectors for the glitch-subtracted analysis of Run 1 and the glitch-marginalized analysis with the slow-scattering model of Run 13. The colormap corresponds to results from simulated signals where we bin the mean statistic $\bar{\chi}_r^2$ for each simulated signal. Glitch-marginalization results in a statistic whose mean is more closely in accordance with the frequentist expectation value in both detectors ($\bar{\chi}_r^2 = 0.96$ and $\bar{\chi}_r^2 = 0.97$ in LHO and LLO respectively) than glitch-subtraction ($\bar{\chi}_r^2 = 1.16$ and $\bar{\chi}_r^2 = 1.24$ in LHO and LLO). Compared to the simulated signals, glitch-marginalization results in $\bar{\chi}_r^2$ more extreme than that of 41% (45%) of simulations in LHO (LLO), while the glitch-subtracted result has a $\bar{\chi}_r^2$ more extreme than 81% (90%) of simulations in LHO (LLO). To produce a meta-statistic, we use Fisher's method [162] to compute the likelihood of these statistics occurring together, assuming that the p-values are uncorrelated. This creates another χ^2 statistic, this time with two degrees of freedom per detector. For the glitch-marginalized result, we obtain 2.25, corresponding to a p-value of 0.69, while for the glitch-subtracted results we have 7.93, giving a p-value of 0.09. Consistent with expectations, glitch-marginalization results in residuals that are more consistent with Gaussian noise after removing the glitch and signal reconstruction.

[8] Two degrees of freedom correspond to the real and imaginary components in each bin, while two are removed since deviations must sum to zero in each of the real and imaginary components.

Chapter 8

WHERE GLITCHES DO NOT MATTER

This chapter contains work from

Sophie Hourihane and Katerina Chatziioannou. Glitches far from transient gravitational-wave events do not bias inference. *In Prep*, 2025. Currently in preparation, presented here in 8. **SH** conceptualized the project, led all analyses, created all figures, and authored the text.

8.1 Abstract

Non-Gaussian noise in gravitational-wave detectors, known as "glitches," can bias the inferred parameters of transient signals when they occur nearby in time and frequency. These biases are addressed with a variety of methods that remove or otherwise mitigate the impact of the glitch. Given the computational cost and human effort required for glitch mitigation, we study the conditions under which it is strictly necessary. We consider simulated glitches and gravitational-wave signals in various configurations that probe their proximity both in time and in frequency. We determine that glitches located outside the time-frequency space spanned by the gravitational-wave model prior and with a signal-to-noise ratio below 100 do not impact estimation of the signal parameters.

8.2 Introduction

The properties of compact binary coalescences (CBCs) observed via gravitational-waves (GWs) [33, 38], offer insights on the astrophysical properties of black holes and neutron stars [35] and test General Relativity [32]. Since the first direct detection of GWs [20] the LIGO-Virgo-Kagra (LVK) detector network [1, 43] has identified 89 more CBCs [38]. The current fourth observing run has so far tallied a further ∼200 event candidates [116], for an approximate event rate of once every other day.

Also present in the data are transient, short-duration bursts of power of terrestrial origin, "glitches," with a rate of 0.5-1.28 per minute *in each detector* [8]. When glitches overlap with a signal, i.e., they are coincident in both time and frequency, they impact the inferred source parameters [1, 1, 1, 4, 110, 214, 232, 253, 265, 278]. This is because glitches violate two fundamental assumptions that underlie GW inference, namely that the noise is stationary and Gaussian. In practice, the impact of glitches ranges from biasing inference of subdominant effects like spin-precession [268] or spin-alignment [1], to mimicking an entirely different signal [18, 265]. Simply put, glitches make the noise non-Gaussian such that the common Whittle likelihood (see Sec. 8.3) no longer describes it.

GW inference is based on a time-frequency "analysis window" that is determined by the detector and signal properties, see Appendix E of Abbott et al. [38] for details. The window frequency extent is defined by a lower bound, $f_{\mathrm{low}} = 20\,\mathrm{Hz}$, set by the detector low-frequency sensitivity, and an upper bound, f_{high}, chosen to contain the merger frequency of the $(\ell, |m|) = (3, 3)$ signal mode. The time extent is designed to enclose the signal from f_{low} to merger (and ringdown) with an additional $2\,\mathrm{s}$ of data post-merger [26]. For reference, a typical analysis window for a $30\,M_{\odot}$+$30\,M_{\odot}$ (detector-frame masses) binary is $4-8\,\mathrm{s}$ whereas for a $1.4\,M_{\odot}$+$1.4\,M_{\odot}$ binary the analysis window is $128-256\,\mathrm{s}$ in length.

Every event candidate is vetted for glitches within the analysis window [8, 138], and if one is identified, it is subtracted [138]. Notably, this procedure flags glitches anywhere in the analysis window, regardless of whether they overlap, with the signal, i.e., whether they intercept the actual signal time-frequency track [8, 138]. For instance, of the 16 events flagged for glitch subtraction [33, 38],[1] in only 12 cases did the glitch overlap with the dominant, $(\ell, |m|) = (2, 2)$, signal mode. Since each event requires considerable compute and person time (including vetting, review, run time, etc. [138]) it is desirable to restrict to only cases were mitigation is necessary to avoid biases.

In this study, we explore the conditions under which glitch mitigation can be avoided. Specifically, we address *whether glitches that do not overlap with signals in time-frequency bias inference*. We focus on high-detector-frame mass events, $m \in (20, 100)\,M_{\odot}$, as glitches disproportionately impact shorter, higher-mass events compared to longer ones containing neutron stars [1]. This is likely due to the fact that inference for short signals hinges on less data, sometimes a single cycle [248, 268], and can thus be affected by short glitches.

[1] An additional 10 were initially flagged by visual inspection, but were ultimately deemed consistent with Gaussian noise [138, 365].

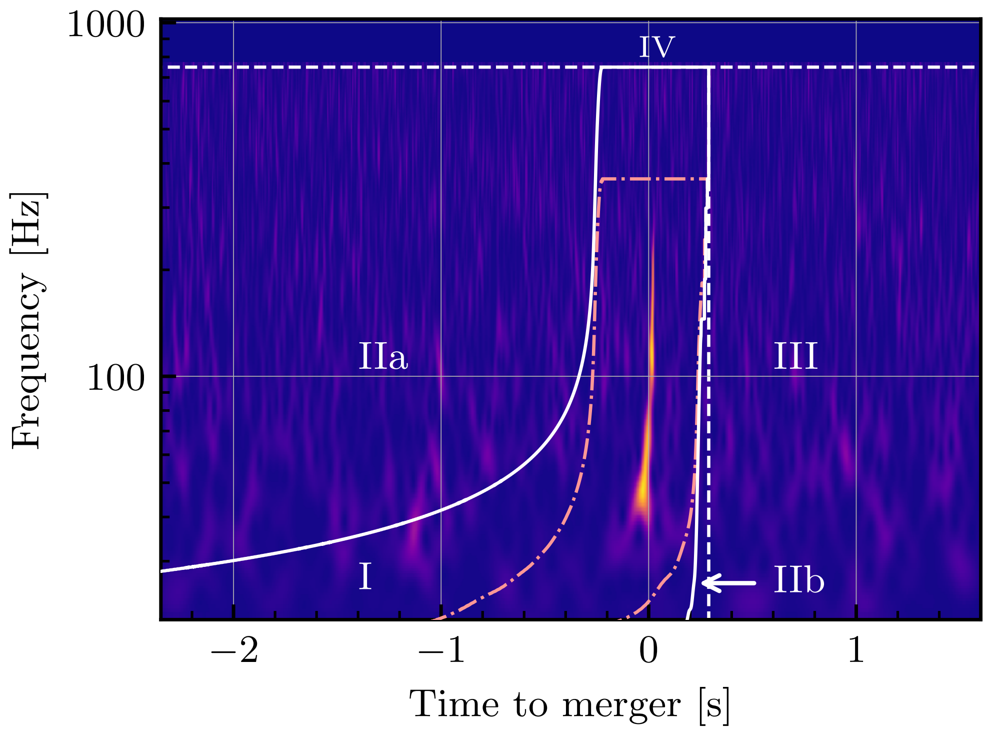

Figure 8.1: Time-frequency breakdown of the CBC analysis window, schematically describing the ways in which glitches can be positioned with respect to signals. Region I (solid line) encloses all time-frequency tracks (including those of higher order modes) within the analysis prior. For reference, in pink we plot the time-frequency content of the (2, 2) mode prior. Region II contains glitches coincident in time and frequency, but never concurrently. We split between Region IIa, those above the (4,4) mode in frequency, and Region IIb, those below the (2,1) mode. Region III contains glitches coincident in frequency, but not in time. Region IV contains glitches not coincident in frequency.

We schematically lay out the potential glitch-signal configurations in Fig. 8.1 that outlines four (non exclusive) regions.

- **Region I** contains glitches that overlap with *at least one signal within the CBC model prior*. Previous studies [1, 1, 1, 4, 110, 214, 232, 253, 265, 278] have shown that such overlaps can lead to biases when not mitigated properly. We confirm these results, primarily using this region for comparison to others.

- **Region II** contains glitches that are coincident in time *and* frequency, but not simultaneously. We further distinguish between Region IIa, glitches above the signal, and Region IIb, glitches below the signal (in frequency). Since we are focusing on high-mass events that evolve faster and merge at comparatively lower frequencies than their lower-mass counterparts, Region IIb is very small. We therefore focus on the comparatively much larger Region IIa. We evaluate differences in CBC parameter posteriors from data with and without glitches within this region. Glitches can induce a bias only when they have SNR above 100 *and* they are close in frequency to the CBC merger frequency.

- **Region III** includes glitches that share frequency content with the signal, but are not coincident in time. We again simulate glitches within that region and vary the distance (in time) from the signal, as well as the glitch signal-to-noise ratio (SNR).[2] Via a standard P-P test, we find that glitches in this region leave no statistical imprint on a population of signals. When considering individual glitches and a GW150914-like signal, glitches with SNR < 50 never induce a bias, neither do glitches with SNR< 100 if more than 0.5 s after from the signal.

- **Region IV** contains glitches that do not share any frequency content with the event. We analytically show that such glitches do not impact inference.

This rest of the paper is organized as follows. In Sec. 8.3 we recap current glitch mitigation techniques and lay out the noise assumptions that are the foundations for CBC parameter estimation. In Sec. 8.4 we lay out the methodology of our study. In Sec. 8.5 we go through each Region in Fig. 8.1 and present our results. In Sec. 8.6 we conclude.

8.3 Background

In this section we recap glitch mitigation in Sec. 8.3 and standard inference under Gaussian noise in Sec. 8.3.

[2] We only consider glitches *after* the signal because, as those before it are typically cut from the analysis window

Glitches and Glitch Mitigation

During the third observing run, glitches occurred more than once every minute in each detector [28, 38]. Of the total 79 events with an astrophysical probability greater than 0.5, 16 contained glitches within their analysis window. In 4 of 16 glitch-mitigated events, the glitch did *not* overlap the signal time-frequency track [26, 33, 38, 138]. Assuming a fixed glitch rate between $1-1.5$ glitches per minute, the number of signals requiring glitch mitigation during the fourth LVK observing run is expected to increase simply due to the increased event rate. Adopting a 4 s analysis window and two detectors, we expect 26-40 ($13-20\%$) of the 203 current candidates to contain glitches within their analysis window [116]. The increased demand for glitch mitigation motivates our detailed look into the conditions that require it.

The amount of glitch-signal overlap will determine how biased inference will be. A glitch and a GW signal overlap in frequency when, under a Fourier decomposition, there are frequencies, f_i, for which they are both nonzero. A glitch and a signal overlap in time when there are times, t_i, for which they are both nonzero. A glitch and a signal overlap in time-frequency if, when decomposed into time-frequency space, there are bins with non-zero content from both. Therefore a glitch and signal can overlap in both time and frequency, but *not* overlap in time-frequency. These definitions apply even for GWs with higher-order modes, for which there is no 1-1 relationship between time and frequency. In this case, we consider all the CBC $(\ell, |m|)$ to determine overlaps, e.g., Fig. 8.1.

Currently, if a glitch cannot be excluded from the analysis by redefinition of the analysis window (subject to the established criteria [38]), it is subject to mitigation. Common mitigation methods include: (i) subtracting a single estimate of the glitch from the data, a process known as "glitch subtraction" [137, 138], (ii) modeling both the signal and the glitch and thus marginalizing over the uncertainty of both models [1, 1, 4, 109], and (ii) removing all affected data by zeroing [393] or by replacing with Gaussian noise [385]. Glitch subtraction requires some (time or frequency) reconstruction of the glitch. Some glitch classes can be described with physical models such as slow and fast scattering [351]. Another option is to utilize a "witness" channel [137] (if one exists). In the absence of physically-motivated models or witness channels, most glitches are targeted phenomenologically with `BayesWave` [121, 124], further described in Sec. 8.4.

Gaussian Noise likelihood

The noise assumptions underlying GW data analysis determine the form of the likelihood [296]. The data, $\mathbf{d}$, are a combination of a GW signal $\mathbf{h}$ and noise $\mathbf{n}$, which is a sum of Gaussian, $\mathbf{n_G}$, and transient, non-Gaussian noise (glitches), $\mathbf{g}$. All quantities are considered in the frequency domain. For stationary $\mathbf{n_G}$, the noise is uncorrelated between frequency bins, meaning that the frequency domain noise-covariance matrix is diagonal. Gaussianity means that the noise is described by a Gaussian distribution at each frequency. The per-frequency Gaussian distribution is then entirely described by the variance, leading to a frequency-domain noise-covariance matrix that is proportional to the noise power spectral density (PSD), $\mathbf{S_n}$.

The likelihood for the GW model with parameters θ, $\mathbf{h}_\theta$ in a single detector is then [296],

$$\mathcal{L}\left(\mathbf{d}|\theta\right) = \exp\left(-\frac{1}{2}|\mathbf{d}-\mathbf{h}_\theta|^2 - \sum_{f_i} \ln\left(2\pi S_{ni}\right)\right) , \tag{8.1}$$

where $|\mathbf{a}|$ is the noise-weighted magnitude

$$|\mathbf{a}| = \sqrt{(\mathbf{a}|\mathbf{a})} , \tag{8.2}$$

$(\mathbf{a}|\mathbf{b})$ is the noise-weighted inner product

$$(\mathbf{a}|\mathbf{b}) = 4\Delta f \sum_{f_i=f_{\text{low}}}^{f_{\text{high}}} \frac{a_i {b_i}^*}{S_{ni}} , \tag{8.3}$$

and Δf is the frequency resolution. The likelihood across all detectors is the product of each individual-detector likelihood.

8.4 Methods

In this section, we describe our methodology. We simulate data containing CBC signals and glitches and analyze them ignoring the presence of the glitch. In Sec. 8.4 we introduce the `BayesWaveCpp` analysis package that is used to sample from the posterior for the CBC parameters, Sec. 8.4, and to simulate glitches, Sec 8.4. We then introduce "glitch reweighting" in Sec. 8.4 as a method to (i) quickly generate posteriors and (ii) to quantify the degree of similarity between two probability distributions, specifically those obtained from data with and without glitches. We introduce the Jensen Shannon Divergence in Sec. 8.4 as another quantity with which to compare two posteriors.

BayesWaveCpp

`BayesWaveCpp` [144] is a rewrite of and upgrade to `BayesWave` [121, 124, 222], a software package used to stochastically sample the posteriors of signals, glitches, and noise PSDs in GW data. GWs can be modeled through coherent sums of sine-Gaussian wavelets, physical CBC waveform models, and combinations thereof. Glitches are modeled as sums of sine-Gaussian wavelets. The noise PSD is modeled via broadband splines and Lorentzians for the spectral lines. For the purposes of this study, `BayesWaveCpp` is only used the sample the CBC posterior under a CBC waveform model. Though it has the capacity to do so, we do not sample the glitch or PSD posterior, assuming no glitch for the former and a known PSD for the latter.

CBC model

We model the CBC signal with `IMRPhenomXPHM` [281], an inspiral-merger-ringdown model that includes precession and higher order modes, but does not include eccentricity.[3] The CBC parameters and their priors are listed in Table 8.1, here we briefly describe some parameters of interest that feature in the following figures. The total mass, M, is the sum of the component masses of the compact binary in the detector frame. The spin angular momentum of the compact binary system can be summarized with "effective" parameters: the effective precessing parameter $\chi_{\rm p}$ (Eq. 3.4 in [316]) and the effective spin-aligned parameter $\chi_{\rm eff}$ (Eq. 2 in [28]). Both parameters have been shown to be susceptible to biases due to the presence of glitches in real data [1, 268].

Simulated glitches

We simulate glitches using `BayesWaveCpp`'s glitch model for convenience, which consists of sums of sine-Gaussian Morlet-Gabor wavelets. Such wavelets constitute an over-complete basis over a smooth function space and are thus flexible enough to mimic most noise transients in GW data. In addition to each of

[3] The ability to model precession is novel to `BayesWaveCpp`, whereas `BayesWave` is restricted to spin-aligned waveforms.

Parameter	**Symbol**	**Prior**
Mass	m_i	U[20, 100] $M_\odot$
Spin amplitude	χ_i	U[0, 1]
Spin in-plane angle	ϕ_i	U[0, 2π]
Spin polar angle	θ_i	$\cos\theta_i \sim$ U[-1, 1]
Hanford time	t_{LHO}	U[-0.05, 0.05] s
Luminosity distance	D_L	$D_L^3 \sim$ U[1, 10000^3] Mpc^3
Inclination	ι	$\cos\iota \sim$ U[0, 1]
Right ascension	α	U[0, 2π]
Declination	δ	$\sin\delta \sim$ U[-1, 1]
Polarization	ψ	U[0, 2π]
Coalescence phase	ϕ	U[0, 2π]

Table 8.1: Parameter definition, notation, and prior distribution for all CBC parameters. Here $i \in \{1, 2\}$ indexes the compact objects; $i = 1$ (2) is the larger (smaller) mass object. The time in the Hanford detector is centered 2 s before the end of the 4 s analysis window.

Parameter	**Symbol**	**Prior**
Dimension	D_g	U[1, 10]
Wavelet central time	t_0	U[-0.1, 0.1]
Wavelet central frequency	f_0	U[16, 1024] Hz
Wavelet quality factor	Q	U[0.1, 40]
Wavelet amplitude	A	$\rho_i \in$ [1, 100]
Wavelet phase	ϕ	U[0, 2π]

Table 8.2: Parameter definition, notation, and prior distribution for the glitch model, used to simulate glitches. Here ρ_i is the approximate SNR of the wavelet (Eq. 12 in [124]). This distribution is only used to simulate glitches. The wavelet central time distribution is centered at $\{0, 0.25, 0.5, 1\}$ s after the center of the CBC time prior.

the five parameters describing each wavelet (time, frequency, quality factor, amplitude, phase), the number of wavelets itself can be varied. We adopt the `BayesWaveCpp` glitch prior as the distribution from which we draw glitches to simulate, listed in Table 8.2.

The use of Morlet-Gabor wavelets allows us to quantify the glitch support in time and frequency space, e.g., Fig. 8.1. Each wavelet is characterized by a central time t_0 and a central frequency f_0, around which its power decreases exponentially, forming a Gaussian envelope in both time and frequency. In the time domain, the decay has an e-folding time scale of

$$\tau = \frac{Q}{2\pi f_0}\,, \tag{8.4}$$

where Q is the wavelet quality factor. Similarly, in the frequency domain, the decay e-folding frequency scale is $1/\tau$. So, a wavelet that is well-localized in time will be poorly localized in frequency and vice-versa. We use these estimates to approximately place the glitch with respect to the signal in subsequent sections.

Glitch reweighting

We quantify the impact of glitches on the CBC posterior by reweighting the posterior from data *without* to data *with* a glitch. Equal-weighted samples from one (reference) probability distribution can be "reweighted" to approximate another (target) probability distribution; this is a form of importance sampling called "reweighting". The exact formalism for reweighting is laid out in Sec. II of **Hourihane** et al. [11], which we adopt here. Reweighting provides a metric to compare how similar or disparate the two distributions are via the efficiency,

$$\mathcal{E} = \frac{n_{\text{eff}}}{N_s} \tag{8.5}$$

where n_{eff} is the effective number of samples (Eq. 11 in Ref.[11]) after reweighting and N_s is the original number of samples. Distributions with an efficiency of 1 are identical whereas disparate distributions have an efficiency of 0.

In this work we use reweighting in a somewhat novel way. While reweighting has been employed to change the likelihood function [11], the model [266, 300–303], the prior, or to evaluate posteriors from a neural network approximate [143], here we change the *data*. That is, we use reweighting to transform a CBC posterior distribution in Gaussian noise, $p(\theta \mid \mathbf{d}_{NG})$, to a posterior distribution of identical data plus a glitch, $p(\theta \mid \mathbf{d}_{\rm G})$. The (unnormalized) log weights are

$$\begin{aligned} \ln w(\theta) &= \ln \mathcal{L}(\mathbf{d}_{\rm G}|\theta) - \ln \mathcal{L}(\mathbf{d}_{\rm NG}|\theta) \\ &= -\frac{1}{2}|\mathbf{d}_{NG} + \mathbf{g} - \mathbf{h}_\theta|^2 + \frac{1}{2}|\mathbf{d}_{NG} - \mathbf{h}_\theta|^2 \\ &= (\mathbf{g}|\mathbf{h}_\theta) - C_g \,, \end{aligned} \tag{8.6}$$

where $C_g = (\mathbf{g}|\mathbf{g})/2 + (\mathbf{d}|\mathbf{g})$ is constant.

Reweighting has a number of advantages. Firstly, it provides a sensitive, direct estimate of the difference between two probability distributions via the efficiency, Eq. (8.5).[4] Secondly, reweighting bypasses expensive stochastic sampling and allows us to consider a wide range of glitches from a single glitch-free set of samples. Thirdly, reweighting is not subject to sampling uncertainty when stochastically sampling from a posterior. That is, $\mathcal{E} \neq 1$ if and only if the standard deviation of the likelihood over the posterior changes between between the approximate and the target distribution. Since glitches far away from the signal have a small impact on the posterior, we expect reweighting to be a particularly effective method for our purposes.

Such glitch-reweighting is possible on simulated data because we have access to glitch-free data. It is still possible to use glitch-reweighting on real data if there is a model for the glitch; for instance, the model used in glitch subtraction. The efficiency of such post-facto glitch-reweighting would serve as an estimate of the impact of that glitch on inference. In cases where the efficiency is low, it might be important to consider the uncertainty of the glitch model, e.g. [1, 1].

Jensen Shannon Divergence

We also quantify the difference between two posteriors using the Jensen Shannon (JS) Divergence [225], see App. A of Abbott et al. [28] for more details. The JS divergence between probability distributions p and q, $D_{JS}(p,q)$, is smoothed, normalized, and symmetrized. We adopt the threshold of 0.007 bit [28]. For a Gaussian, this corresponds to a 20% shift in the mean, which is larger than the 0.002 bit JS divergence expected between two identical posteriors from sampling uncertainty alone [298]. We compute the JS divergence of

[4] Another option is to use the Kullback-Liebler divergence which can likewise be computed from the weights. However, efficiency is normalized so we prefer it.

the 15 marginalized, 1-dimensional CBC posteriors and report the maximum

$$\max_{\theta} D^{\theta}_{\rm JS} = \max_{i \in 1 \ldots 15} D_{\rm JS}\left[p(\theta_i|\mathbf{d}_{\rm G}),\ p(\theta_i|\mathbf{d}_{NG})\right]. \tag{8.7}$$

8.5 Results

In this section we study how glitches impact CBC inference in each region of Fig. 8.1.

Region IV, Glitches above the signal

When the glitch and signal do not overlap in frequency, the CBC posterior is unaffected by the presence of the glitch, even when the glitch frequency is part of the analysis bandwidth. The proof presented below is based on the fact that the noise-covariance is diagonal in the frequency domain, and thus the likelihood is a 1-dimensional integral in frequency. Consider a GW signal only containing frequency content up to frequency F, coincident in time with a glitch that contains only frequency content above F. Then $(\mathbf{g}|\mathbf{h}) = 0$. If additionally the frequency content of $h_\theta(f > F) = 0$ for *all* θ in the CBC prior, π, it follows that, for all $\theta \in \pi$, the likelihood is independent of any glitch-signal cross-terms. That is:

$$\begin{aligned}\mathcal{L}\left(\mathbf{d}|\theta\right) &\propto \exp\left[-\frac{1}{2}\left(|\mathbf{d}_{\rm NG}+\mathbf{g}|^2 + |\mathbf{d}_{\rm NG}-\mathbf{h}_\theta|^2 - 2(\mathbf{g}|\mathbf{h}_\theta)\right)\right] \\ &= \exp\left[-\frac{1}{2}\left(|\mathbf{d}_{\rm NG}-\mathbf{g}|^2 + |\mathbf{d}_{\rm NG}-\mathbf{h}_\theta|^2\right)\right].\end{aligned} \tag{8.8}$$

The $|\mathbf{d}_{\rm NG}+\mathbf{g}|^2$ term is constant and does not depend on θ.[5] Since the likelihood is not normalized θ, *the extra glitch term does not change the shape of the posterior*. It will, however, scale the evidence, the expected value of the likelihood over the prior.

Region III, Glitches after the signal

We now turn to glitches occurring after a GW event that do not overlap with the GW in time. The proof of Sec. 8.5–glitches that do not overlap in frequency do not bias inference–no longer applies. This is because the noise-covariance is no longer diagonal in the time domain, and distinct times are correlated, see Sec. 8.3. Even if a glitch and signal do not overlap in time, the correlations between their times *could* induce a bias. We study this bias for a population of signals and glitches in Sec. 8.5, and for individual glitches on a GW150914-like event in Secs. 8.5 and 8.5.

[5] In analyses where the glitch is also sampled, this term would no longer be constant.

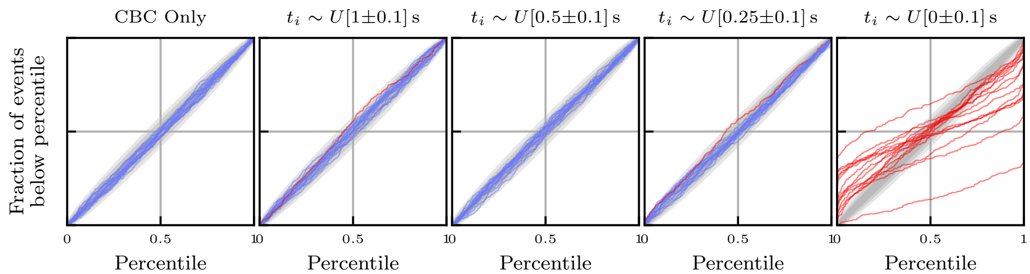

Figure 8.2: Percentile–Percentile (P-P) plots for various simulations, each drawn from the same CBC prior (Table 8.1) but varying in glitch content (Table 8.2). The titles of each plot specify the relative time between the glitch distribution and the CBC time distribution, with the leftmost plot representing data without glitches and subsequent plots showing glitches progressively closer to the CBC. Each plot comprises 400 simulations, with recovery performed using only a CBC model. Each plot includes 15 lines, one for each CBC parameter, displaying the cumulative distribution function of the percentiles of the true values within their marginal posteriors. Lines are color-coded in red (blue) to indicate whether the parameter failed, $p \leq 0.05$ (passed, $p > 0.05$) the P-P test. A failure rejects the null hypothesis: the percentiles of the true values are uniformly distributed across their posteriors. Three-sigma confidence intervals are plotted in gray. Left: P–P plot when only the CBC model is simulated and recovered. All parameters pass the P-P test which serves as a baseline for the test and its implementation. Center (left to right): P–P plots for glitches in Region III. Right: P–P plot for glitches in Region I. When glitches overlap with the signals, all parameters fail the P-P test.

Percentile-Percentile Test

We first introduce the percentile-percentile (P-P) test, a standard method to determine if an ensemble of posteriors are statistically robust [117, 170]. In a P-P test, a set of simulations is generated by drawing parameters from the prior distribution of a model. The posterior is then computed with the same model and prior. If the prior reflects the underlying population (including the noise model) and the posteriors have the correct statistical coverage, the percentiles of the true parameter values within their respective posteriors should follow a uniform distribution. Consequently, the cumulative distribution function (CDF) of these percentiles should form a diagonal line with a slope of 1.

We perform this test with 400 simulated CBC signals (without glitches) in Gaussian noise and the CBC prior of Table 8.1. The leftmost panel of Fig. 8.2

shows the CDF for each of the 15 CBC model parameters. All parameters lie within the 3-sigma confidence region (indicated by the gray outline) as expected. This demonstrates that the `BayesWaveCpp` CBC sampler is unbiased under the conditions of the P-P test.

Varying distance between glitch and signal

To assess the impact of glitches, we perform a variation of the standard P-P test described in Sec. 8.5. Instead of pure Gaussian noise, the simulated data now also contain glitches drawn from the distribution of Table 8.2. The glitch time distribution is progressively moved closer to the signal. Still, just as in Sec. 8.5, the posterior is computed only over the CBC model, leaving the glitch unaccounted for. Five scenarios are compared in Fig. 8.2: a control case with no glitches, and four cases with glitches whose time distributions have a width of 0.2 and are centered at 1 s, 0.5 s, 0.25 s, and 0 s after the center of the CBC time prior. All simulations are performed with a single, LIGO Hanford (LHO) detector (to maximize the impact of glitches), and the Advanced LIGO sensitivity [1]. By leaving the glitch unaccounted for during recovery and varying the timing of the glitch relative to the CBC, we explore how the timing of glitches impacts inference.

Results are shown in Fig. 8.2. The leftmost panel displays the control case with no glitches; all parameters pass the P–P test, indicating unbiased recovery. The center three panels (from left to right) depict results on data containing glitches after, but progressively closer to, the CBC signals. Recovery remains largely unaffected, with most of the 15 CBC parameters passing the test for each distance. A few cases with $p < 0.05$, are not unexpected given the large number of tests performed. In contrast, the rightmost panel, corresponding to glitches coincident in time with the CBC signals, reveals that all parameters fail the P–P test, highlighting significant biases. These result suggest that glitches occurring after a GW signal (Region III) have minimal impact on recovery for an ensemble of signals. However, glitches coincident with the signal (Region I) severely bias the recovered parameters, confirming previous results [1, 1, 1, 4, 110, 214, 232, 253, 265, 278].

Parameter	**Value**
Masses	$m_1 = 38.2\,M_\odot$, $m_2 = 32.9\,M_\odot$
Spin amplitudes	$a_1 = 0.998$, $a_2 = 0.126$
Spin in-plane angle	$\phi_1 = 0.55\pi\,\mathrm{rad}$, $\phi_2 = 0.364\pi\,\mathrm{rad}$
Spin polar angle	$\theta_1 = 0.14\pi\,\mathrm{rad}$ $\theta_2 = 0.49\pi\,\mathrm{rad}$
Hanford time	$1126259462.424\,\mathrm{s}$ $2.42455\,\mathrm{s}$ (in segment)
Luminosity distance	$D_L = 415.66\,\mathrm{Mpc}$
Inclination	$\iota = 0.88\pi\,\mathrm{rad}$
Right ascension	$\alpha = 0.35\pi\,\mathrm{rad}$
Declination	$\delta = -0.36\pi\,\mathrm{rad}$
Polarization	$\psi = 0.42\pi\,\mathrm{rad}$
Coalescence phase	$\phi = 1.30\pi\,\mathrm{rad}$

Table 8.3: Simulated CBC parameters for Sec. 8.5, Sec. 8.5, and Sec. 8.5.

Parameter	**Value**
Glitch dimension	$D_g = 1$
Wavelet central time	$t_1 = 2.38\,\mathrm{s}$
Wavelet central frequency	$f_1 = 100\,\mathrm{Hz}$
Wavelet quality factor	$Q_1 = 28.73$
Wavelet amplitude	A_1: Varied such that $\rho_1 \in \{5, 10, 50, 100,$ $500, 1000, 5000\}$
Wavelet phase	$\phi_1 = 1.61\pi\,\mathrm{rad}$

Table 8.4: Simulated glitch parameters used in Sec. 8.5.

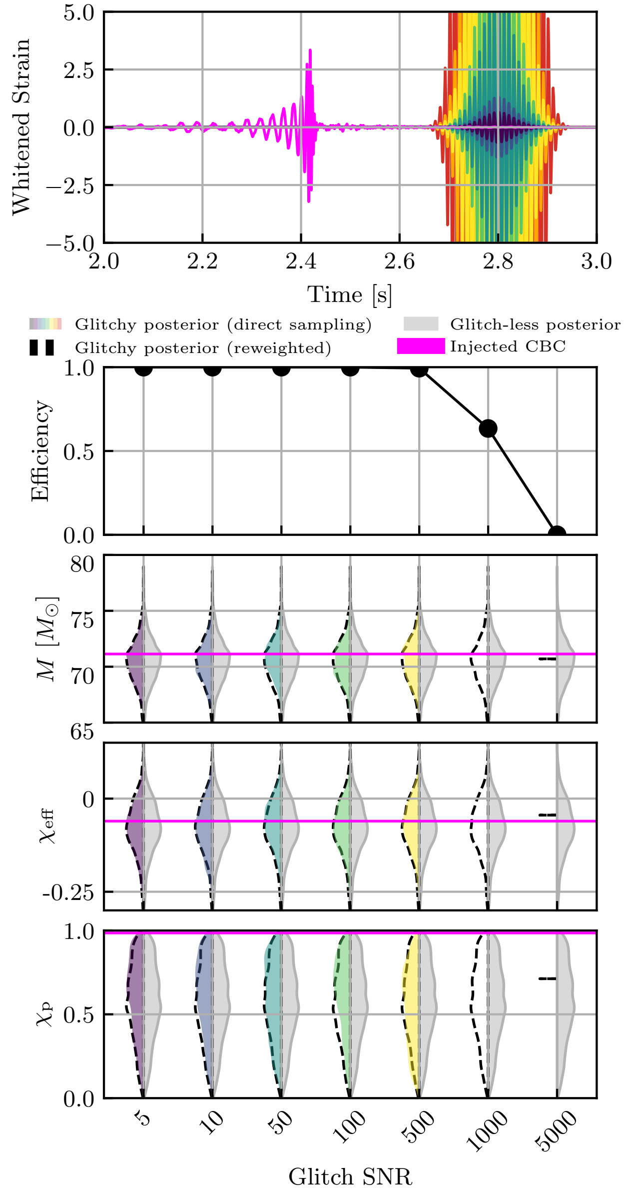

Figure 8.3: Top: Time domain whitened waveforms for the CBC (magenta) and a glitch with increasing SNR from 5 to 5000 (various colors), 0.38 s after the signal, see Table 8.3 and Table 8.4 for details. Second down: Efficiency when reweighting from a posterior on data with no glitch to a posterior on data with a glitch as a function of the glitch SNR. Bottom 3: 1-dimensional posteriors for select CBC parameters as a function of the glitch SNR. True values are marked in magenta. The direct-sampled posterior on glitch-impacted data is colored and the corresponding reweighted posterior is marked with black dashed lines. To the right we show the control dataset, a posterior recovered from data with identical Gaussian noise but without any glitch; all gray posteriors are identical. For glitch SNR ≥ 1000 we omit the direct-sampled posteriors due to nonphysically waveform behavior, further discussed in App. 8.8. The efficiency, indicates that glitches louder than SNR 500 start impacting inference.

Can very loud glitches bias inference?

We now consider a single instance of a CBC and a glitch and study at which SNR a glitch can create a bias. We compare two datasets, again in a single LHO detector:

1. **Control Dataset:** A simulated GW150914-like signal with a network SNR of 25, injected into Gaussian noise. The CBC parameters are detailed in Table 8.3.

2. **Glitch-impacted Dataset:** The glitch-impacted dataset is identical to the control one in terms of the CBC and Gaussian noise realization but includes seven distinct glitch configurations. The glitches share all parameters, Table 8.4, but SNR which we incrementally increase from 5 and 5000.

We analyze both datasets with `BayesWaveCpp`, modeling *only* the CBC signal, *leaving again the glitch power entirely unmodeled*. Results for the control dataset serve as a baseline for comparison; if the glitches have no impact, the posteriors from the glitch-impacted dataset will match those from the glitch-free dataset. We compute glitch-impacted posteriors both with direct sampling (`BayesWaveCpp`) and via reweighting from the control to the glitch-impacted dataset, see Section 8.4. If the glitches have no impact, then the reweighting efficiency will be 1.

We present the results in Fig. 8.3. The top panel shows the whitened time-domain waveforms of the CBC (magenta) and the glitches (various colors) for reference. Below that, we plot the reweighting efficiency and posteriors for select CBC parameters (total mass, $\chi_{\rm eff}$, and $\chi_{\rm p}$) as a function of glitch SNR. The direct-sampled, glitch-impacted, `BayesWaveCpp` posteriors are colored and the corresponding reweighted posteriors are plotted with black dashed lines. The fiducial, glitch-free posterior is displayed in gray. For glitch SNR ≤ 500[6] we find that not only are the glitch-impacted posteriors *visually* indistinguishable to the glitch-free one, but they are *actually* identical; the reweighting efficiency between the two distributions is 100%. As the glitch SNR is increased, the efficiency does drop. At glitch SNR of 1000 the posteriors are still visually

[6]For reference, the glitch on the binary neutron star merger, GW170817, was SNR ~ 800 [265].

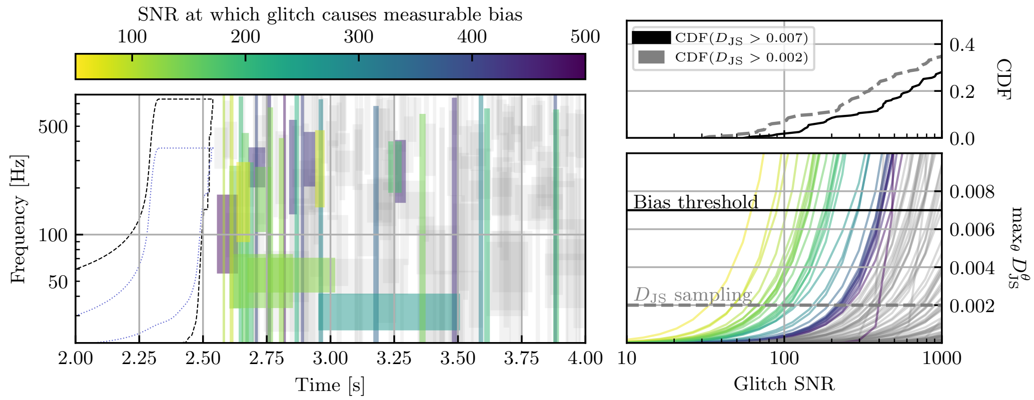

Figure 8.4: Left: Time-frequency locations of 200 glitches simulated in Region III; after the signal. The glitch time-frequency locations are colored by the SNR at which that glitch induces a measurable change in the CBC posterior and are gray if the requisite SNR is above 500. For reference, Region I (black-dashed) and the time-frequency content of the (2,2) mode across the CBC prior (blue dots) are displayed. Bottom right: Maximum Jensen-Shannon divergence across CBC parameters between a CBC posterior on glitch-free data and the posterior on glitch-impacted data, plotted as a function of glitch SNR. Each curve corresponds to a glitch in the left figure. The black horizontal line is the threshold for posteriors considered distinct [28]. The gray dashed line is the JS divergence due only to stochastic sampling uncertainty [298], plotted for reference. As the glitch SNR increases, so does does D_{JS}. Top right: Cumulative distribution function of the number of glitches that induce a requisite divergence as a function of glitch SNR. Glitches below SNR 50 never induce a measurable bias. Higher-SNR glitches might induce a bias if within 0.5 s from the signal merger.

identical, but the efficiency is 80%, suggesting it is a more sensitive test than visual comparison (it also depends on the full 15-dimensional posterior and not select marginal ones). By a glitch SNR of 5000, the efficiency is zero, signaling severe biases.

At what glitch SNR can we expect a bias?

Finally, we consider a random collection of 200 glitches per Table 8.5 in Region III after the same CBC signal from Table 8.3. Since we expect glitch-induced biases to be low, we omit direct sampling, and instead only compute posteriors via reweighting. Since the weights are proportional to the glitch SNR, per Eq. (8.6), the reweighted posterior for a single glitch can be trivially

Parameter	Value
Glitch dimension	$D_g = 1$
Wavelet central time	$t_1 \sim U[0, 4]\,\mathrm{s}$
Wavelet central frequency	$f_1 \sim U[20, 700]\,\mathrm{Hz}$
Wavelet quality factor	$Q_1 \sim U[0.1, 40]$
Wavelet amplitude	A_1: Varied
Wavelet phase	$\phi_1 \sim U[0, 2\pi]$

Table 8.5: Parameter distribution of simulated glitches in Sec. 8.5, additionally restricted to Region III, and in Sec. 8.5, likewise restricted to Region IIa.

scaled in SNR. We leverage this for each of the 200 glitches to compute the maximum JS divergence over all CBC parameters, and determine the glitch SNR at which it exceeds a JS threshold.

Results are displayed in Fig. 8.4. The time-frequency locations of the 200 simulated glitches are displayed in the leftmost panel, corresponding to three e-folds of exponential decay in each direction, see Sec. 8.4). Boxes are colored by the SNR at which that glitch induces a posterior with a JS divergence above 0.007 compared to glitch-free data. Glitches are colored gray if the requisite SNR is above 500. On the bottom right we plot the maximum (among all 15 CBC parameters) JS divergence between the glitch-impacted and glitch-less datasets, plotted as a function of glitch SNR. Each curve corresponds to a glitch in the left figure, colored in the same manner. The top right shows the cumulative distribution of the number of glitches that induce an above-threshold divergence, plotted as a function of glitch SNR.

The majority of glitches do not incur a bias unless their SNR is greater than 500. However, glitches closer to the signal (per the left plot, within 0.5 s after merger) can incur a bias at SNR as low as ~ 56 ~~∼100~~. Those are rare (2% at SNR 100 and 17% at SNR 500) and will likely need mitigation. Even at SNR 1000, fewer than 40% of the glitches induced a bias greater than one would expect from stochastic sampling. Overall, *no glitches below SNR 50 in Region III induce a measurable bias, no matter how close to (but confidently after) the signal.*

Parameter	Value
Glitch dimension	$D_g = 1$
Wavelet central time	$t_1 = 2.04\,\mathrm{s}$
Wavelet central frequency	$f_1 \in \{25, 50, 100, 250, 500, 750\}\,\mathrm{Hz}$
Wavelet quality factor	$Q_1 = 28.73$
Wavelet amplitude	A_1: Varied such that $\rho_1 = 25$
Wavelet phase	$\phi_1 = 1.61\pi\,\mathrm{rad}$

Table 8.6: Simulated glitch parameters in Sec. 8.5.

Region II, Glitches sharing time and frequency content but not concurrently

Now we move to glitches that overlap with the GW signal (or its prior) in either time or frequency, but not concurrently. These are glitches “just above," Region IIa, or “just below," Region IIb, the signal. No pure time- or frequency-domain analysis can exclude those with analysis window redefinition.

Region IIa: Glitches just above the signal

We follow the methodology of Sec. 8.5 and consider two simulated datasets:

1. **Control Dataset:** Identical to that in Sec 8.5, with a GW150914-like signal given in Table 8.3 in Gaussian noise.

2. **Glitch-impacted Dataset:** identical to the control dataset in terms of CBC and Gaussin noise, but it includes six distinct glitch configurations. All glitches have an SNR of 25 and are centered ∼13 cycles before merger; further parameters are given in Table 8.6. The frequency of the glitch is incrementally increased, starting with one that overlaps with the CBC's (2,2) mode and continuing until the glitch is outside the entire CBC model prior.

The left panel of Fig. 8.5 displays the time-frequency tracks for the modes of the injected CBC signal (colored lines), overlaid with the time-frequency support of the simulated glitches (colored boxes). The solid magenta line represents the dominant (2,2) frequency track, while the dashed lines indicate higher-order modes. The dashed black line encloses the CBC model prior,

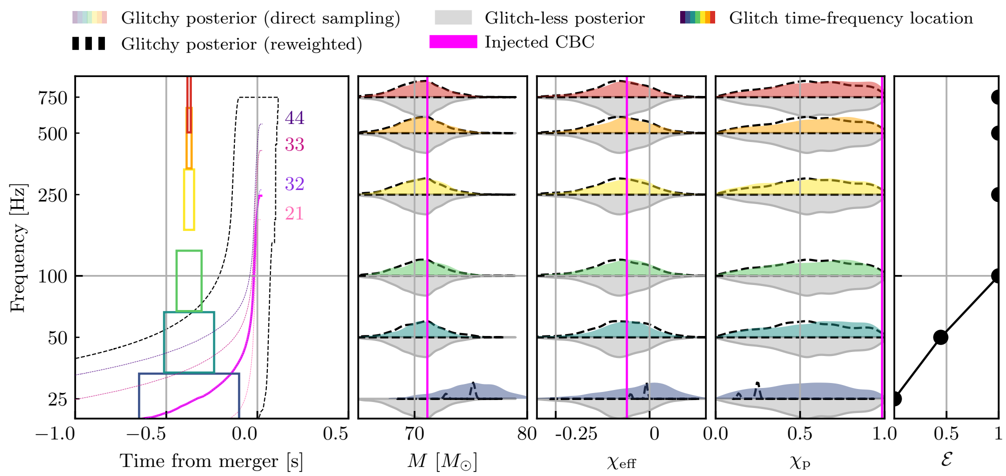

Figure 8.5: Inference results on a GW150914-like signal with SNR 25 glitches of increasing frequency. Left: Time-frequency track of the GW150914-like signal with glitches overlaid. Region I is outlined in black-dashed lines; it includes the time-frequency content of the entire GW model prior. The solid magenta line shows the (2,2) frequency track of the simulated CBC and the dotted lines display the higher-order modes, each labeled with its corresponding color. The location in time-frequency space of the glitches is shown in a colored box, each displaying three e-folds of the exponential time and frequency glitch decay. The only glitches that share time-frequency content with the GW model are the ones at 25 Hz and 50 Hz. Center: 1-dimensional posteriors for select CBC parameters from data with a glitch at the corresponding frequency (y-axis and color). In the lower half of each violin plot we show the posterior from identical data but without a glitch (gray); all such posteriors are identical. On the top of each violin plot are the posteriors recovered in glitch-impacted data; the colored posterior are those recovered with `BayesWaveCpp` (direct sampling), and the black-dashed lines are the posterior obtained via reweighting. Right: Reweighting efficiency as a function of glitch frequency. Where the posteriors differ the most, at glitch frequency 25 Hz, the efficiency drops to 0. At 50 Hz, even though the posteriors are visually similar, the efficiency also dips to 50%, meaning that the glitch is nonetheless impacting the posterior.

including all higher-order modes. The glitches at 25 Hz and 50 Hz overlap with the time-frequency GW prior, but only the glitch at 25 Hz overlaps with the (2,2) mode.

As always, we analyze the data leaving the glitch unmodeled. In the central panels of Fig. 8.5 we plot select marginalized CBC posteriors corresponding to each glitch (same color as the boxes) as well as the fiducial, glitch-less posterior (gray). The y-axis corresponds to the glitch central frequency. The black, dashed lines are again the glitch-reweighted distributions which should be identical to the colored distributions (for sufficiently high efficiency). Above 25 Hz, all distributions agree on visual inspection for all parameters. Glitches that do not overlap with the signal (2,2) mode do not have a noticeable visual effect on the posteriors.

The 25 Hz glitch overlaps the CBC's (2,2) mode in time-frequency space, and thus falls squarely into Region I. It is clearly biasing inference, increasing the total mass and χ_{eff}. The reweighing efficiency (right panel) is close to 0%, and thus reweighting is unable to reconstruct the direct-sampled posteriors.

Glitches with central frequencies 100 Hz and above have efficiencies close to 100%, meaning that posteriors from data with and without the glitch are indistinguishable. This is visually evident for M, χ_{eff}, and χ_{p} (middle panels), but the efficiency is a stronger test that considers the full 15-dimensional posteriors. Thus, glitches that do not overlap in time-frequency with the CBC model prior do not impact inference in this example.

The 50 Hz glitch presents a middle case: identical-looking posteriors but an efficiency of 50%, indicating a measurable bias. This is because that glitch shares time-frequency content with the signal's high-order modes as well as the general CBC model prior (left panel). Glitches that overlap with a signal's high-order modes can therefore still lead to small biases, even if the dominant (2,2) mode is not impacted.

At what glitch SNR can we expect a bias?

Finally, we repeat the analysis of Sec. 8.5 with 200 random glitches in Region IIa. The analysis setup is identical and refer to that discussion for details. We present on our results in Fig. 8.6 in similar format to Fig. 8.6. All glitches

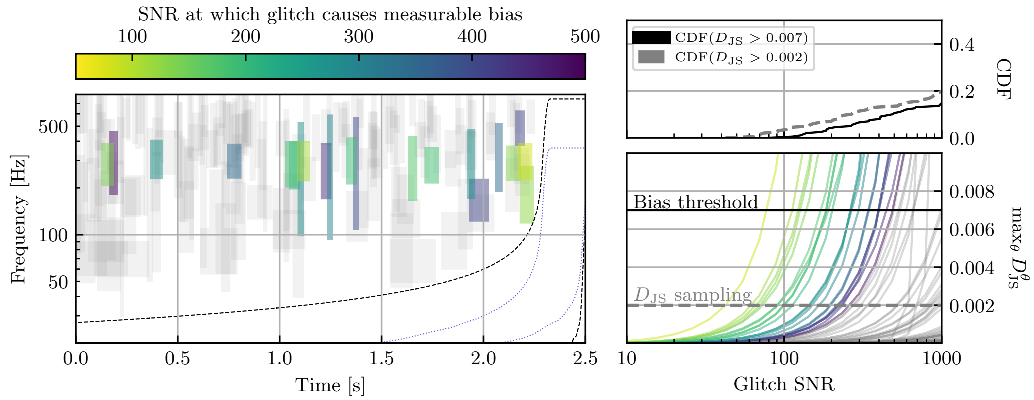

Figure 8.6: Left: Time-frequency locations of 200 glitches simulated in Region IIa. The glitch time-frequency locations are colored by the SNR at which that glitch induces a measurable change in the CBC posterior and are gray if the requisite SNR is above 500. For reference, Region I (black-dashed) and the time-frequency content of the (2,2) mode across the CBC prior (blue dots) are displayed. Bottom right: Maximum Jensen-Shannon divergence across CBC parameters between a CBC posterior on glitch-free data and the posterior on glitch-impacted data, plotted as a function of glitch SNR. Each curve corresponds to a glitch in the left figure, and are colored in the same manner. The black horizontal line is the threshold for posteriors considered distinct [28]. The gray dashed line is the JS divergence expected to arise due to stochastic sampling uncertainty [298], displayed for reference. As the glitch SNR increases, so does D_{JS}. Top right: Cumulative distribution function of the number of glitches that induced a requisite divergence as a function of glitch SNR. Only a single glitch with SNR less than 100 induces a measurable bias. Even at SNR 1000 fewer than 20% of the glitches induced a difference greater than one would expect from stochastic sampling.

that result in a bias for SNR less than 500 have central frequencies close to the CBC merger frequency, 150 Hz to 400 Hz. Even then, biases are rare. *Only a single glitch* (0.005%) causes a bias by SNR 100. Somewhat surprisingly, glitches in this region have, on average, a smaller impact on inference than those in Region III.

Region IIb: Glitches just below the signal

Since we are focusing on high-mass events (where inference is most heavily impacted by glitches [1]), Region IIb is very small. Figure 8.1 shows that there is very little time-frequency space below the waveform and outside the prior. We therefore do not perform any analyses here, instead expecting similar results to Region IIa. Region IIb will be bigger (and could contain glitches) for lower-mass signals.

Region I, Glitches on top of the signal

It has been shown both on real events [1, 1, 265] and simulations [1] that glitches overlapping GWs can be detrimental to inference. We do not present further dedicated studies of this region here, but nonetheless confirm the previous results in the right panel on Fig. 8.2 and Fig. 8.5. The former shows that a population of glitches that overlap the time-frequency prior of a population of CBCs will results in strong biases. The "s-curves" in a P-P plot are indicative of the standard deviation of the posterior being incorrect. The latter shows that glitches that overlap to dominant signal mode can lead to large (visible) biases, but a smaller impact is also expected when the glitch overlaps the higher-order modes.

8.6 Conclusion

The purpose of this study was to determine if there are time-frequency locations where one can ignore glitches nearby a transient GW, even when it is in the analysis window. We split our findings and recommendations into the same Regions as in Fig. 8.1.

- Region I includes glitches that are coincident in time-frequency with the GW event and have been shown to impact inference on real signals [1, 265, 268]. We confirm these results even at the population level; such glitches need to be carefully mitigated.

- Region II includes glitches just above and just below the GW model prior in time-frequency. That is, these glitches overlap in time and in frequency with the signal, but not concurrently. Glitches in this region do not impacting inference and thus do not require mitigation, unless they have SNR above 100 *and* are close in frequency to the CBC merger frequency.

- Region III includes glitches that share frequency content with the CBC and its prior, but occur *after* the GW, and thus do not overlap in time. For a fiducial population of glitches, biases are not expected for a glitch SNR below 50, or an SNR below 100 if the glitch is more than 0.5 s after the merger. Mitigation is again not strictly required, though in that case it is significantly simpler to perform though inpainting [385], gating [393], or modeling solely the glitch with `BayesWaveCpp` with an appropriate analysis window that excludes the CBC [138].

- Region IV includes glitches that do not share frequency content with the GW signal. Even if these frequencies are included in the analysis window, they cannot impact inference.

Throughout this study we have assumed that the noise PSD is perfectly known and can be estimated regardless of the glitch. When estimating the PSD "on source" [108, 124, 226] the glitch model is present to account for non Gaussian noise that could affect the PSD is the glitch is loud. The uncertainty associated with the power spectrum estimate is also not accounted for here, although such uncertainty is a subdominant effect in CBC analyses on *glitch-subtracted* data [7]. Other noise-mitigation techniques such as inpainting do require knowledge of the PSD which could be estimated off-source.

In summary, glitches outside Region I in Fig 8.1 are unlikely to affect parameter estimation unless they are sufficiently loud and close to the merger in time and/or frequency. Though the results quantifying the glitch SNR required for biases were based on a single, representative CBC signal, we probed a large number of potential glitch parameters and time-frequency configurations. Moreover, the population analysis of Sec. 8.5 confirms these expectations for CBC signals drawn from the prior. Quieter glitches (conservatively, below SNR 50) can thus be left in the data. Loud glitches are fortunately much rarer than their low SNR counterparts.

8.7 Acknowledgments

Thank you to Colm Talbot, Reed Essick, Jacob Golomb, Lucy Thomas, and Ryan Magee for useful discussions about the `IMRPhenomXPHM` waveforms systematics. Thank you to Eliot Finch for help with determining the time-frequency location of the waveform ringdown. Thank you to Derek Davis for their expertise on glitches and glitch mitigation. Thank you to Jane Glanzer and Siddharth Soni for help with glitch rate estimates as a function of SNR. SH was supported by the National Science Foundation Graduate Research Fellowship under Grant DGE-1745301. KC was supported by National Science

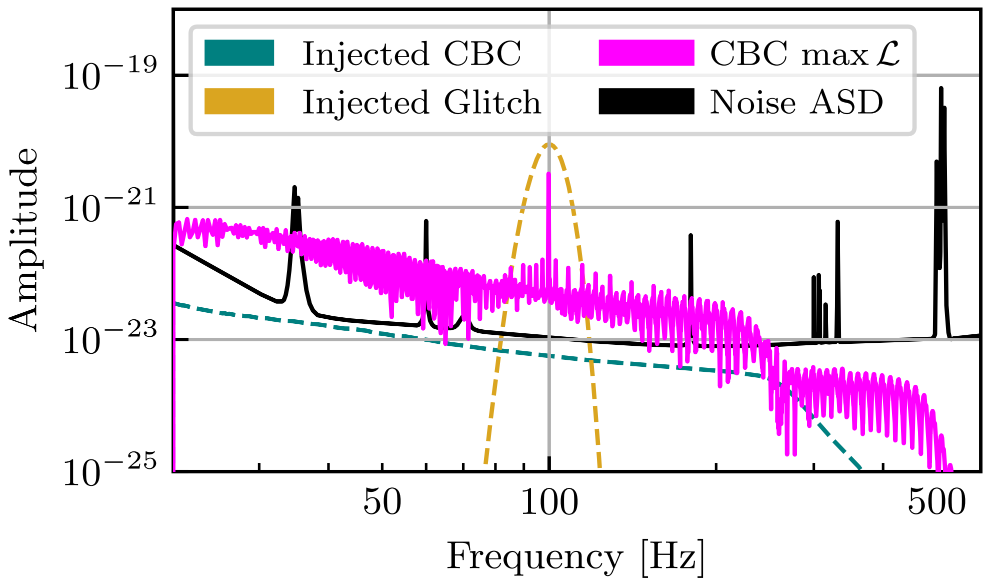

Figure 8.7: Spectral amplitude for the SNR 5000 glitch from Sec. 8.5. We show the noise ASD (square root of the PSD) in black as well as the simulated CBC (gold) and glitch (green). The pink line shows the maximum likelihood posterior sample recovered when directly sampling the posterior with `BayesWaveCpp` and `IMRPhenomXPHM`. The sampler stumbled upon a rare waveform pathology that resulted in a spike in frequency that was able to fit the glitch.

Foundation Grant PHY-2409001. This material is based upon work supported by NSF's LIGO Laboratory which is a major facility fully funded by the National Science Foundation. LIGO was constructed by the California Institute of Technology and Massachusetts Institute of Technology with funding from the National Science Foundation, and operates under cooperative agreement PHY-2309200. The authors are grateful for computational resources provided by the LIGO Laboratory and supported by National Science Foundation Grants PHY-0757058 and PHY-0823459. This work made use of `BayesWaveCpp` [144], `gwpy` [235], `scipy` [370], `matplotlib` [197], `pandas` [263], and `numpy` [182].

8.8 Appendix: Waveform Conditioning and Systematics

In Sec. 8.5, while we showed that the reweighting efficiency decreases for glitch SNRs of 1000 and 5000, we omitted posteriors created by direct-sampling (via `BayesWaveCpp`) on the glitch-impacted data. This was because in these cases the CBC model ended up fitting part of the glitch due to some waveform unphysical behavior. In Fig. 8.7, we show the maximum likelihood waveform recovered on this data, which has a clear, non-physical spike caused by a known [158] failure in the multibanding [281] of the waveform.

Regardless of present waveform systematic, we recover a drop in efficiency due to the presence of these high SNR glitches. This efficiency drop (coputed via reweighting) *cannot* be caused by the aforementioned waveform systematic. On glitch-free data, such a pathological waveform is so disfavored by the likelihood and occupies a minute region of parameter space, that it has a vanishing probability of appearing in a direct-sampled glitch-free posterior. We interpret the fact that the `BayesWaveCpp` sampler located this pathological behavior as a testament to its efficiency.

If such inference was obtained from a real signal, the issue would be immediately obvious. Nonetheless, we recommend that glitches with an SNR above 500 be mitigated.

Part II

Gravitational wave stochastic background: Analysis and challenges

Chapter 9

BACKGROUND ON THE STOCHASTIC GRAVITATIONAL-WAVE BACKGROUND

9.1 What is a stochastic gravitational-wave background

Unlike the LVK transient signals discussed in Part I that can be picked out and analyzed on top of detector noise, a stochastic gravitational-wave background (SGWB) is an incoherent accumulation of every gravitational-wave source adding on top of each other. That is, the SGWB can be characterized only statistically, in terms of expectation values. If a transient gravitational wave is a ripple in a pond, the SGWB is the signal from a rainstorm[1]. In this part of my thesis I discuss specifically the gravitational-wave background for gravitational waves in the nanohertz frequency band, as that is the target of pulsar timing array analyses.

9.2 What could create the nanohertz stochastic gravitational-wave background

Most, if not all, massive galaxies contain a supermassive black-hole (SMBH) at their center [292]. Over-densities in the early universe are the theorized site of galaxy formation. In such a high-density environment, galaxy mergers are common, naturally leading to the formation of supermassive black-hole binaries (SMBHBs) [215, 262, 377]. As these galaxy-mass black holes inspiral around each other they radiate energy as gravitational waves (see Sec 4.1).

The nanohertz frequency hum of all the combination of these SMBHBs out to $z \approx 2$ are the most likely source of a SGWB. However, there are other possible sources of the SGWB including cosmic strings and gravitational waves from the early Universe [49], but here I will limit the discussion to SMBHBs.

9.3 How can the stochastic gravitational-wave background be detected?

In Chapter 4 I discussed how gravitational-wave signals from transient, solar-mass sources can be modeled, found, and their parameters can be estimated

[1] a very quiet rainstorm

using gravitational-waveforms. But how do you model something that is random?

Additionally, a nanohertz signal takes ∼31.7 years to complete a cycle, meaning that to even hope to detect something at that low of a frequency, you have to have been measuring for at least 31.7 years [2].

Such low frequency gravitational waves cannot be studied with ground-based interferometers described in Chapter 4. The primary reason for that being that the mirrors at the end of the arms in the LVK detectors are suspended pendulums and gravitational waves can only be detected above that resonance frequency. Gravitational-waves at or below the resonance cause the pendulums to swing, which make the detection of low-frequency gravitational-waves impossible with this method [311]. In order to have an interferometric gravitational-wave detector sensitive to nanohertz frequencies, the pendulum would need to be[3], 2.5×10^{17} meters; 26.4 light-years long. So an interferometer working on the same principles as LIGO simply will not be feasible.

Pulsar timing arrays were first proposed as a means of detecting a low-frequency SGWB in 1978 [312].

Pulsars

Pulsars are rapidly rotating neutron stars [189] that possess incredibly strong magnetic fields, $\mathcal{O}(10^9 - 10^{15}\,\mathrm{G})$[4], and they can emit a constant beam of light along their magnetic field lines. When a pulsar's magnetic poles are not aligned with its axis of rotation, the pulsar acts like a "lighthouse". If the Earth happens to be aligned just-so, we observe the pulsar as a lighthouse-like signal flashing intermittently as it points towards and away from the Earth.

Pulsars vary in both the frequency of the light (e.g. radio waves vs. X-rays) and the period of their rotation. Millisecond pulsars, first discovered in 1982 [72], have periods of rotation of $\mathcal{O}(\mathrm{ms})$ which remain stable and predictable over very long timescales.

[2] As of the time of writing, I haven't even been alive that long

[3] using $f = \frac{1}{2\pi}\sqrt{\frac{g}{L}}$

[4] The magnetic field of the Earth, for reference, is $\mathcal{O}(0.2 - 0.7\,\mathrm{G})$.

Single gravitational-wave effect on pulse time of arrival

Because millisecond pulsars have remarkably stable rotation periods, we can predict the arrival of an upcoming pulse very accurately ($\mathcal{O}(10 - 100\,\mathrm{ns})$ in some cases) using information gained from studying previous pulses over many years. Here we follow the derivation in Taylor [333].

Let's imagine the pulses from a single pulsar p, located at $\vec{p}$, a distance L from the Earth with period P. A gravitational field $h_{ab}(t, \vec{x})$ will cause the spacetime in between the Earth and the pulsar to shrink or stretch making that pulse arrive early or late respectively. That is the timing difference ΔP between what we expect in flat space and what we observe in the presence of a gravitational-field is given by

$$\Delta P \equiv t_{\mathrm{obs}} - t'_{\mathrm{obs}} - P = \frac{p^a p^b}{2} \int_{t_{\mathrm{em}}}^{t_{\mathrm{em}}+L} dt' \left(h_{ab}[t' + P, \vec{x}_0(t' + P)] - h_{ab}[t', \vec{x}_0(t')] \right) \tag{9.1}$$

where $\vec{x}_0(t') = (t_{\mathrm{em}} + L - t')\ \hat{p}$ is simply a way to parameterize the path traveled by the pulsar pulse (which is necessary because of course the pulse will only be affected by the space it travels through). Since the period is of order milliseconds and the gravitational-wave periods are months to decades, we can Taylor expand leaving

$$\frac{\Delta P}{P} \approx \frac{p^a p^b}{2} \int_{t_{\mathrm{em}}}^{t_{\mathrm{em}}+L} dt' \frac{\partial}{\partial t'} \left(h_{ab}[t', \vec{x}_0(t')] \right). \tag{9.2}$$

If we imagine now the field is caused by some fiducial monochromatic plane wave with frequency ω_{GW} propagating in the $\hat{\Omega}$ direction then we can write

$$h_{ab}(t, \vec{x}) = \mathcal{A}_{ab}(\hat{\Omega}) \cos \Big(\omega_{\mathrm{GW}}(t - \hat{\Omega} \cdot \vec{x}) \Big) . \tag{9.3}$$

Substituting this into equation 9.2,

$$\frac{\Delta P}{P} = \frac{1}{2} \frac{p^a p^b \mathcal{A}_{ab}}{(1 + \hat{\Omega} \cdot \hat{p})} \Big\{ \cos\left(\omega_{\mathrm{GW}} t_{\mathrm{obs}}\right) - \cos \Big(\omega_{\mathrm{GW}}(t_{\mathrm{em}} - L\ \hat{\Omega} \cdot \hat{p}) \Big) \Big\} . \tag{9.4}$$

Here we see that the timing difference will be minimized when the gravitational wave is propagating parallel to the path of the pulse $\hat{\Omega} \cdot \hat{p} = 1$. This is again because gravitational-waves are transverse and will not warp space in the direction of propagation.

We can generalize the gravitational-wave timing delay to be from any kind of plane wave traveling in the $\hat{\Omega}$ direction and define the "GW-induced redshift of the pulse arrival rate", $z(t) := \frac{\nu_0 - \nu(t)}{\nu_0} = \frac{\Delta P}{P}$ giving us

$$z(t, \hat{\Omega}) = \frac{1}{2} \frac{p^a p^b}{(1 + \hat{\Omega} \cdot \hat{p})} \left[h_{ab}(t, \vec{x}_{\text{earth}}) - h_{ab}(t - L, \vec{x}_{\text{pulsar}}) \right] , \tag{9.5}$$

$$= F^A(\hat{\Omega}) \Delta h_A \, , \tag{9.6}$$

$$\text{where} \quad F^A(\hat{\Omega}) = \frac{1}{2} \frac{p^a p^b}{(1 + \hat{\Omega} \cdot \hat{p})} e^A_{ab}(\hat{\Omega}) \tag{9.7}$$

$$\text{and} \quad \Delta h_A = \left[h_A(t, \vec{x}_{\text{earth}}) - h_A(t - L, \vec{x}_{\text{pulsar}}) \right] , \tag{9.8}$$

finding that the timing delay is only a function of the gravitational-field when the pulse was observed at Earth and the field when it was emitted by the pulsar, and nothing in between. We have also replaced the geometric term relating the response of the plane wave to the gravitational-wave as $F^A(\hat{\Omega})$ which details the response of the pulsar to some gravitational-wave (maximal when they are perpendicular, minimal when they are aligned ($p^a p^b h_{ab} = 0$)).

In the frequency domain the gravitational-wave induced redshift becomes

$$\tilde{z}(f, \hat{\Omega}) = (e^{-2\pi i f L(1 + \hat{\Omega} \cdot \hat{p})} - 1) \sum_{A \in \{+ \times\}} h_A(f, \hat{\Omega}) F^A(\hat{\Omega}) \tag{9.9}$$

which is the pulsar-timing array version of equation 4.5.

Correlating signals between pulsars

So now we know how to predict the "timing residual" ΔP from a plane wave. However, the gravitational-wave background is a sum of all of these plane-wave-like sources, piling up in each frequency bin making them impossible to resolve individually. We will only be able to characterize it statistically meaning that we won't be able to predict $\tilde{z}(f)$, but we can predict, for instance $\langle \tilde{z}^2(f) \rangle$. This will end up acting like a noise process in a single pulsar.

However, there are plenty of other processes that could also make the pulse arrive early or late that act stochastically but are not from the SGWB. For example, the radio waves can get dispersed by ions in the interstellar medium, and we need to account for this (radio) frequency-dependent delay. Additionally, internal physics in the neutron stars can cause "intrinsic noise" that make that star's rotation period fluctuate stochastically over long timescales. Many of these noise processes can look like a SGWB when you observe a single pulsar, but they should *not* be correlated between pulsars.

So now that we know that we need to analyze the SGWB through expectation values *and* correlations between pulsars, let's compute the expected sky-averaged correlation of the gravitational-wave induced redshifts between pulsar i and pulsar j,

$$\begin{aligned}\langle \tilde{z}_i(f)\tilde{z}_j^*(f')\rangle = \int_{S^2}\int_{S'^{\,2}} d^2\hat{\Omega}d^2\hat{\Omega}' \left[e^{-2\pi i f L_i(1+\hat{\Omega}\cdot\hat{p}_i)}-1\right]\left[e^{2\pi i f' L_j(1+\hat{\Omega}'\cdot\hat{p}_j)}-1\right] \\ \times \langle \sum_{A\in\{+\times\}} h_A(f,\hat{\Omega})F_i^A(\hat{\Omega}) \sum_{A'\in\{+\times\}} h^*_{A'}(f',\hat{\Omega}')F_j^{A'}(\hat{\Omega}')\rangle\,, \quad (9.10)\end{aligned}$$

which by stationarity, Gaussianity, isotropy can be written

$$\langle \tilde{z}_i(f)\tilde{z}_j^*(f')\rangle = \frac{1}{2}\delta(f-f')S_s(f)_{ij} \quad (9.11)$$

$$= \frac{1}{2}\delta(f-f')S_h(f)\int_{S^2}\frac{d^2\hat{\Omega}}{8\pi}\kappa_{ij}(f,\hat{\Omega})\sum_{A'\in\{+\times\}}F_i^A(\hat{\Omega})F_j^A(\hat{\Omega}) \quad (9.12)$$

where

$$\kappa_{ij}(f,\hat{\Omega}) = \left[e^{-2\pi i f L_i(1+\hat{\Omega}\cdot\hat{p}_i)}-1\right]\left[e^{2\pi i f L_j(1+\hat{\Omega}\cdot\hat{p}_j)}-1\right] \quad (9.13)$$

and $S_h(f)$ is the one-sided power spectral density (PSD) of the Fourier modes of the SGWB. κ_{ij} controls how rapidly the pulsar terms of induced redshift vary. Since $f \approx 10^{-9}$Hz, the closest known millisecond pulsars are > 100 parsecs away, leaves $fL > 10$[5]. This means that the rapidly oscillating terms contribute negligibly to the above integral except where we are comparing the correlations of a pulsar to itself, $i = j$. Then when $i = j$ $\kappa_{ij} \to 2$ and $\kappa_{ij} \to 1$ otherwise.

This leaves us with the integral over the response functions

$$\Gamma_{ij} = \int_{S^2}\frac{d^2\hat{\Omega}}{4\pi}\kappa_{ij}(f,\hat{\Omega})\sum_{A'\in\{+\times\}}F_i^A(\hat{\Omega})F_j^A(\hat{\Omega})\,, \quad (9.14)$$

$$\approx \frac{3}{2}x_{ij}\ln(x_{ij}) - \frac{1}{4}x_{ij} + \frac{1}{2} + \frac{1}{2}\delta_{ij} \quad \text{where} \quad x_{ij} = \frac{1-\cos(\theta_{ij})}{2}\,, \quad (9.15)$$

where θ_{ij} is the angular separation between the positions of pulsars on the sky. Γ_{ij} is the so called "Hellings-Downs" curve [187] which is displayed in Fig. 9.1. It represents how correlated the residuals from two pulsars are. This

[5]In other words, we can ignore this term because the pulsars are much further than the gravitational-wavelengths

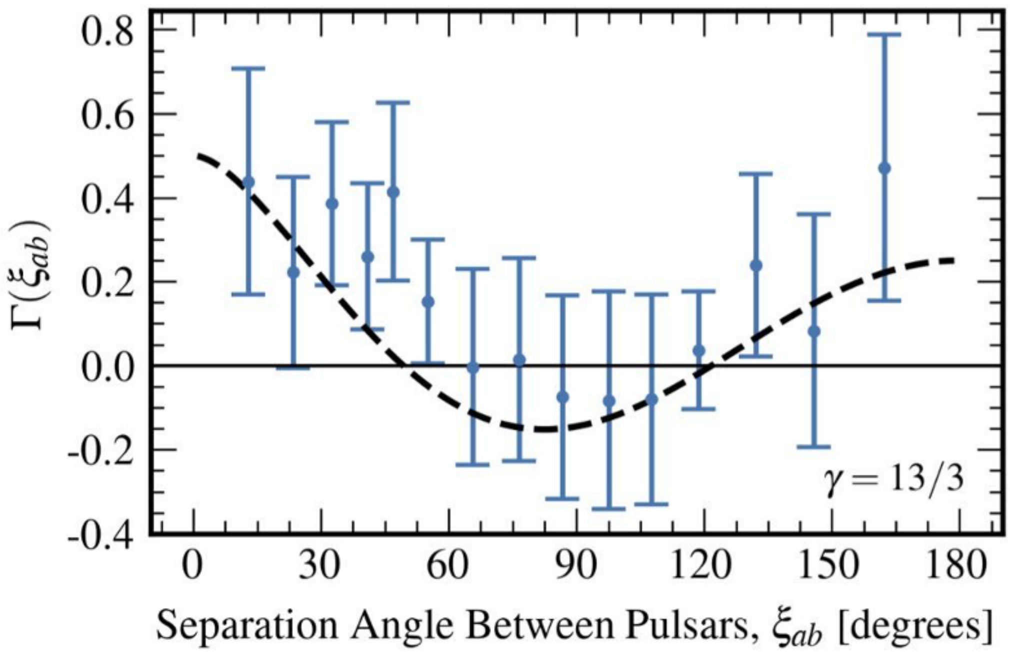

Figure 9.1: Figure from [50]. The figure illustrates the angular-separation-binned interpulsar correlations compared against the Hellings and Downs curve (dashed-line), providing a clear visual comparison between the observed and theoretically expected signals. This serves as a validation step, demonstrating consistency between the measured correlations from the NANOGrav 15-year dataset and the signature of a SGWB. In the absence of gravitational-radiation, the signal is expected to be uncorrelated, following the horizontal line.

curve has notable features. Firstly, the maximal cross-correlation is no greater than 0.5, this is because of the κ_{ij} term; since the pulsars are much further away than a single gravitational-wavelength the contribution of the oscillatory terms all but disappear. Because of this, pulsars at different distances but with 0° degrees between them will also only be correlated a maximum of 50%

[6]. Second, the correlation is not purely quadrupolar; the value at $\theta_{ij} = 180°$ does not equal the value at $\theta_{ij} = 0°$

With this Hellings-Downs function, our expected signal becomes

$$\langle \tilde{z}_i(f)\tilde{z}_j^*(f')\rangle = \frac{1}{3}\Gamma_{ij}S_h(f). \tag{9.16}$$

Correlating pulsar times of arrival

Now pick another pulsar, b, on the sky some angular distance θ_{ab} away from pulsar a.Because the pulses are measured on Earth, they will go through the same gravitational-field, but this pulse will not be at a optimal angle! This means that the effect of a gravitational wave on these two pulsars will be *correlated*, but not the same.

9.4 SGWB spectrum expected from supermassive black hole binaries

Now that we understand how pulsars can be used to find a background we want to understand, what do we *expect* the signal from the SGWB to look like? Here I follow the derivation from Phinney [272].

Since our background is *stochastic* its strain cannot be described like a wave in the same way we did in Chapter 4, it can only be described by expectation values [296]. In this case we choose the second moment (the variance) of the time derivative of the strain field—the energy density [296], $\mathcal{E}_{\rm gw}$. To make this a more immediately meaningful value, we want to compare it to all the other contributions to the energy density of the universe, so we normalize by $\rho_c c^2$, the energy density required to make the universe flat[7].

[6] If two pulsars have timing residuals that are 50% correlated, this implies that 25% of the variability in one pulsar's residuals can be explained by the variability in the other due to their shared response to a gravitational-wave signal. Consequently, knowing the timing of the residuals in one pulsar provides partial information about the timing in the other. Specifically, for two series A and B, the correlation implies that the optimal linear relationship minimizes the residual variance $\langle (A - \alpha B)^2\rangle$, where α is determined by the covariance between A and B. This relationship can be expressed as

$$\langle (A - \alpha B)^2\rangle = 0.5\langle A^2\rangle,$$

highlighting that only a fraction of A's variance is reduced by accounting for its correlation with B.

[7] Cosmology research has found that the universe is flat [47, 52]. So we are finding the contribution of gravitational-waves to the total energy in the universe

Then the total energy density due to gravitational radiation $\mathcal{E}_{\rm GW}$ is then

$$\mathcal{E}_{\rm GW} := \rho_c c^2 \int_0^\infty \Omega_{\rm GW}(f) d\ln(f) = \rho_c c^2 \int_0^\infty \Omega_{\rm GW}(f) \frac{df}{f} \,, \tag{9.17}$$

where we have introduced $\Omega_{\rm GW}(f)$, the dimensionless energy density per logarithmic frequency bin.

If we then take the universe to be isotropic and homogeneous[8] then it should be true that the total energy in gravitational waves in the Universe should be be equal to the sum of densities radiated at each redshift. Let f_r be the gravitational-wave frequency as measured in the rest frame of the source and f be the frequency as measured on Earth. They are then related by the redshift $f_r = (1+z)f$. Then, the total present day energy in gravitational radiation [9]

$$\begin{aligned} \mathcal{E}_{\rm GW} &= \int_0^\infty \int_0^\infty N(z) \frac{dE_{\rm GW}}{df} \frac{f_r}{f_r} df\, dz \,, \\ &= \int_0^\infty \int_0^\infty N(z) \frac{1}{1+z} f_r \frac{dE_{\rm GW}}{df_r} dz \frac{df}{f} \,. \end{aligned} \tag{9.18}$$

Then, we can equate Eq 9.18 to Eq 9.17 to get

$$\rho_c c^2 \Omega_{\rm GW}(f) = \int_0^\infty N(z) f_r \frac{dE_{\rm GW}}{df_r} dz \,. \tag{9.19}$$

If we then assume that our SGWB is made from inspiraling binaries in quasi-circular orbits, as we did in Chap 4, we get

$$\frac{\mathrm{d}E_{\rm GW}}{df_r} = \frac{\pi}{3} \frac{1}{G} \frac{(G\mathcal{M})^{\frac{5}{3}}}{(\pi f_r)^{\frac{1}{3}}} \,, \tag{9.20}$$

which we can derive using only the potential energy of the system, Kepler's laws, their derivatives, and knowing nothing about gravitational-waves other than that the gravitational radiation has twice the orbital frequency [17].

Inserting eq. 9.20 into eq. 9.19 gives us

$$\Omega_{\rm GW}(f) = \frac{\pi^{\frac{2}{3}} (G\,\mathcal{M})^{\frac{5}{3}}}{3\,G} \frac{1}{\rho_c c^2} f^{\frac{2}{3}} \int \frac{N(z)}{(1+z)^{\frac{1}{3}}} dz \,, \tag{9.21}$$

[8]Isotropic, meaning the same in all directions. Homogeneous, meaning the same at all points in space.

[9]which must be modified to account for the redshift taken to reach Earth

giving us the important relationship, $\Omega_{\rm GW}(f) \propto f^{\frac{2}{3}}$. Since pulsar timing arrays measure timing residuals[10] this power is generally instead given as the timing-residual cross-power spectral density[11]

$$S_{ij}(f) = \Gamma_{ij}\frac{A^2_{\rm GWB}}{12\pi^2}\left(\frac{f}{f_{\rm yr}}\right)^{-\gamma} f_{\rm yr}^{-3}\,, \tag{9.22}$$

between pulsars a and b, where $A_{\rm GWB}$ is the amplitude of the SGWB and our spectral index $\gamma = \frac{13}{3}$ [66]. Γ_{ij} is the "overlap reduction function" (AKA our Hellings-Downs function), see equation 9.15 which encodes the expected correlation between the timing residuals from pulsar i and j. $f_{\rm yr} = 1/{\rm yr}$.

9.5 What has been measured so far

Thus far, pulsar timing arrays have found strong evidence of a shared "common process" (essentially equation 9.22 but with the Hellings-Downs correlations $\Gamma_{ij} = \delta_{ij}$). This means that there is support for a shared noise-process in all pulsars that accelerates or delays pulsar timing residuals [50, 66].

Gravitational-waves in the nanohertz frequency band $((10^{-8.75},\ 10^{-7.5})\,{\rm Hz})$ have yet to be conclusively detected by finding these Hellings-Downs correlations, although strong evidence has been found by multiple pulsar timing array collaborations [50, 60, 290, 382].

[10] A timing residual is the offset between the actual arrival of a pulse and its predicted arrival from some deterministic model. There are a few potential contributions to these residuals, including intrinsic noise in the pulsars, detector noise, or gravitational waves (among others) [333].

[11] which essentially is the variance of timing residuals, similar in principal to the PSD in Part I

Chapter 10

LIKELIHOOD REWEIGHTING WITH PULSAR TIMING ARRAYS

Sophie Hourihane, Patrick Meyers, Aaron Johnson, Katerina Chatziioannou, and Michele Vallisneri. Accurate characterization of the stochastic gravitational-wave background with pulsar timing arrays by likelihood reweighting. *Phys. Rev. D*, 107(8):084045, 2023. doi: 10.1103/PhysRevD.107.084045. Printed here as Chapter 10. **SH** conceptualized the project, authored the text, performed all analyses except for P-P tests, and all figures.

10.1 Abstract

An isotropic stochastic background of nanohertz gravitational waves creates excess residual power in pulsar-timing-array datasets, with characteristic interpulsar correlations described by the Hellings-Downs function. These correlations appear as non-diagonal terms in the noise covariance matrix, which must be inverted to obtain the pulsar-timing-array likelihood. Searches for the stochastic background, which require many likelihood evaluations, are therefore quite computationally expensive. We propose a more efficient method: we first compute approximate posteriors by ignoring cross correlations, and then reweight them to exact posteriors via importance sampling. We show that this technique results in accurate posteriors and marginal likelihood ratios, because the approximate and exact posteriors are similar, which makes reweighting especially accurate. The Bayes ratio between the marginal likelihoods of the exact and approximate models, commonly used as a detection statistic, is also estimated reliably by our method, up to ratios of at least 10^6.

10.2 Introduction

The nanohertz stochastic gravitational-wave (GW) background can be detected through the induced delay on the times of arrival of pulses from millisecond pulsars [145, 187, 333]. Recent evidence that the datasets collected by the three major pulsar-timing-array (PTA) consortia all include excess timing noise of common amplitude and spectral shape [60, 66, 113, 172] suggests that we might be getting closer to detection [65, 274]. However, since such common-spectrum noise may arise from a non-GW astrophysical or terrestrial source [172, 392] (even if this seems unlikely in current data [66, 173]), a GW detection claim needs to wait for the finding that the excess noise is correlated across pulsars with the characteristic angular pattern known as the Hellings-Downs curve [187].

In PTA data analysis, timing noise is represented as a Gaussian process with covariance matrix $C_{ai\,bj}$, where a, b range over pulsars and i, j over timing measurements (or equivalently frequency components). For common-spectrum uncorrelated noise, the matrix factorizes as $C_{ij}\delta_{ab}$; for an isotropic GW background, it is given by $C_{ij}\Gamma_{ab}$, with $\Gamma_{ab} = \Gamma(\theta_{ab})$ the Hellings-Downs correlation coefficient, a function of the angular separation θ_{ab} between pairs of pulsars. The PTA data model includes several other stochastic components, but GW detection is usually formulated by comparing a common process (CP) model that includes common-spectrum uncorrelated noise and an "HD" model that includes common-spectrum Hellings-Downs–correlated noise.[1] By contrast, information about the GW amplitude and spectral shape is carried primarily by the autocorrelation terms (the $C_{ai\,bj}$ elements with $a = b$).

Parameter estimation and model selection for the CP and HD models are both typically handled through stochastic sampling, which requires repeated evaluations of the data likelihood. Since the CP excess-noise covariance matrix factorizes across pulsars but the HD matrix does not [337], the likelihood is significantly slower to compute for the latter model (e.g., a factor of $\sim$ 25 for the NANOGrav 12.5yr dataset, which will only grow larger as more pulsars are observed). The number of likelihood evaluations is magnified by the thinning of sample chains (typically by $N_t \sim \mathcal{O}(10^3)$) and by the use of parallel tempering schemes (typically by $N_c \sim \mathcal{O}(10)$ temperatures) which require many likelihood evaluations per CP posterior sample. The overall cost can be prohibitive for the HD model, particularly when multiple background analyses (e.g., "sky scrambles" [123, 332] and "phase shifts" [332]) are required to estimate the significance of a result.

Methods to optimize PTA search strategies in both data acquisition and modeling have been studied extensively. On the data acquisition side, studies found the most impactful observing cadences and radio frequency bands for

[1] In the NANOGrav 12.5yr stochastic background analysis that initially reported the evidence for a common process [66], the CP and HD models are labeled model 2A and 3A, respectively.

detecting a GW background (GWB) [216, 217, 219]. On the modeling side, improvements in computational efficiency have been made by using Fourier basis methods [220, 358] to characterize red-noise processes, as opposed to dense covariance matrix approaches [359]. More recently, the factorized likelihood approach reduces the wall clock time needed to evaluate a CP-only model by a factor proportional to the number of pulsars [337], and Hamiltonian Monte Carlo methods have been implemented to improve sampling efficiency [164].

In this study we propose an approach that further mitigates computational cost of producing posterior samples for the HD model in terms of both CPU and wall clock time. Rather than exploring the HD model stochastically, we reuse parameter-estimation results for the inexpensive CP model and "reweight" them to obtain posteriors and marginal likelihoods under the HD model. Specifically, a thinned set of CP-model samples yields a set of weighted HD-model samples, with weights equal to the ratios of the HD and CP likelihoods. The computational gains are realized by performing only one HD likelihood evaluation per HD posterior sample, and parallelizing the calculation of weights.

The general reweighting formalism can be applied to any combination of models, though convergence and low sampling error depend on stochastic chains for the original posterior having a sufficient number of samples in the support of the target posterior. This is the case for the HD and CP posteriors, since both are dominated by single-pulsar autocorrelation terms. Additionally, the two models share the same parameters and corresponding priors. In this paper we apply the reweighting formalism to simulated PTA data, and compare posteriors and marginal likelihoods obtained by reweighting and by brute-force sampling. We find that (i) the posteriors recovered through reweighting are statistically unbiased, and that (ii) the HD vs. CP Bayes factors (the ratios of marginal likelihoods) agree with the "hypermodel" method typically used in PTA analyses [185] to within 10% uncertainty for Bayes factors $\in [10^{-3}, 10^{7}]$.

The rest of the paper is organized as follows. In Sec. 10.3 we introduce the general reweighting formalism following [266]. In Sec. 10.4 we describe the HD and CP models in more detail. In Sec. 10.5 we present results from simulated data that validate the reweighting approach. In Sec. 10.6 we conclude by discussing the application of our method and its computational gains.

10.3 Posterior reweighting

Samples distributed according to one posterior distribution can, under some circumstances, be reweighted to estimate a second posterior distribution; this is a form of importance sampling. In this section, we present the general methodology behind this posterior reweighting following Ref. [266], and describe how it can be used to also estimate the marginal likelihood of a model and the Bayes factor between models.

The posterior distribution, $p(\theta|d, T)$ for a target model T with parameters θ given data d can be written explicitly in terms of the Bayes theorem,

$$p(\theta|d, T) = \frac{\mathcal{L}(d|\theta, T)\pi(\theta|T)}{\mathcal{Z}_T}\,, \tag{10.1}$$

where $\mathcal{L}(d|\theta, T)$ is the likelihood, $\pi(\theta|T)$ is the prior, and $\mathcal{Z}_T$ is the marginal likelihood (also known as evidence, though we do not use this term here). We rewrite this target posterior distribution in terms of the likelihood and prior for another "approximate" model A,

$$\begin{aligned} p(\theta|d, T) &= \frac{\mathcal{L}(d|\theta, A)\frac{\mathcal{L}(d|\theta,T)}{\mathcal{L}(d|\theta,A)}\pi(\theta|A)\frac{\pi(\theta|T)}{\pi(\theta|A)}}{\mathcal{Z}_T} && (10.2)\\ &= w_{\mathcal{L}}(d|\theta)w_\pi(\theta)\frac{\mathcal{L}(d|\theta, A)\pi(\theta|A)}{\mathcal{Z}_T}\,. && (10.3) \end{aligned}$$

In the last line we have introduced weights given by the ratio of the likelihoods and priors of the two models

$$\begin{aligned} w_{\mathcal{L}}(d|\theta) &= \frac{\mathcal{L}(d|\theta, T)}{\mathcal{L}(d|\theta A)}\,, && (10.4)\\ w_\pi(\theta) &= \frac{\pi(\theta|T)}{\pi(\theta|A)}; && (10.5) \end{aligned}$$

we can also combine the weights to get

$$w(d|\theta) = w_{\mathcal{L}}(d|\theta)w_\pi(\theta)\,. \tag{10.6}$$

Given N_s posterior samples $\theta_s \sim p(\theta|d, A)$ for model A, we can resample them with weights $w(d|\theta_s)$ to obtain a posterior sampling of model T; the marginal likelihood $\mathcal{Z}_T$ can also be estimated as

$$\begin{aligned} \mathcal{Z}_T &= \int d\theta\, \mathcal{L}(d|\theta, T)\pi(\theta|T) && (10.7)\\ &= \mathcal{Z}_A \int d\theta\, w_{\mathcal{L}}(d|\theta)w_\pi(\theta)p(\theta|d, A)\,. && (10.8) \end{aligned}$$

The integral in Eq. (10.8) can be approximated with Monte Carlo integration:

$$\mathcal{Z}_T \approx \frac{\mathcal{Z}_A}{N_s} \sum_{s=1}^{N_s} w_{\mathcal{L}}(d|\theta_s) w_\pi(\theta_s) = \mathcal{Z}_A \bar{w}\,, \tag{10.9}$$

where $\bar{w}$ is the mean of the weights, $w(d|\theta)$. If we are interested in model selection between the approximate and target models, the Bayes factor between them is then simply

$$\mathcal{B}_A^T = \bar{w}\,. \tag{10.10}$$

Though the reweighting procedure is mathematically exact, it is subject to sampling errors, especially if the approximate and target posteriors are too disjoint. We quantify sampling error with the "effective number of samples" n_{eff}—the approximate number of samples drawn independently from the target posterior that would approximate $\mathcal{Z}_T$ as accurately as the reweighting estimate (10.9). Reference [156] estimates n_{eff} as

$$n_{\text{eff}} \approx \frac{\left[\sum_s w_{\mathcal{L}}(d|\theta_s) w_\pi(\theta_s)\right]^2}{\sum_s \left[w_{\mathcal{L}}(d|\theta_s) w_\pi(\theta_s)\right]^2} = \frac{N_s}{1 + \left(\frac{\sigma_w}{\bar{w}}\right)^2}\,, \tag{10.11}$$

where σ_w is the standard deviation of the weights. We also define the efficiency

$$\mathcal{E} \equiv \frac{n_{\text{eff}}}{N_s}\,. \tag{10.12}$$

It follows from Eq. (10.10) that the error $\sigma_{\mathcal{B}}$ on the mean $\mathcal{B}_A^T$ is

$$\sigma_{\mathcal{B}} = \frac{\sigma_w}{\sqrt{n_{\text{eff}}}} = \frac{\sigma_w}{\sqrt{\mathcal{E}\, N_s}}\,. \tag{10.13}$$

If we represent the target posterior by a set of equal-weight samples by performing a weighted redraw from the approximate distribution, then Eq. (10.11) makes intuitive sense. It implies that a few samples with high weights (relative to $\bar{w}$), will result in the same sample being drawn many times and lead to comparatively lower n_{eff}. Equivalently, such high individual weights (relative to $\bar{w}$) increase σ_w and thus decrease n_{eff}. In the limit of vanishing variance, $n_{\text{eff}} \to N_s$, while as variance grows $n_{\text{eff}} \to 0$.

We can also use the weights to estimate the statistical distance between the approximate and target posteriors in the form of the Kullback-Leibler (KL) divergence. That is,

$$\mathcal{D}_{KL}(A||\,T) \equiv \int d\theta\, p_T(\theta) \ln \frac{p_A(\theta)}{p_T(\theta)}\,, \tag{10.14}$$

$$\approx \sum_{\theta \sim p_A} \frac{1}{N_s} \ln \frac{p_A(\theta)}{p_T(\theta)}, \tag{10.15}$$

$$= \ln(\bar{w}) - \overline{\ln(w)}, \tag{10.16}$$

where we have written the posteriors $p_K(\theta) \equiv p(\theta|d, K)$ for $K \in \{A, T\}$ and $\overline{\ln(w)}$ is the average of the log of the weights. This equation can be used in combination with Eq. (10.11) to provide guidance when reweighting leads to low efficiency.

The main reason for low efficiency is that a region of low posterior for the approximate model overlaps with a region of high posterior for the target model. Samples in this region get very high weights and can lead to poor reconstruction of the target posterior. Understanding when and where this can occur is an area of active research, and has led to other forms of importance sampling [366].

10.4 The pulsar-timing-array models for stochastic gravitational waves

In this section we discuss the statistical framework used to detect a stochastic GW background with an array of regularly timed millisecond pulsars. We first introduce a Gaussian likelihood that includes the full interpulsar correlations induced by GWs (the Hellings-Downs model). We then introduce a secondary model that ignores interpulsar correlations and includes GWs as a common (but uncorrelated) power law spectrum in each pulsar's residuals (a common process model). We claim evidence for GWs when a dataset significantly favors HD over CP.

These two models contain the same parameters and priors; furthermore, the posterior distributions of model parameters are not affected strongly by the inclusion of interpulsar correlations. This makes the CP likelihood a good approximate distribution for the HD likelihood. As we discuss below, the CP model is significantly faster to evaluate, and so we will use CP as our approximate likelihood, $\mathcal{L}_A$, while the HD model will be the target $\mathcal{L}_T$. An added bonus of our choice of these models is that, in implementing the reweighting scheme discussed above to speed up computation of the HD posterior, we naturally also calculate Bayes factors that can be used as a GW detection statistic.

Pulsar-timing-array likelihood

A detailed presentation of the PTA likelihood derivation can be found in Refs. [64, 220, 332, 333, 357, 358]; in this section we describe only the relevant details. A reader familiar with PTA analyses can skip to Sec. 10.4.

Pulse arrival times [time(s) of arrival (TOA)] are affected by both deterministic and stochastic processes. The deterministic contribution (described more fully in [332]) contains terms relating to the motion of the pulsar, such as sky location, rotation period, etc., as well as individually resolvable GW sources such as continuous waves. An initial solution for the timing model is subtracted from the measured TOA, leaving behind the fit residuals $\delta\boldsymbol{t}$. The uncertainties on this timing model are described by a Taylor expansion in timing parameters $\boldsymbol{\epsilon}$ with partial derivatives (comprising the design matrix) $\boldsymbol{M}$ evaluated at the initial timing solution.

The stochastic contribution to the TOA is due to a combination of the intrinsic low-frequency spin noise (or "red noise") of individual pulsars (IRN) and a common, stochastic process induced by a GW background. We model both as stationary zero-mean Gaussian random processes with Fourier vector bases. It follows that the Fourier coefficients $\boldsymbol{a}$ (the basis weights) are described entirely by their covariance $\boldsymbol{\phi}|_{\boldsymbol{\eta}} = \langle \boldsymbol{a}_{ai}\boldsymbol{a}_{bj}\rangle$. Here indices a, b index pulsars, i, j index frequencies, brackets indicate the ensemble average, and the $\boldsymbol{\eta}$ are the "hyperparameters" associated with the distribution of $\boldsymbol{a}$.

With both the deterministic and stochastic contributions to the noise modeled, the timing residuals $\boldsymbol{r}$ are

$$\boldsymbol{r} = \delta\boldsymbol{t} - \boldsymbol{M}\boldsymbol{\epsilon} - \boldsymbol{F}\boldsymbol{a} = \delta\boldsymbol{t} - \boldsymbol{T}\boldsymbol{b}\,, \tag{10.17}$$

where the matrix $\boldsymbol{F}$ collects the Fourier basis vectors evaluated at the TOA and where we have introduced

$$\boldsymbol{T} \equiv [\boldsymbol{M}\,\boldsymbol{F}]\,, \quad \boldsymbol{b} \equiv \begin{bmatrix} \boldsymbol{\epsilon} \\ \boldsymbol{a} \end{bmatrix}\,, \tag{10.18}$$

for ease of notation. The residuals $\boldsymbol{r}$ should now be white and Gaussian, with a covariance matrix $\boldsymbol{N}$ that describes the uncertainty associated with each TOA observation. The likelihood is then

$$\mathcal{L}(\delta\boldsymbol{t}|\boldsymbol{b}) = \frac{\exp\left(-\frac{1}{2}\boldsymbol{r}^T\boldsymbol{N}^{-1}\boldsymbol{r}\right)}{\sqrt{\det 2\pi\boldsymbol{N}}}\,. \tag{10.19}$$

We complement the likelihood with the Gaussian-process prior for the Fourier components,

$$\pi(\boldsymbol{a}|\boldsymbol{\eta}) = \frac{\exp(-\frac{1}{2}\boldsymbol{a}^T\boldsymbol{\phi}|_{\boldsymbol{\eta}}^{-1}\boldsymbol{a})}{\sqrt{\det 2\pi\boldsymbol{\phi}|_{\boldsymbol{\eta}}}}\,. \tag{10.20}$$

The Gaussian form of the likelihood and prior means that we can marginalize analytically over the $\boldsymbol{a}$, leaving only the hyperparameters $\boldsymbol{\eta}$. A similar choice is made for the timing model correction prior $\pi(\boldsymbol{\epsilon})$ [332]. The marginalized likelihood is then

$$\mathcal{L}(\delta\boldsymbol{t}|\boldsymbol{\eta}) = \int d\boldsymbol{b}\;\mathcal{L}(\delta\boldsymbol{t}|\boldsymbol{b})\;\pi(\boldsymbol{a}|\boldsymbol{\eta})\;\pi(\boldsymbol{\epsilon})\,, \tag{10.21}$$

$$\propto \frac{\exp\left(-\frac{1}{2}\delta\boldsymbol{t}^T\boldsymbol{C}^{-1}\delta\boldsymbol{t}\right)}{\sqrt{\det\left(2\pi\boldsymbol{C}\right)}}\,, \tag{10.22}$$

where $\boldsymbol{C} = \boldsymbol{N} + \boldsymbol{T}\boldsymbol{B}\boldsymbol{T}^{\boldsymbol{T}}$ is the covariance kernel, and

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{\infty} & 0 \\ 0 & \boldsymbol{\phi} \end{bmatrix}. \tag{10.23}$$

Here $\boldsymbol{\infty}$ represents a formal limit of covariance for a uniform unbounded prior on $\boldsymbol{\epsilon}$.

Pulsar-timing-array stochastic models

We model both the IRN and the GWB as power laws in the frequency domain.[2] The model hyperparameters are then $\boldsymbol{\eta} = (A^a, \gamma^a, A_{\rm GW}, \gamma_{\rm GW})$ where γ^a, $\gamma_{\rm GW}$ and A^a, $A_{\rm GW}$ are the negative spectral indices and amplitudes of the IRN and GW power laws respectively. We split $\boldsymbol{\phi}$ into its two contributions; one from the IRN and the other the common GWB

$$\boldsymbol{\phi} = \boldsymbol{\phi}^{\rm IRN} + \boldsymbol{\phi}^{\rm GW}\,. \tag{10.24}$$

By stationarity, both the IRN and the GWB are uncorrelated between frequencies. Therefore ϕ will contain no cross-frequency terms, and $\boldsymbol{\phi}_{ai,bj} \propto \delta_{ij}$.

By definition, the IRN is an independent process in each pulsar:

$$\boldsymbol{\phi}|_{\boldsymbol{\eta}\;(ai),(bj)}^{\rm IRN} = \kappa_i(\boldsymbol{\eta}_a)\delta_{ab}\delta_{ij}\,. \tag{10.25}$$

[2]Power laws are not the only choice for the distribution of Fourier coefficients. Other choices include (but are not limited to) a free spectral model with independent densities for each Fourier frequency, and a broken power law [66, 308, 334].

The IRN power $\kappa_i(\boldsymbol{\eta}_a)$ in frequency bin i for pulsar a is modeled as the power law

$$\kappa_i(\boldsymbol{\eta}_a) = \kappa_i(A_a, \gamma_a) = \frac{A_a^2}{12\pi^2}\frac{1}{T}\left(\frac{\nu_i}{\mathrm{yr}^{-1}}\right)^{-\gamma_a} \mathrm{yr}^2 \,, \tag{10.26}$$

where T is the total observation time and ν_i is the frequency associated with bin i.

In contrast to the IRN, the GW background is correlated between pulsars:

$$\boldsymbol{\phi}|^{\mathrm{GW}}_{\boldsymbol{\eta}\ (ai),(bj)} = \Gamma_{ab}\kappa_i(\boldsymbol{\eta}_{\mathrm{GW}})\delta_{ij} \,. \tag{10.27}$$

Here κ_i is again given by Eq. (10.26), except that every pulsar has the same amplitude A_{GW} and spectral index γ_{GW}. The function Γ_{ab} describes GW correlations between pulsars a and b and is known as the Hellings-Downs curve [[187], Eq. (5)].

The $a = b$ components of Eq. (10.27) represent the excess-noise power induced by the GWB in each pulsar. Half of this power is caused by the "Earth term" in the pulsar GW response, and contributes to interpulsar correlations; the other half is caused by the "pulsar term" and is statistically uncorrelated among pulsars. The first indications of a GWB in PTA data will appear through these diagonal self-correlations [65, 297], so they could be detected using the CP model as well as the HD model. However, evidence for CP could also be caused by physical effects such as the solar wind [343] or by model misspecification, such as incorrect priors [392] or poor IRN models [173]. Other mechanisms can induce interpulsar correlations that are inconsistent with the Hellings-Downs curve, such as clock errors with monopolar correlations [343, 344] or Solar System ephemeris errors with dipolar correlations [91, 171, 295, 356]. Thus, the detection of HD correlations through the off-diagonal terms of ϕ_{GW} is considered the decisive factor in claiming a GWB detection, and the CP vs. HD Bayes factor is used as a GWB detection statistic [60, 66, 113, 173].

Implementation and computational considerations

The standard PTA likelihood (10.22) requires the inverse noise covariance matrix $\boldsymbol{C}^{-1}$ and therefore the inverse of $\boldsymbol{\phi}|_{\boldsymbol{\eta}}$. Although the PTA analysis software, such as Enterprise [155], is optimized to speed up the likelihood evaluation, inversion becomes the most expensive computation when $\boldsymbol{\phi}|_{\boldsymbol{\eta}}$ is not pulsar diagonal. For instance, for the NANOGrav "12.5yr" dataset each

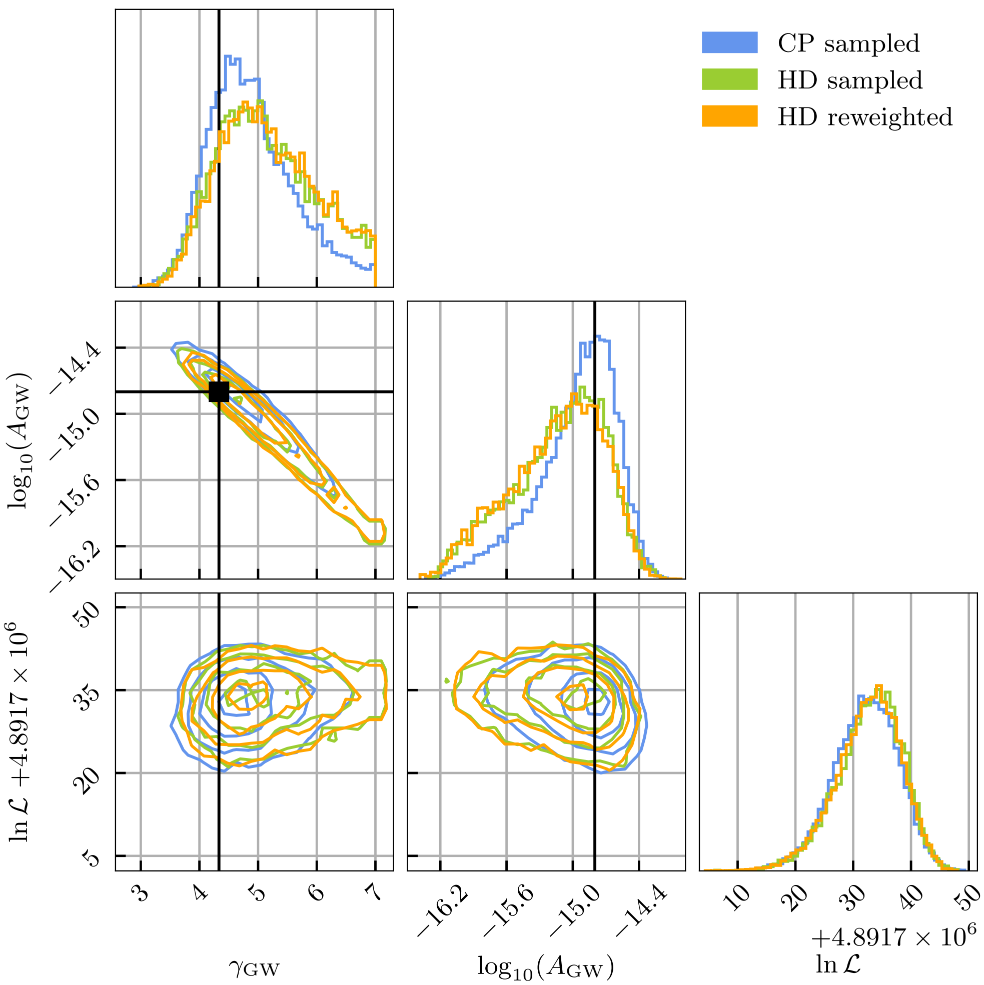

Figure 10.1: Posteriors for $\gamma_{\rm GW}$ and $\log_{10} A_{\rm GW}$ and $\ln\mathcal{L}$ distribution for simulated PTA data with a $\log_{10}(A_{\rm GW}) = -14.8$ GWB. We show histograms for direct sampling of CP (blue), HD (green), and for CP-to-HD reweighting (orange). Black lines indicate the injected values. For this plot we selected one of our simulations with the most visually different CP and HD GW posteriors. Even so, the direct-sampling and reweighted HD posteriors are almost identical. The reweighted posterior is well sampled, with 51% efficiency.

HD-model likelihood is $\sim$ 25 slower than the corresponding CP likelihood. This factor applies to 45 pulsars over a 12.9 year dataset and will increase with the number of pulsars.

The current workhorse method to compute $\mathcal{B}^{\mathrm{HD}}_{\mathrm{CP}}$ is a hypermodel Markov chain Monte Carlo sampler [154, 185, 335]. In such an analysis, a discrete metaparameter tracks the current model (HD or CP) while the sampler jumps between them. The final Bayes factor is the number of samples in the HD model divided by the number of samples in the CP model. The two posteriors are also selected by the value of the metaparameter. As the evidence for a GWB becomes stronger, the HD model will be sampled more often than the CP model, slowing down calculations further.[3]

Despite the difference in likelihood functions and computation time, the posteriors for CP and HD are generally quite similar. Figure 10.1 displays the marginalized one- and two-dimensional γ_{GW} and A_{GW} posteriors and the $\ln \mathcal{L}$ distributions for CP (blue) and HD (green), as recovered by hypermodel sampling. The similarity between the posteriors and the $\sim$25x likelihood speedup suggest that this problem is well suited for the reweighting method introduced in Sec. 10.3.[4] The HD posterior created by reweighting the CP posterior is plotted in orange and is almost identical to the direct-sampling HD posterior. The efficiency of the reweighting method as posterior differences is discussed in the next section.

10.5 Demonstration of the method

To show that we can safely reweight CP to HD, we simulate PTA datasets containing GWBs with different amplitudes, and demonstrate that reweighting yields unbiased Bayes factors and posteriors.

Bayes factors

To test $\mathcal{B}^{\mathrm{HD}}_{\mathrm{CP}}$ recovery, we simulate 100 datasets for 45 pulsars over 12.9 years, using maximum-likelihood red-noise hyperparameters from the 12.5yr NANOGrav

[3] A constant added to the CP log-likelihood can mitigate this particular issue and result in a comparable number of samples in each model. That constant should be close to the Bayes factor, which is unknown *a priori* in real data. In our study we estimated this constant by using the likelihood ratio between the CP and HD models evaluated at the injected parameters. This ensured that both models contained enough samples particularly in high Bayes factor regimes.

[4] Here we use identical priors between the target and the original distribution, meaning the prior weights $w_\pi(\theta) = 1$.

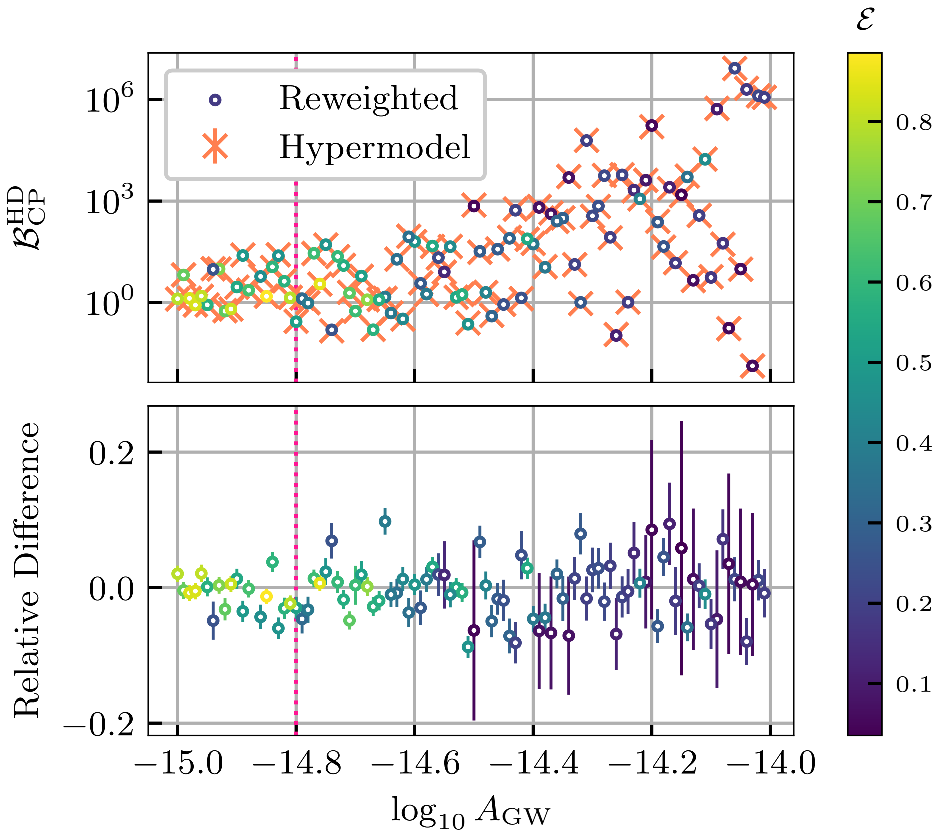

Figure 10.2: Top: $\mathcal{B}_{\rm CP}^{\rm HD}$ vs simulated GWB amplitude. Bayes factors recovered via reweighting, Eq. (10.10), are colored by their efficiency $\mathcal{E}$, Eq. (10.12). Bayes factors recovered via the hypermodel are plotted as coral Xs. The hypermodel error is calculated with a bootstrap method described in [184] whereas the reweighting error is estimated with Eq. (10.13), although both errors are too small to see. Bottom: relative difference in the hypermodel and recovered Bayes factors, again colored by efficiency. The error bars are propagated from the hypermodel and reweighting errors above. As the GWB amplitude increases, the efficiency decreases due to the distribution of the weights broadening as in Eq. (10.11). The relative difference between these Bayes factors is usually small, typically $-0.5 \pm 4\%$, but can be as large as 10%. A 10% difference in $\mathcal{B}_{\rm CP}^{\rm HD}$ is not large enough to change a detection conclusion to a nondetection conclusion or vice versa and therefore we can consider the difference small. For instance, a Bayes factor of 100 would lead to the same qualitative conclusion as a Bayes factor of 110. The pink vertical line in both plots is $\log_{10} A_{GW} = -14.8$, the posterior plotted in Fig. 10.1 to demonstrate that this posterior is typical.

dataset [65, 254].[5] Each simulation includes a power-law GWB with $\log_{10} A_{\mathrm{GW}}$ varying uniformly between -15 and -14. We set γ_{GW} to 13/3, the theoretical value for a GW background from supermassive black-hole binaries [272]. For each simulated dataset, we obtain a thinned set of CP posterior samples using PTMCMCSampler [154]. We reweight the CP posterior sample to the HD model and calculate $\mathcal{B}^{\mathrm{HD}}_{\mathrm{CP}}$ following Eq. (10.10). To verify the accuracy of these reweighted $\mathcal{B}^{\mathrm{HD}}_{\mathrm{CP}}$s, we obtain an independent estimate from hypermodel runs on the same simulations. We compare the reweighted and hypermodel Bayes factors in Fig. 10.2, finding them in excellent agreement. The top panel shows the $\mathcal{B}^{\mathrm{HD}}_{\mathrm{CP}}$ estimates plotted against $\log_{10} A_{\mathrm{GW}}$; the bottom panel shows the relative difference of the $\mathcal{B}^{\mathrm{HD}}_{\mathrm{CP}}$ estimates (reweighted minus hypermodel, divided by their average). Marker colors encode reweighting efficiency. The mean relative difference is $-0.5 \pm 4\%$, so we observe no systematic effect. The maximum relative difference is 10%, small enough that it could not affect a GWB detection claim. Error bars are computed by combining (in quadrature) reweighted Bayes factor errors from Eq. (10.13) and hypermodel Bayes factor errors from the bootstrap method of [184]. Bayes factor differences are not strongly correlated with the injected GW amplitude or the Bayes factor, although the difference uncertainties are inversely correlated with efficiency [see Eq. (10.13)].

Figure 10.2 shows also that as we increase the simulated amplitude, the sampling efficiency tends to decrease. This is expected; as the amplitude of the GWB increases, the off-diagonal terms in Eq. (10.27) become more significant. The likelihood can then change between the two models significantly, which affects $\bar{w}$, and can even be maximized in different parts of parameter space. Such conditions can lead to a large spread in the weights as some points get heavily upweighted and others get downweighted. From Eq. (10.11), a large spread in the weights means that n_{eff} will decrease, and more samples from the CP distribution will be needed in order to faithfully represent the HD posterior and calculate $\mathcal{B}^{\mathrm{HD}}_{\mathrm{CP}}$ accurately: see Eq. (10.13). In our simulated datasets, however, the recovered Bayes factor remains within 10% of that calculated with the hypermodel even in regions where $\mathcal{B}^{\mathrm{HD}}_{\mathrm{CP}} > 10^6$.

In order to study the relation between the model posterior similarity and the efficiency of the reweighting procedure, we compute the Kullback–Leibler diver-

[5] The NANOGrav 12.5yr dataset is actually 12.9 years in length.

gence [213], which quantifies the difference between two distributions. We plot the relationship between the KL divergence and the efficiency in Fig. 10.3. The upper plot shows total KL divergence [Eq. (10.16)] vs. efficiency [Eq. (10.12)] for the CP and HD posteriors. As the KL divergence increases, the posteriors become more distinct and the sampling efficiency decreases. The bottom plot shows the fractional contributions of different model parameters to the total KL divergence.[6] We split the 92 parameters into four sets: the IRN amplitudes and spectral indices (pink and red respectively) and the GW background amplitude and spectral index (blue and gold respectively). We compute the partial KL divergence of the CP and HD marginalized posteriors for each parameter and sum those of the IRN parameters. The fractional contribution is then obtained by dividing those partial KL divergences by the total. The set of all red-noise parameters contributes more to the total divergence than do the GWB parameters individually. The set of all IRN amplitude posteriors is the major contributor to the divergence ($55 \pm 11\%$), followed by the set of all IRN spectral indices ($27 \pm 8\%$); the contribution from $A_{\rm GW}$ and $\gamma_{\rm GW}$ are roughly equivalent at percent level, $9 \pm 8\%$ each.

Posterior recovery

Figure 10.1 offered visual confirmation that the GWB parameter posteriors under the HD model are recovered without bias via reweighting. In this section we confirm these initial findings through a more extensive percentile-percentile (P-P) test [170]. We generate 100 simulations similar to those described in Sec. 10.5, except that each simulated parameter is drawn from its analysis prior, as required to achieve Bayesian coverage. The priors for the spectral indices are $\gamma_{\rm GW}, \gamma^a \in \mathrm{U}\,[2, 6]$, and the priors for the amplitudes are $\log_{10} A_{\rm GW} \in \mathrm{U}\,[-15, -12]$ and $\log_{10} A^a \in \mathrm{U}[-16, -14]$. We recover CP posteriors from these simulations with direct sampling, and then reweight and resample those posteriors to the HD model.

The P-P test is a standard measure of bias in recovered posteriors. Datasets are first simulated by drawing parameters from their priors and adding Gaussian noise. The posterior of each data-set is then sampled. The percentile

[6] The total KL divergences are not directly comparable to the KL divergences of the various marginalized posteriors; the total KL divergence is equal to the sum of the marginalized KL divergences only when parameters are uncorrelated. Although this is not the case in our analysis, the normalized marginal KL divergences still inform us of parameters that most greatly influence the total KL divergence.

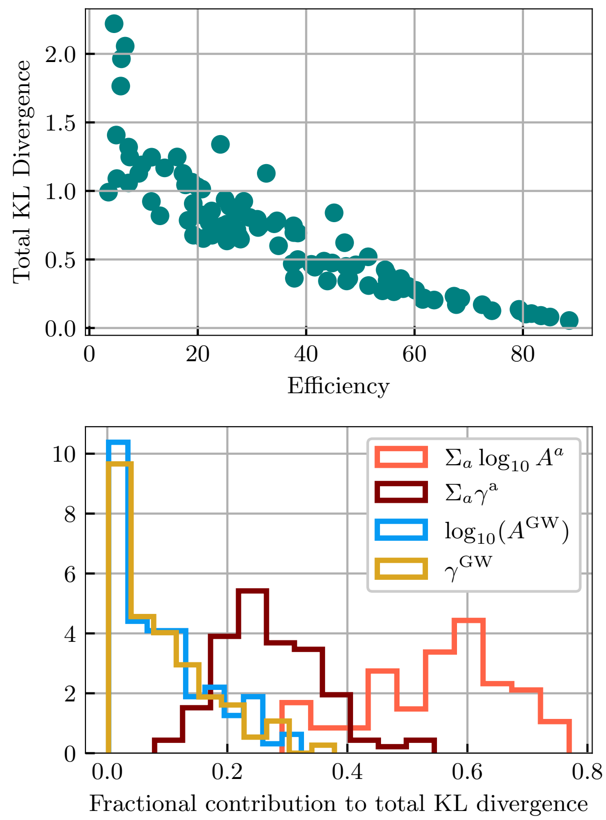

Figure 10.3: Top: total KL divergence, Eq. (10.16), vs efficiency, Eq. (10.12), between the CP and the reweighted HD posterior. As the KL divergence increases, the posteriors become more distinct, and the sampling efficiency decreases. Bottom: fractional contributions to total KL divergence from sets of parameters including all the IRN amplitudes and spectral indices (pink and red, respectively) and the GWB amplitude and spectral index (blue and gold respectively). The set of all IRN parameters contribute more to the total divergence than the GWB parameters individually.

of each of the "true" or injected values is calculated in the marginalized, one dimensional posterior of each parameter. For a set unbiased posteriors, the injected value will be distributed according to each posterior. That is, the percentile of where each injected value lands in a 1-D marginalized posterior will be distributed uniformly between the 0th and 100th percentile, the x-axis of Fig. 10.4. This test of uniformity in posterior space is represented with the cumulative distribution function (CDF) of the posteriors percentile. Since the CDF of a uniform distribution between 0 and 1 (0th and 100th percentile) is a line of slope 1, the P-P plot is usually represented this way. A P-P plot showing a line consistently below (above) the line x = y is indicative of parameter bias of overestimating (underestimating) the parameter value. An S-curve going above (below) then below (above) the diagonal is indicative of a overestimate (underestimate) of the posterior's standard-deviation.

Figure 10.4 shows the corresponding P-P plots. The 92 different parameters (γ^a, A^a for 45 pulsars as well as $\gamma^{\mathrm{GW}}, A^{\mathrm{GW}}$ for the GWB) are plotted in teal. The expected 1-, 2-, and 3-σ confidence intervals are plotted in black. The recovered posteriors agree with expectations; only two lines briefly leave the three-sigma error bars. This suggests that the reweighting method neither over- nor underestimates parameters systematically, as would be the case if some parameters were always above or below the diagonal; nor does it recover incorrect variance, as would be indicated by S-curves around the diagonal.

Bayes factor recovery on extended dataset

To this point, we have demonstrated that likelihood reweighting is a promising tool for recovering accurate Bayes factors and unbiased posteriors in a simulated data-set with 12.9 years of timing data, 45 pulsars, and a range of injected GWB amplitudes. As PTAs continue to collect more data, it becomes natural to ask at what point the reweighting scheme could fail, either by misestimating Bayes factors or by exhibiting low efficiencies. We examine the performance of likelihood reweighting after the addition of additional pulsars and additional observation time to the dataset. We find that even for 80 pulsars and 22 years of data, the efficiency remains above 20% and the errors between the Bayes factor calculated with direct sampling and reweighting are comparable to those in Sec. 10.5.

To create this extended dataset, we simulate realistic pulsars and add ad-

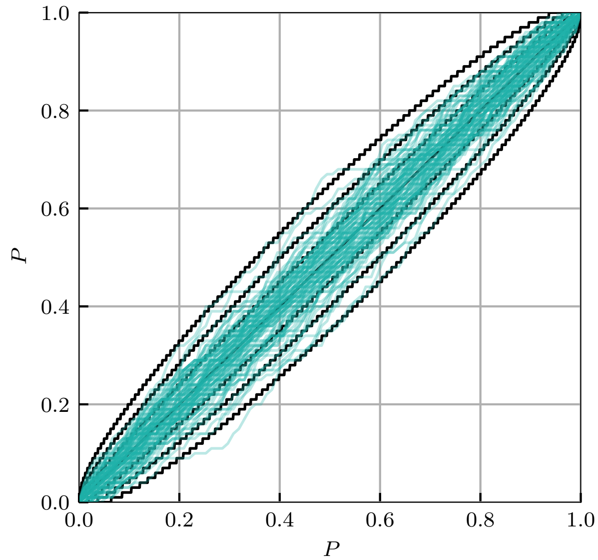

Figure 10.4: P-P plots for all 92 reweighted, HD-model parameters (teal) with the 1-, 2-, and 3-σ standard deviations (black). The y-axis is the percentile of each parameter's injected value in its marginalized posterior. The x-axis is the percentile of the sorted y-axis values. The recovered posteriors are consistent with expectations, suggesting that the posterior recovery is unbiased.

ditional observing time to each pulsar. To create new pulsars, we sample sky locations by fitting existing pulsar locations with a kernel density estimate and sample from it. Each new pulsar is assigned white noise parameters and observing epochs (plus Gaussian scatter) from an existing pulsar. To simulate additional years of data, to each pulsar we add TOAs with Gaussian scatter. The red noise parameters for the new pulsars are drawn from the IRN prior. For existing pulsars, the red-noise amplitudes were set from the maximum-likelihood draw as in Sec. 10.5. The GWB was injected with $A_{\mathrm{GW}} = 1.92 \times 10^{-15}$, the median posterior amplitude of the NANOGrav 12.5yr analysis [66]. In total, we simulated 22 years of data in 80 pulsars; below we present results based subsets of that data.

In Fig. 10.5 we plot $\mathcal{B}^{\mathrm{HD}}_{\mathrm{CP}}$ as a function of the number of pulsars, N, and for different observation durations, T_{obs}. We find that the relative difference between the $\mathcal{B}^{\mathrm{HD}}_{\mathrm{CP}}$ recovered by direct sampling and by reweighting remain within 10% of each other, suggesting that the reweighting scheme remains valid for these extended data-sets. Moreover, we find that while the ratio between the HD and CP likelihood computation times is approximately constant across

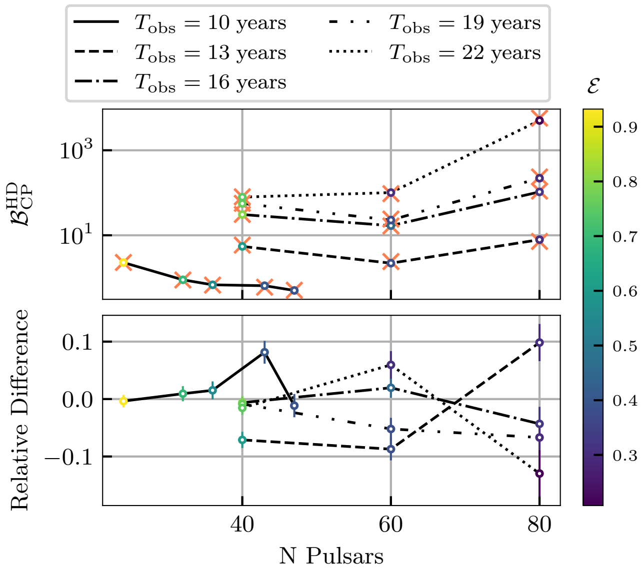

Figure 10.5: Top: $\mathcal{B}^{\rm HD}_{\rm CP}$ vs number of pulsars for sets of fixed observation times between 10 and 22 years, $T_{\rm obs}$. Bayes factors recovered via reweighting, Eq. (10.10), are colored by their efficiency, $\mathcal{E}$, Eq. (10.12). Bayes factors recovered via the hypermodel are plotted as coral Xs. The error estimate of each point is described in the caption of Fig. 10.2. For a fixed number of pulsars, an increase in the observation time leads to a higher Bayes factor. In each case, as N increases, the efficiency decreases. Additionally, as the observation time increases, efficiency tends to decrease, albeit less distinctly. The relative difference between the direct sampling and reweighting Bayes factors remains quite small and is independent of both $T_{\rm obs}$ and N. The small errors and high efficiencies (each greater than 20%) imply that likelihood reweighting remains reliable when additional time and pulsars are added.

extended observation time, the ratio scales with the number of pulsars (ranging between 10 and 40). Thus as more pulsars are added to the dataset, reweighting becomes more important.

10.6 Discussion and Conclusions

We have introduced a reweighting method to efficiently and reliably obtain GW posterior and marginal likelihood for a GWB model in PTA data analysis. We first compute an inexpensive approximate posterior (CP) that omits pulsar-pulsar correlations, then reweight it to a full posterior (HD) that includes them. We have validated this method by comparing reweighted posteriors and Bayes factors with distributions and factors obtained with direct sampling. Reweighting appears to be reliable and unbiased. Even in cases with low reweighting efficiency (as defined by the reduction in the number of effective samples), the reweighted Bayes factor estimate remained robust up to $\mathcal{B}_{\rm CP}^{\rm HD} > 10^6$, far larger than required for a confident GWB detection.

Even though our method requires evaluating the computationally expensive HD likelihood, it is still much more efficient than direct stochastic sampling. This is due to the additional evaluations required for the latter, which do not need to be repeated when reweighting. Direct sampling results in very autocorrelated sample chains, which are thinned [by factors $N_{\rm t} \sim \mathcal{O}(10^3)$, on the order of the chain autocorrelation length] to obtain quasi-independent samples. By contrast, reweighting is applied *after* thinning, reducing the number of HD likelihood evaluations by $N_{\rm t}$. In addition, the weights of Eq. (10.4) can be computed in parallel on multiple cores, allowing a further wall clock speedup (by the number $N_{\rm P}$ of parallel processes). Finally, if parallel tempering was used to sample the approximate model, only samples from the coldest chain should need be reweighted, decreasing the necessary number of computations by a factor of the number of chains $N_{\rm c} \sim \mathcal{O}(10)$.

While the reweighting procedure is mathematically exact, the method is subject to sampling error; reweighted posteriors could have too few effective samples to accurately reflect the true distribution. Constructing generic diagnostic tools for such situations can be challenging, as the effective number of independent samples $n_{\rm eff}$ can vary between applications. In such cases, estimating the Bayes factor sampling error or inspecting posteriors visually can help identify undersampling. If $n_{\rm eff}$ is low, a few strategies are available. The simplest is to increase the number of samples for the approximate model. A more sophisticated option involves the importance sampling of the approximate model by concentrating on the region of parameter space that the target seems to prefer. In the most extreme case, so many approximate-model samples are needed that the method becomes less efficient than direct sampling. This happens when the efficiency drops to the ratio of likelihood computation times (e.g., to 1/25 for the NANOGrav 12.5yr dataset). If parallel tempering is used then "hot" chains, with a correspondingly broader posterior, could be used in situations where efficiency is low due to a lack of samples from the approximate distribution available to estimate tails in the target distribution.

The reweighting formalism is generic and can be applied to any pair of approximate and target distributions. For example, one could model a clock error by including a process with monopolar correlations in addition to the HD correlations. In this situation, extra parameters are added to the target model, which requires drawing samples from some proposal distribution for the new parameters (see [300, 301] for examples of reweighting between models with varying numbers of parameters). In practice, sampling error (efficiency) increases (decreases) if the approximate and target posteriors do not overlap, as quantified in Fig. 10.3 using the Kullback-Leibler divergence.

Throughout this work we have presented examples that are based on the NANOGrav 12.5yr analysis. Although our simulations are consistent with the NANOGrav dataset and our understanding of the stochastic GWB, we have not simulated realistic radio frequency noise such as dispersion measure variations or solar wind fluctuations. More "advanced" noise modeling adds numerous extra parameters to each pulsar to measure chromatic effects [171, 183, 341–343] increasing the complexity of the analysis. Given that most of these additional parameters impact only individual pulsar measurements, a factorized-likelihood approach to estimate the CP model, followed by this reweighting scheme could significantly reduce the wall clock time of an analysis that uses more advanced noise models.

In the context of PTA searches for GWBs, the reweighting formalism introduced in this paper offers an accurate and computationally efficient shortcut to GW posteriors and HD vs. CP Bayes factors. In this paper we tested the method on simulated datasets with increasing GWB amplitudes, which served a proxy for increased observing time and number of pulsars. Our results suggest that reweighting remains robust for PTA datasets with Bayes factors of at least 10^6, orders of magnitude larger than current results. Thus, our method can reliably characterize the GWB from PTA datasets for the foreseeable future and into the detection regime.

10.7 Acknowledgments

We thank Ethan Payne for useful discussions about reweighting and Ken Olum, Steve Taylor, and Paul Baker for comments on the paper draft. Numerical investigations were performed using services provided by the OSG Consortium [275, 317], which is supported by the National Science Foundation Grants No. 2030508 and No. 1836650. Additional computing resources were provided by Caltech's Theoretical AstroPhysics Including Relativity and Cosmology (TAPIR) group. In addition to `enterprise` [155, 336], our software stack included `scipy` [370], `matplotlib` [197], `numpy` [182], `pandas` [263], and `corner` [163]. S.H. and P.M.M. acknowledge the VIPER PTA Summer School at Vanderbilt University, which was funded under NSF CAREER-2146016. A.D.J. and K.C. acknowledge support from the Caltech and Jet Propulsion Laboratory President's and Director's Fund and the Sloan Foundation. S.H. is supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1745301. P.M.M. and M.V. were supported by the NANOGrav Physics Frontiers Center, National Science Foundation (NSF), Grant No. 2020265. Part of this research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (80NM0018D0004).

Epilogue

10.8 Future

The presence of transient, non-Gaussian noise artifacts, or "glitches," significantly impacts the noise environment for gravitational-wave detection and introduces biases into our analyses. While current methods mitigate glitches on individual gravitational-wave events by subtracting single estimates, this *completely* neglects the uncertainty on the glitch model. Glitch subtraction has been shown to bias the estimation of various parameters, with inferred spin proving particularly sensitive to *which* glitch point estimate is taken (Chap 6 and Chap 7).

Throughout this thesis I have shown that glitch mitigation has out-sized effects on the physical measurements of individual events, there has never to-date been a study of the impact of glitch-mitigation on population analyses. This lack of attention is alarming: whereas current detectors observe mergers approximately once per day [116], glitches occur about once every minute; $\sim 20\%$ *of events require glitch mitigation* [28**?**]. As gravitational-wave astronomy matures and becomes more data-driven with future detectors like LISA [76], Cosmic Explorer [291], and Einstein Telescope [283], we will need accurate, automatic, and fast methods to mitigate noise for individual events, and a developed way to incorporate such advancements into population measurements and tests of general relativity.

Chapter 11

BIBLIOGRAPHY

Bibliography

[1] J. Aasi et al. Advanced LIGO. *Class. Quant. Grav.*, 32:074001, 2015. doi: 10.1088/0264-9381/32/7/074001.

[2] R. Abbot et al. GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run — Glitch modelling for events, November 2021. URL https://doi.org/10.5281/zenodo.5546680.

[3] R. Abbot et al. GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run — Parameter estimation data release, October 2023. URL https://doi.org/10.5281/zenodo.8177023.

[4] B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt, and et al. "GW190425: Observation of a Compact Binary Coalescence with Total Mass $\sim$ 3.4 $M_\odot$". *The Astrophysical Journal*, 892(1):L3, Mar 2020. ISSN 2041-8213. doi: 10.3847/2041-8213/ab75f5. URL http://dx.doi.org/10.3847/2041-8213/ab75f5.

[5] B. P. Abbott et al. Characterization of transient noise in Advanced LIGO relevant to gravitational wave signal GW150914. *Class. Quant. Grav.*, 33(13):134001, 2016. doi: 10.1088/0264-9381/33/13/134001.

[6] B. P. Abbott et al. Tests of general relativity with GW150914. *Phys. Rev. Lett.*, 116(22):221101, 2016. doi: 10.1103/PhysRevLett.116.221101. [Erratum: Phys.Rev.Lett. 121, 129902 (2018)].

[7] B.. P.. Abbott et al. GW170608: Observation of a 19-solar-mass Binary Black Hole Coalescence. *Astrophys. J.*, 851(2):L35, 2017. doi: 10.3847/2041-8213/aa9f0c.

[8] B. P. Abbott et al. A gravitational-wave standard siren measurement of the Hubble constant. *Nature*, 551(7678):85–88, 2017. doi: 10.1038/nature24471.

[9] B. P. Abbott et al. Multi-messenger Observations of a Binary Neutron Star Merger. *Astrophys. J. Lett.*, 848(2):L12, 2017. doi: 10.3847/2041-8213/aa91c9.

[10] B. P. Abbott et al. GW170817: Measurements of neutron star radii and equation of state. *Phys. Rev. Lett.*, 121(16):161101, 2018. doi: 10.1103/PhysRevLett.121.161101.

[11] B. P. Abbott et al. Tests of General Relativity with GW170817. *Phys. Rev. Lett.*, 123(1):011102, 2019. doi: 10.1103/PhysRevLett.123.011102.

[12] B. P. Abbott et al. Binary Black Hole Population Properties Inferred from the First and Second Observing Runs of Advanced LIGO and Advanced Virgo. *Astrophys. J. Lett.*, 882(2):L24, 2019. doi: 10.3847/2041-8213/ab3800.

[13] 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. *Phys. Rev. X*, 9(3):031040, 2019. doi: 10.1103/PhysRevX.9.031040.

[14] B. P. Abbott et al. All-Sky Search for Short Gravitational-Wave Bursts in the Second Advanced LIGO and Advanced Virgo Run. *Phys. Rev. D*, 100(2):024017, 2019. doi: 10.1103/PhysRevD.100.024017.

[15] B. P. Abbott et al. Binary black hole population properties inferred from the first and second observing runs of advanced LIGO and advanced virgo. *Astrophys. J.*, 882(2):L24, sep 2019. doi: 10.3847/2041-8213/ab3800. URL http://doi.org/10.3847/2041-8213/ab3800.

[16] Benjamin P. Abbott et al. Prospects for Observing and Localizing Gravitational-Wave Transients with Advanced LIGO, Advanced Virgo and KAGRA. *Living Rev. Rel.*, 19:1, 2013. doi: 10.1007/lrr-2016-1.

[17] Benjamin P. Abbott et al. The basic physics of the binary black hole merger GW150914. *Annalen Phys.*, 529(1-2):1600209, 2017. doi: 10.1002/andp.201600209.

[18] Benjamin P. Abbott et al. GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral. *Phys. Rev. Lett.*, 119(16):161101, 2017. doi: 10.1103/PhysRevLett.119.161101.

[19] Benjamin P Abbott et al. A guide to LIGO–Virgo detector noise and extraction of transient gravitational-wave signals. *Class. Quant. Grav.*, 37(5):055002, 2020. doi: 10.1088/1361-6382/ab685e.

[20] B.P. Abbott et al. Observation of Gravitational Waves from a Binary Black Hole Merger. *Phys. Rev. Lett.*, 116(6):061102, 2016. doi: 10.1103/PhysRevLett.116.061102.

[21] B.P. Abbott et al. Tests of general relativity with GW150914. *Phys. Rev. Lett.*, 116(22):221101, 2016. doi: 10.1103/PhysRevLett.116.221101. [Erratum: Phys.Rev.Lett. 121, 129902 (2018)].

[22] B.P. Abbott et al. Properties of the binary neutron star merger GW170817. *Phys. Rev. X*, 9(1):011001, 2019. doi: 10.1103/PhysRevX.9.011001.

[23] B.P. Abbott et al. Tests of General Relativity with the Binary Black Hole Signals from the LIGO-Virgo Catalog GWTC-1. *Phys. Rev. D*, 100 (10):104036, 2019. doi: 10.1103/PhysRevD.100.104036.

[24] R. Abbott et al. GW190412: Observation of a Binary-Black-Hole Coalescence with Asymmetric Masses. *Phys. Rev. D*, 102(4):043015, 2020. doi: 10.1103/PhysRevD.102.043015.

[25] 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. *Astrophys. J. Lett.*, 896(2):L44, 2020. doi: 10.3847/2041-8213/ab960f.

[26] R. Abbott et al. GWTC-2.1: Deep Extended Catalog of Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run, 8 2021.

[27] R. Abbott et al. All-sky search for short gravitational-wave bursts in the third Advanced LIGO and Advanced Virgo run. *Phys. Rev. D*, 104 (12):122004, 2021. doi: 10.1103/PhysRevD.104.122004.

[28] R. Abbott et al. GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run. *Phys. Rev. X*, 11:021053, 2021. doi: 10.1103/PhysRevX.11.021053.

[29] R. Abbott et al. Population Properties of Compact Objects from the Second LIGO-Virgo Gravitational-Wave Transient Catalog. *Astrophys. J. Lett.*, 913(1):L7, 2021. doi: 10.3847/2041-8213/abe949.

[30] R. Abbott et al. Tests of general relativity with binary black holes from the second LIGO-Virgo gravitational-wave transient catalog. *Phys. Rev. D*, 103(12):122002, 2021. doi: 10.1103/PhysRevD.103.122002.

[31] R. Abbott et al. All-sky search for short gravitational-wave bursts in the third Advanced LIGO and Advanced Virgo run. *Phys. Rev. D*, 104 (12):122004, 2021. doi: 10.1103/PhysRevD.104.122004.

[32] R. Abbott et al. Tests of General Relativity with GWTC-3. *arXiv e-prints*, art. arXiv:2112.06861, December 2021. doi: 10.48550/arXiv.2112.06861.

[33] R. Abbott et al. GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run. *Phys. Rev. X*, 11:021053, 2021. doi: 10.1103/PhysRevX.11.021053.

[34] R. Abbott et al. Population Properties of Compact Objects from the Second LIGO–Virgo Gravitational-Wave Transient Catalog. *Astrophys. J. Lett.*, 913(1):L7, May 2021. ISSN 2041-8205, 2041-8213. doi: 10.3847/2041-8213/abe949. URL http://doi.org/10.3847/2041-8213/abe949.

[35] R. Abbott et al. Population of Merging Compact Binaries Inferred Using Gravitational Waves through GWTC-3. *Phys. Rev. X*, 13(1):011048, 2023. doi: 10.1103/PhysRevX.13.011048.

[36] R. Abbott et al. Open Data from the Third Observing Run of LIGO, Virgo, KAGRA, and GEO. *Astrophys. J. Suppl.*, 267(2):29, 2023. doi: 10.3847/1538-4365/acdc9f.

[37] R. Abbott et al. Population of Merging Compact Binaries Inferred Using Gravitational Waves through GWTC-3. *Phys. Rev. X*, 13(1):011048, 2023. doi: 10.1103/PhysRevX.13.011048.

[38] R. Abbott et al. GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo during the Second Part of the Third Observing Run. *Phys. Rev. X*, 13(4):041039, 2023. doi: 10.1103/PhysRevX.13.041039.

[39] Rich Abbott et al. Open data from the first and second observing runs of Advanced LIGO and Advanced Virgo. *SoftwareX*, 13:100658, 2021. doi: 10.1016/j.softx.2021.100658.

[40] Rich Abbott et al. Search for intermediate-mass black hole binaries in the third observing run of Advanced LIGO and Advanced Virgo. *Astron. Astrophys.*, 659:A84, 2022. doi: 10.1051/0004-6361/202141452.

[41] T. Accadia et al. Noise from scattered light in Virgo's second science run data. *Class. Quant. Grav.*, 27:194011, 2010. doi: 10.1088/0264-9381/27/19/194011.

[42] F. Acernese. The Advanced Virgo detector. *J. Phys. Conf. Ser.*, 610(1):012014, 2015. doi: 10.1088/1742-6596/610/1/012014.

[43] F. Acernese et al. Advanced Virgo: a second-generation interferometric gravitational wave detector. *Class. Quant. Grav.*, 32(2):024001, 2015. doi: 10.1088/0264-9381/32/2/024001.

[44] F. Acernese et al. Virgo Detector Characterization and Data Quality during the O3 run. 5 2022.

[45] Christian Adamcewicz, Paul D. Lasky, and Eric Thrane. Evidence for a Correlation between Binary Black Hole Mass Ratio and Black Hole Spins. *Astrophys. J.*, 958(1):13, 2023. doi: 10.3847/1538-4357/acf763.

[46] T. Adams, D. Buskulic, V. Germain, G. M. Guidi, F. Marion, M. Montani, B. Mours, F. Piergiovanni, and G. Wang. Low-latency analysis pipeline for compact binary coalescences in the advanced gravitational wave detector era. *Class. Quant. Grav.*, 33(17):175012, 2016. doi: 10.1088/0264-9381/33/17/175012.

[47] P. A. R. Ade et al. Planck 2015 results. XIII. Cosmological parameters. *Astron. Astrophys.*, 594:A13, 2016. doi: 10.1051/0004-6361/201525830.

[48] C. Affeldt et al. Advanced techniques in GEO 600. *Class. Quant. Grav.*, 31(22):224002, 2014. doi: 10.1088/0264-9381/31/22/224002.

[49] Adeela Afzal et al. The NANOGrav 15 yr Data Set: Search for Signals from New Physics. *Astrophys. J. Lett.*, 951(1):L11, 2023. doi: 10.3847/2041-8213/acdc91. [Erratum: Astrophys.J.Lett. 971, L27 (2024), Erratum: Astrophys.J. 971, L27 (2024)].

[50] Gabriella Agazie et al. The NANOGrav 15 yr Data Set: Evidence for a Gravitational-wave Background. *Astrophys. J. Lett.*, 951(1):L8, 2023. doi: 10.3847/2041-8213/acdac6. Performed analysis comparing Hellings-Downs to Hellings-Downs + sinusoid and computed detection Bayes Factors.

[51] Gabriella Agazie et al. The NANOGrav 15 yr Data Set: Constraints on Supermassive Black Hole Binaries from the Gravitational-wave Background. *Astrophys. J. Lett.*, 952(2):L37, 2023. doi: 10.3847/2041-8213/ace18b.

[52] N. Aghanim et al. Planck 2018 results. VI. Cosmological parameters. *Astron. Astrophys.*, 641:A6, 2020. doi: 10.1051/0004-6361/201833910. [Erratum: Astron.Astrophys. 652, C4 (2021)].

[53] P. Ajith et al. Inspiral-merger-ringdown waveforms for black-hole binaries with non-precessing spins. *Phys. Rev. Lett.*, 106:241101, 2011. doi: 10.1103/PhysRevLett.106.241101.

[54] T. Akutsu et al. KAGRA: 2.5 Generation Interferometric Gravitational Wave Detector. *Nature Astron.*, 3(1):35–40, 2019. doi: 10.1038/s41550-018-0658-y.

[55] T. Akutsu et al. Overview of KAGRA: Detector design and construction history. *PTEP*, 2021(5):05A101, 2021. doi: 10.1093/ptep/ptaa125.

[56] Bruce Allen. χ^2 time-frequency discriminator for gravitational wave detection. *Phys. Rev. D*, 71:062001, 2005. doi: 10.1103/PhysRevD.71.062001.

[57] Bruce Allen, Wen-sheng Hua, and Adrian C. Ottewill. Automatic cross talk removal from multichannel data. 9 1999.

[58] T. W. Anderson and D. A. Darling. A test of goodness of fit. *Journal of the American Statistical Association*, 49(268):765–769, 1954. doi: 10.1080/01621459.1954.10501232. URL https://www.tandfonline.com/doi/abs/10.1080/01621459.1954.10501232.

[59] Andrea Antonelli, Konstantinos Kritos, Ken K. Y. Ng, Roberto Cotesta, and Emanuele Berti. Classifying the generation and formation channels of individual LIGO-Virgo-KAGRA observations from dynamically formed binaries. *Phys. Rev. D*, 108(8):084044, 2023. doi: 10.1103/PhysRevD.108.084044.

[60] J. Antoniadis et al. The International Pulsar Timing Array second data release: Search for an isotropic gravitational wave background. *Mon. Not. Roy. Astron. Soc.*, 510(4):4873–4887, 2022. doi: 10.1093/mnras/stab3418.

[61] T.A. Apostolatos. Search templates for gravitational waves from precessing, inspiraling binaries. *Phys.Rev.*, D52:605–620, 1995. doi: 10.1103/PhysRevD.52.605.

[62] Theocharis A. Apostolatos, Curt Cutler, Gerald J. Sussman, and Kip S. Thorne. Spin induced orbital precession and its modulation of the gravitational wave forms from merging binaries. *Phys.Rev.*, D49:6274–6297, 1994. doi: 10.1103/PhysRevD.49.6274.

[63] K. G. Arun, Alessandra Buonanno, Guillaume Faye, and Evan Ochsner. Higher-order spin effects in the amplitude and phase of gravitational waveforms emitted by inspiraling compact binaries: Ready-to-use gravitational waveforms. *Phys. Rev. D*, 79:104023, 2009. doi: 10.1103/PhysRevD.79.104023. [Erratum: Phys.Rev.D 84, 049901 (2011)].

[64] Z. Arzoumanian, A. Brazier, S. Burke-Spolaor, S. J. Chamberlin, S. Chatterjee, B. Christy, J. M. Cordes, N. J. Cornish, K. Crowter, P. B. Demorest, and et al. The NANOGrav Nine-year Data Set: Limits on the Isotropic Stochastic Gravitational Wave Background. *The Astrophysical Journal*, 821(1):13, April 2016. doi: 10.3847/0004-637X/821/1/13.

[65] Z. Arzoumanian et al. The NANOGrav 11-year Data Set: Pulsar-timing Constraints On The Stochastic Gravitational-wave Background. *Astrophys. J.*, 859(1):47, 2018. doi: 10.3847/1538-4357/aabd3b.

[66] Zaven Arzoumanian et al. The NANOGrav 12.5 yr Data Set: Search for an Isotropic Stochastic Gravitational-wave Background. *Astrophys. J. Lett.*, 905(2):L34, 2020. doi: 10.3847/2041-8213/abd401.

[67] Gregory Ashton and Eric Thrane. The astrophysical odds of GW151216. *Mon. Not. Roy. Astron. Soc.*, 498(2):1905–1910, 2020. doi: 10.1093/mnras/staa2332.

[68] Gregory Ashton, Sarah Thiele, Yannick Lecoeuche, Jess McIver, and Laura K. Nuttall. Parameterised population models of transient non-Gaussian noise in the LIGO gravitational-wave detectors. *Class. Quant. Grav.*, 39(17):175004, 2022. doi: 10.1088/1361-6382/ac8094.

[69] Gregory Ashton et al. BILBY: A user-friendly Bayesian inference library for gravitational-wave astronomy. *Astrophys. J. Suppl.*, 241(2):27, 2019. doi: 10.3847/1538-4365/ab06fc.

[70] 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, 2013. doi: 10.1103/PhysRevD.88.043007.

[71] F. Aubin et al. The MBTA pipeline for detecting compact binary coalescences in the third LIGO–Virgo observing run. *Class. Quant. Grav.*, 38(9):095004, 2021. doi: 10.1088/1361-6382/abe913.

[72] D. Backer, S. Kulkarni, C. Heiles, M. Davis, and W. M. Goss. A millisecond pulsar. *Nature*, 300:615–618, 1982. doi: 10.1038/300615a0.

[73] Sharan Banagiri, Michael W. Coughlin, James Clark, Paul D. Lasky, M. A. Bizouard, Colm Talbot, Eric Thrane, and Vuk Mandic. Constraining the Gravitational-Wave Afterglow From a Binary Neutron Star Coalescence. *Mon. Not. Roy. Astron. Soc.*, 492(4):4945–4951, 2020. doi: 10.1093/mnras/staa181.

[74] Simone S. Bavera, Tassos Fragos, Ying Qin, Emmanouil Zapartas, Coenraad J. Neijssel, Ilya Mandel, Aldo Batta, Sebastian M. Gaebel, Chase Kimball, and Simon Stevenson. The origin of spin in binary black holes: Predicting the distributions of the main observables of Advanced LIGO. *Astron. Astrophys.*, 635:A97, 2020. doi: 10.1051/0004-6361/201936204.

[75] Simone S. Bavera et al. The impact of mass-transfer physics on the observable properties of field binary black hole populations. *Astron. Astrophys.*, 647:A153, 2021. doi: 10.1051/0004-6361/202039804.

[76] Emanuele Berti, Vitor Cardoso, and Clifford M. Will. On gravitational-wave spectroscopy of massive black holes with the space interferometer LISA. *Phys. Rev. D*, 73:064030, 2006. doi: 10.1103/PhysRevD.73.064030.

[77] Sylvia Biscoveanu, Carl-Johan Haster, Salvatore Vitale, and Jonathan Davies. Quantifying the Effect of Power Spectral Density Uncertainty on Gravitational-Wave Parameter Estimation for Compact Binary Sources. *Phys. Rev. D*, 102(2):023008, 2020. doi: 10.1103/PhysRevD.102.023008.

[78] Sylvia Biscoveanu, Maximiliano Isi, Vijay Varma, and Salvatore Vitale. Measuring the spins of heavy binary black holes. *Phys. Rev. D*, 104(10): 103018, 2021. doi: 10.1103/PhysRevD.104.103018.

[79] Jonathan Blackman, Scott E. Field, Chad R. Galley, Béla Szilágyi, Mark A. Scheel, Manuel Tiglio, and Daniel A. Hemberger. Fast and Accurate Prediction of Numerical Relativity Waveforms from Binary Black Hole Coalescences Using Surrogate Models. *Phys. Rev. Lett.*, 115(12): 121102, 2015. doi: 10.1103/PhysRevLett.115.121102.

[80] Jonathan Blackman, Scott E. Field, Mark A. Scheel, Chad R. Galley, Daniel A. Hemberger, Patricia Schmidt, and Rory Smith. A Surrogate Model of Gravitational Waveforms from Numerical Relativity Simulations of Precessing Binary Black Hole Mergers. *Phys. Rev. D*, 95(10): 104023, 2017. doi: 10.1103/PhysRevD.95.104023.

[81] Jonathan Blackman, Scott E. Field, Mark A. Scheel, Chad R. Galley, Christian D. Ott, Michael Boyle, Lawrence E. Kidder, Harald P. Pfeiffer, and Béla Szilágyi. Numerical relativity waveform surrogate model for generically precessing binary black hole mergers. *Phys. Rev. D*, 96(2): 024058, 2017. doi: 10.1103/PhysRevD.96.024058.

[82] Luc Blanchet, Thibault Damour, Bala R. Iyer, Clifford M. Will, and Alan G. Wiseman. Gravitational radiation damping of compact binary systems to second postNewtonian order. *Phys. Rev. Lett.*, 74:3515–3518, 1995. doi: 10.1103/PhysRevLett.74.3515.

[83] Alejandro Bohé, Sylvain Marsat, and Luc Blanchet. Next-to-next-to-leading order spin–orbit effects in the gravitational wave flux and orbital phasing of compact binaries. *Class. Quant. Grav.*, 30:135009, 2013. doi: 10.1088/0264-9381/30/13/135009.

[84] Alejandro Bohé, Guillaume Faye, Sylvain Marsat, and Edward K. Porter. Quadratic-in-spin effects in the orbital dynamics and gravitational-wave energy flux of compact binaries at the 3PN order. *Class. Quant. Grav.*, 32(19):195010, 2015. doi: 10.1088/0264-9381/32/19/195010.

[85] Michael Boyle et al. The SXS Collaboration catalog of binary black hole simulations. *Class. Quant. Grav.*, 36(19):195006, 2019. doi: 10.1088/1361-6382/ab34e2.

[86] A. Buonanno and T. Damour. Effective one-body approach to general relativistic two-body dynamics. *Phys. Rev.*, D59:084006, 1999. doi: 10.1103/PhysRevD.59.084006.

[87] Alessandra Buonanno and Thibault Damour. Transition from inspiral to plunge in binary black hole coalescences. *Phys. Rev.*, D62:064015, 2000. doi: 10.1103/PhysRevD.62.064015.

[88] Alessandra Buonanno, Yan-bei Chen, and Michele Vallisneri. Detecting gravitational waves from precessing binaries of spinning compact objects: Adiabatic limit. *Phys. Rev.*, D67:104025, 2003. doi: 10.1103/PhysRevD.67.104025,10.1103/PhysRevD.74.029904. [Erratum: Phys. Rev.D74,029904(2006)].

[89] Alessandra Buonanno, Yan-bei Chen, Yi Pan, and Michele Vallisneri. A Quasi-physical family of gravity-wave templates for precessing binaries of spinning compact objects. 2. Application to double-spin precessing binaries. *Phys. Rev. D*, 70:104003, 2004. doi: 10.1103/PhysRevD.74.029902. [Erratum: Phys.Rev.D 74, 029902 (2006)].

[90] Alessandra Buonanno, Yanbei Chen, and Thibault Damour. Transition from inspiral to plunge in precessing binaries of spinning black holes. *Phys. Rev. D*, 74:104005, 2006. doi: 10.1103/PhysRevD.74.104005.

[91] R. N. Caballero et al. Studying the solar system with the International Pulsar Timing Array. *Mon. Not. Roy. Astron. Soc.*, 481(4):5501–5516, 2018. doi: 10.1093/mnras/sty2632.

[92] M Cabero, A Lundgren, A H Nitz, T Dent, D Barker, E Goetz, J S Kissel, L K Nuttall, P Schale, R Schofield, and D Davis. Blip glitches in advanced LIGO data. *Classical and Quantum Gravity*, 36(15):155010, jul 2019. doi: 10.1088/1361-6382/ab2e14. URL https://doi.org/10.1088%2F1361-6382%2Fab2e14.

[93] Juan Calderón Bustillo, Sascha Husa, Alicia M. Sintes, and Michael Pürrer. Impact of gravitational radiation higher order modes on single aligned-spin gravitational wave searches for binary black holes. *Phys. Rev. D*, 93(8):084019, 2016. doi: 10.1103/PhysRevD.93.084019.

[94] Juan Calderón Bustillo, Pablo Laguna, and Deirdre Shoemaker. Detectability of gravitational waves from binary black holes: Impact of precession and higher modes. *Phys. Rev. D*, 95(10):104038, 2017. doi: 10.1103/PhysRevD.95.104038.

[95] Juan Calderón Bustillo, James A. Clark, Pablo Laguna, and Deirdre Shoemaker. Tracking black hole kicks from gravitational wave observations. *Phys. Rev. Lett.*, 121(19):191102, 2018. doi: 10.1103/PhysRevLett.121.191102.

[96] Thomas Callister, M. W. Coughlin, and J. B. Kanner. Gravitational-wave Geodesy: A New Tool for Validating Detection of the Stochastic Gravitational-wave Background. *Astrophys. J. Lett.*, 869(2):L28, 2018. doi: 10.3847/2041-8213/aaf3a5.

[97] Manuela Campanelli, C. O. Lousto, and Y. Zlochower. Spinning-black-hole binaries: The orbital hang up. *Phys. Rev. D*, 74:041501, 2006. doi: 10.1103/PhysRevD.74.041501.

[98] Kipp Cannon et al. GstLAL: A software framework for gravitational wave discovery. 10 2020.

[99] Collin D. Capano, Miriam Cabero, Julian Westerweck, Jahed Abedi, Shilpa Kastha, Alexander H. Nitz, Yi-Fan Wang, Alex B. Nielsen, and Badri Krishnan. Multimode Quasinormal Spectrum from a Perturbed Black Hole. *Phys. Rev. Lett.*, 131(22):221402, 2023. doi: 10.1103/PhysRevLett.131.221402.

[100] Gregorio Carullo, Walter Del Pozzo, and John Veitch. Observational Black Hole Spectroscopy: A time-domain multimode analysis of GW150914. *Phys. Rev. D*, 99(12):123029, 2019. doi: 10.1103/PhysRevD.99.123029. [Erratum: Phys.Rev.D 100, 089903 (2019)].

[101] Koustav Chandra, V. Villa-Ortega, T. Dent, C. McIsaac, Archana Pai, I. W. Harry, G. S. Cabourn Davies, and K. Soni. An optimized PyCBC search for gravitational waves from intermediate-mass black hole mergers. *Phys. Rev. D*, 104:042004, 2021. doi: 10.1103/PhysRevD.104.042004.

[102] S. Chatterji, L. Blackburn, G. Martin, and E. Katsavounidis. Multiresolution techniques for the detection of gravitational-wave bursts. *Class. Quant. Grav.*, 21:S1809–S1818, 2004. doi: 10.1088/0264-9381/21/20/024.

[103] Katerina Chatziioannou. Uncertainty limits on neutron star radius measurements with gravitational waves. *Phys. Rev. D*, 105(8):084021, 2022. doi: 10.1103/PhysRevD.105.084021.

[104] Katerina Chatziioannou, Neil Cornish, Antoine Klein, and Nicolas Yunes. Detection and Parameter Estimation of Gravitational Waves from Compact Binary Inspirals with Analytical Double-Precessing Templates. *Phys. Rev. D*, 89:104023, 2014. doi: 10.1103/PhysRevD.89.104023.

[105] Katerina Chatziioannou, James Alexander Clark, Andreas Bauswein, Margaret Millhouse, Tyson B. Littenberg, and Neil Cornish. Inferring the post-merger gravitational wave emission from binary neutron star coalescences. *Phys. Rev. D*, 96(12):124035, 2017. doi: 10.1103/PhysRevD.96.124035.

[106] Katerina Chatziioannou, Antoine Klein, Neil Cornish, and Nicolas Yunes. Analytic Gravitational Waveforms for Generic Precessing Binary Inspirals. *Phys. Rev. Lett.*, 118(5):051101, 2017. doi: 10.1103/PhysRevLett.118.051101.

[107] Katerina Chatziioannou, Antoine Klein, Nicolás Yunes, and Neil Cornish. Constructing Gravitational Waves from Generic Spin-Precessing

Compact Binary Inspirals. *Phys. Rev. D*, 95(10):104004, 2017. doi: 10.1103/PhysRevD.95.104004.

[108] Katerina Chatziioannou, Carl-Johan Haster, Tyson B. Littenberg, Will M. Farr, Sudarshan Ghonge, Margaret Millhouse, James A. Clark, and Neil Cornish. Noise spectral estimation methods and their impact on gravitational wave measurement of compact binary mergers. *Phys. Rev. D*, 100:104004, Nov 2019. doi: 10.1103/PhysRevD.100.104004. URL https://link.aps.org/doi/10.1103/PhysRevD.100.104004.

[109] Katerina Chatziioannou, Carl-Johan Haster, Tyson B. Littenberg, Will M. Farr, Sudarshan Ghonge, Margaret Millhouse, James A. Clark, and Neil Cornish. Noise spectral estimation methods and their impact on gravitational wave measurement of compact binary mergers. *Phys. Rev. D*, 100(10):104004, 2019. doi: 10.1103/PhysRevD.100.104004.

[110] Katerina Chatziioannou, Neil Cornish, Marcella Wijngaarden, and Tyson B. Littenberg. Modeling compact binary signals and instrumental glitches in gravitational wave data. *Phys. Rev. D*, 103(4):044013, 2021. doi: 10.1103/PhysRevD.103.044013.

[111] Katerina Chatziioannou, Maximiliano Isi, Carl-Johan Haster, and Tyson B. Littenberg. Morphology-independent test of the mixed polarization content of transient gravitational wave signals. *Phys. Rev. D*, 104(4):044005, 2021. doi: 10.1103/PhysRevD.104.044005.

[112] Katerina Chatziioannou et al. On the properties of the massive binary black hole merger GW170729. *Phys. Rev. D*, 100(10):104015, 2019. doi: 10.1103/PhysRevD.100.104015.

[113] S. Chen et al. Common-red-signal analysis with 24-yr high-precision timing of the European Pulsar Timing Array: inferences in the stochastic gravitational-wave background search. *Mon. Not. Roy. Astron. Soc.*, 508 (4):4970–4993, 2021. doi: 10.1093/mnras/stab2833.

[114] Qi Chu et al. SPIIR online coherent pipeline to search for gravitational waves from compact binary coalescences. *Phys. Rev. D*, 105(2):024023, 2022. doi: 10.1103/PhysRevD.105.024023.

[115] Robert E. Colgan, Zsuzsa Márka, Jingkai Yan, Imre Bartos, John N. Wright, and Szabolcs Márka. Detecting and diagnosing terrestrial gravitational-wave mimics through feature learning. *Phys. Rev. D*, 107 (6):062006, 2023. doi: 10.1103/PhysRevD.107.062006.

[116] LIGO Scientific Collaboration and Virgo Collaboration. Ligo/virgo public alerts (o4). https://gracedb.ligo.org/superevents/public/O4/, 2024. Accessed: 2024-10-09.

[117] Samantha R. Cook, Andrew Gelman, and Donald B. Rubin. Validation of software for bayesian models using posterior quantiles. *Journal of Computational and Graphical Statistics*, 15(3):675–692, 2006. ISSN 10618600. URL http://www.jstor.org/stable/27594203.

[118] Neil J. Cornish. Time-Frequency Analysis of Gravitational Wave Data. 8 2020. doi: 10.1103/PhysRevD.102.124038.

[119] Neil J. Cornish. Heterodyned likelihood for rapid gravitational wave parameter inference. *Phys. Rev. D*, 104(10):104054, 2021. doi: 10.1103/PhysRevD.104.104054.

[120] Neil J. Cornish. Rapid and Robust Parameter Inference for Binary Mergers. *Phys. Rev. D*, 103(10):104057, 2021. doi: 10.1103/PhysRevD.103.104057.

[121] Neil J. Cornish and Tyson B. Littenberg. BayesWave: Bayesian Inference for Gravitational Wave Bursts and Instrument Glitches. *Class. Quant. Grav.*, 32(13):135012, 2015. doi: 10.1088/0264-9381/32/13/135012.

[122] Neil J. Cornish and Joseph D. Romano. Towards a unified treatment of gravitational-wave data analysis. *Phys. Rev. D*, 87(12):122003, 2013. doi: 10.1103/PhysRevD.87.122003.

[123] Neil J. Cornish and Laura Sampson. Towards Robust Gravitational Wave Detection with Pulsar Timing Arrays. *Phys. Rev. D*, 93(10):104047, 2016. doi: 10.1103/PhysRevD.93.104047.

[124] Neil J. Cornish, Tyson B. Littenberg, Bence Bécsy, Katerina Chatziioannou, James A. Clark, Sudarshan Ghonge, and Margaret Millhouse. BayesWave analysis pipeline in the era of gravitational wave observations. *Phys. Rev. D*, 103(4):044006, 2021. doi: 10.1103/PhysRevD.103.044006.

[125] Michael W. Coughlin et al. Measurement and subtraction of Schumann resonances at gravitational-wave interferometers. *Phys. Rev. D*, 97(10):102007, 2018. doi: 10.1103/PhysRevD.97.102007.

[126] Pep Covas, Anamaria Effler, Rick Savage, and Robert Schofield. aLIGO LHO Logbook. 37630, 2017.

[127] C. Cutler and J. Harms. BBO and the neutron-star-binary subtraction problem. *Phys. Rev. D*, 73:042001, 2006. doi: 10.1103/PhysRevD.73.042001.

[128] Curt Cutler and Eanna E. Flanagan. Gravitational waves from merging compact binaries: How accurately can one extract the binary's parameters from the inspiral wave form? *Phys. Rev. D*, 49:2658–2697, 1994. doi: 10.1103/PhysRevD.49.2658.

[129] Tito Dal Canton, Alexander H. Nitz, Bhooshan Gadre, Gareth S. Cabourn Davies, Veronica Villa-Ortega, Thomas Dent, Ian Harry, and Liting Xiao. Real-time Search for Compact Binary Mergers in Advanced LIGO and Virgo's Third Observing Run Using PyCBC Live. *Astrophys. J.*, 923(2):254, 2021. doi: 10.3847/1538-4357/ac2f9a.

[130] Gergely Dálya, Peter Raffai, and Bence Bécsy. Bayesian Reconstruction of Gravitational-wave Signals from Binary Black Holes with Nonzero Eccentricities. *arXiv e-prints*, art. arXiv:2006.06256, June 2020.

[131] Thibault Damour. Coalescence of two spinning black holes: an effective one-body approach. *Phys. Rev. D*, 64:124013, 2001. doi: 10.1103/PhysRevD.64.124013.

[132] Thibault Damour and Alessandro Nagar. An Improved analytical description of inspiralling and coalescing black-hole binaries. *Phys. Rev. D*, 79:081503, 2009. doi: 10.1103/PhysRevD.79.081503.

[133] Thibault Damour, Piotr Jaranowski, and Gerhard Schaefer. On the determination of the last stable orbit for circular general relativistic binaries at the third postNewtonian approximation. *Phys. Rev. D*, 62: 084011, 2000. doi: 10.1103/PhysRevD.62.084011.

[134] Thibault Damour, Piotr Jaranowski, and Gerhard Schaefer. Effective one body approach to the dynamics of two spinning black holes with next-to-leading order spin-orbit coupling. *Phys. Rev. D*, 78:024009, 2008. doi: 10.1103/PhysRevD.78.024009.

[135] Thibault Damour, Bala R. Iyer, and Alessandro Nagar. Improved resummation of post-Newtonian multipolar waveforms from circularized compact binaries. *Phys. Rev. D*, 79:064004, 2009. doi: 10.1103/PhysRevD.79.064004.

[136] Gareth S. Davies, Thomas Dent, Márton Tápai, Ian Harry, Connor McIsaac, and Alexander H. Nitz. Extending the PyCBC search for gravitational waves from compact binary mergers to a global network. *Phys. Rev. D*, 102(2):022004, 2020. doi: 10.1103/PhysRevD.102.022004.

[137] D. Davis, T. J. Massinger, A. P. Lundgren, J. C. Driggers, A. L. Urban, and L. K. Nuttall. Improving the Sensitivity of Advanced LIGO Using Noise Subtraction. *Class. Quant. Grav.*, 36(5):055011, 2019. doi: 10.1088/1361-6382/ab01c5.

[138] D. Davis, T. B. Littenberg, I. M. Romero-Shaw, M. Millhouse, J. McIver, F. Di Renzo, and G. Ashton. Subtracting glitches from gravitational-wave detector data during the third LIGO-Virgo observing run. *Class. Quant. Grav.*, 39(24):245013, 2022. doi: 10.1088/1361-6382/aca238.

[139] D. Davis et al. Data Quality Vetoes Applied to the Analysis of LIGO Data from the Third Observing Run. Technical Report DCC-T2100045, LIGO, 2021. URL https://dcc.ligo.org/LIGO-T2100045/public.

[140] Derek Davis and Marissa Walker. Detector Characterization and Mitigation of Noise in Ground-Based Gravitational-Wave Interferometers. *Galaxies*, 10(1):12, 2022. doi: 10.3390/galaxies10010012.

[141] Derek Davis, Laurel V. White, and Peter R. Saulson. Utilizing aLIGO Glitch Classifications to Validate Gravitational-Wave Candidates. *Class. Quant. Grav.*, 37(14):145001, 2020. doi: 10.1088/1361-6382/ab91e6.

[142] Derek Davis et al. LIGO detector characterization in the second and third observing runs. *Class. Quant. Grav.*, 38(13):135014, 2021. doi: 10.1088/1361-6382/abfd85.

[143] Maximilian Dax, Stephen R. Green, Jonathan Gair, Nihar Gupte, Michael Pürrer, Vivien Raymond, Jonas Wildberger, Jakob H. Macke, Alessandra Buonanno, and Bernhard Schölkopf. Real-time inference for binary neutron star mergers using machine learning. *Nature*, 639(8053): 49–53, 2025. doi: 10.1038/s41586-025-08593-z.

[144] Howard Deshong, Sophie Hourihane, Katerina Chatziioannou, Aaron Johnson, Maven Holst, Meg Millhouse, Tyson Littenberg, and Neil Cornish. Bayeswave c++ repository, 2025. URL https://git.ligo.org/bayeswave/bayeswave-cpp. GitLab repository, accessed May 21, 2025.

[145] S. Detweiler. Pulsar timing measurements and the search for gravitational waves. *The Astrophysical Journal*, 234:1100–1104, December 1979. doi: 10.1086/157593.

[146] Tim Dietrich, Sebastiano Bernuzzi, and Wolfgang Tichy. Closed-form tidal approximants for binary neutron star gravitational waveforms constructed from high-resolution numerical relativity simulations. *Phys. Rev. D*, 96(12):121501, 2017. doi: 10.1103/PhysRevD.96.121501.

[147] Tim Dietrich, Anuradha Samajdar, Sebastian Khan, Nathan K. Johnson-McDaniel, Reetika Dudi, and Wolfgang Tichy. Improving the NRTidal model for binary neutron star systems. *Phys. Rev. D*, 100(4): 044003, 2019. doi: 10.1103/PhysRevD.100.044003.

[148] Tim Dietrich et al. Matter imprints in waveform models for neutron star binaries: Tidal and self-spin effects. *Phys. Rev.*, D99(2):024029, 2019. doi: 10.1103/PhysRevD.99.024029.

[149] K. L. Dooley et al. GEO 600 and the GEO-HF upgrade program: successes and challenges. *Class. Quant. Grav.*, 33:075009, 2016. doi: 10.1088/0264-9381/33/7/075009.

[150] J. C. Driggers et al. Improving astrophysical parameter estimation via offline noise subtraction for Advanced LIGO. *Phys. Rev. D*, 99(4):042001, 2019. doi: 10.1103/PhysRevD.99.042001.

[151] Matthew C. Edwards, Renate Meyer, and Nelson Christensen. Bayesian semiparametric power spectral density estimation with applications in gravitational wave data analysis. *Phys. Rev. D*, 92(6):064011, 2015. doi: 10.1103/PhysRevD.92.064011.

[152] Oliver Edy, Andrew Lundgren, and Laura K. Nuttall. Issues of mismodeling gravitational-wave data for parameter estimation. *Physical Review D*, 103(12):124061, June 2021. doi: 10.1103/PhysRevD.103.124061.

[153] Albert Einstein. Über Gravitationswellen. *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften*, pages 154–167, January 1918.

[154] Justin Ellis and Rutger van Haasteren. jellis18/ptmcmcsampler: Official release, October 2017. URL https://doi.org/10.5281/zenodo.1037579.

[155] Justin A. Ellis, Michele Vallisneri, Stephen R. Taylor, and Paul T. Baker. Enterprise: Enhanced numerical toolbox enabling a robust pulsar inference suite. Zenodo, September 2020. URL https://doi.org/10.5281/zenodo.4059815.

[156] Víctor Elvira, Luca Martino, and Christian P. Robert. Rethinking the effective sample size. *International Statistical Review*, 90(3):525–550, 2022. doi: https://doi.org/10.1111/insr.12500. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/insr.12500.

[157] Reed Essick. Calibration uncertainty's impact on gravitational-wave observations. *Phys. Rev. D*, 105(8):082002, 2022. doi: 10.1103/PhysRevD.105.082002.

[158] Reed Essick. Compact binary coalescence sensitivity estimates with injection campaigns during the ligo-virgo-kagra collaborations' fourth observing run. *In Preparation*, 2025.

[159] Stephen Fairhurst, Rhys Green, Charlie Hoy, Mark Hannam, and Alistair Muir. Two-harmonic approximation for gravitational waveforms from precessing binaries. *Phys. Rev. D*, 102(2):024055, 2020. doi: 10.1103/PhysRevD.102.024055.

[160] R. Farmer, M. Renzo, S. E. de Mink, P. Marchant, and S. Justham. Mind the gap: The location of the lower edge of the pair instability supernovae black hole mass gap. 10 2019. doi: 10.3847/1538-4357/ab518b.

[161] Lee Samuel Finn and David F. Chernoff. Observing binary inspiral in gravitational radiation: One interferometer. *Phys. Rev. D*, 47:2198–2219, 1993. doi: 10.1103/PhysRevD.47.2198.

[162] Ronald Aylmer Fisher. *Statistical methods for research workers*. Number 5. Oliver and Boyd, 1928.

[163] Daniel Foreman-Mackey. corner.py: Scatterplot matrices in python. *The Journal of Open Source Software*, 24, 2016. doi: 10.21105/joss.00024. URL http://dx.doi.org/10.5281/zenodo.45906.

[164] Gabriel E. Freedman, Aaron D. Johnson, Rutger van Haasteren, and Sarah J. Vigeland. Efficient gravitational wave searches with pulsar timing arrays using Hamiltonian Monte Carlo. *Phys. Rev. D*, 107(4): 043013, 2023. doi: 10.1103/PhysRevD.107.043013.

[165] Jim Fuller and Linhao Ma. Most Black Holes are Born Very Slowly Rotating. *Astrophys. J. Lett.*, 881(1):L1, 2019. doi: 10.3847/2041-8213/ab339b.

[166] 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, 2020. doi: 10.1103/PhysRevD.102.064002.

[167] Davide Gerosa, Michael Kesden, Ulrich Sperhake, Emanuele Berti, and Richard O'Shaughnessy. Multi-timescale analysis of phase transitions in precessing black-hole binaries. *Phys. Rev. D*, 92:064016, 2015. doi: 10.1103/PhysRevD.92.064016.

[168] Davide Gerosa, Emanuele Berti, Richard O'Shaughnessy, Krzysztof Belczynski, Michael Kesden, Daniel Wysocki, and Wojciech Gladysz. Spin orientations of merging black holes formed from the evolution of stellar binaries. *Phys. Rev. D*, 98(8):084036, 2018. doi: 10.1103/PhysRevD.98.084036.

[169] Sudarshan Ghonge, Katerina Chatziioannou, James A. Clark, Tyson Littenberg, Margaret Millhouse, Laura Cadonati, and Neil Cornish. Reconstructing gravitational wave signals from binary black hole mergers with minimal assumptions. *Phys. Rev. D*, 102(6):064056, 2020. doi: 10.1103/PhysRevD.102.064056.

[170] Jean Dickinson Gibbons and Subhabrata Chakraborti. *Nonparametric Statistical Inference, Fourth Edition*. Marcel Dekker, Inc, 270 Madison Avenue, New York, NY 10016, 2003. ISBN 0-8247-4052-1.

[171] Boris Goncharov et al. Identifying and mitigating noise sources in precision pulsar timing data sets. *Mon. Not. Roy. Astron. Soc.*, 502(1): 478–493, 2021. doi: 10.1093/mnras/staa3411.

[172] Boris Goncharov et al. On the Evidence for a Common-spectrum Process in the Search for the Nanohertz Gravitational-wave Background with the Parkes Pulsar Timing Array. *Astrophys. J. Lett.*, 917(2):L19, 2021. doi: 10.3847/2041-8213/ac17f4.

[173] Boris Goncharov et al. Consistency of the Parkes Pulsar Timing Array Signal with a Nanohertz Gravitational- wave Background. *Astrophys. J.*, 932(2):L22, 2022. doi: 10.3847/2041-8213/ac76bb.

[174] Peter J. Green. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4):711–732, 12 1995. ISSN 0006-3444. doi: 10.1093/biomet/82.4.711. URL https://dx.doi.org/10.1093/biomet/82.4.711.

[175] Nihar Gupte et al. Evidence for eccentricity in the population of binary black holes observed by LIGO-Virgo-KAGRA. 4 2024.

[176] Roland Haas et al. Simulations of inspiraling and merging double neutron stars using the Spectral Einstein Code. *Phys. Rev. D*, 93(12):124062, 2016. doi: 10.1103/PhysRevD.93.124062.

[177] Chad Hanna et al. Fast evaluation of multidetector consistency for real-time gravitational wave searches. *Phys. Rev. D*, 101(2):022003, 2020. doi: 10.1103/PhysRevD.101.022003.

[178] Mark Hannam, Patricia Schmidt, Alejandro Bohé, Leïla Haegel, Sascha Husa, Frank Ohme, Geraint Pratten, and Michael Pürrer. Simple Model of Complete Precessing Black-Hole-Binary Gravitational Waveforms. *Phys. Rev. Lett.*, 113(15):151101, 2014. doi: 10.1103/PhysRevLett.113.151101.

[179] Mark Hannam et al. General-relativistic precession in a black-hole binary. *Nature*, 610(7933):652–655, 2022. doi: 10.1038/s41586-022-05212-z.

[180] O. A. Hannuksela, K. Haris, K. K. Y. Ng, S. Kumar, A. K. Mehta, D. Keitel, T. G. F. Li, and P. Ajith. Search for gravitational lensing signatures in LIGO-Virgo binary black hole events. *Astrophys. J. Lett.*, 874(1):L2, 2019. doi: 10.3847/2041-8213/ab0c0f.

[181] K. Haris, Ajit Kumar Mehta, Sumit Kumar, Tejaswi Venumadhav, and Parameswaran Ajith. Identifying strongly lensed gravitational wave signals from binary black hole mergers. 7 2018.

[182] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020. doi: 10.1038/s41586-020-2649-2. URL https://doi.org/10.1038/s41586-020-2649-2.

[183] Jeffrey S. Hazboun et al. Bayesian Solar Wind Modeling with Pulsar Timing Arrays. *Astrophys. J.*, 929(1):39, 2022. doi: 10.3847/1538-4357/ac5829.

[184] Daniel W. Heck, Antony M. Overstall, Quentin F. Gronau, and Eric-Jan Wagenmakers. Quantifying uncertainty in transdimensional markov chain monte carlo using discrete markov models. *Statistics and Computing*, 29(4):631–643, aug 2018. doi: 10.1007/s11222-018-9828-0. URL https://doi.org/10.1007%2Fs11222-018-9828-0.

[185] S. Hee, W. J. Handley, M. P. Hobson, and A. N. Lasenby. Bayesian model selection without evidences: Application to the dark energy equation-of-state. 455(3):2461–2473, 2016. ISSN 0035-8711, 1365-2966. doi: 10.1093/mnras/stv2217. URL https://academic.oup.com/mnras/article-lookup/doi/10.1093/mnras/stv2217.

[186] Jack Heinzel, Colm Talbot, Gregory Ashton, and Salvatore Vitale. Inferring the astrophysical population of gravitational wave sources in the presence of noise transients. *Mon. Not. Roy. Astron. Soc.*, 523(4):5972–5984, 2023. doi: 10.1093/mnras/stad1823.

[187] R. W. Hellings and G. S. Downs. Upper limits on the isotropic gravitational radiation background from pulsar timing analysis. *The Astrophysical Journal*, 265:L39, 1983. ISSN 0004-637X, 1538-4357. doi: 10.1086/183954. URL http://adsabs.harvard.edu/doi/10.1086/183954.

[188] Edward P. J. van den Heuvel. Formation of Double Neutron Stars, Millisecond Pulsars and Double Black Holes. *J. Astrophys. Astron.*, 38:45, 2017. doi: 10.1007/S12036-017-9458-5.

[189] A. Hewish, S. J. Bell, J. D. H Pilkington, P. F. Scott, and R. A. Collins. Observation of a rapidly pulsating radio source. *Nature*, 217:709–713, 1968. doi: 10.1038/217709a0.

[190] Yoshiaki Himemoto and Atsushi Taruya. Impact of correlated magnetic noise on the detection of stochastic gravitational waves: Estima-

tion based on a simple analytical model. *Phys. Rev. D*, 96(2):022004, 2017. doi: 10.1103/PhysRevD.96.022004.

[191] Yoshiaki Himemoto and Atsushi Taruya. Correlated magnetic noise from anisotropic lightning sources and the detection of stochastic gravitational waves. *Phys. Rev. D*, 100(8):082001, 2019. doi: 10.1103/PhysRevD.100.082001.

[192] Sophie Hourihane, Katerina Chatziioannou, Marcella Wijngaarden, Derek Davis, Tyson Littenberg, and Neil Cornish. Accurate modeling and mitigation of overlapping signals and glitches in gravitational-wave data. *Phys. Rev. D*, 106(4):042006, 2022. doi: 10.1103/PhysRevD.106.042006.

[193] Charlie Hoy, Cameron Mills, and Stephen Fairhurst. Evidence for subdominant multipole moments and precession in merging black-hole-binaries from GWTC-2.1. *Phys. Rev. D*, 106(2):023019, 2022. doi: 10.1103/PhysRevD.106.023019.

[194] Qian Hu and John Veitch. Assessing the model waveform accuracy of gravitational waves. *Phys. Rev. D*, 106(4):044042, 2022. doi: 10.1103/PhysRevD.106.044042.

[195] Edwin Hubble. A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae. *Proceedings of the National Academy of Science*, 15(3):168–173, March 1929. doi: 10.1073/pnas.15.3.168.

[196] R. A. Hulse and J. H. Taylor. Discovery of a pulsar in a binary system. *The Astrophysical Journal Letters*, 195:L51–L53, January 1975. doi: 10.1086/181708.

[197] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing In Science & Engineering*, 9(3):90–95, 2007. doi: 10.1109/MCSE.2007.55.

[198] Sascha Husa, Sebastian Khan, Mark Hannam, Michael Pürrer, Frank Ohme, Xisco Jiménez Forteza, and Alejandro Bohé. Frequency-domain gravitational waves from nonprecessing black-hole binaries. I. New numerical waveforms and anatomy of the signal. *Phys. Rev. D*, 93(4):044006, 2016. doi: 10.1103/PhysRevD.93.044006.

[199] Maximiliano Isi and Will M. Farr. Analyzing black-hole ringdowns. 7 2021.

[200] Maximiliano Isi, Matthew Giesler, Will M. Farr, Mark A. Scheel, and Saul A. Teukolsky. Testing the no-hair theorem with GW150914. *Phys. Rev. Lett.*, 123(11):111102, 2019. doi: 10.1103/PhysRevLett.123.111102.

[201] Maximiliano Isi, Will M. Farr, Matthew Giesler, Mark A. Scheel, and Saul A. Teukolsky. Testing the Black-Hole Area Law with GW150914. *Phys. Rev. Lett.*, 127(1):011103, 2021. doi: 10.1103/PhysRevLett.127.011103.

[202] Tousif Islam, Vijay Varma, Jackie Lodman, Scott E. Field, Gaurav Khanna, Mark A. Scheel, Harald P. Pfeiffer, Davide Gerosa, and Lawrence E. Kidder. Eccentric binary black hole surrogate models for the gravitational waveform and remnant properties: comparable mass, nonspinning case. *Phys. Rev. D*, 103(6):064022, 2021. doi: 10.1103/PhysRevD.103.064022.

[203] Tousif Islam, Avi Vajpeyi, Feroz H. Shaik, Carl-Johan Haster, Vijay Varma, Scott E. Field, Jacob Lange, Richard O'Shaughnessy, and Rory Smith. Analysis of GWTC-3 with fully precessing numerical relativity surrogate models. 9 2023.

[204] Kamiel Janssens et al. Correlated 1–1000 Hz magnetic field fluctuations from lightning over Earth-scale distances and their impact on gravitational wave searches. *Phys. Rev. D*, 107(2):022004, 2023. doi: 10.1103/PhysRevD.107.022004.

[205] Vassiliki Kalogera. Spin orbit misalignment in close binaries with two compact objects. *Astrophys. J.*, 541:319–328, 2000. doi: 10.1086/309400.

[206] Michael Kesden, Davide Gerosa, Richard O'Shaughnessy, Emanuele Berti, and Ulrich Sperhake. Effective potentials and morphological transitions for binary black-hole spin precession. *Phys. Rev. Lett.*, 114(8):081103, 2015. doi: 10.1103/PhysRevLett.114.081103.

[207] Sebastian Khan, Sascha Husa, Mark Hannam, Frank Ohme, Michael Pürrer, Xisco Jiménez Forteza, and Alejandro Bohé. Frequency-domain gravitational waves from nonprecessing black-hole binaries. II. A phenomenological model for the advanced detector era. *Phys. Rev. D*, 93(4):044007, 2016. doi: 10.1103/PhysRevD.93.044007.

[208] Sebastian Khan, Katerina Chatziioannou, Mark Hannam, and Frank Ohme. Phenomenological model for the gravitational-wave signal from precessing binary black holes with two-spin effects. *Phys. Rev. D*, 100(2):024059, 2019. doi: 10.1103/PhysRevD.100.024059.

[209] Sebastian Khan, Frank Ohme, Katerina Chatziioannou, and Mark Hannam. Including higher order multipoles in gravitational-wave models for precessing binary black holes. *Phys. Rev. D*, 101(2):024056, 2020. doi: 10.1103/PhysRevD.101.024056.

[210] Lawrence E. Kidder. Coalescing binary systems of compact objects to postNewtonian 5/2 order. 5. Spin effects. *Phys.Rev.*, D52:821–847, 1995. doi: 10.1103/PhysRevD.52.821.

[211] Lawrence E. Kidder, Clifford M. Will, and Alan G. Wiseman. Spin effects in the inspiral of coalescing compact binaries. *Phys. Rev. D*, 47 (10):R4183–R4187, 1993. doi: 10.1103/PhysRevD.47.R4183.

[212] S. Klimenko et al. Method for detection and reconstruction of gravitational wave transients with networks of advanced detectors. *Phys. Rev. D*, 93(4):042004, 2016. doi: 10.1103/PhysRevD.93.042004.

[213] S. Kullback and R. A. Leibler. On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1):79 – 86, 1951. doi: 10.1214/aoms/1177729694. URL https://doi.org/10.1214/aoms/1177729694.

[214] Jack Y. L. Kwok, Rico K. L. Lo, Alan J. Weinstein, and Tjonnie G. F. Li. Investigation of the effects of non-Gaussian noise transients and their mitigation in parameterized gravitational-wave tests of general relativity. *Phys. Rev. D*, 105(2):024066, 2022. doi: 10.1103/PhysRevD.105.024066.

[215] Cedric Lacey and Shaun Cole. Merger rates in hierarchical models of galaxy formation. *Mon. Not. Roy. Astron. Soc.*, 262(3):627–649, 1993. doi: 10.1093/mnras/262.3.627.

[216] M. T. Lam. Optimizing Pulsar Timing Array Observational Cadences for Sensitivity to Low-frequency Gravitational-wave Sources. *Astrophys. J.*, 868(1):33, 2018. doi: 10.3847/1538-4357/aae533.

[217] M. T. Lam, M. A. McLaughlin, J. M. Cordes, S. Chatterjee, and T. J. W. Lazio. Optimal Frequency Ranges for Submicrosecond Precision Pulsar Timing. *Astrophys. J.*, 861(1):12, 2018. doi: 10.3847/1538-4357/aac48d.

[218] Jacob Lange, Richard O'Shaughnessy, and Monica Rizzo. Rapid and accurate parameter inference for coalescing, precessing compact binaries. 5 2018.

[219] K. J. Lee, C. G. Bassa, R. Karuppusamy, M. Kramer, R. Smits, and B. W. Stappers. The optimal schedule for pulsar timing array observations. *Mon. Not. Roy. Astron. Soc.*, 423:2642, 2012. doi: 10.1111/j.1365-2966.2012.21070.x.

[220] Lindley Lentati, P. Alexander, M. P. Hobson, S. Taylor, J. Gair, S. T. Balan, and R. van Haasteren. Hyper-efficient model-independent Bayesian method for the analysis of pulsar timing data. *Physical Review D*, 87(10):104021, May 2013. doi: 10.1103/PhysRevD.87.104021.

[221] Nathaniel Leslie, Liang Dai, and Geraint Pratten. Mode-by-mode relative binning: Fast likelihood estimation for gravitational waveforms with spin-orbit precession and multiple harmonics. *Phys. Rev. D*, 104 (12):123030, 2021. doi: 10.1103/PhysRevD.104.123030.

[222] LIGO Scientific Collaboration and Virgo Collaboration. BayesWave, https://git.ligo.org/lscsoft/bayeswave, 2018. URL https://git.ligo.org/lscsoft/bayeswave.

[223] LIGO Scientific Collaboration and Virgo Collaboration. Bilby, https://git.ligo.org/lscsoft/bilby, 2018. URL https://git.ligo.org/lscsoft/bilby.

[224] LIGO Scientific Collaboration and Virgo Collaboration. *GCN*, 26926, 2020. URL https://gcn.gsfc.nasa.gov/other/S200129m.gcn3.

[225] J. Lin. Divergence measures based on the shannon entropy. *IEEE Transactions on Information Theory*, 37(1):145–151, 1991. doi: 10.1109/18.61115.

[226] Tyson B. Littenberg and Neil J. Cornish. Bayesian inference for spectral estimation of gravitational wave detector noise. *Phys. Rev.*, D91(8):084034, 2015. doi: 10.1103/PhysRevD.91.084034.

[227] Tyson B. Littenberg, Michael Coughlin, Benjamin Farr, and Will M. Farr. Fortifying the characterization of binary mergers in LIGO data. *Physical Review D*, 88(8):084044, October 2013. doi: 10.1103/PhysRevD.88.084044.

[228] Lionel London, Sebastian Khan, Edward Fauchon-Jones, Cecilio García, Mark Hannam, Sascha Husa, Xisco Jiménez-Forteza, Chinmay Kalaghatgi, Frank Ohme, and Francesco Pannarale. First higher-multipole model of gravitational waves from spinning and coalescing black-hole binaries. *Phys. Rev. Lett.*, 120(16):161102, 2018. doi: 10.1103/PhysRevLett.120.161102.

[229] Alessandro Longo, Stefano Bianchi, Wolfango Plastino, Nicolas Arnaud, Antonino Chiummo, Irene Fiori, Bas Swinkels, and Michal Was. Scattered light noise characterisation at the Virgo interferometer with tvf-EMD adaptive algorithm. *Class. Quant. Grav.*, 37(14):145011, 2020. doi: 10.1088/1361-6382/ab9719.

[230] Alessandro Longo, Stefano Bianchi, Guillermo Valdes, Nicolas Arnaud, and Wolfango Plastino. Daily monitoring of scattered light noise due to microseismic variability at the Virgo interferometer. *Class. Quant. Grav.*, 39(3):035001, 2022. doi: 10.1088/1361-6382/ac4117.

[231] Ronaldas Macas and Andrew Lundgren. Sensitive test of non-Gaussianity in gravitational-wave detector data. *Phys. Rev. D*, 108(6):063016, 2023. doi: 10.1103/PhysRevD.108.063016.

[232] Ronaldas Macas, Joshua Pooley, Laura K. Nuttall, Derek Davis, Martin J. Dyer, Yannick Lecoeuche, Joseph D. Lyman, Jess McIver, and

Katherine Rink. Impact of noise transients on low latency gravitational-wave event localization. *Phys. Rev. D*, 105(10):103021, 2022. doi: 10.1103/PhysRevD.105.103021.

[233] Ronaldas Macas, Andrew Lundgren, and Gregory Ashton. Revisiting the evidence for precession in GW200129 with machine learning noise mitigation. *Phys. Rev. D*, 109(6):062006, 2024. doi: 10.1103/PhysRevD.109.062006.

[234] D. M. Macleod, J. S. Areeda, S. B. Coughlin, T. J. Massinger, and A. L. Urban. GWpy: A Python package for gravitational-wave astrophysics. *SoftwareX*, 13:100657, 2021. ISSN 2352-7110. doi: 10.1016/j.softx.2021.100657. URL https://www.sciencedirect.com/science/article/pii/S2352711021000029.

[235] Duncan Macleod, Alex L. Urban, Scott Coughlin, Thomas Massinger, Matt Pitkin, paulaltin, Joseph Areeda, Eric Quintero, The Gitter Badger, Leo Singer, and Katrin Leinweber. gwpy/gwpy: 1.0.1, January 2020. URL https://doi.org/10.5281/zenodo.3598469.

[236] Duncan Macleod, Evan Goetz, Derek Davis, Jeffrey Bidler, Joshua Smith, Michael Lowry, siddharth soni, Alexandra Macedo, Jack Lynam, katerinaleman, Scott Coughlin, Andy Lundgren, Luke Cotnoir, and Thomas Massinger. gwdetchar/gwdetchar: 2.2.7, July 2024. URL https://doi.org/10.5281/zenodo.12786150.

[237] Duncan M. Macleod, Joseph S. Areeda, Scott B. Coughlin, Thomas J. Massinger, and Alexander L. Urban. GWpy: A Python package for gravitational-wave astrophysics. *SoftwareX*, 13:100657, January 2021. doi: 10.1016/j.softx.2021.100657.

[238] Michele Maggiore. *Gravitational Waves: Volume 1: Theory and Experiments*. Oxford University Press, Oxford, online edn edition, 2007. doi: 10.1093/acprof:oso/9780198570745.001.0001. URL https://doi.org/10.1093/acprof:oso/9780198570745.001.0001. Accessed 17 Apr. 2025.

[239] Ilya Mandel and Selma E. de Mink. Merging binary black holes formed through chemically homogeneous evolution in short-period stellar binaries. *Mon. Not. Roy. Astron. Soc.*, 458(3):2634–2647, 2016. doi: 10.1093/mnras/stw379.

[240] Ilya Mandel and Alison Farmer. Merging stellar-mass binary black holes. *Phys. Rept.*, 955:1–24, 2022. doi: 10.1016/j.physrep.2022.01.003.

[241] Pablo Marchant, Norbert Langer, Philipp Podsiadlowski, Thomas M. Tauris, and Takashi J. Moriya. A new route towards merging massive black holes. *Astron. Astrophys.*, 588:A50, 2016. doi: 10.1051/0004-6361/201628133.

[242] LIGO Scientific Collaboration, Virgo Collaboration. LALSuite. https://git.ligo.org/lscsoft/lalsuite, 2018. URL `https://doi.org/10.7935/GT1W-FZ16`.

[243] Grant David Meadors, Keita Kawabe, and Keith Riles. Increasing LIGO sensitivity by feedforward subtraction of auxiliary length control noise. *Class. Quant. Grav.*, 31:105014, 2014. doi: 10.1088/0264-9381/31/10/105014.

[244] Jonathan Merritt, Ben Farr, Rachel Hur, Bruce Edelman, and Zoheyr Doctor. Transient glitch mitigation in Advanced LIGO data. *Phys. Rev. D*, 104(10):102004, 2021. doi: 10.1103/PhysRevD.104.102004.

[245] Cody Messick et al. Analysis Framework for the Prompt Discovery of Compact Binary Mergers in Gravitational-wave Data. *Phys. Rev. D*, 95 (4):042001, 2017. doi: 10.1103/PhysRevD.95.042001.

[246] Patrick M. Meyers, Katarina Martinovic, Nelson Christensen, and Mairi Sakellariadou. Detecting a stochastic gravitational-wave background in the presence of correlated magnetic noise. *Phys. Rev. D*, 102(10):102005, 2020. doi: 10.1103/PhysRevD.102.102005.

[247] M. Coleman Miller and Vanessa M. Lauburg. Mergers of Stellar-Mass Black Holes in Nuclear Star Clusters. *Astrophys. J.*, 692:917–923, 2009. doi: 10.1088/0004-637X/692/1/917.

[248] Simona J. Miller, Maximiliano Isi, Katerina Chatziioannou, Vijay Varma, and Ilya Mandel. GW190521: Tracing imprints of spin-precession on the most massive black hole binary. *Phys. Rev. D*, 109 (2):024024, 2024. doi: 10.1103/PhysRevD.109.024024.

[249] Margaret Millhouse, Neil J. Cornish, and Tyson Littenberg. Bayesian reconstruction of gravitational wave bursts using chirplets. *Phys. Rev.*, D97(10):104057, 2018. doi: 10.1103/PhysRevD.97.104057.

[250] Cameron Mills and Stephen Fairhurst. Measuring gravitational-wave higher-order multipoles. *Phys. Rev. D*, 103(2):024042, 2021. doi: 10.1103/PhysRevD.103.024042.

[251] Chandra Kant Mishra, Aditya Kela, K. G. Arun, and Guillaume Faye. Ready-to-use post-Newtonian gravitational waveforms for binary black holes with nonprecessing spins: An update. *Phys. Rev. D*, 93(8):084054, 2016. doi: 10.1103/PhysRevD.93.084054.

[252] Kentaro Mogushi. Reduction of transient noise artifacts in gravitational-wave data using deep learning, 2021.

[253] Simone Mozzon, Gregory Ashton, Laura K. Nuttall, and Andrew R. Williamson. Does nonstationary noise in LIGO and Virgo affect the estimation of H0? *Phys. Rev. D*, 106(4):043504, 2022. doi: 10.1103/PhysRevD.106.043504.

[254] NANOGrav Scientific Collaboration. Nanograv data. URL https://data.nanograv.org/.

[255] Alex Nitz, Ian Harry, Duncan Brown, Christopher M. Biwer, Josh Willis, Tito Dal Canton, Collin Capano, Thomas Dent, Larne Pekowsky, Gareth S Cabourn Davies, Soumi De, Miriam Cabero, Shichao Wu, Andrew R. Williamson, Bernd Machenschalk, Duncan Macleod, Francesco Pannarale, Prayush Kumar, Steven Reyes, dfinstad, Sumit Kumar, Márton Tápai, Leo Singer, Praveen Kumar, veronica villa, maxtrevor, Bhooshan Uday Varsha Gadre, Sebastian Khan, Stephen Fairhurst, and Arthur Tolley. gwastro/pycbc: v2.3.3 release of pycbc, January 2024. URL https://doi.org/10.5281/zenodo.10473621.

[256] Alexander H. Nitz, Thomas Dent, Tito Dal Canton, Stephen Fairhurst, and Duncan A. Brown. Detecting binary compact-object mergers with gravitational waves: Understanding and improving the sensitivity of the PyCBC search. *The Astrophysical Journal*, 849(2):118, nov 2017. doi: 10.3847/1538-4357/aa8f50. URL https://doi.org/10.3847%2F1538-4357%2Faa8f50.

[257] Alexander H. Nitz, Tito Dal Canton, Derek Davis, and Steven Reyes. Rapid detection of gravitational waves from compact binary mergers with PyCBC Live. *Phys. Rev. D*, 98(2):024050, 2018. doi: 10.1103/PhysRevD.98.024050.

[258] J. R. Oppenheimer and G. M. Volkoff. On massive neutron cores. *Phys. Rev.*, 55:374–381, Feb 1939. doi: 10.1103/PhysRev.55.374. URL https://link.aps.org/doi/10.1103/PhysRev.55.374.

[259] Rich Ormiston, Tri Nguyen, Michael Coughlin, Rana X. Adhikari, and Erik Katsavounidis. Noise Reduction in Gravitational-wave Data via Deep Learning. *Phys. Rev. Res.*, 2(3):033066, 2020. doi: 10.1103/PhysRevResearch.2.033066.

[260] R. O'Shaughnessy, L. London, J. Healy, and D. Shoemaker. Precession during merger: Strong polarization changes are observationally accessible features of strong-field gravity during binary black hole merger. *Phys. Rev. D*, 87(4):044038, 2013. doi: 10.1103/PhysRevD.87.044038.

[261] Serguei Ossokine et al. Multipolar Effective-One-Body Waveforms for Precessing Binary Black Holes: Construction and Validation. *Phys. Rev. D*, 102(4):044055, 2020. doi: 10.1103/PhysRevD.102.044055.

[262] J. P. Ostriker and M. A. Hausman. Cannibalism among the galaxies: dynamically produced evolution of cluster luminosity functions. *The Astrophysical Journal Letters*, 217:L125–L129, November 1977. doi: 10.1086/182554.

[263] The pandas development team. pandas-dev/pandas: Pandas, February 2020. URL https://doi.org/10.5281/zenodo.3509134.

[264] Peter T. H. Pang, Juan Calderón Bustillo, Yifan Wang, and Tjonnie G. F. Li. Potential observations of false deviations from general relativity in gravitational wave signals from binary black holes. *Phys. Rev. D*, 98 (2):024019, 2018. doi: 10.1103/PhysRevD.98.024019.

[265] Chris Pankow et al. Mitigation of the instrumental noise transient in gravitational-wave data surrounding GW170817. *Phys. Rev.*, D98(8): 084016, 2018. doi: 10.1103/PhysRevD.98.084016.

[266] Ethan Payne, Colm Talbot, and Eric Thrane. Higher order gravitational-wave modes with likelihood reweighting. *Phys. Rev. D*, 100(12):123017, 2019. doi: 10.1103/PhysRevD.100.123017.

[267] Ethan Payne, Colm Talbot, Paul D. Lasky, Eric Thrane, and Jeffrey S. Kissel. Gravitational-wave astronomy with a physical calibration model. *Phys. Rev. D*, 102:122004, 2020. doi: 10.1103/PhysRevD.102.122004.

[268] Ethan Payne, Sophie Hourihane, Jacob Golomb, Rhiannon Udall, Derek Davis, and Katerina Chatziioannou. Curious case of GW200129: Interplay between spin-precession inference and data-quality issues. *Phys. Rev. D*, 106(10):104017, 2022. doi: 10.1103/PhysRevD.106.104017.

[269] Ethan Payne, Sophie Hourihane, Jacob Golomb, Rhiannon Udall, Derek Davis, and Katerina Chatziioannou. Data release for the curious case of gw200129: interplay between spin-precession inference and data-quality issues, 2022. URL https://zenodo.org/record/7259655.

[270] Larne Pekowsky, James Healy, Deirdre Shoemaker, and Pablo Laguna. Impact of Higher-order Modes on the Detection of Binary Black Hole Coalescences. *Phys.Rev.*, D87:084008, 2013. doi: 10.1103/PhysRevD.87.084008.

[271] P. C. Peters. Gravitational Radiation and the Motion of Two Point Masses. *Phys. Rev.*, 136:B1224–B1232, 1964. doi: 10.1103/PhysRev.136.B1224.

[272] E. S. Phinney. A Practical theorem on gravitational wave backgrounds. 7 2001.

[273] Eric Poisson and Clifford M. Will. Gravitational waves from inspiraling compact binaries: Parameter estimation using second postNewtonian wave forms. *Phys. Rev. D*, 52:848–855, 1995. doi: 10.1103/PhysRevD.52.848.

[274] Nihan Pol, Stephen R. Taylor, and Joseph D. Romano. Forecasting pulsar timing array sensitivity to anisotropy in the stochastic gravitational wave background. (arXiv:2206.09936), 2022. URL http://arxiv.org/abs/2206.09936.

[275] Ruth Pordes, Don Petravick, Bill Kramer, Doug Olson, Miron Livny, Alain Roy, Paul Avery, Kent Blackburn, Torre Wenaus, Frank Würthwein, Ian Foster, Rob Gardner, Mike Wilde, Alan Blatecky, John McGee, and Rob Quick. The open science grid. In *J. Phys. Conf. Ser.*, volume 78 of *78*, page 012057, 2007. doi: 10.1088/1742-6596/78/1/012057.

[276] Simon Portegies Zwart, Steve McMillan, and Mark Gieles. Young massive star clusters. *Ann. Rev. Astron. Astrophys.*, 48:431–493, 2010. doi: 10.1146/annurev-astro-081309-130834.

[277] Konstantin A. Postnov and Lev R. Yungelson. The Evolution of Compact Binary Star Systems. *Living Rev. Rel.*, 17:3, 2014. doi: 10.12942/lrr-2014-3.

[278] Jade Powell. Parameter Estimation and Model Selection of Gravitational Wave Signals Contaminated by Transient Detector Noise Glitches. *Class. Quant. Grav.*, 35(15):155017, 2018. doi: 10.1088/1361-6382/aacf18.

[279] Geraint Pratten, Sascha Husa, Cecilio Garcia-Quiros, Marta Colleoni, Antoni Ramos-Buades, Hector Estelles, 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. *Phys. Rev. D*, 102(6):064001, 2020. doi: 10.1103/PhysRevD.102.064001.

[280] Geraint Pratten, Patricia Schmidt, Riccardo Buscicchio, and Lucy M. Thomas. Measuring precession in asymmetric compact binaries. *Phys. Rev. Res.*, 2(4):043096, 2020. doi: 10.1103/PhysRevResearch.2.043096.

[281] Geraint Pratten et al. Computationally efficient models for the dominant and subdominant harmonic modes of precessing binary black holes. *Phys. Rev. D*, 103(10):104056, 2021. doi: 10.1103/PhysRevD.103.104056.

[282] Adrian M. Price-Whelan et al. The Astropy Project: Sustaining and Growing a Community-oriented Open-source Project and the Latest Major Release (v5.0) of the Core Package. *The Astrophysical Journal*, 935 (2):167, August 2022. doi: 10.3847/1538-4357/ac7c74.

[283] M. Punturo et al. The Einstein Telescope: A third-generation gravitational wave observatory. *Class. Quant. Grav.*, 27:194002, 2010. doi: 10.1088/0264-9381/27/19/194002.

[284] Michael Pürrer. Frequency domain reduced order model of aligned-spin effective-one-body waveforms with generic mass-ratios and spins. *Phys. Rev.*, D93(6):064041, 2016. doi: 10.1103/PhysRevD.93.064041.

[285] Michael Pürrer and Carl-Johan Haster. Gravitational waveform accuracy requirements for future ground-based detectors. *Phys. Rev. Res.*, 2(2): 023151, 2020. doi: 10.1103/PhysRevResearch.2.023151.

[286] Y. Qin, T. Fragos, G. Meynet, J. Andrews, M. Sørensen, and H. F. Song. The spin of the second-born black hole in coalescing binary black holes. *Astron. Astrophys.*, 616:A28, 2018. doi: 10.1051/0004-6361/201832839.

[287] Etienne Racine. Analysis of spin precession in binary black hole systems including quadrupole-monopole interaction. *Phys. Rev.*, D78:044021, 2008. doi: 10.1103/PhysRevD.78.044021.

[288] Antoni Ramos-Buades, Patricia Schmidt, Geraint Pratten, and Sascha Husa. Validity of common modeling approximations for precessing binary black holes with higher-order modes. *Phys. Rev. D*, 101(10):103014, 2020. doi: 10.1103/PhysRevD.101.103014.

[289] Nayyer Raza, Jess McIver, Gergely Dálya, and Peter Raffai. Prospects for reconstructing the gravitational-wave signals from core-collapse supernovae with Advanced LIGO-Virgo and the BayesWave algorithm. *Phys. Rev. D*, 106(6):063014, 2022. doi: 10.1103/PhysRevD.106.063014.

[290] Daniel J. Reardon et al. Search for an Isotropic Gravitational-wave Background with the Parkes Pulsar Timing Array. *Astrophys. J. Lett.*, 951(1):L6, 2023. doi: 10.3847/2041-8213/acdd02.

[291] David Reitze et al. Cosmic Explorer: The U.S. Contribution to Gravitational-Wave Astronomy beyond LIGO. *Bull. Am. Astron. Soc.*, 51(7):035, 2019.

[292] D. Richstone et al. Supermassive black holes and the evolution of galaxies. *Nature*, 395:A14–A19, 1998.

[293] Florent Robinet, Nicolas Arnaud, Nicolas Leroy, Andrew Lundgren, Duncan Macleod, and Jessica McIver. Omicron: a tool to characterize transient noise in gravitational-wave detectors. *SoftwareX*, 12:100620, 2020. doi: 10.1016/j.softx.2020.100620.

[294] Carl L. Rodriguez, Meagan Morscher, Bharath Pattabiraman, Sourav Chatterjee, Carl-Johan Haster, and Frederic A. Rasio. Binary Black Hole Mergers from Globular Clusters: Implications for Advanced LIGO. *Phys.*

Rev. Lett., 115(5):051101, 2015. doi: 10.1103/PhysRevLett.115.051101. [Erratum: Phys.Rev.Lett. 116, 029901 (2016)].

[295] Elinore Roebber. Ephemeris errors and the gravitational wave signal: Harmonic mode coupling in pulsar timing array searches. *Astrophys. J.*, 876(1):55, 2019. doi: 10.3847/1538-4357/ab100e.

[296] Joseph D. Romano and Neil J. Cornish. Detection methods for stochastic gravitational-wave backgrounds: a unified treatment. *Living Rev. Rel.*, 20(1):2, 2017. doi: 10.1007/s41114-017-0004-1.

[297] Joseph D. Romano, Jeffrey S. Hazboun, Xavier Siemens, and Anne M. Archibald. Common-spectrum process versus cross-correlation for gravitational-wave searches using pulsar timing arrays. *Phys. Rev. D*, 103(6):063027, 2021. doi: 10.1103/PhysRevD.103.063027.

[298] I. M. Romero-Shaw et al. Bayesian inference for compact binary coalescences with bilby: validation and application to the first LIGO–Virgo gravitational-wave transient catalogue. *Mon. Not. Roy. Astron. Soc.*, 499(3):3295–3319, 2020. doi: 10.1093/mnras/staa2850.

[299] I. M. Romero-Shaw et al. Bayesian inference for compact binary coalescences with bilby: validation and application to the first LIGO–Virgo gravitational-wave transient catalogue. *Mon. Not. Roy. Astron. Soc.*, 499(3):3295–3319, 2020. doi: 10.1093/mnras/staa2850.

[300] Isobel M. Romero-Shaw, Nicholas Farrow, Simon Stevenson, Eric Thrane, and Xing-Jiang Zhu. On the origin of GW190425. *Mon. Not. Roy. Astron. Soc.*, 496(1):L64–L69, 2020. doi: 10.1093/mnrasl/slaa084.

[301] Isobel M. Romero-Shaw, Paul D. Lasky, Eric Thrane, and Juan Calderon Bustillo. GW190521: orbital eccentricity and signatures of dynamical formation in a binary black hole merger signal. *Astrophys. J. Lett.*, 903 (1):L5, 2020. doi: 10.3847/2041-8213/abbe26.

[302] Isobel M. Romero-Shaw, Paul D. Lasky, and Eric Thrane. Signs of Eccentricity in Two Gravitational-wave Signals May Indicate a Subpopulation of Dynamically Assembled Binary Black Holes. *Astrophys. J. Lett.*, 921 (2):L31, 2021. doi: 10.3847/2041-8213/ac3138.

[303] Isobel M. Romero-Shaw, Paul D. Lasky, and Eric Thrane. Four Eccentric Mergers Increase the Evidence that LIGO–Virgo–KAGRA's Binary Black Holes Form Dynamically. *Astrophys. J.*, 940(2):171, 2022. doi: 10.3847/1538-4357/ac9798.

[304] Pablo A. Rosado, Alberto Sesana, and Jonathan Gair. Expected properties of the first gravitational wave signal detected with pulsar timing arrays. *Mon. Not. Roy. Astron. Soc.*, 451(3):2417–2433, 2015. doi: 10.1093/mnras/stv1098.

[305] Christian Rover. A Student-t based filter for robust signal detection. *Phys. Rev. D*, 84:122004, 2011. doi: 10.1103/PhysRevD.84.122004.

[306] Christian Rover, Renate Meyer, and Nelson Christensen. Modelling coloured residual noise in gravitational-wave signal processing. *Class. Quant. Grav.*, 28:015010, 2011. doi: 10.1088/0264-9381/28/1/015010.

[307] Surabhi Sachdev et al. The GstLAL Search Analysis Methods for Compact Binary Mergers in Advanced LIGO's Second and Advanced Virgo's First Observing Runs. 1 2019.

[308] Laura Sampson, Neil J. Cornish, and Sean T. McWilliams. Constraining the Solution to the Last Parsec Problem with Pulsar Timing. *Phys. Rev. D*, 91(8):084055, 2015. doi: 10.1103/PhysRevD.91.084055.

[309] L. Santamaria et al. Matching post-Newtonian and numerical relativity waveforms: systematic errors and a new phenomenological model for non-precessing black hole binaries. *Phys. Rev.*, D82:064016, 2010. doi: 10.1103/PhysRevD.82.064016.

[310] B. S. Sathyaprakash and S. V. Dhurandhar. Choice of filters for the detection of gravitational waves from coalescing binaries. *Phys. Rev. D*, 44:3819–3834, 1991. doi: 10.1103/PhysRevD.44.3819.

[311] Peter R. Saulson. *Fundamentals of Interferometric Gravitational Wave Detectors*. World Scientific, 2nd. ed. edition, 2017. ISBN 978-981-314-307-4, 978-981-314-620-4. doi: 10.1142/10116.

[312] Mikhail V. Sazhin. Opportunities for detecting ultralong gravitational waves. *Sov. Astron.*, 22:36–38, 1978.

[313] Patricia Schmidt, Mark Hannam, Sascha Husa, and P. Ajith. Tracking the precession of compact binaries from their gravitational-wave signal. *Phys. Rev. D*, 84:024046, 2011. doi: 10.1103/PhysRevD.84.024046.

[314] Patricia Schmidt, Mark Hannam, and Sascha Husa. Towards models of gravitational waveforms from generic binaries: A simple approximate mapping between precessing and non-precessing inspiral signals. *Phys. Rev. D*, 86:104063, 2012. doi: 10.1103/PhysRevD.86.104063.

[315] Patricia Schmidt, Frank Ohme, and Mark Hannam. Towards models of gravitational waveforms from generic binaries II: Modelling precession effects with a single effective precession parameter. *Phys. Rev.*, D91(2): 024043, 2015. doi: 10.1103/PhysRevD.91.024043.

[316] Patricia Schmidt, Frank Ohme, and Mark Hannam. Towards models of gravitational waveforms from generic binaries: Ii. modelling precession effects with a single effective precession parameter. *Physical Review D*,

91(2), Jan 2015. ISSN 1550-2368. doi: 10.1103/physrevd.91.024043. URL http://dx.doi.org/10.1103/PhysRevD.91.024043.

[317] Igor Sfiligoi, Daniel C Bradley, Burt Holzman, Parag Mhashilkar, Sanjay Padhi, and Frank Wurthwein. The pilot way to grid resources using glideinwms. In *2009 WRI World Congress on Computer Science and Information Engineering*, volume 2 of *2*, pages 428–432, 2009. doi: 10.1109/CSIE.2009.950.

[318] Feroz H. Shaik, Jacob Lange, Scott E. Field, Richard O'Shaughnessy, Vijay Varma, Lawrence E. Kidder, Harald P. Pfeiffer, and Daniel Wysocki. Impact of subdominant modes on the interpretation of gravitational-wave signals from heavy binary black hole systems. *Phys. Rev. D*, 101 (12):124054, 2020. doi: 10.1103/PhysRevD.101.124054.

[319] Steinn Sigurdsson and Lars Hernquist. Primordial black holes in globular clusters. *Nature*, 364:423–425, 1993. doi: 10.1038/364423a0.

[320] Rory J. E. Smith, Gregory Ashton, Avi Vajpeyi, and Colm Talbot. Massively parallel Bayesian inference for transient gravitational-wave astronomy. *Mon. Not. Roy. Astron. Soc.*, 498(3):4492–4502, 2020. doi: 10.1093/mnras/staa2483.

[321] Kentaro Somiya. Detector configuration of KAGRA: The Japanese cryogenic gravitational-wave detector. *Class. Quant. Grav.*, 29:124007, 2012. doi: 10.1088/0264-9381/29/12/124007.

[322] S. Soni et al. Reducing scattered light in LIGO's third observing run. *Class. Quant. Grav.*, 38(2):025016, 2020. doi: 10.1088/1361-6382/abc906.

[323] S. Soni et al. Discovering features in gravitational-wave data through detector characterization, citizen science and machine learning. *Class. Quant. Grav.*, 38(19):195016, 2021. doi: 10.1088/1361-6382/ac1ccb.

[324] Joshua S. Speagle. dynesty: a dynamic nested sampling package for estimating Bayesian posteriors and evidences. *Mon. Not. Roy. Astron. Soc.*, 493(3):3132–3158, 2020. doi: 10.1093/mnras/staa278.

[325] Leo C. Stein. qnm: A Python package for calculating Kerr quasinormal modes, separation constants, and spherical-spheroidal mixing coefficients. *J. Open Source Softw.*, 4(42):1683, 2019. doi: 10.21105/joss.01683.

[326] Nathan Steinle and Michael Kesden. Pathways for producing binary black holes with large misaligned spins in the isolated formation channel. *Phys. Rev. D*, 103(6):063032, 2021. doi: 10.1103/PhysRevD.103.063032.

[327] Benjamin Steltner, Maria Alessandra Papa, and Heinz-Bernd Eggenstein. Identification and removal of non-Gaussian noise transients for gravitational-wave searches. *Phys. Rev. D*, 105(2):022005, 2022. doi: 10.1103/PhysRevD.105.022005.

[328] Ling Sun et al. Characterization of systematic error in Advanced LIGO calibration. *Class. Quant. Grav.*, 37(22):225008, 2020. doi: 10.1088/1361-6382/abb14e.

[329] Colm Talbot and Eric Thrane. Gravitational-wave astronomy with an uncertain noise power spectral density. *Phys. Rev. Res.*, 2(4):043298, 2020. doi: 10.1103/PhysRevResearch.2.043298.

[330] Andrea Taracchini et al. Effective-one-body model for black-hole binaries with generic mass ratios and spins. *Phys. Rev.*, D89(6):061502, 2014. doi: 10.1103/PhysRevD.89.061502.

[331] J. H. Taylor and J. M. Weisberg. A new test of general relativity - Gravitational radiation and the binary pulsar PSR 1913+16. *The Astrophysical Journal*, 253:908–920, February 1982. doi: 10.1086/159690.

[332] S. R. Taylor, L. Lentati, S. Babak, P. Brem, J. R. Gair, A. Sesana, and A. Vecchio. All correlations must die: Assessing the significance of a stochastic gravitational-wave background in pulsar-timing arrays. 95(4):042002, 2017. ISSN 2470-0010, 2470-0029. doi: 10.1103/PhysRevD.95.042002. URL http://arxiv.org/abs/1606.09180.

[333] Stephen R. Taylor. The Nanohertz Gravitational Wave Astronomer. *arXiv e-prints*, art. arXiv:2105.13270, May 2021.

[334] Stephen R. Taylor, Joseph Simon, and Laura Sampson. Constraints On The Dynamical Environments Of Supermassive Black-hole Binaries Using Pulsar-timing Arrays. *Phys. Rev. Lett.*, 118(18):181102, 2017. doi: 10.1103/PhysRevLett.118.181102.

[335] Stephen R. Taylor, Rutger van Haasteren, and Alberto Sesana. From Bright Binaries To Bumpy Backgrounds: Mapping Realistic Gravitational Wave Skies With Pulsar-Timing Arrays. *Phys. Rev. D*, 102(8):084039, 2020. doi: 10.1103/PhysRevD.102.084039.

[336] Stephen R. Taylor, Paul T. Baker, Jeffrey S. Hazboun, Joseph Simon, and Sarah J. Vigeland. enterprise extensions, 2021. URL https://github.com/nanograv/enterprise_extensions. v2.3.3.

[337] Stephen R. Taylor, Joseph Simon, Levi Schult, Nihan Pol, and William G. Lamb. A parallelized Bayesian approach to accelerated gravitational-wave background characterization. *Phys. Rev. D*, 105(8):084049, 2022. doi: 10.1103/PhysRevD.105.084049.

[338] E. Thrane, N. Christensen, R. M. S. Schofield, and A. Effler. Correlated noise in networks of gravitational-wave detectors: subtraction and mitigation. *Phys. Rev. D*, 90(2):023013, 2014. doi: 10.1103/PhysRevD.90.023013.

[339] Eric Thrane and Colm Talbot. An introduction to Bayesian inference in gravitational-wave astronomy: parameter estimation, model selection, and hierarchical models. *Publ. Astron. Soc. Austral.*, 36:e010, 2019. doi: 10.1017/pasa.2019.2. [Erratum: Publ.Astron.Soc.Austral. 37, e036 (2020)].

[340] Eric Thrane, Nelson Christensen, and Robert Schofield. Correlated magnetic noise in global networks of gravitational-wave interferometers: observations and implications. *Phys. Rev. D*, 87:123009, 2013. doi: 10.1103/PhysRevD.87.123009.

[341] C. Tiburzi et al. On the usefulness of existing solar wind models for pulsar timing corrections. *Mon. Not. Roy. Astron. Soc.*, 487(1):394–408, 2019. doi: 10.1093/mnras/stz1278.

[342] C. Tiburzi et al. The impact of solar wind variability on pulsar timing. *Astron. Astrophys.*, 647:A84, 2021. doi: 10.1051/0004-6361/202039846.

[343] Caterina Tiburzi, George Hobbs, Matthew Kerr, William Coles, Shi Dai, Richard Manchester, Andrea Possenti, Ryan Shannon, and Xiaopeng You. A study of spatial correlations in pulsar timing array data. *Mon. Not. Roy. Astron. Soc.*, 455(4):4339–4350, 2016. doi: 10.1093/mnras/stv2143.

[344] Massimo Tinto. Gravitational wave searches with pulsar timing arrays: Cancellation of clock and ephemeris noises. *Phys. Rev. D*, 97(8):084047, 2018. doi: 10.1103/PhysRevD.97.084047.

[345] Vaibhav Tiwari et al. Regression of Environmental Noise in LIGO Data. *Class. Quant. Grav.*, 32(16):165014, 2015. doi: 10.1088/0264-9381/32/16/165014.

[346] Arthur E. Tolley, Gareth S. Cabourn Davies, Ian W. Harry, and Andrew P. Lundgren. ArchEnemy: removing scattered-light glitches from gravitational wave data. *Class. Quant. Grav.*, 40(16):165005, 2023. doi: 10.1088/1361-6382/ace22f.

[347] Hui Tong, Shanika Galaudage, and Eric Thrane. Population properties of spinning black holes using the gravitational-wave transient catalog 3. *Phys. Rev. D*, 106(10):103019, 2022. doi: 10.1103/PhysRevD.106.103019.

[348] Christopher Torrence and Gilbert P. Compo. A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, 79(1):61 – 78, 1998. doi: 10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2. URL https://journals.ametsoc.org/view/journals/bams/79/1/1520-0477_1998_079_0061_apgtwa_2_0_co_2.xml.

[349] Andoni Torres-Rivas, Katerina Chatziioannou, Andreas Bauswein, and James Alexander Clark. Observing the post-merger signal of GW170817-like events with improved gravitational-wave detectors. *Phys. Rev. D*, 99(4):044014, 2019. doi: 10.1103/PhysRevD.99.044014.

[350] A. V. Tutukov and L. R. YungelSon. The merger rate of neutron star and black hole binaries. *Mon. Not. Roy. Astron. Soc.*, 260(3):675–678, 1993. doi: 10.1093/mnras/260.3.675.

[351] Rhiannon Udall and Derek Davis. Bayesian modeling of scattered light in the LIGO interferometers. *Appl. Phys. Lett.*, 122(9):094103, 2023. doi: 10.1063/5.0136896.

[352] Rhiannon Udall, Sophie Hourihane, Simona Miller, Derek Davis, Katerina Chatziioannou, Max Isi, and Howard Deshong. Data for "anti-aligned spin of gw191109: Glitch mitigation and its implications", January 2025. URL https://doi.org/10.5281/zenodo.14611118.

[353] Samantha A. Usman et al. The PyCBC search for gravitational waves from compact binary coalescence. *Class. Quant. Grav.*, 33(21):215004, 2016. doi: 10.1088/0264-9381/33/21/215004.

[354] Gabriele Vajente, Yiwen Huang, Maximiliano Isi, Jenne C. Driggers, Jeffrey S. Kissel, Marek J. Szczepanczyk, and Salvatore Vitale. Machine-learning nonstationary noise out of gravitational-wave detectors. *Phys. Rev. D*, 101(4):042003, 2020. doi: 10.1103/PhysRevD.101.042003.

[355] Avi Vajpeyi, Rory Smith, Eric Thrane, Gregory Ashton, Thomas Alford, Sierra Garza, Maximiliano Isi, Jonah Kanner, T. J. Massinger, and Liting Xiao. A follow-up on intermediate-mass black hole candidates in the second LIGO–Virgo observing run with the Bayes Coherence Ratio. *Mon. Not. Roy. Astron. Soc.*, 516(4):5309–5317, 2022. doi: 10.1093/mnras/stac2332.

[356] M. Vallisneri et al. Modeling the uncertainties of solar-system ephemerides for robust gravitational-wave searches with pulsar timing arrays. 1 2020. doi: 10.3847/1538-4357/ab7b67.

[357] Rutger van Haasteren and Yuri Levin. Understanding and analysing time-correlated stochastic signals in pulsar timing. *mnras*, 428(2):1147–1159, January 2013. doi: 10.1093/mnras/sts097.

[358] Rutger van Haasteren and Michele Vallisneri. New advances in the Gaussian-process approach to pulsar-timing data analysis. *Physical Review D*, 90(10):104012, November 2014. doi: 10.1103/PhysRevD.90.104012.

[359] Rutger van Haasteren, Yuri Levin, Patrick McDonald, and Tingting Lu. On measuring the gravitational-wave background using Pulsar Timing Arrays. *Mon. Not. Roy. Astron. Soc.*, 395:1005, 2009. doi: 10.1111/j.1365-2966.2009.14590.x.

[360] Vijay Varma and Parameswaran Ajith. Effects of nonquadrupole modes in the detection and parameter estimation of black hole binaries with nonprecessing spins. *Phys. Rev. D*, 96(12):124024, 2017. doi: 10.1103/PhysRevD.96.124024.

[361] Vijay Varma, Parameswaran Ajith, Sascha Husa, Juan Calderon Bustillo, Mark Hannam, et al. Gravitational-wave observations of binary black holes: Effect of nonquadrupole modes. *Phys.Rev.*, D90(12):124004, 2014. doi: 10.1103/PhysRevD.90.124004.

[362] Vijay Varma, Scott E. Field, Mark A. Scheel, Jonathan Blackman, Davide Gerosa, Leo C. Stein, Lawrence E. Kidder, and Harald P. Pfeiffer. Surrogate models for precessing binary black hole simulations with unequal masses. *Phys. Rev. Research.*, 1:033015, 2019. doi: 10.1103/PhysRevResearch.1.033015.

[363] Vijay Varma, Davide Gerosa, Leo C. Stein, François Hébert, and Hao Zhang. High-accuracy mass, spin, and recoil predictions of generic black-hole merger remnants. *Phys. Rev. Lett.*, 122(1):011101, 2019. doi: 10.1103/PhysRevLett.122.011101.

[364] Vijay Varma, Sylvia Biscoveanu, Tousif Islam, Feroz H. Shaik, Carl-Johan Haster, Maximiliano Isi, Will M. Farr, Scott E. Field, and Salvatore Vitale. Evidence of Large Recoil Velocity from a Black Hole Merger Signal. *Phys. Rev. Lett.*, 128(19):191102, 2022. doi: 10.1103/PhysRevLett.128.191102.

[365] Leah Vazsonyi and Derek Davis. Identifying glitches near gravitational-wave signals from compact binary coalescences using the Q-transform. *Class. Quant. Grav.*, 40(3):035008, 2023. doi: 10.1088/1361-6382/acafd2.

[366] Aki Vehtari, Daniel Simpson, Andrew Gelman, Yuling Yao, and Jonah Gabry. Pareto smoothed importance sampling, 2015. URL https://arxiv.org/abs/1507.02646.

[367] J. Veitch et al. Parameter estimation for compact binaries with ground-based gravitational-wave observations using the LALInference software

library. *Phys. Rev. D*, 91(4):042003, 2015. doi: 10.1103/PhysRevD.91.042003.

[368] Aaron Viets and Madeline Wade. Subtracting Narrow-band Noise from LIGO Strain Data in the Third Observing Run. Technical report, https://dcc.ligo.org/LIGO-T2100058/public, February 2021.

[369] Justin Vines, Éanna É. Flanagan, and Tanja Hinderer. Post-1-Newtonian tidal effects in the gravitational waveform from binary inspirals. *Phys. Rev. D*, 83:084051, 2011. doi: 10.1103/PhysRevD.83.084051.

[370] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17: 261–272, 2020. doi: 10.1038/s41592-019-0686-2.

[371] C. V. Vishveshwara. Stability of the schwarzschild metric. *Phys. Rev. D*, 1:2870–2879, May 1970. doi: 10.1103/PhysRevD.1.2870. URL https://link.aps.org/doi/10.1103/PhysRevD.1.2870.

[372] Salvatore Vitale, Walter Del Pozzo, Tjonnie G. F. Li, Chris Van Den Broeck, Ilya Mandel, Ben Aylott, and John Veitch. Effect of calibration errors on Bayesian parameter estimation for gravitational wave signals from inspiral binary systems in the Advanced Detectors era. *Phys. Rev.*, D85:064034, 2012. doi: 10.1103/PhysRevD.85.064034.

[373] Salvatore Vitale, Ryan Lynch, John Veitch, Vivien Raymond, and Riccardo Sturani. Measuring the spin of black holes in binary systems using gravitational waves. *Phys. Rev. Lett.*, 112(25):251101, 2014. doi: 10.1103/PhysRevLett.112.251101.

[374] Salvatore Vitale, Davide Gerosa, Carl-Johan Haster, Katerina Chatziioannou, and Aaron Zimmerman. Impact of Bayesian Priors on the Characterization of Binary Black Hole Coalescences. *Phys. Rev. Lett.*, 119(25):251103, 2017. doi: 10.1103/PhysRevLett.119.251103.

[375] Salvatore Vitale, Ryan Lynch, Vivien Raymond, Riccardo Sturani, John Veitch, and Philp Graff. Parameter estimation for heavy binary-black holes with networks of second-generation gravitational-wave detectors. *Phys. Rev. D*, 95(6):064053, 2017. doi: 10.1103/PhysRevD.95.064053.

[376] Salvatore Vitale, Carl-Johan Haster, Ling Sun, Ben Farr, Evan Goetz, Jeff Kissel, and Craig Cahillane. Physical approach to the marginalization of LIGO calibration uncertainties. *Phys. Rev. D*, 103(6):063016, 2021. doi: 10.1103/PhysRevD.103.063016.

[377] Simon D. M. White. Mixing processes in galaxy mergers. *Monthly Notices of the Royal Astronomical Society*, 191(1):1P–4P, 05 1980. ISSN 0035-8711. doi: 10.1093/mnras/191.1.1P. URL https://doi.org/10.1093/mnras/191.1.1P.

[378] Peter Whittle. *Hypothesis Testing in Time Series Analysis*. Almqvist & Wiksells Boktryckeri AB, Uppsala, 1951.

[379] Marcella Wijngaarden, Katerina Chatziioannou, Andreas Bauswein, James A. Clark, and Neil J. Cornish. Probing neutron stars with the full premerger and postmerger gravitational wave signal from binary coalescences. *Phys. Rev. D*, 105(10):104019, 2022. doi: 10.1103/PhysRevD.105.104019.

[380] S. E. Woosley and Alexander Heger. The Pair-Instability Mass Gap for Black Holes. *Astrophys. J. Lett.*, 912(2):L31, 2021. doi: 10.3847/2041-8213/abf2c4.

[381] D. Wysocki, R. O'Shaughnessy, Jacob Lange, and Yao-Lung L. Fang. Accelerating parameter inference with graphics processing units. *Phys. Rev. D*, 99(8):084026, 2019. doi: 10.1103/PhysRevD.99.084026.

[382] Heng Xu et al. Searching for the Nano-Hertz Stochastic Gravitational Wave Background with the Chinese Pulsar Timing Array Data Release I. *Res. Astron. Astrophys.*, 23(7):075024, 2023. doi: 10.1088/1674-4527/acdfa5.

[383] Shunsei Yamamura, Hirotaka Yuzurihara, Takahiro Yamamoto, and Takashi Uchiyama. Extension of the characterization of non-Gaussianity in gravitational wave detectors with a statistical hypothesis test. *Class. Quant. Grav.*, 41(20):205008, 2024. doi: 10.1088/1361-6382/ad7665.

[384] Hang Yu and Rana X. Adhikari. Nonlinear Noise Cleaning in Gravitational-Wave Detectors With Convolutional Neural Networks. *Front. Artif. Intell.*, 5:811563, 2022. doi: 10.3389/frai.2022.811563.

[385] Barak Zackay, Tejaswi Venumadhav, Javier Roulet, Liang Dai, and Matias Zaldarriaga. Detecting gravitational waves in data with non-stationary and non-Gaussian noise. *Phys. Rev. D*, 104(6):063034, 2021. doi: 10.1103/PhysRevD.104.063034.

[386] M Zevin, S Coughlin, S Bahaadini, E Besler, N Rohani, S Allen, M Cabero, K Crowston, A K Katsaggelos, S L Larson, and et al. Gravity spy: integrating advanced ligo detector characterization, machine

learning, and citizen science. *Classical and Quantum Gravity*, 34(6): 064003, Feb 2017. ISSN 1361-6382. doi: 10.1088/1361-6382/aa5cea. URL http://dx.doi.org/10.1088/1361-6382/aa5cea.

[387] Michael Zevin, Christopher P. L. Berry, Scott Coughlin, Katerina Chatziioannou, and Salvatore Vitale. You Can't Always Get What You Want: The Impact of Prior Assumptions on Interpreting GW190412. *Astrophys. J. Lett.*, 899(1):L17, 2020. doi: 10.3847/2041-8213/aba8ef.

[388] Michael Zevin, Simone S. Bavera, Christopher P. L. Berry, Vicky Kalogera, Tassos Fragos, Pablo Marchant, Carl L. Rodriguez, Fabio Antonini, Daniel E. Holz, and Chris Pankow. One Channel to Rule Them All? Constraining the Origins of Binary Black Holes Using Multiple Formation Pathways. *Astrophys. J.*, 910(2):152, 2021. doi: 10.3847/1538-4357/abe40e.

[389] Michael Zevin, Isobel M. Romero-Shaw, Kyle Kremer, Eric Thrane, and Paul D. Lasky. Implications of Eccentric Observations on Binary Black Hole Formation Channels. *Astrophys. J. Lett.*, 921(2):L43, 2021. doi: 10.3847/2041-8213/ac32dc.

[390] Michael Zevin et al. Gravity Spy: Integrating Advanced LIGO Detector Characterization, Machine Learning, and Citizen Science. *Class. Quant. Grav.*, 34(6):064003, 2017. doi: 10.1088/1361-6382/aa5cea.

[391] Rachel C. Zhang, Giacomo Fragione, Chase Kimball, and Vicky Kalogera. On the Likely Dynamical Origin of GW191109 and Binary Black Hole Mergers with Negative Effective Spin. *Astrophys. J.*, 954(1): 23, 2023. doi: 10.3847/1538-4357/ace4c1.

[392] Andrew Zic, George Hobbs, R. M. Shannon, Daniel Reardon, Boris Goncharov, N. D. Ramesh Bhat, Andrew Cameron, Shi Dai, J. R. Dawson, Matthew Kerr, and et al. Evaluating the prevalence of spurious correlations in pulsar timing array data sets. *mnras*, 516(1):410–420, October 2022. doi: 10.1093/mnras/stac2100.

[393] John Zweizig and Keith Riles. Information on self-gating of $h(t)$ used in O3 continuous-wave and stochastic searches. Technical report, https://dcc.ligo.org/LIGO-T2000384, January 2021.