Probing Astrophysics, Cosmology, and Nuclear Physics with Gravitational Waves from Black Holes and Neutron Stars

Thesis by
Jacob Matthew Golomb

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

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2025
Defended May 22, 2025

Jacob Matthew Golomb
ORCID: 0000-0002-6977-670X

ACKNOWLEDGEMENTS

My time at Caltech has been among the most rewarding experiences of my life. Though I began graduate school amid the uncertainty and disruption of the start of the Covid-19 pandemic, I leave with an education, a depth of experience, and lasting friendships that have prepared me to face the world ahead.

Thank you to my family:

To Mom, Dad, Nikki (and Beasley), whose constant support had allowed me to come here and complete my PhD. It is because of your encouragement of my curiosity and offers of support throughout my life that I am able to reach this stage.

To my grandmother, Florence Golomb, who unfortunately passed away during my time in graduate school, but would still be claiming responsiblity for sparking my interest in astrophysics to this day.

To Uncle Jeff for spending time with me, having me over, and going on hikes while living in the LA area.

Thank you to the faculty I have worked with:

To my advisor, Alan Weinstein, who roped me into the field of gravitational wave astrophysics during our very first meeting in 2020 when I was a prospective graduate student. Your contagious enthusiasm for this field and your seemingly endless list of research ideas has been a source of inspiration for me from the start. Thank you for supporting me as your graduate student for the past five years, giving me the freedom (maybe even too much freedom!) to work on pretty much any topic I am interested in, and for tolerating my constant stream of questions for you. Your commitment to your students as a teacher and advisor is reflected in all of our interactions, and I sincerely appreciate your dedication to giving the group a positive and social atmosphere.

To the other faculty members I have been able to work with here at Caltech, for being a source of guidance, support, and inspiration throughout my time here. In particular, to Katerina Chatziiaonnou, with whom I have worked very closely throughout my time at Caltech, for pitching me ideas, mentoring me on several papers and projects, and always making time to discuss various research topics. Working with you has made me a better scientist and you have taught me so much

about black holes, neutron stars, gravitational waves, data analysis, and technical writing and communication.

To Yanbei Chen for allowing me to attend your relativity group meetings, which have featured so many creative ideas and discussions related to a vast array of topics in theoretical physics, always stressing the intuition and first-principles thinking for thinking about problems.

To Will Farr at the Flatiron Institute for always carrying creative project ideas in your back pocket and leading enthusiastic discussions about technical approaches for solving problems. Thank you for your standing invitation to visit Flatiron; my times visiting there have been some of my most rewarding and productive times of graduate school.

Thank you to all of the postdocs I have had the priviledge to work with and learn from these past five years:

To Colm Talbot, for mentoring me since my very first term at Caltech and continuing to do so even after leaving Caltech. Thank you for introducing me to parameter estimation and population inference and involving me in countless projects over the years. You have certainly taught me most of what I know about gravitational wave data analysis.

To Derek Davis, for being a daily source of support, scientific inspiration, and good vibes for my entire time here. Whether from wandering into each other's offices, having lunch at Red Door, or discussing work in a meeting, you have played an enormous role in both my development as a scientist and making my time in graduate school a really fun five years. I have been able to rely on you for thorough discussions of research topics and to always help me to clarify things with me when I am confused. Our group here (as well as the entire LIGO-Virgo-KAGRA Collaboration) would not be the same without you. I am jealous of your future students.

To Arianna Renzini, Ryan Magee, Pat Meyers, and Lucy Thomas for regularly taking time to work with me, discuss research, and hang out at social hours, lunches, and while traveling for conferences. I have learned so much from all of you and you all have made such a great impact on my graduate school career.

To Sophie Bini and Jane Glazer who joined the group during my final year but I have fortunately had the chance to get to know and work with.

To Max Isi, for being a mentor at Flatiron, and adopting several of us Caltech graduate students. Working with you has taught me so much about black holes, gravitational waves, data analysis (you introduced me to JAX!), and how to be an effective researcher.

Thank you to my fellow graduate students I have had the pleasure of working with, particularly those in the LIGO Lab and TAPIR, for being great colleagues and friends at Caltech:

To Isaac Legred, Ethan Payne, Rhiannon Udall, Sophie Hourihane, Alvin Li, Rico Lo, and Simona Miller, I could not have asked for a better group to be my West Bridge office mates. Thank you for putting up with my constant ramblings and engaging in impromptu discussions, some of which have turned into impactful science and others just amusing distractions. Together, our office has reached the combined state of being productive and entertaining.

To Brian Seymour, Colin Weller, James Gardner, Su Direkci, Andrew Laeguer in TAPIR for insightful and amusing discussions at group meetings and pizza lunches.

Thank you to the staff scientists, particularly Stuart Anderson and Jonah Kanner, for their advice and support.

Thank you to all of my friends I have made during graduate school. You have all really made these five years so much fun and a greater experience than I could have ever imagined. Special shoutout to my roommates of three years: Andrew Lu, Victoria Tobin, Ethan Payne, and the cats Alan and Humphrey. I could not have asked for better roommates, even when you steal my food and break into my room.

To Hershey, Sheva, and the Stolik kids for hosting Shabbat dinners and events on holidays. I cannot express enough how much I appreciate your successful efforts at building a community at Caltech. Thank you for inviting me over so regularly and continuing to invite me even when I can be bad about checking my messages.

Thank you to the undergraduates, both Caltech students and external SURF students, I have had the opportunity to work with: Noah Wolfe, Jingyi Zhang, Daniela Yano, April Cheng, Sterling Scarlett, Jake Summers, and Olivia Durrett. I have been sincerely impressed working with all of you. I enjoyed every project and have learned so much working together.

But see that the imagination of nature is far, far greater than the imagination of man.

– Richard Feynman

What we think of as smooth simple space is really a wiggly business.

– John A. Wheeler

ABSTRACT

Gravitational waves now serve as a powerful tool for studying physics of compact objects, including black holes and neutron stars. When two compact objects merge, they emit gravitational waves that encode information about their masses, spins, and orbital dynamics. Ground-based detectors capture these signals, allowing us not only to measure the properties of individual mergers but also to characterize the population properties of black holes and neutron stars. In this thesis, I present a collection of works using real and simulated gravitational wave observations of compact binary coalescences to study the physics of black holes and neutron stars, and the implications these observations have on our broader understanding of astrophysics and fundamental physics.

The first part of this thesis is background material reviewing some of the theory behind gravitational waves. The second part focuses on measuring the physical properties of a compact binary coalescence detected in gravitational wave data. This includes the methods and models used in parameter estimation and a presentation of the properties of detections in the fourth Gravitational Wave Transient Catalog (GWTC-4). The third part of this thesis turns to measuring and extracting astrophysical information from the population properties of compact binaries. This features the astrophysical distributions of binary black holes as inferred from GWTC-3 and GWTC-4. I also present studies measuring specific aspects of the binary black hole mass and spin distributions, and the implications these results have for understanding binary black hole formation channels and stellar astrophysics. This section additionally features applications of population inference to studies of large-scale structure and predictions for the gravitational wave stochastic background, as well as technical discussions of the methods and custom libraries used to implement population analyses and potential biases associated with commonly-used methods. The fourth part explores how properties of dense nuclear matter are encoded in observations of neutron stars. This section includes studies using our knowledge of the nuclear equation of state to classify low-mass compact binary mergers, and results from using gravitational waves and electromagnetic observations of neutron stars to measure the equation of state and neutron star population properties.

PUBLISHED CONTENT AND CONTRIBUTIONS

[1] A. G. Abac et al. "GWTC-4: Compact Binary Coalescences Observed by LIGO, Virgo, and Kagra During the First Half of the Fourth Observing Run". In: *Astrophys. J.* (2025 in prep.).
J.G. contributed to the development of the analysis pipelines and methods, and produced and reviewed parameter estimation results for the catalog.

[2] A. G. Abac et al. "GWTC-4: Methods for identifying and characterizing gravitational-wave transients". In: *Astrophys. J.* (2025 in prep.).
J.G. contributed to the development of the analysis pipelines and methods.

[3] A. G. Abac et al. "Population Properties of Merging Compact Objects from the Fourth LIGO-Virgo-KAGRA Gravitational-Wave Transient Catalog". In: *Astrophys. J.* (2025 in prep.).
J.G. contributed to the development of the analysis pipelines and methods, produced results for several parametric population models, and conducted reviews.

[4] A. Makai Baker, Paul D. Lasky, Eric Thrane, et al. "Significant challenges for astrophysical inference with next-generation gravitational-wave observatories". In: (Mar. 2025). arXiv: 2503.04073 [gr-qc].
J.G. wrote the code for the detector response in the presence of effects from Earth rotation and finite arm length (free spectral range effects).

[5] Jacob Golomb, Isaac Legred, Katerina Chatziioannou, et al. "Interplay of astrophysics and nuclear physics in determining the properties of neutron stars". In: *Phys. Rev. D* 111.2 (2025), p. 023029. DOI: 10.1103/PhysRevD.111.023029. arXiv: 2410.14597 [astro-ph.HE].
J.G. participated in conceiving the project, collecting data, developing and implementing analysis methods, and interpreting and writing results.

[6] Jacob Golomb, Maximiliano Isi, and Will M. Farr. "Physical Models for the Astrophysical Population of Black Holes: Application to the Bump in the Mass Distribution of Gravitational-wave Sources". In: *Astrophys. J.* 976.1 (2024), p. 121. DOI: 10.3847/1538-4357/ad8572. arXiv: 2312.03973 [astro-ph.HE].
J.G. led simulations and analyses, as well as results interpretations and presentations and paper writing.

[7] Jacob Golomb, Isaac Legred, Katerina Chatziioannou, et al. "Using equation of state constraints to classify low-mass compact binary mergers". In: *Phys. Rev. D* 110.6 (2024), p. 063014. DOI: 10.1103/PhysRevD.110.063014. arXiv: 2403.07697 [astro-ph.HE].

[8] Arianna I. Renzini and Jacob Golomb. "Projections of the uncertainty on the compact binary population background using popstock". In: *Astron.*

Astrophys. 691 (2024), A238. DOI: 10.1051/0004-6361/202451374. arXiv: 2407.03742 [astro-ph.CO].
J.G. developed code for the analyses in this work and participated in writing, presenting, and interpreting results.

[9] Colm Talbot, Amanda Farah, Shanika Galaudage, et al. "GWPopulation: Hardware agnostic population inference for compact binaries and beyond". In: *J. Open Source Softw.* (Sept. 2024). DOI: 10.21105/joss.07753. arXiv: 2409.14143 [astro-ph.IM].
J.G. is a core developer of the associated codebase GWPopulation.

[10] R. Abbott et al. "Population of Merging Compact Binaries Inferred Using Gravitational Waves through GWTC-3". In: *Phys. Rev. X* 13.1 (2023), p. 011048. DOI: 10.1103/PhysRevX.13.011048. arXiv: 2111.03634 [astro-ph.HE].
J.G. conducted analyses with parametric population models for masses, spins, and redshift distributions, as well as participated in paper writing, review, and made contributions to various sections of the paper.

[11] Jacob Golomb and Colm Talbot. "Searching for structure in the binary black hole spin distribution". In: *Phys. Rev. D* 108.10 (2023), p. 103009. DOI: 10.1103/PhysRevD.108.103009. arXiv: 2210.12287 [astro-ph.HE].
J.G. led coding for this project, as well as simulations, data analysis, results presentation, and paper writing.

[12] Colm Talbot and Jacob Golomb. "Growing pains: understanding the impact of likelihood uncertainty on hierarchical Bayesian inference for gravitational-wave astronomy". In: *Mon. Not. Roy. Astron. Soc.* 526.3 (2023), pp. 3495–3503. DOI: 10.1093/mnras/stad2968. arXiv: 2304.06138 [astro-ph.IM].
J.G. participated in derivations, paper writing, and presentation of results.

[13] Noah E. Wolfe, Colm Talbot, and Jacob Golomb. "Accelerating tests of general relativity with gravitational-wave signals using hybrid sampling". In: *Phys. Rev. D* 107.10 (2023), p. 104056. DOI: 10.1103/PhysRevD.107.104056. arXiv: 2208.12872 [gr-qc].
J.G. contributed to results presentations, and discussions about method devleopment and implementation.

[14] Yanyan Zheng, Nikolaos Kouvatsos, Jacob Golomb, et al. "Angular Power Spectrum of Gravitational-Wave Transient Sources as a Probe of the Large-Scale Structure". In: *Phys. Rev. Lett.* 131.17 (2023), p. 171403. DOI: 10.1103/PhysRevLett.131.171403. arXiv: 2305.02652 [astro-ph.CO].
J.G. led development of simulation campaign and parameter estimation pipeline and contributed to analysis methods and implementation along with presentation in the paper.

[15] Jacob Golomb and Colm Talbot. "Hierarchical Inference of Binary Neutron Star Mass Distribution and Equation of State with Gravitational Waves". In: *Astrophys. J.* 926.1 (2022), p. 79. DOI: 10.3847/1538-4357/ac43bc.

arXiv: 2106.15745 [astro-ph.HE].
J.G. led the development of the analysis code, results, and paper writing.

[16] Ethan Payne, Sophie Hourihane, Jacob Golomb, et al. "Curious case of GW200129: Interplay between spin-precession inference and data-quality issues". In: *Phys. Rev. D* 106.10 (2022), p. 104017. DOI: 10.1103/PhysRevD.106.104017. arXiv: 2206.11932 [gr-qc].
J.G. participated in formulating the project, implementing parameter estimation, interpreting and presenting results, and paper writing.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

Number *Page*

LIST OF TABLES

This work does not reflect the scientific opinion of the LIGO Scientific Collaboration and it has not been reviewed by the collaboration.

Part I

Background

Chapter 1

MOTIVATION FOR THIS THESIS

The development of general relativity came with predictions for unique phenomena that were absent in the old Newtonian theory of gravity. Einstein's theory of special relativity already revealed that Newtonian mechanics broke down when describing the behavior of objects moving at velocities approaching the speed of light. General relativity (GR), on the other hand, revealed the limitations of Newtonian gravity when describing objects in strong gravitational fields, sufficiently close to a massive, gravitating body. As a result, the relativistic and Newtonian descriptions diverge when describing an object that moves very fast or travels deep into a gravitational potential—regimes now directly accessible through gravitational wave observations.

The coupling between mass and the gravitational field is extremely weak; the gravitational force obtained from G coupling to typical masses is orders of magnitude smaller than the electromagnetic coupling obtained from typical electric charges. Given the relative weakness of gravity, the effects of GR become most apparent in astrophysical environments, where we observe the motion of objects close to massive bodies. Crucially, these interactions can happen at small enough distances just outside the massive source, such that they probe the physics of the strong gravitational field. Classic tests—explaining Mercury's perihelion precession and observing starlight deflection by the Sun in 1919—were early examples; compact binaries, systems composed of two dense stellar remnants in orbit, probe far stronger fields and involve objects that are far more exotic and unfamiliar.

A major consequence from replacing Newtonian gravity with GR is that gravity is not treated like a force in the usual sense, but as a manifestation of curvature in spacetime. This leads to predictions absent in the Newtonian treatment: as it turns out, GR says that the spacetime manifold itself can carry wavelike perturbations (ripples) which carry energy and may be detected by distant observers with sufficiently sensitive instruments. These ripples, known as gravitational waves, can be generated in environments where the effects of GR become important: massive bodies with close gravitational interactions, exerting asymmetric motion. Two-body systems composed of compact remnants of dead stars, such as neutron stars, black holes, and white dwarfs, are ideal candidates for producing gravitational waves, as they

are massive and dense, and can interact deeply in the gravitational potential of the other.

Gravitational waves encode information about these extreme environments offering a fundamentally new way to study the Universe—one that complements and in many cases surpasses other methods. Direct probes of strong gravity environments are difficult to develop; it is obviously not feasible to send a probe to the vicinity of such an object, given the distances from Earth and the extreme conditions involved. Black holes do not directly emit their own electromagnetic radiation; one must rely on poorly-understood mechanisms of emission and propagation (possibly through intervening matter) based on interactions with their environments to study them in traditional astronomy. Gravitational waves are sourced purely by mass-energy dynamics and travel to Earth virtually unimpeded. This makes them ideal candidates for probing the strong gravity around compact objects. Detecting them, however, requires both powerful sources and exquisitely sensitive detectors.

Originally attempted by Joseph Weber in the 1960s, modern gravitational wave detection was pioneered by the LIGO (Laser Interferometer Gravitational Wave Observatory) detectors, developed under the leadership of Rainer Weiss, Kip Thorne, and Ronald Drever. These detectors, located in Livingston, Louisiana and Hanford, Washington, use laser interferometry to measure the minute changes in distance between two mirrors suspended in vacuum, separated by 4 km. These make some of the most precise measurements of any manmade instrument ever created; for example, the LIGO detectors measure relative changes on the order of 10^{-23}, or 10^{-19} meters over the distance of its 4 km arms ($\sim 1/10,000$th of the width of a proton).

I began graduate school in 2020, five years after the LIGO detectors made their first detection of gravitational waves, from the merger of two black holes approximately 30 times the mass of our sun, and located around 400 Mpc (~ 1.3 billion light-years) from us. This event, known as GW150914, welcomed a brand new messenger into astronomy, sending information from the most extreme grasp of spacetime previously inaccessible to humans. Since then, LIGO, now joined by its European partner Virgo, has detected hundreds of signals from compact binary coalescences, most of which have been from black hole mergers. The first two observing runs (O1-O2) detected 11 events. I joined the field shortly after the end of the third observing run (O3), participating in the subsequent data analysis, extracting the astrophysical information contained in the ~ 100 events observed through then.

Since then, I have been heavily involved in the analysis of the data collected during the (ongoing) fourth observing run (O4), which has expanded the total number of high-significance gravitational wave events to $\sim$ 300, enriching our census of the previously-unseen parts of the Universe.

Treating these signals as probes of their sources and of the Universe through which they propagate allows us to learn information previously out of reach for humans. Motivated to participate in the development of this new field, my research has been focused on how to use gravitational waves from black holes and neutron stars as a tool to learn astrophysics, cosmology, and fundamental physics. In addition to exploiting new data, while taking advantage of techniques now afforded to us by modern computation, I have adopted related methods and developed new tools for conducting analyses on gravitational wave data. My contributions to the field can largely be fit into addressing the following questions:

- *What physical properties of a gravitational wave source do we learn from a signal appearing in the data collected by a detector?*
- *What does the population of gravitational wave sources reveal about the astrophysical properties of the progenitors of compact objects?*
- *What do gravitational waves tell us about fundamental physics, cosmology, and the behavior of dense matter?*
- *How can we best apply modern and novel computational and mathematical techniques to answer these questions?*

I address these questions in this thesis, organized in the following parts:

Background overviews the basic concepts of GR and gravitational waves, including their generation and detection.

Measuring Properties of Black Holes and Neutron Stars Using Gravitational Waves from Compact Binary Coalescences covers the methods we use to measure the physical properties of gravitational wave sources, and includes results from analyses from O4 that I contributed to.

Astrophysics and Cosmology from the Population of Compact Binaries discusses how population inference works and what the population of binary black hole mergers reveals about astrophysics and cosmology.

Matter Effects and Equation of State Implications is related to learning and exploiting effects from nuclear physics in neutron star observations.

Conclusions summarizes the main results of my work and discusses future directions for the field.

Chapter 2

INTRODUCTION TO GRAVITATIONAL WAVE THEORY

2.1 General Relativity

In 1905, Albert Einstein upended the Newtonian paradigm that had been the accepted description of mechanics for hundreds of years, resolving the tension between mechanics and Maxwell's equations of electromagnetism. In doing so, he showed that the correct transformation that can relate inertial reference frames preserves physical laws and the consistent speed of light, c. Crucially, special relativity was applicable for high velocities (relative to c) and reduced to Newtonian mechanics at the more familiar lower velocities typical in everyday experience. However, it was not clear how this described the experience of observers under the effects of gravity, which then would imply acceleration and therefore being in non-inertial reference frames. Therefore, it became necessary to figure out how the new theory of relativity generalized to include effects of gravity. In this section, I review the principles and consequences of general relativity (GR), under assumption of suitable mathematical foundations. Introductory reviews can be found in, e.g., [383, 492, 132]

Isaac Newton, Galileo Galilei, and others knew that objects in an uniform gravitational field fall with the same acceleration, regardless of their mass. This is known as the *weak equivalence principle*. This is implied by recognizing that *inertial mass* (m_i), that is, the mass that is accelerated by a force in Newton's second law, $F = m_i a$, is equal to the gravitational mass m_g in Newton's law of gravitation. In other words, the gravitational mass of the source M_g, sources the field that couples to the gravitational mass to the probe m_g as:

$$F = m_i a = \frac{G m_g M_g}{r^2}, \tag{2.1}$$

if $m_i = m_g$, then the acceleration a of an object in a gravitational field is independent of its mass, consistent with the results of Galileo's famous experiments.

This has an interesting and important implication: one can recreate the effects of gravitation by being in an accelerated frame. An observer situated in a small, sealed room feeling an acceleration $g = 9.81\mathrm{m/s}^2$ could not tell by dropping objects if they are on Earth or accelerating at g in empty space (say, in a rocket with its engine on); objects in free fall behave the same, regardless of whether the acceleration

is caused by gravity or other means. If inertial and gravitational masses were not equivalent, the object's acceleration due to gravity would depend on its mass, so the observer would be able to drop objects of different masses to determine if they are in a gravitational field.

Taking this a step further, Einstein postulated that there should be no local experiment an observer can perform to determine whether they are in a gravitational field or in a uniformly accelerating rocket. In both cases, all local measurements—such as the ticking rate of a clock or the speed of light—yield the same results.

Because an observer in free fall experiences weightlessness (i.e., they feel no support force), they do not feel what we would call "acceleration due to gravity". In Einstein's theory, this means that a freely falling observer is actually in an inertial frame, even though they would appear to accelerate in a Newtonian description. Being in an inertial frame and unable to determine whether they are under the influence of gravity, the freely falling observer finds that all local experiments obey the laws of special relativity.

Using this principle, one can derive the existence of a novel effect — gravitational redshift. Consider a tower of height h on Earth. An observer at the bottom of the tower emits a light pulse of frquency f_0, and an observer at the top of the tower receives the light pulse. Assume h is sufficiently small that the gravitational field is uniform, so this counts as a "local measurement". Our goal is to determine the frequency f the observer at the top sees. We can determine how light is affected by gravity in relativity by examining the system in an equivalent configuration: replace the tower with a rocket of height h accelerating at g in empty space.

The time $\Delta t = h/c$ elapses between the emission of the light pulse and its arrival at the top of the tower. In the time between it is emitted and received, the rocket has changed its velocity by $\Delta v = g\Delta t = gh/c$. Imagine that the observer at the top needs to receive at least one wavelength of light (equivalently, observe it for one full period) in order to detect it. If we assume that the velocity of the rocket is approximately constant over one period of the light wave, then $g/f \ll gh/c \to \lambda \ll h$, which is almost certainly the case for practical light sources. In this case, we can enter a reference frame that is momentarily comoving with the receiver at the top and we can therefore use regular special relativity to describe their observations during that moment. In that case, we can recognize that the frequency shift observed by the receiver is simply the normal special relativistic Doppler effect, which depends only on the instantaneous velocity. This reduces to the usual Doppler shift for $v \ll c$, so

$f = (1 - v/c) f_0$, where we assume $v \ll c$, so we can use the normal, nonrelativistic Doppler expression. If the light was emitted when the rocket was at rest, then $\Delta v = v$ and $f = (1 - gh/c^2) f_0$ is observed by the observer at the top of the rocket. Under the Einstein equivalence principle, this is equivalent to the case of the tower of height h on Earth. Therefore, we conclude gravity should cause redshifting along a vertical height, given by:

$$f = (1 - gh/c^2) f_0, \tag{2.2}$$

when h is small enough that g is treated as constant ($gh/c^2 \ll 1$).

Now, we can take a step further and show that the equivalence between accelerating and gravitating systems breaks down for non-local experiments, hinting at the geometric nature of gravity. Consider the configuration in Figure 2.1. Imagine a platform (grey box) above a planet of mass M. Pretend the platform is attached to the planet, so it is not in free-fall and anything that stands on the platform is accelerating. An observer standing on the platform, unable to see through it, will not be able to do any local experiments to determine if they are above the planet with mass M or if the platform is accelerating through space at the corresponding acceleration. Let's see what happens when we do an experiment that we will term non-local; that is, we will conduct an experiment involving two balls (which will be used as probes), where the distance between the probes is non-negligible.

An observer on the platform holds two red balls the same distance dy above the platform. He holds them a horizontal distance δx apart. We will label the ball in the center of the platform as 1 and on the right as 2. If he is in a uniformly accelerating system accelerating at a, then the balls will take the same time to fall in the observer coordinate time: $dt_1^2 = dt_2^2 = \frac{2dy}{a}$. However, what happens when the observer is actually under the influence of gravity? In that case the acceleration is non-uniform (i.e., tidal effects) and the observer will notice differences in the behavior of the two balls falling if his arms are long enough and δx is large.

Specifically, let $\vec{g}_1$ and $\vec{g}_2$ be the acceleration due to gravity of the two balls, respectively. We know $g = GM/r^2$, so $\frac{g_2}{g_1} = (\frac{R_1}{R_2})^2$, determined by the ratio of the distances between the balls and the planet. Ball 1 is dropped directly above the planet and thus has no x-component of acceleration. Ball 2 is dropped a horizontal distance δx away, so $\vec{g}_2$ has both x and y components, labeled g_x and g_y (suppressing the subscript 2, as it is clear this is not referring to components of $\vec{g}_1$).

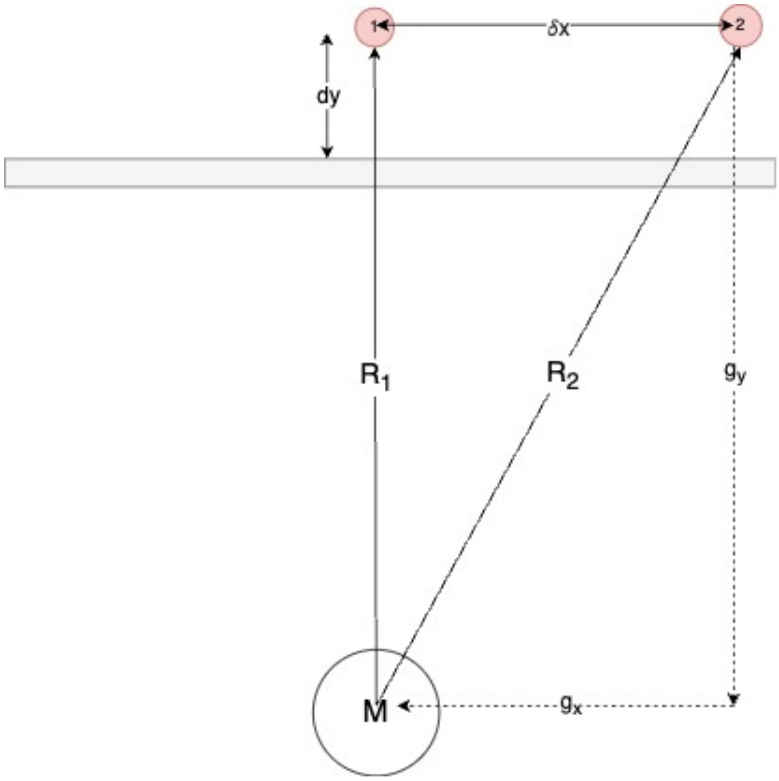

Figure 2.1: Thought experiment to illustrate the break down of the equivalence principle for non-local experiments. An observer on a platform above a planet of mass M holds two balls the same distance dy above the platform. He holds them a horizontal distance δx apart. The ball in the center of the platform is labeled 1 and the one on the right is labeled 2. R_i is the radial distance from the center of the planet to the ball i. The observer cannot see through the platform and therefore cannot tell visually if they are standing above a planet or if the platform is accelerating.

From trigonometry, one can see that:

$$g_x = g_2 \frac{\delta x}{R_2} \tag{2.3}$$

and

$$g_y = g_2 \frac{R_1}{R_2} = g_1 \left(\frac{R_1}{R_2}\right)^3. \tag{2.4}$$

With $R_2^2 = R_1^2 + \delta x^2$, we can expand:

$$R_2 = \sqrt{R_1^2 + \delta x^2} = R_1 \sqrt{1 + \frac{\delta x^2}{R_1^2}} \approx R_1 \left(1 + \frac{\delta x^2}{2R_1^2}\right), \tag{2.5}$$

so,

$$\left(\frac{R_1}{R_2}\right)^3 \approx \left(1 + \frac{\delta x^2}{2R_1^2}\right)^{-3} \approx 1 - \frac{3\delta x^2}{2R_1^2}. \tag{2.6}$$

This immediately leads us to see that the two balls will fall at different rates, as kinematics tells us:

$$\frac{dt_2^2}{dt_1^2} = \frac{g_1}{g_y} \approx (1 - \frac{3\delta x^2}{2R_1^2})^{-1} \approx 1 + \frac{3\delta x^2}{2R_1^2}. \tag{2.7}$$

Therefore, if the observer is sensitive to changes on the order $O(\frac{\delta x}{R_1})^2$, they will be able to tell that $dt_2^2 \neq dt_1^2$, and they are therefore not in a uniformly accelerating frame. Now, to make a tenuous connection to spacetime geometry, consider the metric distance the observer measures while the ball falls.

The metric distance between two events at (t_1, x_1, y_1, z_1) and (t_2, x_2, y_2, z_2) is given by:

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu = g_{tt}dt^2 + g_{xx}dx^2 + g_{yy}dy^2 + g_{zz}dz^2, \tag{2.8}$$

where $(x^0, x^1, x^2, x^3) = (t, x, y, z)$ are coordinates of the events. For flat spacetime, the metric is the Minkowski metric, $g_{\mu\nu} = \eta_{\mu\nu}$ = diag(-1,1,1,1). Geometry tell us that locally a (pseudo-)Riemannian manifold can be approximated by a flat metric, which physically means small enough regions look like flat spacetime with mechanics described by special relativity. An observer of a local experiment can always boost into a reference frame where the change of coordinates make ds^2 the canonical Minkowski result (where constant terms of $g_{\mu\nu}$ are $(-1, 1, 1, 1)$ and first derivatives of $g_{\mu\nu}$ vanish).

Locally, this observer approximate himself in an inertial frame, seeing flat spacetime and therefore calculate ds_1^2 using the familiar Minkowski metric from special relativity:

$$ds_1^2 = -dt_1^2 + dy^2, \tag{2.9}$$

where dt and dy are the coordinate time and vertical distances measured by the observer on the platform. He now attempts to do the same for the second ball, which will also include displacement in the x-direction $dx = \frac{1}{2}g_x dt_2^2$. Using the expressions for g_x and dt_2 above, we find $dx^2 = \left(\frac{\delta x}{R_1}\right)^2 dy^2$. Therefore, for the second ball dropping:

$$ds_2^2 = -dt_2^2 + dy^2 + dx^2 = -\left(1 + \frac{3}{2}(\frac{\delta x}{R_1})^2\right)dt_1^2 + \left(1 + (\frac{\delta x}{R_1})^2\right)dy^2, \tag{2.10}$$

which differs from ds_1^2 by terms of order $O(\frac{\delta x}{R_1})^2$. One can show that *no coordinate transformation can introduce these quadratic coordinate terms in the metric necessary to make the ds^2 terms equivalent*. This is unlike the gravitational redshift case (see Eq. 2.2), which shows that the gravitational redshift is a first-order effect

in $\Delta x = h$ for the observer in the small tower. The first order effect can be gauged away by boosting into a freely-falling frame momentarily, meaning there is a local inertial frame that does not see any redshift. In this new case, once the tidal terms can be measured, the observer does not have enough gauge freedom to transform into a frame that makes the metric agree with the flat ds^2. Because δx is finite, one cannot choose a single locally inertial frame that sets $g_{\mu\nu} = \eta_{\mu\nu}$ and $\partial g = 0$ at both events.

Thus, an experiment sensitive to second order deviations in the separation of the two balls (relative to the distance to the gravitating body), will notice that the spacetime metric is inconsistent with the flat, Minkowski metric from special relativity. Clearly, this reduces to the result from special relativity as the observer gets farther from the gravitating body or drops ball 2 closer to ball 1. The difference between special relativity and relativity in gravity has something to do with the geometry of spacetime. Specifically, because in an inertial frame, we can expand the metric (around a point) as $g_{\mu\nu} = \eta_{\mu\nu} - \frac{1}{3}R_{\mu\sigma\nu\rho}x^\sigma x^\rho + O(x^3)$, where $R_{\mu\sigma\nu\rho}$ the existence of the terms that deciate from Minkowski by $O(x^2)$ indicates a nonvanishing Riemman tensor. This signals curvature in the spacetime we are modeling as a manifold.

For the rest of this chapter, I will assume the reader is familiar with other basic aspects of GR.

2.2 Gravitational Waves

We now examine what happens when the gravitational field varies with time. To do this we will find solutions to the Einstein field equations $G_{\mu\nu} = 8\pi T_{\mu\nu}$ (where $c = G = 1$) for linear order perturbations. More thorough derivations can be found in GR references [132, 383, 492]. We begin by expanding the metric to first order about the Minkowski metric:

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}. \tag{2.11}$$

After a lot of manipulation, we find:

$$G^{(1)}_{\mu\nu} = \frac{1}{2}\Big(\partial_\sigma\partial_\mu h^\sigma{}_\nu + \partial_\sigma\partial_\nu h^\sigma{}_\mu - \partial_\mu\partial_\nu h - \Box\, h_{\mu\nu} - \eta_{\mu\nu}\,\partial_\alpha\partial_\beta h^{\alpha\beta} + \eta_{\mu\nu}\,\Box\, h\Big). \tag{2.12}$$

We now consider our coordinate freedom and choose to work in a gauge that makes it simpler to make physical predictions. We first define the trace-reversed metric perturbation:

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}h, \tag{2.13}$$

where $\bar{h} = -h$. In electromagnetism, we can write Maxwell's equations in a particularly simple and convenient form by recognizing the gauge freedom in specifying the vector potential, specifically by specifying the Lorenz gauge condition. Similarly, we can impose the equivalent Lorenz gauge[1]:

$$\partial^\mu \bar{h}_{\mu\nu} = 0, \tag{2.14}$$

which reduces Eq. 2.12 to:

$$G_{\mu\nu} = -\frac{1}{2}\Box \bar{h}_{\mu\nu}. \tag{2.15}$$

This makes the (linearized) Einstein equations look like:

$$\Box \bar{h}_{\mu\nu} = -16\pi T_{\mu\nu}. \tag{2.16}$$

In a vacuum, $T_{\mu\nu} = 0$, so $\Box \bar{h}_{\mu\nu} = 0$, which admits plane-wave solutions. Expressing the solution $\bar{h}_{\mu\nu} = A_{\mu\nu} e^{ik_\alpha x^\alpha}$ (for some tensor components $A_{\mu\nu}$), $k^\alpha k_\alpha = 0$, which means that the wavevector k^α is null. This means that the gravitational waves propagate at the speed of light, and, combined with the Lorenz gauge condition, the oscillations will be orthogonal to the propagation direction: $k^\mu A_{\mu\nu} = 0$.

There is still remaining gauge freedom, as certain infinitesimal coordinate transformations will leave the Lorenz gauge condition satisfied. We can now use up all of our remaining gauge freedom and impose conditions to express the components of the perturbation in a particularly convenient form. These conditions are:

$$\begin{aligned} \eta^{\mu\nu} A_{\mu\nu} &= A^\mu{}_\mu = 0, \\ U^\mu A_{\mu\nu} &= 0, \end{aligned} \tag{2.17}$$

where U^μ is a constant, time-like unit vector, which, for example, could represent the 4-velocity of an observer in their rest frame (in which case $U^\mu = (1, 0, 0, 0)$). The first condition is the "traceles" condition, which sets $\bar{h}^\mu{}_\mu = 0$ meaning $\bar{h}_{\mu\nu} = h_{\mu\nu}$ in this gauge. Considering an observer's four velocity in their rest frame, we see that the second condition enforces $A_{0\nu} = 0$, so $h_{0\nu} = h_{\mu 0} = 0$, meaning the perturbation is only spatial. If we now align the z-axis of the observer with the propagation direction, $k^\mu = (\omega, 0, 0, \omega)$, and recalling $k^\mu A_{\mu\nu} = 0$, we find that $h_{3\nu} = 0$, so the only non-zero perturbations are in the spatial plane orthogonal to the direction of the wave propagation.

[1] This seems to be commonly called the Lorentz gauge in much of the literature, but this is likely due to an unfortunate naming coincidence between the original inventor of the gauge condition, Ludwig Lorenz, and the better-known physicist Hendrik Lorentz.

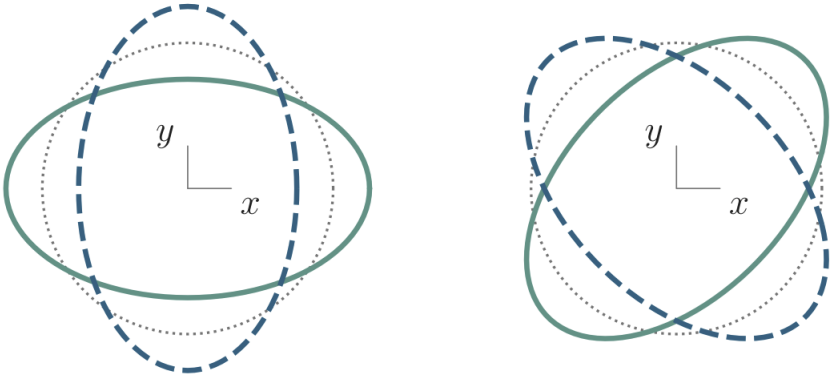

Figure 2.2: Effect on a circular ring of free-falling test particles (grey dotted) due to a the plus (left) and cross (right) polarized components of a gravitational wave incident in the z-direction. Solid and dashed patterns happen half a period apart. From [292].

Having used up our gauge freedom setting the conditions of the *transverse-traceless* (TT) gauge, we are left with two remaining degrees of freedom, which can immediately associate with the independent physical components of $h_{\mu\nu}$. The transverse-traceless conditions allow us to express the metric perturbation in the form:

$$h_{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & h_{11} & h_{12} & 0 \\ 0 & h_{12} & -h_{11} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \tag{2.18}$$

where h_{11} and h_{12} are our two degrees of freedom. Examining the response of a ring of test particles to a gravitational wave incident from overhead reveals the effect of the two polarizations. Figure 2.2 shows the distortion patterns the ring of test particles will see from each of the two linearly independent polarization states. The h_{11} mode causes the particles oscillate between squeezing along the x-axis and squeezing along the y-axis, tracing out a "+" pattern over time. The h_{12} mode causes the particles to oscillate between squeezing along the diagonals, tracing out a "×" pattern over time. These two modes are known as the "plus" and "cross" polarizations of gravitational waves, respectively, so we define:

$$h_+ = h_{11}, \quad h_\times = h_{12}. \tag{2.19}$$

In Chapter 3 I discuss the effect that this perturbation will have on an gravitational wave interferometer such as LIGO.

2.3 Compact Binaries as a Source of Gravitational Waves

Having shown how gravitational wave are transmitted through spacetime, we now turn to how they are generated. In general, radiation comes from the acceleration of a charge, or the time variation of a multipole moment of a configuration of charges. Since the energy flux through a spherical surface is proportional to the square of the amplitude (A) of the field, then the total luminosity carried by radiation to a distance r is simply $L \sim A(r)^2 r^2$. Therefore, the amplitude of the field $A(r)$ must decrease as $1/r$ in order to carry energy to arbitrarily large r. We will consider fields that achieve this property as radiative.

Recall the case for an electric charge. The potential scales like $\Phi(r,t) \propto \frac{Q(t)}{r}$ (i.e., the Coulomb potential). The amplitude of the corresponding electric field E goes like $E \propto \nabla\Phi \sim \frac{1}{r}\nabla Q(t) + O(1/r^2)Q(t)$. The second term scales like $1/r^2$, so it decays too fast to carry energy to infinity, and therefore does not contribute to what we are calling radiation. Now, if we consider the fact that we are a finite distance from the source and changes in the field propagate at a finite speed ($v = c = 1$), then the potential we see is really a function of the retarded time $t_r \approx t - r$, where we assume $r \ggg$ size of the source, and we work only in one spatial dimension. Then we rewrite the first time of the field amplitude as:

$$\frac{1}{r}\nabla Q(t_r) = \frac{\partial Q(t_r)}{\partial t_r}\frac{\partial t_r}{\partial r}\frac{1}{r} = \frac{\dot{Q}(t_r)}{r}. \tag{2.20}$$

The above argument tells us two important facts about radiation. First, the static configuration decays too fast with distance to carry energy to infinity. Generically, the potential from static higher-order terms (dipole, quadrupole, etc.) decay even faster, with $\Phi_\ell^{(static)} \sim \frac{1}{r^{\ell+1}}$, so *static charge configurations do not radiate*. Second, we learn that if the monopole has a nonzero first time derivative, then you can get radiation. However, because charge is conserved, $\dot{Q} = 0$, so the monopole cannot radiate.

In general, you can get a radiative term in the potential from the ℓ-th moment of the charge distribution:

$$\Phi_\ell^{\rm rad}(t_r, r) = \frac{1}{r}\frac{\partial^\ell Q_\ell(t_r)}{\partial t^\ell}, \tag{2.21}$$

which gives a radiative part of the field amplitude A:

$$A_\ell \propto \frac{1}{r}\frac{d^{\ell+1} Q_\ell(t_r)}{dt^{\ell+1}}, \tag{2.22}$$

where we used the product rule and only kept the part that scaled like $\frac{1}{r}$ and includes an extra time derivative.

From this, we can see that electric dipoles radiate because as there is no conservation law preventing $\dot{Q}_1 \neq 0$, where $Q_1 = \int \rho(\vec{x}, t)\vec{x}d^3x$ is the dipole moment.

Now we can take this analogy to the gravitational field. At $\ell = 0, Q_0 = \int \rho(\vec{x}, t)d^3x = M$ is simply the total mass-energy of the system, which is conserved, preventing monopole radiation. At $\ell = 1$, we have the dipole moment Q_1, the first moment of the mass distribution, which is just the center of mass(-energy). Conservation of momentum conserves this and therefore prevents dipole gravitational radiation. At $\ell = 2$, we have the quadrupole moment, which has no corresponding conservation law. We therefore can expect gravitational radiation to come from changes in the quadrupole moment of the mass distribution, to leading order.

Using the Green's function to solve Eq. 2.16 in vacuum, we find a solution for $h(t, \vec{x})$, the strain perturbation at some point $\vec{x}$ at time t:

$$\bar{h}_{\mu\nu}(t, \vec{x}) = 4G \int d^3x' \frac{T_{\mu\nu}(t - |\vec{x} - \vec{x}'|, \vec{x}')}{|\vec{x} - \vec{x}'|}, \tag{2.23}$$

where it is common to define $t_r = t - |\vec{x} - \vec{x}'|$ as the retarded time. Thus, in order to get the strain perturbation at $(t, \vec{x})$, we sum over all of the energy-momentum contributions from sources at $(t_r, \vec{x}')$ in the past light cone.

We assume we are far enough from the source, such that our distance to the source is much larger than the physical size of the source, so we are summing over contributions from a very small region compared to our distance, which we denote r. Combined with the assumption that the energy-momentum does not change much over the light crossing time, we can approximate $|\vec{x} - \vec{x}'| \approx |\vec{x}| = r$, allowing us to write:

$$\bar{h}_{ij}(t, \vec{x}) = \frac{4}{r} \int d^3x' T^{ij}(t - r, \vec{x}'), \tag{2.24}$$

where we additionally restrict ourselves to the spatial components i, j of the energy-momentum tensor, as these are the only components that contribute to the gravitational wave strain in the TT gauge.

Enforcing energy-momentum conservation $\partial^\mu T_{\mu\nu} = 0$, one can show that (e.g., [233]):

$$\frac{4}{r} \int d^3x' T^{ij} = \frac{2}{r} \frac{\partial^2}{\partial t^2} \int d^3x' T^{00} x'^i x'^j = \frac{2}{r} \frac{\partial^2}{\partial t^2} \int d^3x' \rho x'^i x'^j, \tag{2.25}$$

where we make the association of $T^{00} = \rho$ as the configuration of matter (the mass density). With the second moment of the mass distribution as:

$$I^{ij}(t) = \int d^3x' \rho(t, \vec{x}') x'^i x'^j, \tag{2.26}$$

the strain perturbation can be expressed as:

$$\bar{h}_{ij}(t, \vec{x}) = \frac{2}{r} \frac{\partial^2 I_{ij}(t_r)}{\partial t^2}. \tag{2.27}$$

To get the TT part of the strain, we simply need to project out the non transverse-traceless part. We do this by defining the reduced quadrupole moment:

$$\bar{I}_{ij}(t) = \int d^3x \rho(t, \vec{x})(x^i x^j - \frac{1}{3}\delta_{ij}|\vec{x}|^2) = I_{ij}(t) - \frac{1}{3}\delta_{ij} I(t)^{ii}, \tag{2.28}$$

which subtracts off the trace. Then, we define the projection tensor:

$$P_{ij,kl} = P_{ik}P_{jl} - \frac{1}{2}P_{ij}P_{kl}, \tag{2.29}$$

where $P_{ij} = \delta_{ij} - n_i n_j$. This projects out the components parallel to n^i, leaving us with:

$$\bar{h}^{TT}_{ij}(t, \vec{x}) = \frac{2}{r} P_{ij,kl} \frac{\partial^2 \bar{I}^{kl}(t_r)}{\partial t^2}. \tag{2.30}$$

If we now consider a binary system made of two stars or black holes with masses m_1 and m_2, we can calculate the gravitational wave emission over a circular orbit with angular frequency Ω. Kepler's laws tell us this will be:

$$\Omega^2 = \frac{M}{a^3}, \tag{2.31}$$

when the two bodies are separated by a distance a. Here, we define $M = m_1 + m_2$ as the total mass of the system. The reduced mass is $\mu = \frac{m_1 m_2}{M}$. If we orient the system on the x-y plane, so it is face-on with respect to the observer, then we can write the reduced quadrupole moment as the matrix:

$$\bar{I}_{ij} = \mu a^2 \begin{pmatrix} \cos^2(\Omega t) - 1/3 & \cos\Omega t \sin\Omega t & 0 \\ \cos\Omega t \sin\Omega t & \cos^2(\Omega t) - 1/3 & 0 \\ 0 & 0 & -1/3 \end{pmatrix}, \tag{2.32}$$

which has the second time derivative:

$$\frac{\partial^2 \bar{I}_{ij}}{\partial t^2} = -2\mu a^2 \Omega^2 \begin{pmatrix} \cos(2\Omega t) & -\sin(2\Omega t) & 0 \\ -\sin(2\Omega t) & -\cos(2\Omega t) & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{2.33}$$

With this orientation, the projection into TT simply projects out the z-component, so we can easily get the strain:

$$\bar{h}_{ij}^{TT}(t) = \frac{4M^{2/3}\mu\Omega^{2/3}}{r}\begin{pmatrix}\cos(2\Omega t_r) & \sin(2\Omega t_r) & 0\\ \sin(2\Omega t_r) & -\cos(2\Omega t_r) & 0\\ 0 & 0 & 0\end{pmatrix}, \tag{2.34}$$

where a is replaced using Kepler's law. Replacing factors of G and c, we can read off the polarizations:

$$h_+ = \frac{4(G\mathcal{M})^{5/3}\Omega^{2/3}}{c^4 r}\cos(2\Omega t_r), \quad h_\times = \frac{4(G\mathcal{M})^{5/3}\Omega^{2/3}}{c^4 r}\sin(2\Omega t_r), \tag{2.35}$$

where we defined the chirp mass $\mathcal{M} \equiv \frac{(m_1 m_2)^{3/5}}{(m_1+m_2)^{1/5}} = \mu^{3/5}M^{2/5}$. From here, we learn that the quadrupole gravitational radiation has a frequency of twice the system orbital frequency, and that the amplitude of the radiation is controlled by the chirp mass and our distance to the system. In general, h_+ and $h_\times$ will additionally depend on the inclination of the line-of-sight to the observer.

While this gives a good idea for the amplitude of a very simple configuration of a binary system, it leaves out critical details related to the evolution of the system dynamics which will affect the strain. In particular, since gravitational waves carry energy, the orbit will shrink, increasing the orbital frequency, making Ω evolve as a function of time. This is typically solved for by considering the energy carried away by the gravitational waves at the current frequency. At the leading order, this is given from Einstein's quadrupole formula for luminosity when the binary has a separation a:

$$\left(\frac{dE_{\rm GW}}{dt}\right)_0 = \frac{32}{5}\frac{G^4}{c^5}\frac{\mu^2 M^3}{a^5}. \tag{2.36}$$

Balancing this with the energy loss from the orbit results in the evolution of the orbital frequency. For example, the oribital energy is $E_{\rm orb} = -\frac{GM\mu}{2a}$, so, with $\frac{dE_{\rm orb}}{dt} = -\frac{dE_{\rm GW}}{dt}$, we find the orbital frequency evolves like:

$$\frac{df_{\rm orb}}{dt} = \frac{df_{\rm orb}}{dE_{\rm orb}}\frac{dE_{\rm orb}}{dt} = \frac{96\pi^{8/3}}{5c^5}(G\mathcal{M})^{5/3}f_{\rm orb}^{11/3}, \tag{2.37}$$

where $\Omega = 2\pi f_{\rm orb}$. Thus, one can evolve of the orbital frequency by tracking the energy emitted in gravitational waves at that quasi-circular orbit, over which we assume the frequency remains stable. This allows for one to solve for the time-frequency relationship to plug into Eq. 2.35 to get a simple function for strain as a

function of time, $h_{+,\times}(t)$. A similar estimate can be made in the Fourier domain, assuming a stationary phase and using the frequency evolution to solve for the phase.

Whereas this may be a good approximation for a binary system at low velocity (equivalently, low frequency), we still used the Newtonian/Keplerian formula to relate the orbital frequency and separation, which we then used in the quadrupole formula for the luminosity (Eq. 2.36). However, when the bodies get closer to one another, they undergo stronger gravitational interactions and move at faster velocities, necessitating relativistic treatment of their equations of motion. This leads to corrections which show up in powers of the velocity (or, equivalently, powers of $(M/a)^{1/2} = (M\Omega)^{1/3}$). The coefficients of these terms beyond the leading order depend on the other properties of the compact binary system, such as the mass ratios and the spin configuration. Therefore, signals with content from higher-order terms (later in the inspiral) give information beyond the system's chirp mass and can be used to extract the component masses, spins, and other properties of the system.

Part II

Measuring Properties of Black Holes and Neutron Stars Using Gravitational Waves from CBCs

Chapter 3

MEASURING SOURCE PROPERTIES OF COMPACT BINARY COALESCENCES

3.1 The Gravitational Wave Likelihood

After detecting a signal in the data, the goal now turns to determining the properties of the source that generated the signal. In this section, I describe the process of how we measure physical properties from the data collected by a gravitational wave detector assuming it saw a signal from a compact binary coalescence (CBC). Several works in the literature offer thorough reviews of this procedure (e.g., [536, 137, 221, 163]). In practice, we begin by making the following assumptions:

- The (calibrated) data **d** consist of a signal **h** and additive noise **n**.
- The realization of **n** is from a stationary, Gaussian distribution with zero mean.

I denote the data, noise, and signal strain as vectors in boldface, as they typically appear as a time- or frequency-series. The description of the noise process is motivated by the central limit theorem: when there are many small contributions to the noise (e.g., contributions from seismic motion, instrumental effects, quantum shot noise, etc.), a noise amplitude is a realization from a Gaussian distribution. We work under the assumption that we have enough data to describe the noise with a Gaussian, as the higher moments die off as we collect more data; the maximum-entropy distribution for a given specified mean and variance is a Gaussian.

The contribution from **n** is Gaussian with the probability density:

$$p(\mathbf{n}|\mathbf{C}) = \frac{1}{\sqrt{|2\pi C|}} \exp\Big(-\frac{1}{2} n^i C^{-1}_{ij} n^j\Big) \tag{3.1}$$

where $C_{ij} = E[n_i n_j]$ is the autocorrelation function: the expected product of two noise amplitudes, which depends on the difference in times of the two data points. "||" denotes a determinant, and Einstein summation is implied over contracted indices.

Generically, the exponential term in Eq. 3.1 costs $O(N^2)$ to compute, for a vector of N elements. Assuming stationarity and periodicity, we can recast the distribution $n(t)$ into the frequency domain $\tilde{n}(t)$ where the Wiener-Kinchin theorem states that

the covariance matrix becomes diagonal $\tilde{C}_{ij} = \frac{TS(|f_i|)}{2}\delta_{ij}$ (see [293, 475] for more details), allowing the likelihood to be computed $\mathcal{O}(N)$. Here, $S(f_i)$ is the noise Power Spectral Density (PSD) at the ith frequency bin and T is the duration of the data. Typically, $\mathbf{S}$ is calculated by taking the Fourier transform of data nearby the signal and then taking a Welch-like average, or by using an on-source estimation method such as BayesLine [343]. Typically, the noise PSD is considered fixed for the PE analysis [18].

In the frequency domain, the likelihood for the (complex-valued noise) in the ith frequency bin obtained from data with duration T becomes (suppressing explicit frequency-domain notation of $\tilde{n}$, as it is understood this will all be done in the frequency domain):

$$p(n_i|S_i) = \frac{2}{\pi T S_i}\exp\big(-\frac{2|n_i|^2}{TS_i}\big), \tag{3.2}$$

where we lose the square root in Eq. 3.1 because we enforce that both the real and imaginary components of n_i follow the same Gaussian distribution (leading to a Whittle likelihood [523, 469]).

Define the noise-weighted inner product as (assuming a one-sided PSD):

$$\langle \mathbf{a}|\mathbf{b}\rangle \equiv 4\mathrm{Re}\int_0^\infty \frac{\tilde{a}^*(f)\tilde{b}(f)}{S(f)}df, \tag{3.3}$$

which we approximate by evaluating on a grid of N frequency bins spaced Δf apart:

$$\langle \mathbf{a}|\mathbf{b}\rangle = 4\Delta f\sum_i^N \mathrm{Re}\Big[\frac{a_i^* b_i}{S_i}\Big]. \tag{3.4}$$

We can now rewrite Eq. 3.2 into:

$$p(\mathbf{n}|\mathbf{S}) = \frac{1}{|\frac{\pi T}{2}\mathbf{S}|}\exp\big(-\frac{1}{2}\langle \mathbf{n}|\mathbf{n}\rangle\big). \tag{3.5}$$

We now have a probability distribution for the data residuals in the presence of a signal, $n(f) = d(f) - h(f)$. Assuming the signal model $h(f)$ is parameterized by θ, which includes masses, spins, and location of the binary system (and any other measurable property controlling the predicted signal), we can predict the noise in the data for a signal with parameters θ: $n(f,\theta) = d(f) - h(\theta, f)$. The log likelihood is then:

$$\log p(\mathbf{d}|\theta) = -\frac{1}{2}\Big\langle(\mathbf{d}-\mathbf{h})\Big|(\mathbf{d}-\mathbf{h})\Big\rangle - \log\Big(|\frac{\pi T}{2}\mathbf{S}|\Big). \tag{3.6}$$

If we assume the noise PSD is fixed, we need only keep the terms that have dependence on the signal model **h**:

$$\log p(\mathbf{d}|\theta) = -\frac{1}{2}\big(\langle \mathbf{h}|\mathbf{h}\rangle - 2\langle \mathbf{d}|\mathbf{h}\rangle\big) + ..., \tag{3.7}$$

where the ... denotes terms independent of θ.

When measuring the source parameters θ, we typically want to obtain the probability for any value of θ given the data we collected. In the Bayesian inference framework, we begin with our *prior distribution* reflecting our prior beliefs on the values of θ before seeing the data, and then incorporate the likelihood of the data under the possible values of θ, as calculated in Eq. 3.7. This is the *posterior probability* of parameters θ given the data (Bayes' theorem):

$$p(\theta|\mathbf{d}) = \frac{p(\mathbf{d}|\theta)\pi(\theta)}{\mathcal{Z}}, \tag{3.8}$$

where $\pi(\theta)$ is the prior and $\mathcal{Z} = \int p(\mathbf{d}|\theta)\pi(\theta)d\theta$ is the normalization factor known as the *evidence* and is sometimes expressed as $p(\mathbf{d})$, the probability of the data given the specified model and prior distribution. Computing the posterior probability for parameters θ is known as *parameter estimation* (PE).

Typically, we conduct PE under a relatively agnostic prior distribution, where we try to impart minimal assumptions in $\pi(\theta)$. For example, we assume uniform prior distributions for the component masses in the frame of the detector[1]. Priors on angular coordinates, such as spin orientations and sky location of the source, are isotropic. The prior on the magnitudes of the spin vectors is typically uniform just up to the Kerr limit of 1, with lower prior limits occastionally adopted based on the source or assumed waveform model. It is worth noting, however, the agnostic choice that we make is not necessarily the best (or even truly "agnostic") given the prior information we actually have from the population or astrophysical expectations. As explained in Chapter 6, we *reweight* this prior to a population-informed prior for many downstream analyses.

Because for CBCs Eq. 3.8 typically involves a vector of θ with 15+ dimensions, it is intractable to simply compute the posterior on a multidimensional grid of θ. If we were only interested in the probability over just the ith parameter, we would still need to integrate $p(\theta_k|\mathbf{d}) = \int p(\theta|\mathbf{d})d\tilde{\theta}$ where $\tilde{\theta}$ refers to the subset of $\theta \neq \theta_k$.

[1] Technically, for computational efficiency, we sample in chirp mass and mass ratio coordinates with a Jacobian that sets a prior effectively uniform in the component masses.

For example, we may want to know "what is the posterior probability on the spins, regardless of the mass?"

Stochastic sampling techniques are commonly used to overcome this computational challenge. These methods work by drawing samples and computing weights for each sample, ending up with a set of representative samples of $\{\theta\} \sim p(\theta|\mathbf{d})$. This is in contrast to returning the posterior evaluated at pre-set possible values of θ (i.e., as an alternative to evaluating on a grid of θ). Common stochastic sampling methods include Markov Chain Monte Carlo (MCMC) sampling and nested sampling. While nested sampling was originally introduced to compute evidences [500], it also returns samples from the posterior after it reaches some convergence criteria. Typically, the nested sampler Dynesty [507] (as implemented in Bilby [67]) is used for sampling the posterior with the gravitational wave likelihood in PE analyses within the LVK.

Bilby, and the associated automation pipeline Bilby-Pipe, provides a convenient and user-friendly package for conducting PE analyses, with special features and focus on gravitational wave data analysis [470, 67]. In particular, the user can specify analysis settings such as frequency bounds, data duration, priors, and much more, and Bilby will read data and construct Interferometer objects to represent each chosen gravitational wave detector in its correct location. It will then accumulate samples from the posterior by calling a waveform approximant (via, e.g., LALSimulation), and projecting the signal onto the interferometer to get the predicted strain in the detector, and calculating the likelihood (Eq. 3.7).

It still remains to be seen how $\mathbf{h}(f, \theta)$ is calculated. Typically, this is done in two parts:

- A *waveform approximant* predicts the complex strain in the basis of the allowed polarizations. In general relativity, these are + and ×, so this corresponds to generating two vectors $\mathbf{h}_+(\theta), \quad \mathbf{h}_\times(\theta)$, for parameters θ. This is the waveform that reaches Earth prior to accounting for detector effects.
- The strain is projected onto the antenna response to each polarization j, given a line-of-sight vector $\vec{n}$ and polarization angle ψ: $\mathbf{h} = F^j(\vec{n}, \psi) h_j(\theta)$

The functions F^j are the antenna pattern functions, quantifying the response of the detector to a signal with polarization j. This depends on the orientation between the line-of-sight vector to the source and the detector arms. While it is common in the literature to say F is a function of only the celestial coordinates (i.e., right

ascension α, declination δ) and polarization angle ψ, this does not fully specify $\vec{n}$ with respect to the detector, as the celestial sphere rotates over time, requiring the time of the signal to additionally be specified. Specifically, F is a function of either $(\alpha, \delta, t, \psi)$ or the spherical coordinates of the source in the detector's frame.

3.2 The Response of the Detector to a Gravitational Wave

It is convenient to express $h_{\mu\nu}$ in a basis where the two propagating degrees of freedom contribute only spatial perturbations orthogonal to the wave propagation vector which we take to be along the 3-axis. In this *transverse-traceless* (TT) gauge (see Chapter 2),

$$h_{12} = h_{21}$$

$$h_{11} = -h_{22}$$

$$h_{0\mu} = h_{\nu 0} = h_{3\alpha} = h_{\beta 3} = 0.$$

After assigning our two degrees of freedom as $h_{12} = h_\times$ and $h_{11} = h_+$ we can write the basis of each polarization in the transverse spatial subspace (i.e., xy) as:

$$\mathbf{e}_+ = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \mathbf{e}_\times = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Which is generated by

$$\mathbf{e}_+ = \hat{e}_1\hat{e}_1^T - \hat{e}_2\hat{e}_2^T = \hat{e}_1 \otimes \hat{e}_1 - \hat{e}_2 \otimes \hat{e}_2 \tag{3.9}$$

$$\mathbf{e}_\times = \hat{e}_1\hat{e}_2^T + \hat{e}_2\hat{e}_1^T = \hat{e}_1 \otimes \hat{e}_2 + \hat{e}_2 \otimes \hat{e}_1, \tag{3.10}$$

where $\hat{e}_i$ is the standard ith cartesian basis vector for the frame of the radiation.

Putting this together, in this gauge, the metric perturbation can be expressed as the matrix

$$\mathbf{h} = h_+\mathbf{e}_+ + h_\times\mathbf{e}_\times = \begin{pmatrix} h_+ & h_\times \\ h_\times & -h_+ \end{pmatrix}, \tag{3.11}$$

specifying the geometry of perturbations from the incoming radiation. It is worth reiterating here that when expressed in the form above, the components of $h_{\mu\nu}$ refer to the components in the basis defined by the polarization plane.

We now consider how the detector responds to the presence of a gravitational wave signal passing.[2] We begin by establishing a detector-centered frame whose origin

[2] See, e.g., [59, 567, 392] for reviews of this procedure.

is at the beamsplitter, with basis vectors $\hat{X}$ and $\hat{Y}$, pointing along the X and Y arms of the detector, respectively. For this section, the basis vectors are denoted $\hat{X}^i = (\hat{X}, \hat{Y}, \hat{Z})$. Note that, in general, $\hat{X}$ and $\hat{Y}$ are not necessarily aligned with $\hat{e}_x$ and $\hat{e}_y$ and these two frames will be related by a three-dimensional rotation in order to express $h_{\mu\nu}$ in terms of its components in $\hat{X}$, $\hat{Y}$, and $\hat{Z}$.

In our gauge choice, the coordinates are defined with respect to the positions of the test masses, fixing their positions. While the coordinate position of the test masses at the end of either arm does not change due to the passage of a gravitational wave, the proper length of the arm will change, which is an observable effect. If each arm has a rest length L, then the vector pointing to the test mass on the X arm is $X^{(1)} = X = (L, 0, 0)$, and, for the Y arm, $X^{(2)} = Y = (0, L, 0)$.

With $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$, light travels along the null geodesic (with $c = 1$):

$$ds^2 = 0 = g_{\mu\nu}dx^\mu dx^\nu = -dt^2 + g_{ij}dX^i dX^j = -dt^2 + (1 + h_{ij})dX^i dX^j. \quad (3.12)$$

So, for example, for light traveling along the X-arm,

$$-dt^2 + (1 + h_{XX})dX^2 = 0, \quad (3.13)$$

implies

$$dt = \sqrt{1 + h_{XX}}dX \approx (1 + \frac{1}{2}h_{XX})dX, \quad (3.14)$$

which we can simply integrate from the light traveling from 0 to L and back to 0 to get the round-trip coordinate time in the X-arm:

$$T_X = \int_0^L (1 + \frac{1}{2}h_{XX})dX + \int_L^0 (1 + \frac{1}{2}h_{XX})dX = 2L(1 + \frac{1}{2}h_{XX}), \quad (3.15)$$

where h_{XX} is the component of $h_{\mu\nu}$ that lies along the X arm, and we recall that the test masses are at fixed coordinate distance L.

Here, we assumed the *long-wavelength* approximation, so that h_{XX} is constant over the length of the arm because $\lambda_h \gg 2L$ (equivalently, the frequency of the gravitational wave is much less than the frequency of the round trip light time). This allowed us to simplify Eq. 3.15 and factor h_{XX} out of the integral. This is a good approximation for the frequencies of interest for LIGO and Virgo, which are typically $\sim$ 10 Hz to $\sim$ 1 kHz, meaning $\lambda_h \gtrapprox 10^4$ km, much larger than the arm length of LIGO ($L = 4$ km). See [75] for my contributions to a study testing parameter estimation on simulated signals in third generation detectors where the

long-wavelength approximation is not valid and additional time- and frequency-dependent effects need to be considered.

As we have not yet needed to make an assumption about what coordinate system we represent $h_{\mu\nu}$, it is worth expressing this in abstract notation, where h_{ab} represents the components of $h_{\mu\nu}$ without assuming a basis. In abstract notation, with $h_{ij} = h_{ab}\hat{X}^{ia}\hat{X}^{jb}$, the round-trip time for light moving down the arm aligned with $\hat{X}^i$ is $T_i \approx 2L(1 + \frac{1}{2}h_{ab}\hat{X}^{ia}\hat{X}^{ib})$. If we assume a representation where the basis elements of $h_{\mu\nu}$ are aligned with those of the detector, then the round-trip time along the X^i arm is simply $T_i \approx 2L(1 + \frac{1}{2}h_{ii})$, as before.

In LIGO, we measure gravitational wave strain in the difference in light-travel time down the two arms; this is known as the Differential ARM (DARM) response, which contains the least laser frequency noise. The difference in roundtrip light times down the arm aligned with $\hat{X}^i$ and the arm aligned with $\hat{X}^j$ is simply $\Delta T \equiv T_i - T_j = L\big(h_{ab}\hat{X}^{ia}\hat{X}^{ib} - h_{ab}\hat{X}^{ja}\hat{X}^{jb}\big)$. The strain is a fractional change, so:

$$h = \frac{\Delta T}{2L} = h_{ab}\frac{\hat{X}^{ia}\hat{X}^{ib} - \hat{X}^{ja}\hat{X}^{jb}}{2} = h_{\mu\nu}\frac{\hat{X}^i \otimes \hat{X}^i - \hat{X}^j \otimes \hat{X}^j}{2}. \tag{3.16}$$

The final term multiplying h_{ab} is commonly denoted as the detector tensor $d^{ab} = \frac{1}{2}\big(\hat{X}^a\hat{X}^b - \hat{Y}^a\hat{Y}^b\big)$. Stated again, this can be understood as the object which gives the fractional difference between the times (or proper lengths) measured by the round trip of two photons exiting from the beamsplitter simultaneously, with one traveling up and down the X^1 arm and the other up and down the X^2 arm (e.g., $X^1 = X$ and $X^2 = Y$).

In order to actually calculate the strain from the passing gravitational wave signal, we project the signal into the detector response by contracting h_{ab} and d_{ab}:

$$h = h_{ab}\, d^{ab} = \big(h_+\, e_{+\,ab} + h_\times\, e_{\times\, ab}\big)\, d^{ab} = h_+\, F^+ + h_\times\, F^\times, \tag{3.17}$$

where $F^{+/\times}$ is the antenna response factor for the $+/\times$ polarization. In order to actually calculate $F^+ = e_{+\,ab}d^{ab}$ and $F^\times = e_{\times\,ab}d^{ab}$, we need to represent $\mathbf{e}_{+/\times}$ in the same basis as d.

In practice, this can be done by expressing the components of these tensors in a geocentric, Earth-fixed coordinate system, whose x, y, and z axes intersect the Earth surface at $(0, 0)$, $(0, 90°\text{E})$, and $(90°\text{N}, 0)$, respectively.

Now we express the components of detector tensor $\mathbf{d}$ in this coordinate system. For a detector with a vertex at latitude φ and longitude λ, the unit vector for the i arm

with a tilt angle ω_i above the horizontal and angle Ψ_i North of East is (e.g., [59]):

$$\hat{X}^i = (\sin\omega_i \cos\varphi \cos\lambda - \cos\omega_i \cos\Psi_i \sin\lambda - \cos\omega_i \sin\Psi_i \sin\varphi \cos\lambda)\hat{x} \quad (3.18)$$

$$+ (\cos\omega_i \cos\Psi_i \cos\lambda - \sin\varphi \sin\lambda \cos\omega_i \sin\Psi_i + \sin\omega_i \cos\varphi \sin\lambda)\hat{y} \quad (3.19)$$

$$+ (\cos\omega_i \sin\Psi_i \cos\varphi + \sin\omega_i \sin\varphi)\hat{z}, \quad (3.20)$$

where each unit vector points from the origin at the geocenter in the direction of $\hat{X}^i$.

One can then generate the elements of the gravitational wave polarization tensors in this same basis by simply expressing polarization plane basis vectors $\hat{e}_x$ and $\hat{e}_y$ in the geocentric coordinate system, which has the elements:

$$\hat{e}_x = (\sin\phi \cos\psi - \sin\psi \cos\phi \cos\theta)\hat{x} \quad (3.21)$$

$$- (\cos\phi \cos\psi + \sin\psi \sin\phi \cos\theta)\hat{y} \quad (3.22)$$

$$+ \sin\psi \sin\theta\hat{z} \quad (3.23)$$

$$\hat{e}_y = (-\sin\phi \sin\psi - \cos\psi \cos\phi \cos\theta)\hat{x} \quad (3.24)$$

$$+ (\cos\phi \sin\psi - \cos\psi \sin\phi \cos\theta)\hat{y} \quad (3.25)$$

$$+ \sin\theta \cos\psi\hat{z}, \quad (3.26)$$

which simply rotates the wavevector a gravitational wave signal aligned with $\hat{e}_z$ by the polar angle θ (not to be confused with the general single-event source parameters θ) and the azimuthal angle ϕ, placing $\hat{k}$ in the correct point over Earth. Then the polarization plane of the gravitational wave signal is further rotated by ψ, aligning the polarization basis with this coordinate system. So we can now express $\mathbf{e}_{+/\times}$ in terms of the geocentric coordinates by plugging the expressions in Eq. 3.21 into Eq. 3.9. Note that these coordinates are time-dependent as points on the celestial sphere rotate around the fixed Earth coordinate basis. Given the Greenwhich Mean Sidereal Time (GMST) of the arrival of the signal, we can relate the coordinates on the celestial sphere (for local longitude ϕ and polar angle–colatitude θ)[3]:

$$\phi = \alpha - \text{GMST} \quad (3.27)$$

$$\theta = \pi/2 - \delta. \quad (3.28)$$

The final step for computing the waveform in the detector is simply to perform the contraction in Eq. 3.17 given the polarizations h_+ and $h_\times$.

[3]This follows because the GMST is the Right Ascension on the Greenwich Meridian. Therefore, at a given time, the position of a star counterclockwise from the x-axis is the difference between the Right Ascension and GMST. Declination is measured from the equator so it is offset by standard spherical coordinate θ by 90°.

3.3 Waveform Approximants

Given a set of intrinsic parameters (masses and spins) plus a reference phase and inclination, a waveform approximant is used to compute the predicted strain polarizations h_+ and $h_\times$. We use a waveform "approximant" to predict the strain h because there are no generic, closed-form solutions for the CBC strain from a binary system in general relativity, and solving for the strain in real-time using numerical simulations is infeasible.

That being said, *numerical relativity* (NR) simulations solve the Einstein field equations using numerical methods, giving the most accurate solution currently available for the strain produced by a CBC in general relativity. Since the first complete successful BBH merger simulations in 2005 [432, 128, 77], numerical relativity simulations have provided critical information about CBCs in the highly nonlinear, strong-field regime, particularly in the final merger and ringdown phases. Each simulation begins by specifying the mass ratio and the spins of each black hole (three total parameters for aligned-spin simulations, seven for in-plane precessing spin components, total mass acts as an overall scale factor). Conducting these simulations is extremely costly and requires solving coupled differential equations for the metric and its derivatives on a grid along hypersurface slices. The largest and most accurate catalogs of NR simulations are produced by the Simulating Extreme Spacetimes (SXS) collaboration, whose Spectral Einstein Code evolves compact binary systems via integrating the Einstein equations using spectral methods with exponential convergence [110]. These simulations are critical for modeling the strain in the merger and ringdown phases of a CBC, where analytical approaches valid for low speeds (or large separations) have broken down.

For a fixed initial binary separation, the cost of these simulations scales with the mass ratio and increases with precessing spin configurations. This can take $O(\text{weeks})$ to complete a single simulation running in parallel on thousands of cores. However, during the course of a PE analysis, the stochastic sampler will call the likelihood typically for more than $O(10^6)$ configurations of parameters, motivating the use of faster, but less accurate approximants to compute the strain $h(\theta)$ for arbitrary θ.

Waveform approximants can generally be classified in three broad categories based on the strategy used to approximate $h(\theta)$: numerical relativity surrogate models (NRSur), effective-one-body models (EOB), and phenomenological inspiral-

merger-ringdown (IMRPhenom) models.[4]

Achieving mismatches of $\sim 10^{-4} - 10^{-3}$ to NR simulations, NRSurrogate models such as `NRSur7dq4` [550] are considered the state-of-the-art waveform model where they are valid and have the least systematic biases. Constructed natively in the time domain and including higher order mode content, these models essentially quickly predict h by interpolating from a catalog of NR simulations. They work by creating a reduced basis of waveforms from hundreds of NR simulations, then interpolating or fitting coefficients of these basis elements. As an interpolant-based model, the region of parameter space where these are valid is limited by the coverage of existing NR simulations the model was trained on and the length of these simulations. For the precessing surrogate `NRSur7dq4`, this region of parameter space extends to precessing spins with magnitudes $a_{1,2} \lesssim 0.8$ and mass ratios $1/4 < q < 1$, although the authors of this model find that the extrapolation of the model valid down to $q \approx 1/6$ [550]. Crucially, the parameter space where this model is valid is also limited by the length of the NR simulations in its training set. For example, the NR simulations used in training `NRSur7dq4` begin $4300M$, ~ 20 orbits, before merger. With the mass configuration of the simulation determining the physical separation or gravitational wave frequency at this time, h is only generated above a minimum frequency depending on the masses. For `NRSur7dq4`, the $(\ell, m) = (2, 2)$ mode starts at or below 20 Hz for total masses $M \gtrsim 60M_\odot$. Only some of the events through the third Graviational Wave Transient Catalog (GWTC3) are confidently within this region of parameter space, meaning this model is not generally applicable to events seen by the LVK detectors.

The `NRHybSur3dq8` waveform model is an NR surrogate model that circumvents the total mass restriction by hybridizing, or smoothly stitching, the NR surrogate waveform at the merger-ringdown to a combined post-Newtonian (PN) and EOB waveform at the earlier inspiral [549]. This makes this waveform and its higher order mode content valid for the entire BBH mass range when starting at 20 Hz. However, this comes at the cost of efficiency and limitations for where in spin parameter space the model can be used. The hybridization procedure makes `NRHybSur3dq8` relatively slow to evaluate, and its limitation to aligned spins makes it not generally usable for LVK catalog analyses.

The EOB approach models the inspiral-plunge waveform by replacing the two-body problem with an effective one-body problem of the equivalent reduced mass

[4] See [137, 349] for recent reviews of waveform approximant families.

[118]. The dynamics of the system are then evolved according to a Hamiltonian describing a test mass in a background described by the EOB metric. The evolution process requires numerically solving a system of differential equations based on this formulation in order to generate a waveform for a given configuration of compact binary.

The most popular models in the EOB family are the Spinning Effective One Body Numerical Relativity (SEOBNR) models [408, 108], which are constructed in the time domain and include spin effects. The models with "HM" and "P" in their names include higher order mode content and precession effects, respectively. These models explicitly separate the waveform into two distinct phases: the inspiral-plunge phase (computed using the EOB approach described above) followed by a merger-ringdown phase, which begins at the time of the peak amplitude of the $(2, 2)$ mode. In `SEOBNR`, as the inspiral approaches the plunge phase, the waveform additionally relies on non-quasi-circular corrections and fits to NR. The merger-ringdown phase is described by a phenomenological model motivated by NR and includes quasi-normal mode content whose frequencies and damping times are computed from the remnant properties. The waveform in this phase relies on initially free parameters, some of which are fit to NR during the waveform construction process and are constrained to give a smooth and continuous waveform at the transition between the two phases.

The fitting coefficients used in some of the `SEOB` models allow for extrapolation outside the region of parameter space corresponding to the NR simulations used for fitting. This, combined with the inclusion of perturbative training waveforms for highly-asymetric systems, allows `SEOB` models to offer complete coverage across parameter space for modern ground-based gravitational wave catalogs [408, 444, 427]. While `SEOBNRv5PHM` and `SEOBNRv4PHM` give low mismatches to NR across parameter space, the relatively slow speed of evaluating each waveform can be burdensome for PE analyses. However, with modern PE pipelines and accelerations, this model has become more computationally reasonable to adopt. As discussed in Chapter 4, `SEOBNRv5PHM` was employed for all events in GWTC4 which passed the threshold for conducting PE analyses.

The IMR-phenomenological approach separates the frequency domain of a waveform into three phases, each with a phenomenological ansatz intended to produce a faithful representation or approximation of the strain in that phase. These ansatz functions include fitting parameters which are optimized during training, so the

model can be used to interpolate through parameter space. The general approach of these models is to construct a fast, phenomenological waveform that is hybridized between analytic (PN-based) and NR information where appropriate.

For the `IMRPhenom` family of models (e.g., [305, 429, 246]), the ansatz for the earliest phase, the inspiral, is primarily based on the TaylorF2 approximant. TaylorF2 is a simple, analytic PN approximant implemented as a closed-form frequency-domain expression with terms through 3.5PN order [166, 66, 121, 103]. As a PN approximation, TaylorF2 is only valid at low velocities ($v/c \ll 1$) and therefore is only appropriate for the earliest part of the waveform. The inspiral ansatz then includes additional higher-order terms beyond 3.5PN with coefficients fit to EOB inspiral waveforms during the construction of the waveform. This offers more accurate approximations of the strain during the later inspiral, where we do not have information from PN approximants.

The final parts of the waveform are purely phenomenological fits to NR simulations. The end of the waveform, the merger-ringdown, consists of a combination of power laws (for the phase or phase derivative) and a Lorentzian (to capture the quasi-normal mode behavior), all of which are multiplied by fitting coefficients. Between the inspiral and merger-ringdown is the intermediate region, which is a simple function of mostly power laws meant to smoothly stitch the late inspiral to the early merger-ringdown.

The coefficients are fit at the points in parameter space which have NR simulations and use a polynomial interpolation (in terms of PN-motivated physical parameters for mass and spin) to compute the value of the coefficients at arbitrary points in parameter space [305, 51]. The result of this process is an efficient, frequency-domain waveform approximant constructed from a mixture of NR, EOB, and PN information. As with `SEOBNR` models, the `IMRPhenom` models with "P" and "HM" suffixes include precession and higher order mode effects, respectively. Commonly-used aligned-spin models in this family are `IMRPhenomD` [305] and `IMRPhenomXAS` [429], the latter of which includes additional NR calibration and treatment of unequal-spin effects. The more complete approximant `IMRPhenomXPHM` [246, 429] has been widely used for most or all events in LVK catalogs, including those in GWTC3 and GWTC4. Time domain versions of IMRPhenom approximants, such as `IMRPhenomTPHM`, have been developed and used in the literature [202, 201]. While not generally as accurate as SEOB and NRSur, these approximants are typically fast to evaluate (O(1-10ms)) and are commonly used in PE analyses for the LVK. They are crucially employed in

the low-latency PE analyses (discussed in Chapter 4).

As modeling precession (by including non-aligned spin components) increases the parameter space for the waveform, dramatically increasing the size of the necessary training (i.e., NR simulation) data, waveform models take advantage of an approximate mapping between aligned and precessing systems as a shortcut for computing generic, precessing waveforms. The approximation specifically takes advantage of the fact that precession has a negligible effect on the rate of the inspiral and the precession timescale is much slower than the orbital timescale. Therefore, precession can be treated as a separate effect (a simpler transformation) on top of the inspiral of an aligned-spin system.

The process of modeling a waveform with precession in this manner relies on first computing the equivalent aligned-spin waveform, using, for example, `IMRPhenomD` or `IMRPhenomXAS`, which are used for this first step in some of the precessing waveforms in the `IMRPhenom` family. Then, this waveform is "twisted up", in which a time-dependent rotation is applied to (actively) transform the non-precessing waveform into the waveform with the correct in-plane (precessing) spin components.

The z-axis of the waveform in the first frame is fixed to track the orbital plane of the binary and only has spin components aligned with this axis; this frame is known as the co-precessing frame. The in-plane components of the spins, not included in the waveform in this step, are tracked in terms of the Euler angles (α, β, and γ, all functions of time) which prescribe the rotation between the inertial frame (with precessing spins) and the co-precessing frame. In the inertial frame, the z-axis is aligned with the total angular momentum and the waveform exhibits the effects of precession. The transformation of the waveform from the coprecessing to the inertial frame can be expressed in terms of Wigner-d matrices and the Euler angles (time dependence is suppressed) [428, 489]:

$$h_{\ell m}^{\text{inertial}} = \sum_{m'=-\ell}^{\ell} e^{-im\alpha} e^{-im'\gamma} d^{\ell}_{mm'}(\beta) h_{\ell m'}^{\text{coprecessing}}. \tag{3.29}$$

Frequency-domain waveforms assume the stationary phase approximation (SPA), which is formally only valid at lower velocities (lower frequencies, inspiral regime), in order to calculate the above as a function of gravitational wave frequency, introducing possible biases when using precessing, frequency-domain waveforms to analyze high-mass signals containing merger-ringdown content [267, 201]. Time domain waveforms suffer less of a bias when modeling precessing signals in the merger-ringdown regime, as they do not rely on SPA [201].

Several prescriptions exist for calculating the evolution of the Euler angles $(\alpha(t), \beta(t), \gamma(t))$. The simplest evolution prescription is a single-spin PN expansion considering only spin-orbit interactions with $\mathbf{S}_2 = 0$, with all of the spin on the larger black hole and the spin-orbit contributions calculated to next-to-next-to-leading-order (this is employed in `IMRPhenomPv2`) [305, 267, 490]. A newer *multiple scale analysis* approach derives a prescription for evolving the Euler angles by employing an expansion about the ratios of characteristic timescales in the system, which allows for solving PN precession equations and includes the ability to model spins on both black holes [144, 250, 304]. This is the default method employed in `IMRPhenomXPHM`. A final, newer method evolves the Euler angles by numerically integrating the expressions for the motion of the two spin vectors and the binary phasing up to 3PN and 3.5PN, respectively [152]. This approach, known as `SpinTaylor` evolution, tends to be more accurate and is the new default prescription used in the parameter estimation analyses in GWTC4.

Chapter 4

PROPERTIES OF GRAVITATIONAL WAVE SOURCES DETECTED IN GWTC-4

This chapter is intentionally redacted and will be embargoed until August 2025.

Chapter 5

THE CURIOUS CASE OF GW200129: INTERPLAY BETWEEN SPIN-PRECESSION INFERENCE AND DATA-QUALITY ISSUES

Ethan Payne, Sophie Hourihane, Jacob Golomb, et al. "Curious case of GW200129: Interplay between spin-precession inference and data-quality issues". In: *Phys. Rev. D* 106.10 (2022), p. 104017. DOI: 10.1103/PhysRevD.106.104017. arXiv: 2206.11932 [gr-qc].

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

5.1 Introduction

GW200129_065458 (henceforth GW200129) is a gravitational wave (GW) candidate reported in GWTC-3 [26]. The signal was observed by all three LIGO-Virgo detectors [1, 38] 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 [28], 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 [26]. The signal temporally overlapped with a glitch in the LIGO Livingston detector, which was subtracted using information from auxiliary channels [169]. The detection and glitch mitigation procedures for this event are recapped in Appendix 5.A.1.

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` [428] and `SEOBNRv4PHM` [408] waveform models using the parameter inference algorithms `Bilby` [67, 470] and `RIFT` [580] 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 [435]. 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 [428, 430, 408]. Data quality issues could further lead to evidence for spin-precession [69]. 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 [26].

Stronger conclusions in favor of spin-precession [266] and a merger remnant that experienced a large recoil velocity [548] were put forward by means of a third waveform model. `NRSur7dq4` [550] 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 [550]. This fact was exploited in Hannam *et al.* [266] 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 [65, 306]. 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., [120, 119, 489, 488, 267, 143, 144, 250, 303, 445]. The emitted GW signal is modulated in amplitude and phase, and morphologically resembles the beating between two spin-aligned waveforms [204] or a spin-aligned waveform that has been "twisted-up" [489, 488]. As the binary reaches merger, numerical simulations suggest that the direction of peak emission continues precessing [401]. Parameter estimation analyses using `NRSur7dq4` find that spins and spin-precession can be measured from merger-dominated signals for certain spin configurations [98], 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. 5.2 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 Appendix 5.A.1 and first identified in GWTC-3 [26]. 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 $q - \chi_p$ 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 [26]. 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. 5.3 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` [154, 343, 155] to simultaneously model a CBC signal and glitches with CBC waveform models and sine-Gaussian wavelets respectively [139, 286]. 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. 5.4 we revisit the LIGO Livingston data quality issues and compare the original glitch-subtraction based on `gwsubtract` [170, 169] 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. 5.5 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

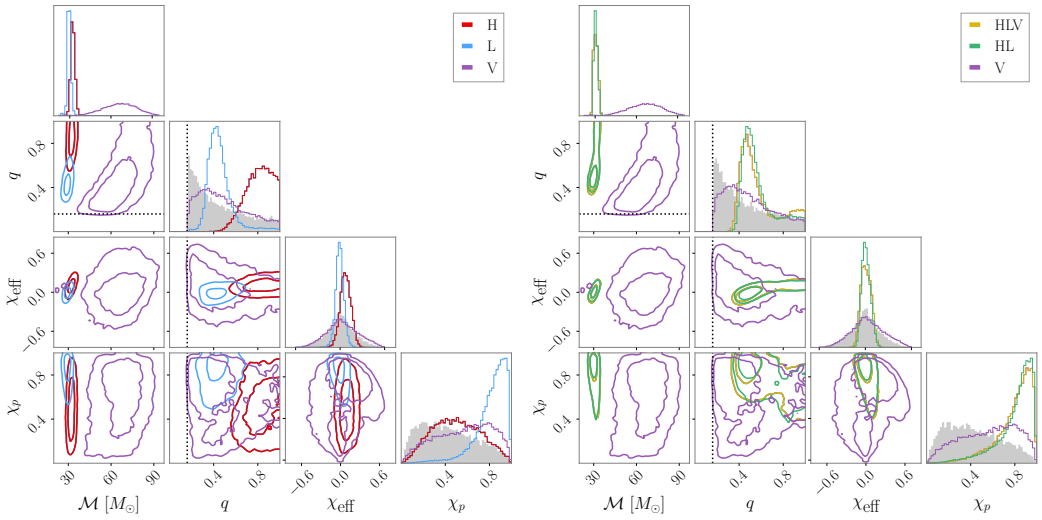

Figure 5.1: One- and two-dimensional marginalized posteriors for select intrinsic binary parameters: detector frame chirp-mass $\mathcal{M}$, mass ratio q, effective spin $\chi_{\rm eff}$, and precessing spin χ_p. See Table 5.1 for analysis settings and Appendix 5.A.2 for detailed parameter definitions. 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, $\chi_{\rm eff}$, χ_p. The $\mathcal{M}$ prior increases monotonically to the maximum allowed value (see Appendix 5.A.2 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 χ_p posteriors. Additionally, the binary masses inferred based on Virgo only are inconsistent with those from the LIGO data.

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.

5.2 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 Figure 5.1. Analysis settings and details are provided in Appendix 5.A.2 and in all cases we use the same LIGO Livingston data as GWTC-3 [26] 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 posterior is much closer to its prior when only LIGO Hanford or Virgo data are considered. A similar conclusion was reached in Hannam *et al.* [266]. 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 $q - \chi_p$ 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 χ_p, the left panel points to a bigger issue 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.

The role of Virgo data on the inferred binary extrinsic parameters is explored in

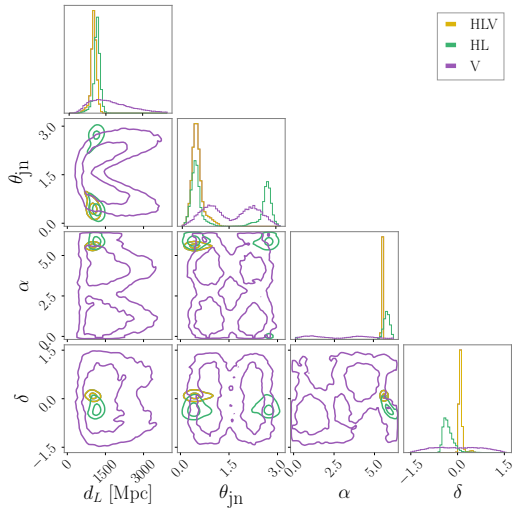

Figure 5.2: Similar to the right panel of Figure 5.1 but for select extrinsic parameters: luminosity distance d_L, angle between total angular momentum and line of sight $\theta_{\rm jn}$, right ascension α, and declination δ. For reference, the median optimal SNR for each run is HLV: 27.6, HL: 26.9, V: 6.7.

Figure 5.2. 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 5.3 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. 5.1 and 5.2, 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 Figure 5.1.

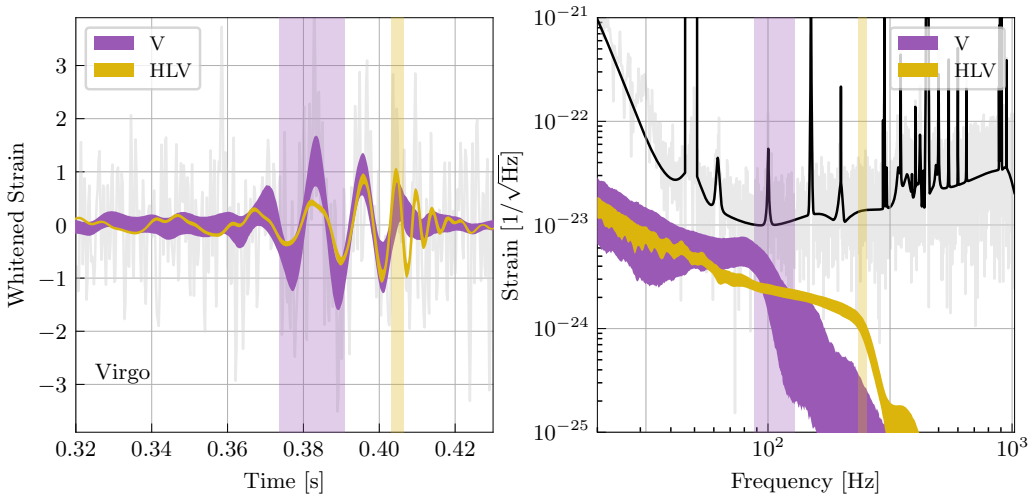

Figure 5.3: 90% credible intervals for the whitened time-domain reconstruction (left) and spectrum (right) of the signal in Virgo from a Virgo-only (purple; V) and a full 3-detector (yellow; HLV) analysis, see Table 5.1 for analysis settings. The data are shown in gray and the noise PSD in black. The 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 [26]. 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` [509], merger remnant properties were computed with `surfinBH` [551]). The Virgo data point to a heavier binary that merges $\sim$ 20ms earlier than the full 3-detector results that are dominated by the LIGO detectors.

The two reconstructions in Figure 5.3 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 ~0.382 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 ~0.382 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 ~2 cycles. As also concluded from the inferred masses in Figure 5.1, 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

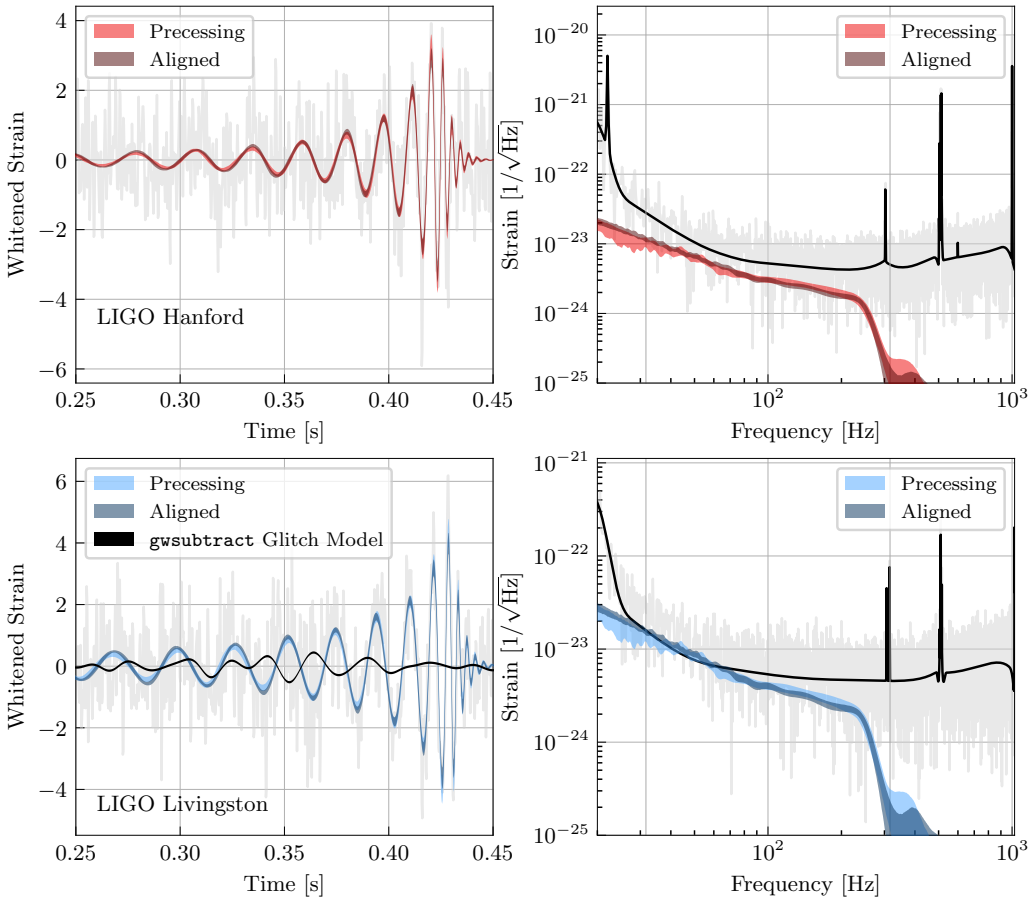

Figure 5.4: Whitened time-domain reconstruction (left) and spectrum (right) of GW200129 in LIGO Hanford (top) and LIGO Livingston (bottom). Shaded regions show the 90% credible intervals for the signal using a spin-precessing (light blue and red) and a spin-aligned (dark blue and red) analysis based on `NRSur7dq4`, see Table 5.1 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.

order to avoid further ambiguities for the remainder of this section we restrict to data from the two LIGO detectors unless otherwise noted. In Figure 5.1 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 Appendix 5.A.2 for details. Figure 5.4 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 ∼30–50 Hz and slightly smaller for $\gtrsim$ 100 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 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 5.5 shows the effect on the posterior for χ_p, q, and $\chi_{\rm eff}$. When we use the full data bandwidth, $f_{\rm low}(L) = 20\,$Hz, we find that q and χ_p are correlated and their two-dimensional posterior appears similar to the combination of the individual-detector posteriors from Figure 5.1. 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 $f_{\rm low}(L) = 50\,$Hz, χ_p is similar to its prior and further increasing $f_{\rm low}(L)$ 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 $f_{\rm low}(L) = 50\,$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 $\chi_{\rm eff}$ posterior is generally only minimally affected, with a small shift to higher values driven by the $q - \chi_{\rm eff}$

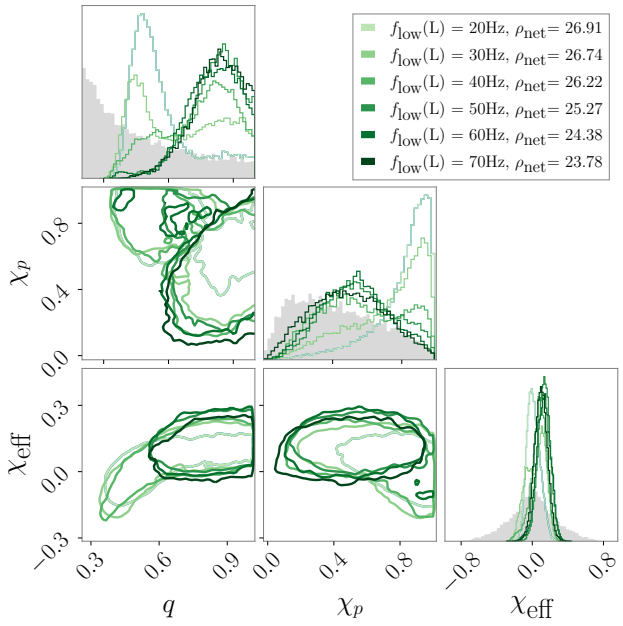

Figure 5.5: One- and two-dimensional marginalized posterior for the mass ratio q, the precession parameter χ_p, and the effective spin parameter $\chi_{\rm eff}$ for analyses using a progressively increasing low frequency cutoff in LIGO Livingston but all the LIGO Hanford data; see Table 5.1 for details. 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 χ_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.

correlation [163]. 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 $q - \chi_p$ inconsistency between LIGO Hanford and LIGO Livingston identified in Figure 5.1. 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

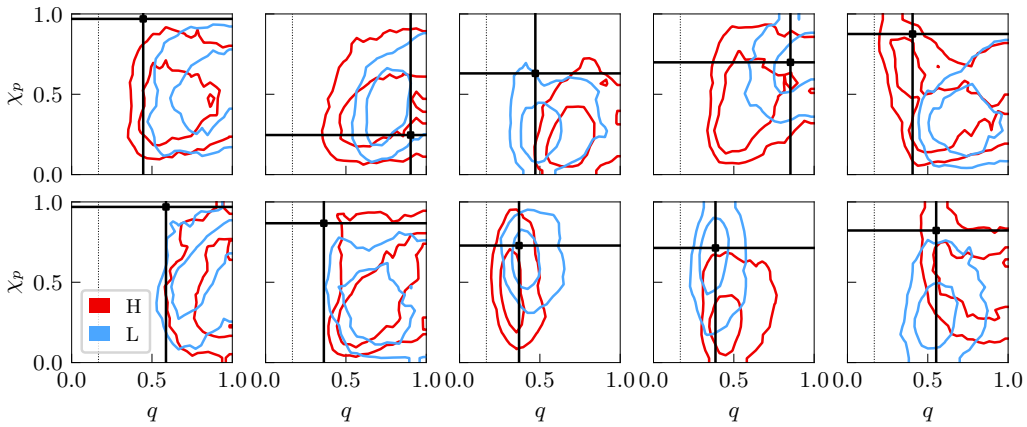

Figure 5.6: 90% contours for the two-dimensional marginalized posteriors for the mass ratio q and the precessing parameter χ_p obtained from analyzing data from each LIGO detector separately for 10 simulated signals. The signal parameters are drawn from the posterior for GW200129 when using LIGO Livingston 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 Figure 5.1 in $\mathcal{O}(5/100)$ injections.

compute the Bayes factor for overlapping posterior distributions relative to if the two distributions do not overlap [269, 268],

$$\mathcal{B}^{\text{overlapping}}_{\text{not overlapping}} = \iint \mathrm{d}\chi_p \mathrm{d}q \, \frac{p_L(\chi_p, q|d) p_H(\chi_p, q|d)}{\pi(\chi_p, q)}, \tag{5.1}$$

where we compute the overlap within the $q-\chi_p$ plane, $p_L(\chi_p, q|d)$ and $p_H(\chi_p, q|d)$ are the LIGO Livingston and LIGO Hanford posteriors, and $\pi(\chi_p, q)$ 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 GW200129), we find $\mathcal{O}(5/100)$ injections have a similar overlap as GW200129 (Figure 5.1). Figure 5.6 shows a selection of $q-\chi_p$ 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. 5.3 and argue that they should be removed from subsequent analyses. We reprise the spin-precession investigations in Sec. 5.4.

5.3 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 5.7 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 [465] and `BayesWave` using its glitch model that fits the data with sine-Gaussian wavelets; see Table 5.2 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 $1.11/\mathrm{min}$, with significant variation versus time [26]. When we consider the hour of data around the event, the rate of glitches with Omicron SNR > 6.5 is $10.2/\mathrm{min}$. Most of the glitches in Virgo at this time are due to scattered light [37, 346, 347, 40, 506]. While Figure 5.7 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

[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 $\sqrt{2}$ [465]. To highlight this difference, we use the phrase "Omicron SNR" whenever a reported result uses this normalization.

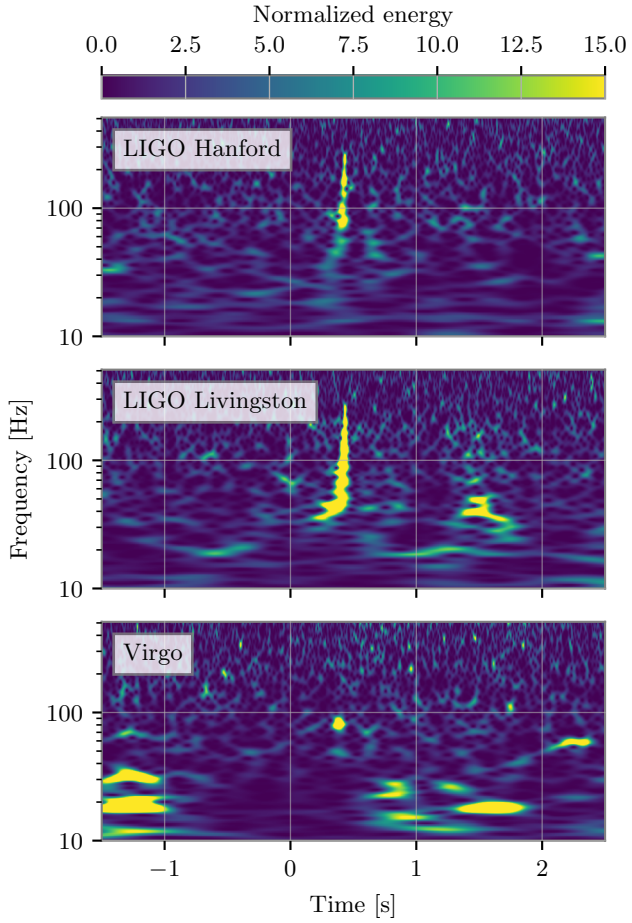

Figure 5.7: Spectrogram of the data in each detector, plotted using plotted using the Q-transform [136, 350]. Listed times are with respect to GPS 1264316116. Besides the clear chirp morphology in LIGO, there is visible excess power $\sim$ 1 s after the signal in LIGO Livingston. Virgo demonstrates a high rate of excess power, though most is due to scattered light and concentrated at frequencies < 30 Hz. The excess power in Virgo that is coincident with GW200129 does not have a chirp morphology.

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

Another option is that the Virgo trigger is a combination of a genuine signal and a detector glitch. We explore this possibility using `BayesWave` [154, 343, 155] to simultaneously model a potential CBC signal that is coherent across the detector network and overlapping glitches that are incoherent [139, 286]. In this "CBC+glitch" analysis, `BayesWave` models the CBC signal with the `IMRPhenomD` waveform [290, 305] and glitches with sine-Gaussian wavelets. Details about the models and run settings are provided in Appendix 5.A.3. 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.* [286] 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 Figure 5.1.

Figure 5.8 compares `BayesWave`'s reconstruction in Virgo with the one obtained with the `NRSur7dq4` analysis from Figure 5.3 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 Figure 5.1 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

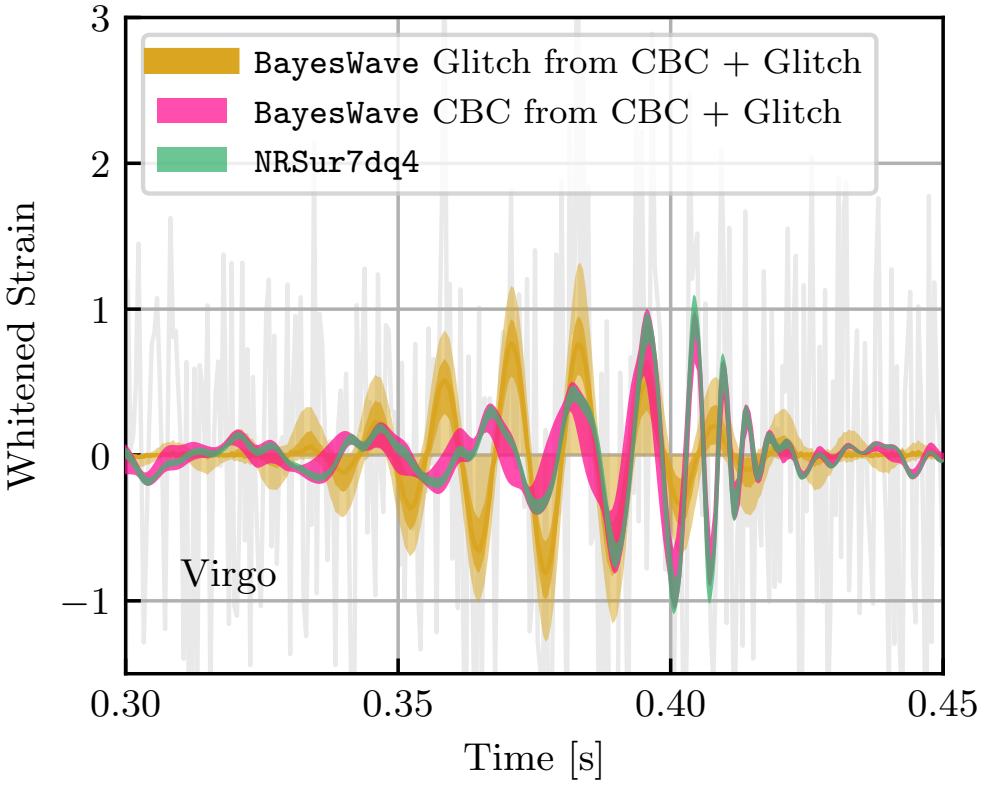

Figure 5.8: Whitened time-domain reconstruction of the signal in Virgo obtained after analysis of data from all three detectors relative to GPS 1264316116. Shaded regions correspond to 90% and 50% (where applicable) credible intervals. Green corresponds to the same 3-detector result obtained with `NRSur7dq4` as Figure 5.3, while pink and gold correspond to the CBC and glitch part of the "CBC+glitch" analysis with `BayesWave`. See Tables 5.1 and 5.2 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.

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 5.9 summarizes the various SNR estimates for the excess power in 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 Figure 5.8 with SNR > 4.6 is $0.31/\text{min}$ and the probability of overlap with a signal in Virgo is $O(10^{-3})$. 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 $O(0.1)$.

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.

5.4 Data quality issues: LIGO Livingston

The data quality issues in LIGO Livingston were identified and mitigated in GWTC-3 [26] through use of information from auxiliary channels [170, 169] and the `gwsubtract` pipeline as also described in Appendix 5.A.1. The comparison of

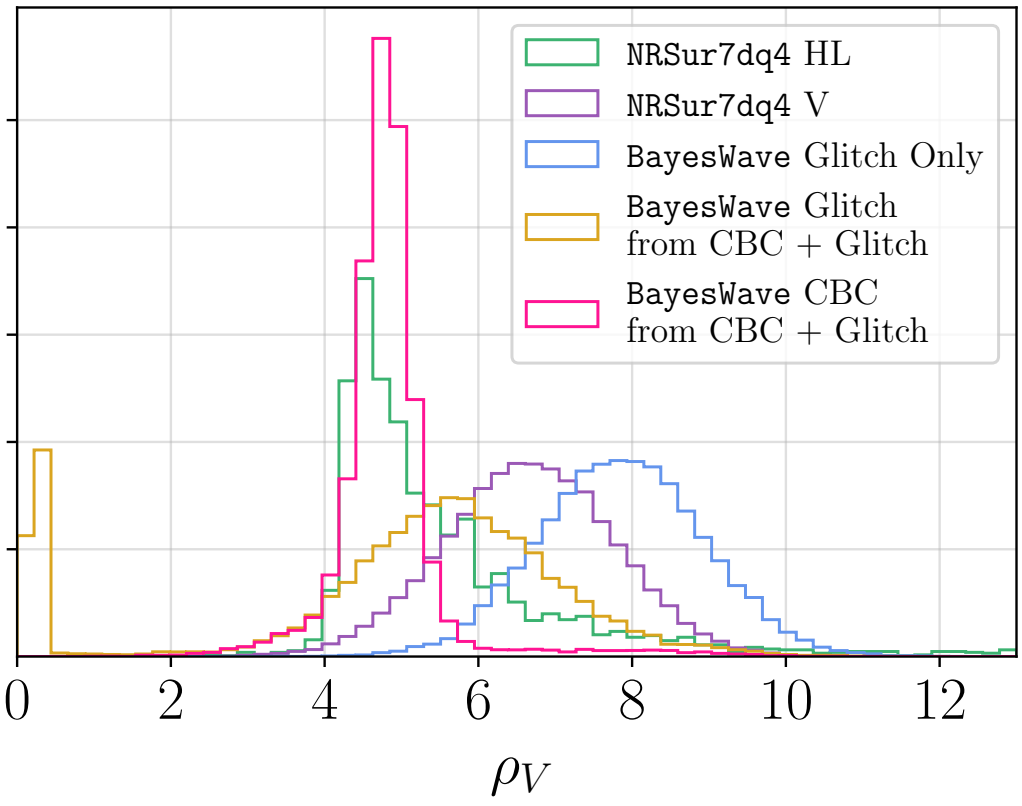

Figure 5.9: Comparison of optimal SNR estimates for Virgo from different analyses. In green is the posterior for the expected SNR in Virgo from just the LIGO data using the `NRSur7dq4` waveform (HL analysis of Figure 5.1), while purple corresponds to the SNR from an analysis of the Virgo data only (V analysis of Figure 5.1). The CBC and glitch SNR posterior from `BayesWave`'s full "CBC+glitch" model (Figure 5.8) 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 Figure 5.8). The SNR posterior from a "glitchOnly" `BayesWave` is shown in blue.

Figs. 5.1 and 5.6, however, suggest that residual data quality issues might remain, as the two LIGO detectors result in inconsistent inferred $q - \chi_p$ 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 Appendix 5.A.1.

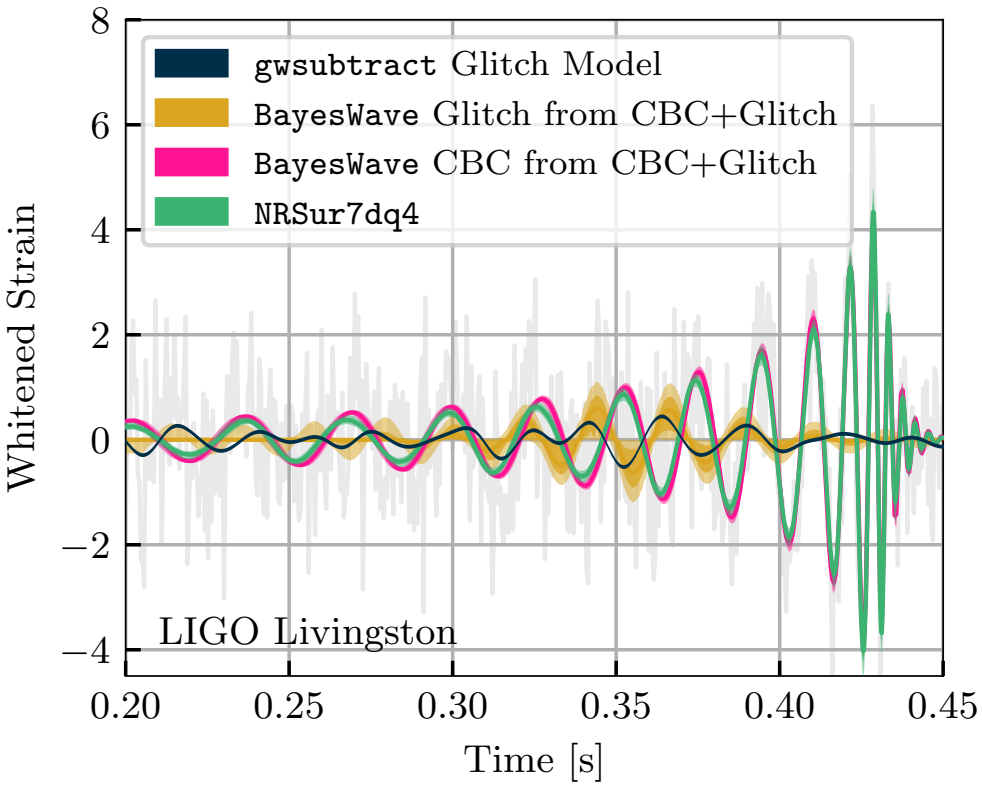

Figure 5.10: Whitened time-domain reconstruction of the data in LIGO Livingston obtained after analysis of data from the two LIGO detectors. Shaded regions correspond to 90% and 50% (where applicable) credible intervals and gray gives the original data without any glitch mitigation. Green corresponds to the same 2-detector result obtained with `NRSur7dq4` as Figure 5.4, 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.

Figure 5.10 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 Figure 5.4. 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 [154, 343, 155]. This approach has been previously utilized for glitch subtraction [26]. 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 in Figure 5.10) 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 Figure 5.10, 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 [416]. For reference, we also analyze the original unmitigated data (no glitch subtraction whatsoever). Figure 5.11 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 5.12 compares the corresponding $q - \chi_p$ posterior inferred from LIGO Hanford and LIGO Livingston separately under each different estimate for the glitch. Each of the 3 `BayesWave` glitch draws results in single-detector posteriors that fully overlap, thus resolving the inconsistency seen in $q - \chi_p$ when using the `gwsubtract` glitch estimate. Due to the lack of spin-precession modeling in the "CBC+glitch" analysis of Figure 5.10, 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 Figure 5.6 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 Figure 5.10 is smaller than the uncertainty in the glitch. This tests suggests that the glitch model is not strongly 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 Figure 5.7. 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 Figure 5.13. For the `gwsubtract` reconstruction we also include 90% confidence intervals, as described in Appendix 5.A.1.

The two estimates of the glitch are broadly similar but they do not always overlap

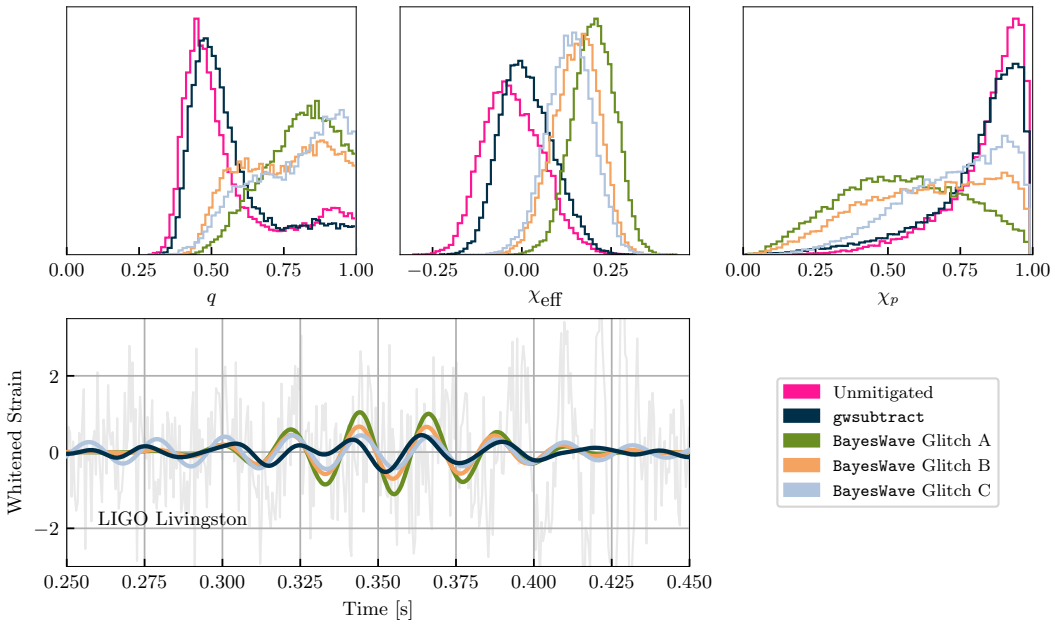

Figure 5.11: Bottom: Whitened, time domain reconstructions of various glitch reconstructions subtracted from LIGO Livingston data. The green line corresponds to the glitch reconstruction obtained from auxiliary data using `gwsubtract`. The rest are glitch posterior draws from the `BayesWave` "CBC+Glitch" 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 χ_p posteriors and eliminate the $q - \chi_p$ inconsistency between LIGO Hanford and LIGO Livingston.

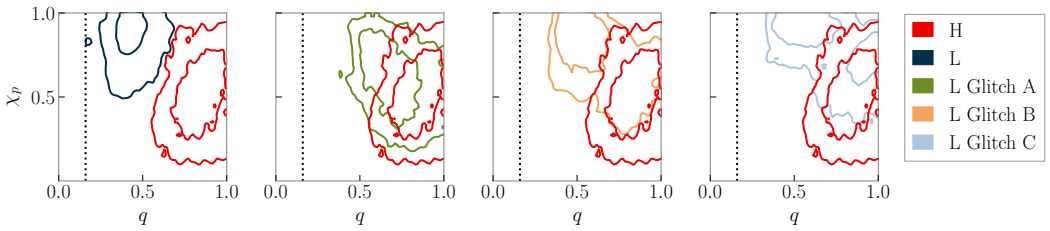

Figure 5.12: Two-dimensional posterior distributions for χ_p and q (50% and 90% contours) from single-detector parameter estimation runs. The far left panel shows the same tension as the LIGO Hanford and LIGO Livingston data plotted in Figure 5.1 when using the `gwsubtract` estimate for the glitch. Subsequent figures show inferred posterior distributions using data where the same three different `BayesWave` glitch models as Figure 5.11 have been subtracted. These results show less tension between the two posterior distributions.

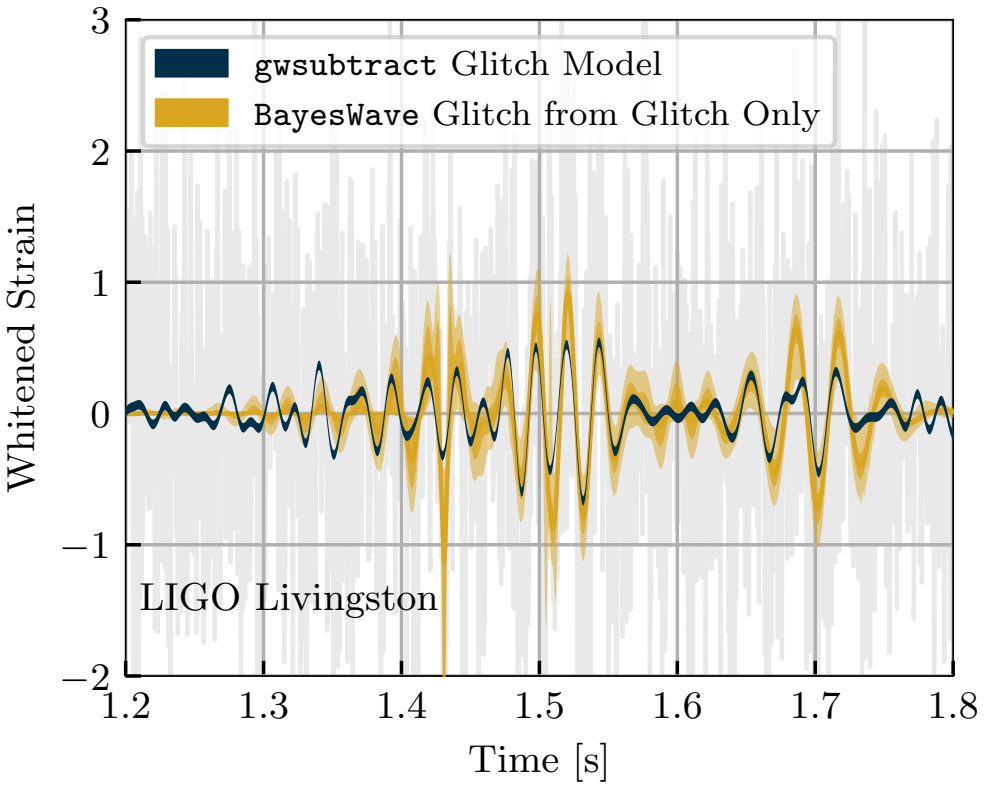

Figure 5.13: Comparison between the two glitch reconstruction and subtraction methods for a glitch in LIGO Livingston $\sim$ 1 s after GW200129, see the middle panel of Figure 5.7. We plot the original data with no glitch mitigation (grey), the glitch reconstruction obtained from auxiliary channels with 90% confidence intervals (black), and the 50% and 90% credible intervals for the glitch obtained with `BayesWave` that uses only the strain data (gold).

within their uncertainties. The main disagreement comes from the sharp data "spike" at $t = 1.43$ 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 [169]. 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 5.10 and 5.13 broadly suggest that `BayesWave` recovers a higher-amplitude glitch. Figure 5.11 shows that the evidence for spin-precession is indeed reduced, the LIGO Hanford-LIGO Livingston inconsistency is alleviated (Figure 5.12), and the LIGO Livingston data become more consistent across low and high frequencies

(Figure 5.5) 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.

5.5 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 [266] 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 [7, 29, 28]. However, the fact that there is no evidence for spin-precession if $f_{\rm low}(L) > 50\,{\rm Hz}$ 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* $q-\chi_p$ posterior compared to LIGO Livingston. Using simulated signals, we find that the latter, i.e., the $q-\chi_p$ inconsistency, is larger than $O(95\%)$ 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 Appendix 5.A.2 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., Figure 5.10, whereas the calibration uncertainty within 20 to 70 Hz is only $\sim$ 5% in amplitude and 5° in phase [513]. 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 [550, 266]. We repeated select analyses with `IMRPhenomXPHM` which also favored a spin-precessing interpretation for GW200129 [26]. 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 Appendix 5.B shows some example results. Nonetheless, our results are directly comparable to the ones of [266, 548] 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, Figure 5.11 shows that a difference in the glitch amplitude of $< 0.5\sigma$ can make the difference between an uninformative χ_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 [17, 33, 34]. 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, (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.

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` [350], `matplotlib` [289], `numpy` [270], `pandas` [526], `scipy` [556], `qnm` [509], `surfinBH` [551], `Bilby` [67], `LALSuite` [335], `BayesWave` [340].

Appendix 5.A 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 [36]. Throughout we use geometric units, $G = c = 1$.

Appendix 5.A.1 Detection and Glitch-subtracted data

GW200129 was identified in low latency [339] by GstLAL [375, 265], cWB [313], PyCBC Live [395, 165], MBTAOnline [41], and SPIIR [150]. 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 [26].

Multiple data quality issues were identified in the data surrounding GW200129. As a part of the rapid response procedures, scattered light noise [37, 40] was identified in the Virgo data, as seen in Figure 5.7 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 [8] 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 [168]. These data quality vetoes are used by some pipelines to veto any candidates identified during the data quality flag time segments [171]. 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 [26]. Due to the data quality veto, the signal was not identified in LIGO Livingston by the PyCBC [396, 167] MBTA [71], and cWB [313] 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 [477, 130] 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 [26, 169]

using the `gwsubtract` algorithm [170]. 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 [170], 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.* [169].

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 [169]. 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.

Appendix 5.A.2 `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, $i \in \{1, 2\}$. 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 $\mathcal{M} = (m_1 m_2)^{3/5}(m_1 + m_2)^{-1/5}$ [426, 103, 222], effective orbit-aligned spin parameter [440, 486, 49]

$$\chi_{\rm eff} = \frac{\vec{\chi}_1 \cdot \vec{L} + q\vec{\chi}_2 \cdot \vec{L}}{1 + q}\,, \tag{5.2}$$

Figure(s)	Waveform Model	Detector Network	Glitch mitigation	$f_{\rm low}$ (Hz)
5.1, 5.12	`NRSur7dq4`	H	`gwsubtract`	20
5.1, 5.12	`NRSur7dq4`	L	`gwsubtract`	20
5.1, 5.2, 5.3	`NRSur7dq4`	V	`gwsubtract`	20
5.1, 5.2, 5.3, 5.8	`NRSur7dq4`	HLV	`gwsubtract`	20
5.1, 5.2, 5.4, 5.10, 5.11, 5.14	`NRSur7dq4`	HL	`gwsubtract`	20
5.4	`NRSur7dq4` spin-aligned	HL	`gwsubtract`	20
5.5	`NRSur7dq4`	HL	`gwsubtract`	{20,30,40,50,60,70} in L, 20 in H
5.11	`NRSur7dq4`	HL	No mitigation	20
5.11	`NRSur7dq4`	HL	`BayesWave` fair draws	20
5.12	`NRSur7dq4`	L	`BayesWave` fair draws	20
5.14	`IMRPhenomXPHM`	HL	`gwsubtract`	20

Table 5.1: 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 5.6 also presents results for a set of 10 injections drawn from the LIGO Livingston only posterior distribution with $f_{\rm low}(L) = 20\,$Hz. These analyses use the same settings as above with $f_{\rm low}(L) = 20\,$Hz.

where $\vec{L}$ is the Newtonian orbital angular momentum, and effective precession spin parameter [267, 490]

$$\chi_p = \max\left(\chi_{1\perp}, q\chi_{2\perp}\frac{3q+4}{4q+3}\right), \tag{5.3}$$

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` [470, 502, 67] using the nested sampler, `Dynesty` [507]. The numerical relativity surrogate, `NRSur7dq4` [550], 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 [110] to $q > 1/4$ and component spin magnitudes $\chi < 0.8$, though it maintains reasonable accuracy when extrapolated to $q > 1/6$ and $\chi < 1$ [550].

The majority of our analyses use the publicly released strain data, including the aforementioned glitch subtraction in LIGO Livingston [169], and noise power spectral densities (PSDs) [26]. The exception to the publicly released data was the construction of glitch-subtracted strain data using `BayesWave` for LIGO Livingston, as discussed in Sec. 5.4. 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 [558, 418, 562, 190]. 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 [26], 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

Figure(s)	Models	Detector Network
5.8, 5.9	CBC+glitch	HLV
5.10, 5.11	CBC+glitch	HL
5.9	glitch	V
5.13	glitch	L

Table 5.2: 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. Appendix 5.A.1. 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. 5.4.

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 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. 5.1.

Appendix 5.A.3 `BayesWave` CBC and glitch analyses

`BayesWave` [154, 343, 155] 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 [286, 139, 570]. 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 Appendix 5.A.2) are sampled via a fixed-dimension Markov Chain Monte Carlo sampler (MCMC) using the priors described in Wijngaarden *et al.* [570]. 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 [154]) 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. 5.2.

Though `BayesWave` typically marginalizes over uncertainty in the noise PSD [343], in this work we use the same fixed PSD as the `Bilby` runs for more direct comparisons. Additionally, we use identical data as Appendix 5.A.2 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.

Appendix 5.B Select results with `IMRPhenomXPHM`

In this appendix, we present select results obtained with the `IMRPhenomXPHM` [428] waveform model that also resulted in evidence for spin-precession in GWTC-3 [26]. Even though `IMRPhenomXPHM` and `NRSur7dq4` both support spin-precesion, in contrast to `SEOBNRv4PHM`, there are still noticeable systematic differences between them. Figure 5.14 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. [26, 266, 288]), which motivates utilizing `NRSur7dq4` for all of our main text results.

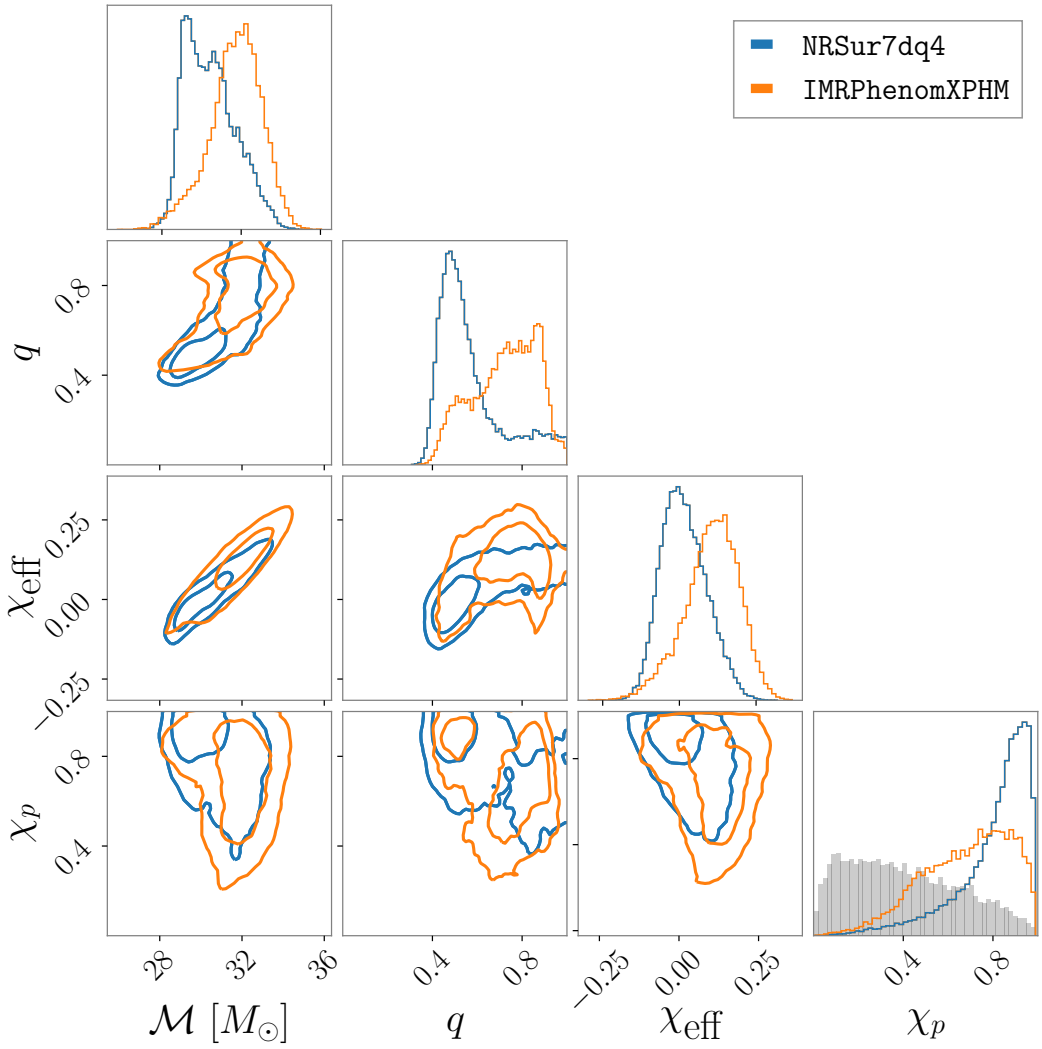

Figure 5.14: Similar to Figure 5.1, using data from LIGO Livingston and LIGO Hanford. The comparison shows slight tension between results when using `NRSur7dq4` and `IMRPhenomXPHM`, though qualitatively `IMRPhenomXPHM` also seems to support the evidence for spin-precession.

Part III

Astrophysics and Cosmology from the Population of Compact Binaries

Chapter 6

OVERVIEW OF POPULATION INFERENCE

6.1 Measuring Astrophysical and Cosmological Properties from the Population of Compact Binaries

Over the past decade, observations of compact binary mergers by the advanced LIGO and Virgo detectors [1, 38] have provided insight into the processes governing their formation and evolution, as well as constraints on fundamental physics. By combining astrophysical information from individual events, we can uncover the environmental and evolutionary characteristics of compact binary systems and leverage population properties as tracers of fundamental physics, stellar physics and cosmology.

Recognizing that our catalog of observations from the past decade forms a biased sample from a larger, unseen *astrophysical* population of compact objects, the process of *population inference* seeks to combine our observed dataset of CBCs and correct for known selection effects to constrain the astrophysical distribution of compact objects. Commonly this is done by learning the distribution of masses, spins, and redshifts of Compact Binary Coalescences (CBCs) as informed by data and selection effects.

Measuring population properties requires modeling the astrophysical distribution (or population model) that individual events are drawn from and estimating the detectability of the events in the underlying astrophysical distribution. This is expressed as a probability distribution over the single-event parameters (e.g., masses, spins) θ, given some hyperparameters Λ (e.g., power law index, cutoffs) controlling the shape of the population distribution, $p(\theta|\Lambda)$. Broadly speaking, the population model can be categorized as either low-dimensional and strongly parametric or as flexible and highly dimensional (so-called nonparametric, despite the higher number of hyperparameters), encoding different amounts of prior assumptions in the population.

6.1.1 Model Flexibility

In the *strongly parametric* approach to population inference, $p(\theta|\Lambda)$ makes relatively strong assumptions about the population, represented with a functional form with limited flexiblity controlled by few hyperparameters. For example, in a very simple and unrealistic case, the population model might be a truncated power law distribution (i.e., for the binary black hole mass distribution) with hyperparameters controlling the slope and truncation bounds, or it may be a Gaussian whose hyperparameters control the mean and standard deviation of the population. Strongly parametric models are useful in two main regimes: when the data are scarce or uniformative, or when hyperparameters Λ are meant to be interpreted as describing underlying physical processes in the context of a particular theory. Populations measured using strongly parametric models are of course subject to assumptions made about the functional form of the population and/or the validity of the description of the underlying physics controlling the population properties. For example, in GWTC-1 the BBH mass distribution was informed only by 10 events observed in O1 and O2, and was described as a simple truncated power law, motivated by the stellar initial mass function (IMF) [7]. Without enough observations to detect finer features, the approach becomes useful to encode reasonable assumptions (such as the BBH mass distribution should follow that of the stellar IMF) to learn broader, global features. As additional observations are made, more features may become resolvable, and the population model can be made more complex. As the catalog of gravitational wave sources grew to $\mathcal{O}(10-100)$ in GWTC-2, GWTC-3, and GWTC-4, parametric models with additional features have been adopted [29, 28]; Powerlaw + Peak featuring a power law with a Gaussian peak, and Broken Powerlaw, featuring two power laws which transition at a break mass, are examples of widely-used parametric mass models [524, 28, 29]. Examples of strongly parametric models used in the population analyses of GWTC-2 and GWTC-3 are shown in Figure 6.1.

As more event data become available, it becomes reasonable to loosen prior assumptions and rely more heavily on data-driven methods for population modeling. *Nonparametric* approaches offer a flexible alternative, representing the population model with a function that includes many adjustable parameters. Since strongly parametric models directly encode the possible features in the distribution, data-driven models can resolve new features whose shape and presence were not specifically outlined in the prior. For example, a strongly parametric approach to modeling a bump or

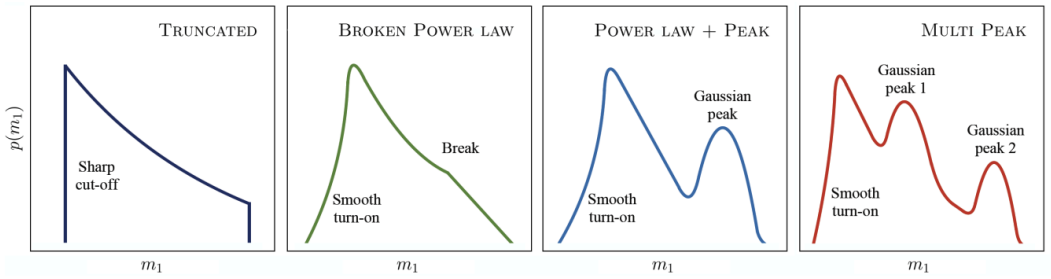

Figure 6.1: Examples of four strongly parametric models used in the population analsyses of GWTC-2 and GWTC-3 [29, 28]. Adopted from [29].

overdensity in the mass distribution may allow for the presence of a Gaussian with variable location and width, whereas a data-driven fit to the mass distribution would simply find a rise at the location of the overdensity. Examples of flexible functions used in the literature for nonparametric population inference include cubic splines [186, 187, 257], histograms [447], and autoregressive Gaussian process methods [125].

Whereas data-driven models offer *flexibility*, strongly-parametric models offer *interpretability*. Strong priors on the shape of the distribution are commonly physically motivated; for example, the powerlaw in Powerlaw+Peak is motivated by the stellar IMF, and the Gaussian peak and cutoff was originally introduced in [524] as a phenomenological signature of the pulsational pair instability supernova mechanism [578, 576, 210]. When the hyperparameters controlling the population features are closely associated with underlying motivating physical mechanisms, it is easier to use population measurements to constrain underlying physics. The downside of strongly parametric modeling is the risk for model mispecification. Since the assumption is that the underlying population truly follows the prescribed functional form for the distribution, if the true population is not well-represented by the chosen model, the resulting measurements and interpretations may be biased. It is therefore crucial to weigh the need for interpretability versus potential bias due to inflexibility and model misspecification when choosing a population model.

As discussed in detail in several works in this thesis, (i.e., Chapters 14,10,16), the other downside of flexible models is the risk of biased likelihood calculations due to systematic error from poorly-converged Monte Carlo integrals. In essence, the likelihood used in hierarchical Bayesian inference involves using numerical methods (Monte Carlo approximations) for solving mutlidimensional integrals. As any Monte Carlo estimator involves uncertainty, using a finite number of samples to

estimate the integrals in the likelihood propagates an uncertainty into the resulting likelihood estimate; this uncertainty is essentially ignored by a stochastic sampler during inference leading to a resulting systematic *bias*. For reasons explained in more detail in Chapter 14, as a population model allows for finer features (relative to the uncertainty on single-event measurements), the risk of a resulting bias in the likelihood increases. One way this can be mitigated on the model-level is by enforcing a correlation length or smoothness of a population model, reducing its flexibility *a priori*. Similarly, this can be mitigated for strongly parametric models by limiting the ability of certain features (e.g., "width" parameters, σ of a Gaussian feature, etc.) from getting too narrow.

6.2 Constructing the Hierarchical Likelihood

A population model quantifies the probability of observing the physical parameters (e.g., masses, spins) associated with an event assuming some functional form for the population model and population hyperparameters governing its shape. In population inference, we seek to contain the probability of different values for (i.e., probability of different population configurations), accounting for measurement uncertainties of single-event parameters and observational Malmquist biases or selection effects. In this section, I construct the likelihood function for the hierarchical Bayesian framework we use for population inference which incorporates these effects. See, e.g., [359, 560] for more details.

I first define some notation in this section:

- $p(a|b)$: The conditional probability of a given b. This probability distribution gives the probability density of a (over which it is normalized), assuming some parameter(s) b which control the conditional shape of the distribution.
- d_i: The data from the ith event. This is the strain data from the detector.
- $\{d_i\}$: The set of data from all observations in the catalog.
- θ: Physical (single-event) parameters describing the properties of a source. This typically refers to quantities such as masses, spins, distances, or locations, which control the gravitational wave signal and therefore parameterizes the per-event likelihoods $p(d_i|\theta)$.
- $N_{\rm det}$: The number of events in the catalog.

- Λ: The hyperparameters of the population model. This is a set of parameters which control the shape of the population distribution, such as a power law index, cutoff masses and truncation bounds, etc.
- R: The astrophysical merger rate (density), with dimensions of inverse volume-time.
- N_{exp}: The number of detections one would expect to make given Λ and the detection efficiency.

We start by assuming that *detection statistics* have been assigned to segments of data throughout an observation period, quantifying the relative chance that a segment of data contains a real signal. In reality, detection pipelines continually scan the data and trigger on segments of data that sufficiently match a signal template. The pipeline then assigns a False Alarm Rate (FAR) to each trigger, quantifying how often detector noise mimicking a signal would provide at least as good of a match to a signal template as the trigger did; in other words, the FAR of a trigger says how often triggers of instrumental origin would be found to be as "real" or "significant" as a given trigger.

For population analyses, the catalog of observed events is formed from all of the triggers which are more significant than some threshold ρ_{thresh}.

While ρ_{thresh} can technically be any detection statistic, such as Signal-to-Noise Ratio (SNR), we typically use a threshold on FAR, allowing us to control the purity (or noise contamination) of the catalog by approximating the number of noise triggers as $\approx \text{T}_{\text{obs}}\ \rho_{\text{thresh}} N_{\text{pipelines}}$ for a catalog formed from a FAR threshold of ρ_{thresh} and $N_{\text{pipelines}}$ operating over an observing time of T_{obs}. For example, in the GWTC-3 Astrophysical Distributions paper [28], the BBH population analysis used $\rho_{\text{thresh}} = 1\text{yr}^{-1}$, resulting in ~ 4 noise triggers in the catalog of 69 BBH observations, with the ~ 1 year observing time and ~ 4 (semi-)independent detection pipelines.

Consider a catalog of N_{det} observations. The likelihood of observing the data from the ith event (d_i) given population hyperparameters Λ is simply the posterior probability and probability of detecting d_i, marginalized over θ:

$$p(\text{det}, d_i|\Lambda) = \int p(d_i|\theta)p(\theta|\Lambda)p(\text{det}|d_i)d\theta. \tag{6.1}$$

Note that the product of the first two terms in the integrand is proportional to the posterior probability where the population model takes the role of the prior. The first

term is simply the (Gaussian noise) likelihood from Chapter 3, the probability of observing the strain data given θ. In our detection scheme, $p(\mathrm{det}|d_i)$, the probability of making a detection given d_i, is 1 for all observations meeting the threshold for detection ρ_{thresh} and is 0 otherwise, so it is not technically a properly normalized probability distribution.[1] The measurement uncertainty is explicitly accounted for in Eq. 6.1 by marginalizing the single-event posterior over all possible values for θ.

Because θ is typically multidimensional, integral over θ is not typically computationally feasible to calculate on a grid over θ for any given population model $p(\theta|\Lambda)$. Instead, we use a Monte Carlo approximation to the integral, in which we *reweight* the posterior samples output from the original parameter estimation analysis of each event to the prior given by $p(\theta|\Lambda)$. This is generally a reasonable approach when the posterior samples for each observation were obtained under some relatively broad, agnostic prior $p_0(\theta)$, giving the posterior non-zero support across relevant regions of parameter space, and $p(\theta|\Lambda)$ is a function that is easy to evaluate. Recalling that the posterior samples from the per-event parameter estimation are drawn from $p_0(\theta|d_i) \propto p(d_i|\theta)p_0(\theta)$, we then reweight these posterior samples to calculate Eq. 6.1 as:

$$\begin{aligned} p(\mathrm{det}, d_i|\Lambda) &= \int \frac{p(d_i|\theta)p_0(\theta)}{p_0(\theta)} p(\theta|\Lambda)p(\mathrm{det}|d_i)d\theta \\ &\approx \frac{1}{S} \sum_{\theta_j \sim p_0(\theta|d_i)} \frac{p(\theta_j|\Lambda)}{p_0(\theta_j)} \end{aligned} \tag{6.2}$$

where S is the number of posterior samples drawn from the original parameter estimation analysis of the event, and j indexes over the samples.

By the rules of conditional probabilities, the likelihood of the data d_i given Λ is

$$p(d_i|\Lambda) = \frac{p(\mathrm{det}, d_i|\Lambda)}{p(\mathrm{det}|\Lambda)}. \tag{6.3}$$

The normalization factor in the denominator of Eq. 6.3 is simply the joint probability marginalized over d:

$$p(\mathrm{det}|\Lambda) = \int p(\mathrm{det}, d|\Lambda)dd = \int p(d|\theta)p(\theta|\Lambda)p(\mathrm{det}|d)d\theta dd. \tag{6.4}$$

Eq. 6.4 is the probability of making a detection from the population characterized by Λ, or equivalently the fraction of detectable events in that population. This term

[1] One could express this as $p(\mathrm{det}|d_i) = \Theta(\rho(d_i) - \rho_{\mathrm{thresh}})$ by identifying ρ with a detection statistic that *increases* with detection significance, such as Inverse-FAR. Here, Θ is the Heaviside step function.

corrects for the selection effects or Malmquist bias, ensuring that we are inferring the *astrophysical* rather than the *observed* distribution. See Appendix 6.A for a more detailed discussion of how to compute this term in practice.

The likelihood of the dataset from the $N_{\rm det}$ observations is simply the joint likelihood of the individual observations:

$$p(\{d_i\}|\Lambda) = \prod_i^{N_{\rm det}} p(d_i|\Lambda). \tag{6.5}$$

Note that since Eq. 6.4 depends only on Λ and not the properties of an individual event i, it is common to calculate it once for each value of Λ considered and express Eq. 6.5 in this equivalent factorized form:

$$p(d_i|\Lambda) = p(\text{det}|\Lambda)^{-N_{\rm det}} \prod_i^{N_{\rm det}} p(\text{det}, d_i|\Lambda). \tag{6.6}$$

While Eq. 6.6 looks sufficient for measuring the shape of the population, there is a missing normalization term necessary for measuring the *rate* of mergers in the population of interest. To measure this, we need the likelihood of observing the data in the catalog *and* the catalog containing $N_{\rm det}$ observations.

With the overall (observed and unobserved) number of mergers N_Λ, the number of detections we expect to make is simply $N_{\rm exp}(\Lambda) = N_\Lambda p(\text{det}|\Lambda) = R\langle VT\rangle(\Lambda)$, where $\langle VT\rangle$ is the sensitive four-volume, the detectable four-volume of the search averaged over sources in the population Λ. We now recognize that the size of our catalog of $N_{\rm det}$ discrete observations is a realization from a Poisson distribution about $N_{\rm exp}$. The corresponding likelihood for $N_{\rm det}$ is therefore:

$$p\big(N_{\rm det}|N_{\rm exp}, \Lambda\big) = \frac{N_{\rm exp}^{N_{\rm det}} e^{-N_{\rm exp}}}{N_{\rm det}!}. \tag{6.7}$$

Where I suppressed the dependence of $N_{\rm exp}$ on Λ. Note that because $\langle VT\rangle$ is completely determined by Λ and the detector behavior, only either N_Λ or the merger rate itself (usually expressed in dimensions of inverse volume-time) needs to be the free parameter controlling the normalization term/merger rate. They are simply related by the spacetime volume of the search (known *a priori*); typically the merger rate is used as the free parameter in population inference. So we are free to replace Eq. 6.7 with $p(N_{\rm det}|R, \Lambda)$.

The total likelihood is the joint likelihood of Eq. 6.5 and Eq. 6.7:

$$p(\{d_i\}, N_{\rm det}|\Lambda, R) = p(N_{\rm det}|R, \Lambda) p(\text{det}|\Lambda)^{-N_{\rm det}} \prod_i^{N_{\rm det}} p(\text{det}, d_i|\Lambda), \tag{6.8}$$

where it is understood that R is used in place of N_Λ, as they share a simple relation. It is common to not explicitly include the $p(\text{det}|d_i)$ term, as it contributes a factor of 1 for all detections (i.e., see Eq. 6.1).

If we assume a prior on R or N_Λ that is uniform-in-log, i.e. $p(R) \propto 1/R$, and that $p(R)$ will decouple from $p(\Lambda)$, we can analytically marginalize over the dependence on R and obtain a simpler likelihood that has one fewer degree of freedom:

$$p(\{d_i\}, N_{\text{det}}|\Lambda) \propto p(\text{det}|\Lambda)^{-N_{\text{det}}} \prod_i^{N_{\text{det}}} p(\text{det}, d_i|\Lambda). \tag{6.9}$$

See Appendix 6.B for a derivation of this form. Note that this effectively reduces to the form of the likelihood in Eq. 6.6, but it is due to a convenient choice of prior on R. Eq. 6.9 is the likelihood serving as a typical starting point in the population inference literature.

Appendix 6.A Computing $p(\mathbf{det}|\Lambda)$

The "selection effects" term of the likelihood, Eq. 6.4, acts as a normalization equal to the fraction of detectable events from the population Λ. The integrand has three key ingredients: the likelihood of the data (noise model), the likelihood of the single-event physical parameters θ assuming the population Λ (population model), and the detection probability. This is clearly not solvable analytically. For one, while $p(\theta|\Lambda)$ may be chosen to be a simple function, θ is typically multidimensional, and $p(d|\theta)$ depends in a very complicated way on θ, due to the details of the detector response and waveform approximants. We therefore resort to numerical methods such as Monte Carlo importance sampling to approximate Eq. 6.4 [538, 213].

A Monte Carlo approximation relies on the fact that the average of a quantity x assuming some distribution $p(x)$ can be written as $\langle x \rangle_{f(x)} = \int x f(x) dx$. Therefore, Eq. 6.4 says that $p(\text{det}|\Lambda)$ is the expectation value of $p(\text{det}|d)$ under the joint distribution $p(d|\theta)p(\theta|\Lambda)$. Just as $\langle x \rangle_{f(x)}$ can be estimated with S samples (like any average) as $\langle x \rangle_{f(x)} \approx \frac{1}{S} \sum_{x_i \sim f(x)} p(x)$, we approximate $p(\text{det}|\Lambda)$ as a sample-based average of S samples drawn from the population:

$$p(\text{det}|\Lambda) \approx \frac{1}{S} \sum_{d_i \sim p(d|\theta)p(\theta|\Lambda)} p(\text{det}|d_i). \tag{6.10}$$

The above Monte Carlo approximation involves the following procedure:

1. Draw S (typically a very large number) samples of θ from the population model $p(\theta|\Lambda)$, $\theta \sim p(\theta|\Lambda)$.

2. Add the corresponding simulated signal $h(\theta)$ to detector noise, real or simulated. As $d = n + h$, this is equivalent to drawing $d \sim p(d|\theta)$.

3. Calculate the detection probability $p(\text{det}|d_i)$ for each of S samples.

4. Take the average of $p(\text{det}|d)$ across samples. This is $p(\text{det}|\Lambda)$.

Each step above has varying degrees of computational burden. Step 1 can be accomplished using rejection sampling or inverse transform sampling from the population. Step 2 involves generating a waveform S times (for each θ) and generating (or reading) noise. Step 3 requires the most computational burden for realistic scenarios, as the detection statistics must be assigned to the simulated samples consistent with how it was done for the real data; this therefore requires running detection pipelines

over the simulated data. With data corresponding to a real observing run, this is a highly complicated procedure, typically taking months at a time and significant coordination between teams of detection pipelines. Again, the complexity here is necessary to k exactly the method used for the real data, or else the $p(\mathrm{det}|d)$ term in Eq. 6.10 will not match the detection procedure used to make the real catalog (i.e., Eq. 6.1). As we need to calculate $p(\mathrm{det}|\Lambda)$ for arbitrary values of Λ during population inference, it is clearly unfeasible to run the above procedure generically.

We therefore use *importance sampling* or *reweighting* as a shortcut for approximating Eq. 6.10 for arbitrary Λ. Assume we have conducted Steps 1-3 above for some fiducial population Λ_0. Multiplying the integrand in Eq. 6.4 by $\frac{p(\mathrm{det}|\Lambda_0)}{p(\mathrm{det}|\Lambda_0)}$, we recognize:

$$\begin{aligned} p(\mathrm{det}|\Lambda) &= \int p(d|\theta)p(\theta|\Lambda)p(\mathrm{det}|d)d\theta dd \\ &= \int p(d|\theta)\frac{p(\theta|\Lambda)}{p(\theta|\Lambda_0)}p(\theta|\Lambda_0)p(\mathrm{det}|d)d\theta dd \\ &= \left\langle \frac{p(\theta|\Lambda)}{p(\theta|\Lambda_0)}p(\mathrm{det}|\Lambda)\right\rangle_{p(d|\theta)p(\theta|\Lambda_0)} \end{aligned} \tag{6.11}$$

which we approximate as:

$$p(\mathrm{det}|\Lambda) \approx \frac{1}{S}\sum_{d_i,\theta_i\sim p(d|\theta)p(\theta|\Lambda_0)} \frac{p(\theta_i|\Lambda)}{p(\theta_i|\Lambda_0)}p(\mathrm{det}|d_i), \tag{6.12}$$

or, in short,

$$p(\mathrm{det}|\Lambda) \approx \frac{1}{S}\sum_{\theta_{\mathrm{found}}} \frac{p(\theta_i|\Lambda)}{p(\theta_i|\Lambda_0)}, \tag{6.13}$$

where we denote θ_{found} as the samples from $p(\theta|\Lambda_0)$ which, when passed through the data generation step, $p(\mathrm{det}|d_i) = 1$, meaning they were assigned as "found" by the detection process. In other words, we take samples of θ from our fiducial population $p(\theta|\Lambda_0)$, simulate signals and sample (real or simulated) noise, run our detection procedure and store the values of θ which pass our detection threshold. Then, for an arbitrary target population $p(\theta|\Lambda)$, we simply need to calculate the "weight", the ratio of the target to fiducial population.

The key advantage of the above procedure is that the simulation and detection procedure is only done once, and we can recycle these results to efficiently compute $p(\mathrm{det}|\Lambda)$ forother values of Λ. This is the exact procedure we use to calculate the selection effects for rates and population properties from LIGO-Virgo-KAGRA observations.

It is worth noting, however, that the above reweighting procedure will return a biased result for $p(\text{det}|\Lambda)$ if there is not sufficient overlap in the support between the fiducial and target distributions. This can be understood as being due to estimating the average of a distribution with effectively too few samples. We therefore must be clever in choosing a Λ_0 with broad enough support and generate enough detectable sources to make Eq. 6.13 a sufficient approximation. See Chapter 14 for an in-depth study of how these poor Monte Carlo approximations translate into biases in the inferred population.

Appendix 6.B Marginalization over rate

The fact that all terms involving R or N_{exp} in Eq. 6.8 are in a Poisson distribution allows these terms to be analytically marginalized with a suitable choice of prior [228]. Consider the two equivalent forms for the likelihood of N_{det}, and let $\alpha(\Lambda) \equiv \langle VT \rangle = p(\text{det}|\Lambda)VT$:

$$\begin{aligned} p(N_{\text{det}}|N_{\text{exp}}) &= N_{\text{exp}}^{N_{\text{det}}} e^{-N_{\text{exp}}} \\ p(N_{\text{det}}|R,\Lambda) &= \big(R\alpha(\Lambda)\big)^{N_{\text{det}}} e^{-R\alpha(\Lambda)}. \end{aligned} \tag{6.14}$$

If we assume separable priors on R and Λ, then the part of the posterior that includes R can be written independently as

$$p(R, N_{\text{det}}|\Lambda) = p(R)p(N_{\text{det}}|R,\Lambda). \tag{6.15}$$

With a convenient choice of prior $p(R)$, we can analytically marginalize over R in Eq. 6.15, removing the need to include the R degree of freedom in the population inference stochastic sampling process. The distribution for $p(R|\Lambda, N_{\text{det}})$ can be reconstructed afterward.

If we choose a uniform-in-log prior such that $p(R) \propto 1/R$, then:

$$p(R, N_{\text{det}}|\Lambda) \propto \alpha(\Lambda)^{N_{\text{det}}} R^{N_{\text{det}}-1} e^{-R\alpha(\Lambda)}, \tag{6.16}$$[2]

which has an antiderivative $-\Gamma(N_{\text{det}}, \alpha(\Lambda)R) + C$, meaning

$$\int_0^\infty p(R, N_{\text{det}}|\Lambda) dR \propto \Gamma(N_{\text{det}}, 0) = \text{const.} \tag{6.17}$$

The fact that the result of the above marginalization is *independent* of Λ, means we can choose the prior $p(R) \propto 1/R$ and simplify the explicit form of Eq. 6.8 into Eq. 6.9.

[2] One may recognize this as proportional to the gamma distribution, with "shape" N_{det} and "rate" $\alpha(\Lambda)$.

Chapter 7

BINARY BLACK HOLE POPULATION THROUGH GWTC-3

This chapter is my own overview of the population properties of compact binary systems as presented in the LIGO-Virgo-KAGRA GWTC-3 Astrophysical Distributions paper. I have adapted the results from the publication with emphasis on my contributions.

> R. Abbott et al. "Population of Merging Compact Binaries Inferred Using Gravitational Waves through GWTC-3". In: *Phys. Rev. X* 13.1 (2023), p. 011048. DOI: 10.1103/PhysRevX.13.011048. arXiv: 2111.03634 [astro-ph.HE].

7.1 Introduction

The third observing run (O3) of the Advanced LIGO [117] and Advanced Virgo [38] gravitational wave observatories ran from April 2019 to March 2020, during which signals were observed from compact binary coalescences comprised of black holes and neutron stars. The third Gravitational Wave Transient Catalog (GWTC-3) [26] combines the observations from the second half of O3 (O3b) with those from O1, O2, and O3a, resulting in a catalog consisting of a cumulative total of 90 gravitational wave events from compact binary coalescences (CBCs) with significance surpassing a particular chosen significance threshold [24, 25, 26]. In this paper, we present the population properties of black holes and neutron stars in binaries as inferred from the gravitational wave data collected through GWTC-3.

The observations in O3 contain several novel events which have significant properties in their own right. For example, O3 observations contain the first observations of a binary system consisting of a neutron star and a black hole (NSBH), the most massive binary black hole system (BBH) observed to date, and systems made of binaries of highly asymmetric masses [23, 27, 22]. The presence of these sources enrich our dataset and provide crucial information for understanding the formation and evolution of compact binary systems.

By combining the individual observations and correcting for the observational bias inherent in the detection process, we can measure the astrophysical distribution of

BBH, NSBH, and binary neutron star (BNS) systems, allowing us to draw conclusions on a population level. These population properties include the merger rate distributions of masses, spins, and redshifts of CBC sources. By encoding physical parameterizations and motivations into models, we can constrain underlying astrophysical processes and formation channels of compact binary systems, as well as leverage these constraints to measure global environmental and cosmological properties.

The population analyses in GWTC-2 identified a few features in the BBH mass distribution such as an overdensity or break in the power law and a nonzero merger rate for masses $\gtrapprox 45M_\odot$ [29]. Additionally, it was found that some fraction of BBHs have spins aligned against the orbital angular momentum and that the spin distribution has evidence for systems undergoing precession. In this work, we further explore these findings, utilizing new data and more flexible models which allow us to search for new features. Furthermore, the inclusion of NSBH events allows us to conduct the first joint analysis of the entire population of CBCs of masses between $1 - 100M_\odot$.

By including the new observations in GWTC-3, we identify or confirm several features of the BBH population. We identify the presence of new over- and underdensities in the BBH mass distribution, begging new astrophysical explanations [180, 70, 84, 508, 174, 390, 467, 310, 371, 63, 593, 79, 278, 496, 363, 405, 113, 357]. We confirm and strengthen earlier findings that the spin components of BBHs tend to be small but nonzero, with no evidence of a significant rapidly spinning population. We also find evidence of support for anti-aligned spins in the population, as well as a correlation between mass ratio and spins of BBH systems; these findings are unexpected given typical astrophysical models for BBH systems originating from the collapse of massive stars [126]. Furthermore, for the first time, we confidently measure an evolution of the BBH merger rate with redshift, a finding which may be expected from observations of the star formation rate history of the universe.

For analysis which include NSBH and BNS systems, we adopt a threshold of FAR $< 0.25\text{yr}^{-1}$, resulting in a total of 67 events, 4 of which contain at least one NS. Due to the higher number of BBH observations, we loosen the threshold to FAR $< 1\text{yr}^{-1}$ for the BBH-focused analyses, resulting in 69 confident BBH events in the bulk population.

7.2 Data and Inference Methods

7.2.1 Data Collection

The detection pipelines GSTLAL [477, 265, 375], PyCBC [394, 56, 55, 114, 546, 396], MBTA [41], and CWB [314, 313] were used to identify possible signals in the data, assigning significance estimates to each trigger. By selecting events that surpass a significance threshold, we put an upper bound on the number of noise events in the dataset, setting the purity of our catalog. A FAR $< 1\mathrm{yr}^{-1}$ threshold gives a cumulative total of 76 events, 6 of which contain NSs and one outlier with ambiguous classification, GW190814. We expect $\lesssim 5$ noise events contaminating dataset given the ~ 4 independent detection pipelines operating in the ~ 1 year observing time.

For each event in the dataset, we perform Bayesian parameter estimation to measure its properties. Posterior samples for the properties of each event are obtained using Bilby [67, 470], LALInference [335], or RIFT [328, 580]. For each event, we equally-weight posterior samples from published parameter estimation analyses with waveforms including effects of precession, and using higher order modes where available. For events in O1, this includes analyses with SEOBNRv3 [525, 411] and IMRPhenomPv2 [267]. For events detected in O3a reported in GWTC-2, we use results tagged as PrecessingSpinIMRPHM, which contains results from waveforms modeling effects of preession and higher order modes. For new events in O3b, we use equally-weighted results obtained from analyses using IMRPhenomXPHM [428] and SEOBNRv4PHM [108, 408]. All analyses are performed with a prior uniform in detector frame component masses, uniform in spin magnitude and isotropic on the sphere, and proportional to the square of the luminosity distance.

7.2.2 Population Inference

We follow the procedure outlined in Chapter 6 and other chapters in this thesis in order to infer the parameters describing the population models. Additional background can be found in [536, 359, 560, 348]. Assuming a uniform-in-log prior over the rate R, we obtain the likelihood for the dataset $\{d\}$ given the population hyperparameters Λ:

$$\mathcal{L}(\{d\}|\Lambda) \propto \prod_{i=1}^{N} \frac{\int \mathcal{L}(d_i|\theta)\pi(\theta|\Lambda)d\theta}{\xi(\Lambda)}, \tag{7.1}$$

where $\xi(\Lambda)$ is the detection efficiency for the population Λ. See Chapter 6 of this thesis for a detailed description of the likelihood (i.e., Eq. 6.9) and these terms. Given that, for each event, we have samples from the posterior distribution which assumed some fiducial "PE prior" $\pi_0(\theta)$, we use importance sampling of these samples to obtain an estimate for Eq. 7.1:

$$\mathcal{L}(\{d\}|\Lambda) \propto \prod_{i=1}^{N} \frac{1}{\xi(\Lambda)} \langle \frac{\pi(\theta|\Lambda)}{\pi_0(\theta)} \rangle, \tag{7.2}$$

where the average is taken over the samples drawn from the posterior of each event. Several packages implement the functionality of computing Eq. 7.2, but I focus on reuslts obtained using the package GWPopulation [517, 520, 518].

The detection deficiency quantifies the fraction of sources in Λ that would be deteactable according to the detection process used to form the dataset of real events. In order to efficiently evaluate $\xi(\Lambda)$ for different values of Λ, this is calculated by importance sampling over samples (or "reweighting") from an injection-recovery campaign. In this procedure, we simulate $N_{\rm inj}$ signals from millions of CBC sources drawn from a fiducial population model Λ_0, add them to real detector noise, and recover them using the same detection pipeline used to form the dataset. The ratio of the number of detected injections (i.e., those that pass the FAR threshold) to the total number of injections gives the detection efficiency $\xi(\Lambda_0)$. We obtain $\xi(\Lambda)$ for arbitrary Λ by importance sampling over the detected injections [213, 538]:

$$\xi(\Lambda) \approx \frac{1}{N_{\rm inj}} \sum_{i=1}^{N_{\rm found}} \frac{\pi(\theta_i|\Lambda)}{\pi(\theta_i|\Lambda_0)}, \tag{7.3}$$

where i runs over the "found", or detected, injections. Note that the denominator in Eq. 7.3 is the draw probability of the ith found injection under the fiducial population Λ_0. See Chapter 6 for details of this component of the likelihood.

The Monte Carlo-based estimate of each integral above has an associated uncertainty due to using a finite number of samples used in the sum [256, 213, 257, 519]. This uncertainty is always present when using samples to compute an estimator. To mitigate the bias from likelihoods with unconverged integrals, we only assign nonzero values of $\mathcal{L}(\{d\}|\Lambda)$ to values of Λ for which we believe the likelihood integrals are properly converged, as determined by the effective sample size ($N_{\rm eff}$) of each Monte Carlo sum. Specifically, we enforce that $N_{\rm eff} > N_{\rm det}$ for the average in Eq. 7.2 and $N_{\rm eff} > 4N_{\rm det}$ for Eq. 7.3, where $N_{\rm det}$ is the number of observed events in the catalog.

7.2.3 Population Models

In this section, I will summarize a select few of the population models employed in the GWTC-3 populations analysis [28], with a particular emphasis on the parametric population models I personally contributed to. All models in this section are assumed to be factorizable in the following sense:

$$p(m_1, q, \vec{s}_1, \vec{s}_2, z|\Lambda) = p(m_1|\Lambda)p(q|m_1, \Lambda)p(\vec{s}_1|\Lambda)p(\vec{s}_2|\Lambda)p(z|\Lambda). \tag{7.4}$$

We adopt several models for the BBH mass distribution $p(m_1, q|\Lambda)$. The exact functional forms of each model and priors on the population hyperparameters can be found in Appendix B of [28]. These parametric mass models follow the same conditional mass ratio distribution:

$$p(q|m_1, \beta_q, m_{\rm min}) \propto \begin{cases} q^{\beta_q} & \text{if } m_{\rm min} < m_2 < m_1 \\ 0 & \text{otherwise.} \end{cases} \tag{7.5}$$

The following mass models describe the mass distribution for the primary masses, $p(m_1|\Lambda)$.

The simplest model is a truncated power law (Truncated [29, 226]), controlled with $\Lambda = (\alpha, m_{\rm min}, m_{\rm max})$, the (negative) power law slope and the minimum and maximum truncation masses, respectively. Physically, this is motivated by the power law nature of the stellar initial mass function (IMF) and the expectation of minimum and maximum BH mass cutoffs.

The Broken Power Law model is similar to the Truncated model, but with two power law slopes, which transition at a break location $m_{\rm break}$.

The default, fiducial mass model adopted in this work (and commonly adopted in the literature) is the Powerlaw+Peak model [524, 228], which is a mixture of two components: a power law component and a Gaussian peak. The power law component is similar to the Truncated model, but features a smooth turn-on (with width δ_m) at low masses to avoid an abrupt, and possibly unphysical, truncation. The peak component is a Gaussian centered at μ_m with width σ_m, and contains a fraction λ_{peak} of the mass distribution. This model was first motivated by the expected presence of a peak and cutoff in the mass distribution due to the pulsational pair instability mechanism. While a peak-like feature has been confidently found, its location determined by this work likely makes it inconsistent with being a signature from pulsational pair instability remnants.

The MULTIPEAK model is similar to POWERLAW+PEAK, but includes an additional Gaussian peak, allowing more flexibility in the model to capture additional features.

We define the DEFAULT spin model [521] in terms of the magnitudes ($\chi_i \equiv |\vec{s}_i|$) and tilt angles ($\theta_i$) of the spin vectors of each object in a binary. We model spin magnitudes as a beta distribution [582]:

$$p(\chi_i|\alpha_\chi, \beta_\chi) \propto \chi_i^{\alpha_\chi - 1}(1-\chi_i)^{\beta_\chi - 1}, \tag{7.6}$$

where we restrict $\alpha_\chi > 1$ and $\beta_\chi > 1$ to avoid singular distributions.

Tilt angles are assumed to come from a mixture of two distributions: a uniform distribution in $\cos(\theta_i)$ and a (truncated) Gaussian distribution centered at $\cos\theta_i = 1$ with width σ_t:

$$p(\cos\theta_1, \cos\theta_2|\sigma_t) = \frac{1}{4}(1-\zeta) + \zeta \mathcal{N}_{[-1,1]}(\cos\theta_1, \cos\theta_2; 1, \sigma_t), \tag{7.7}$$

where ζ is the mixing fraction between the two distributions and $\mathcal{N}_{[-1,1]}$ is a Gaussian truncated to $-1 \leq \cos\theta_i \leq 1$. This is meant to be a phenomenological representation of two formation channels which may contribute to the population. The isotropic (uniform) component corresponds to the subpopulation of BBHs formed dynamically, with spins randomly oriented with respect to the orbit. The Gaussian component is the subpopulation of BBHs formed under common (isolated) evolution and therefore are expected to have spin tilts preferentially aligned with the orbial angular momentum. This form assumes that both components of a given binary are drawn from the same subpopulation.

Motivated by the power law nature of the star formation rate (SFR) [351], we assume the merger rate (density) $R(z)$ follows [228]:

$$R(z|\kappa) = R_0(1+z)^\kappa, \tag{7.8}$$

where R_0 is the local ($z = 0$) merger rate in dimensions of inverse volume-time, with time defined *in the frame of the source*. The parameter κ describes the evolution of the merger rate with redshift, and is expected to be positive due to the increase in star formation rate with redshift for $z \lesssim 2$. Recognizing that $R(z) = \frac{dN}{dV_c t_s}$, we express redshift probability distribution as:

$$p(z|\kappa) \propto \frac{dN}{dz} \propto \frac{dN}{dV_c dt_s}\frac{dt_s}{dt_d}\frac{dV_c}{dz} \propto (1+z)^\kappa \frac{1}{1+z}\frac{dV_c}{dz}, \tag{7.9}$$

where $dt_s/dt_d = \frac{1}{1+z}$ is the time dilation factor and dV_c/dz is the differential comoving volume element. The proportionality factor is set by the overall normalization of $p(z|\kappa)$.

7.2.4 Other Population Models

Results for the BBH population presented in [28] also include other population models, which I did not personally contribute to. Here is a short summary of some of those models.

Power Law + Spline [186] parameterizes deviations to underlying power law using cubic splines at knots placed uniformly in $log m_1$-space. By inferring the spline heights, we can measure over- and under-densities with respect to the power law across m_1-space. This semiparametric approach allows us to find certain features that could not be captured with the lower-dimensional parametric models whose allowable features are fewer and encoded *a priori*.

Flexible Mixtures [539] models the chirp mass, mass ratio, and aligned component spins as a weighted mixture of 11 Gaussians.

Binned Gaussian Process [238, 358] is a histogram-based approach for modeling the mass distribution. The heights of each bin with predefined locations, are inferred and correlated with neighnoring bins via a Gaussian process.

7.3 Population Properties of Black Holes in Binaries

Unless otherwise stated, results in this section were obtained by simultaneously inferring the population according to the Power Law + Peak mass model, Default spin model, and Power Law redshift model. We use 69 events with FAR $< 1\mathrm{yr}^{-1}$, not including the outliers GW190814 ($q \sim 0.11$) and GW190917 (possible NSBH).

7.3.1 Mass Distribution

The mass distribution of BBH systems is one of the most important properties to measure, encoding information about compact binary and stellar evolution, formation and evolution. We find confidently that the mass distribution has substructures beyond a simple power law, and the merger rate is monotonically decreasing and continuous for some range beyond $m > 50M_\odot$.

We present the results of the Power Law + Peak model in Figure 7.1. Comparing the blue and black credible regions, we note broad consistency between the new results and those obtained from GWTC-2.

We find a power law slope of $\alpha = 3.5^{+0.6}_{-0.56}$ with a Gaussian peak at $\mu_m = 34^{+2.6}_{-4.0}\,M_\odot$, consistent with the findings from GWTC-2 ($\alpha = 2.6^{+0.79}_{-0.63}$ and $\mu_m = 33^{+4.0}_{-5.6}\,M_\odot$). The

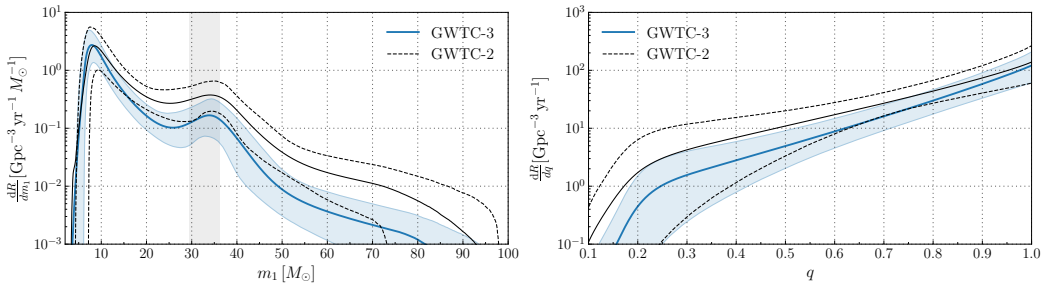

Figure 7.1: Astrophysical distribution for primary mass (left) and mass ratio (right) obtained with the Power Law + Peak model. The solid lines are the posterior population distribution (PPD) and the dashed lines envelope the 90% credible region. The black and blue are GWTC-2 and GWTC-3 results, respectively. From [28].

Gaussian peak has a width of $\sigma_m = 4.6^{+4.1}_{-2.5}$ and contains a fraction of $\lambda = 0.038^{+0.058}_{-0.026}$ sources. While consistent, this peak is slighly less pronounced than what was found from GWTC-2 $\left(\lambda = 0.1^{+0.14}_{-0.071}\right.$ and $\left.\sigma_m = 5.7^{+3.8}_{-3.6}\right)$.

With a higher fraction of low-mass sources in O3b, we find that the mass distribution decays much faster than what was found in GWTC-2. Specifically, we find the mass of the 99th percentile is $m_{99\%} = 44^{+9.2}_{-5.1}\ M_\odot$, which is considerably lower than the 99th percentile found in GWTC-2 (which found $m_{99\%} = 60^{+14}_{-13}\ M_\odot$).

We measure a power law slope for q which is just slightly less peaked toward equal-mass binaries than in GWTC-2 ($\beta_q = 1.1^{+1.7}_{-1.3}$ vs $\beta_q = 1.3^{+2.4}_{-1.5}$).

With the broader set of parametric models, we find that the BBH population is not consistent with simply a single power law, requiring the presence of a feature at $\sim 35M_\odot$. With the Broken Power Law model, this feature can be accomodated with a break around the same location that Power Law + Peak places the Gaussian peak. With similar evidences between the two cases, we cannot confidently determine whether the feature at this location is best described by a peak or a break.

The additional observations collected through GWTC-3 help us resolve new substructures in the BBH mass distribution. With the more data-driven models we find peaks at $m_1 \approx 10M_\odot$ and the familiar peak at $m_1 \approx 35M_\odot$, as well as a possible, less-significant feature at $m_1 \approx 17M_\odot$. This additional structure is clearly illustrated in the mass distribution recoveries shown in Figure 7.2. The significance of these features can be quantified according to the deviation parameters in Power Law + Spline, which favors overdensities at $m_1 \approx 10M_\odot$ and $m_1 \approx 35M_\odot$ with 97.8% and $> 99.9\%$ credibility, respectively. This model also finds a potential dip at

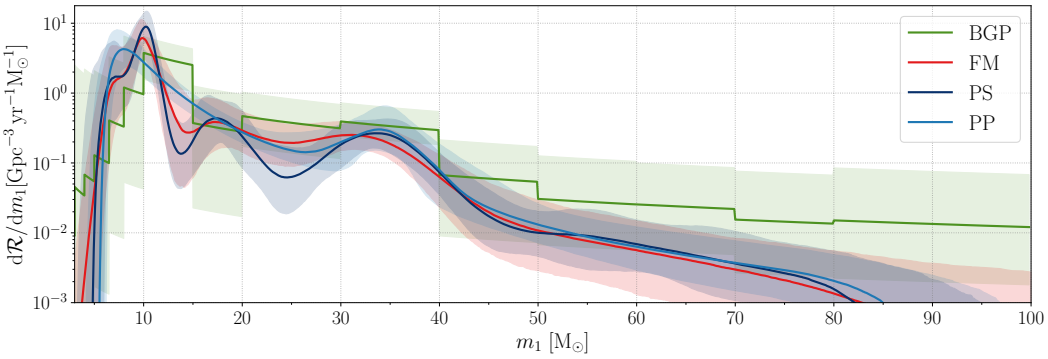

Figure 7.2: Differential merger rate for the primary mass with Power Law + Peak (PP) as well as three more data-driven models Binned Gaussian Process (BGP), Flexible Mixture (FM), and Power Law + Spline (PS). The solid lines are the posterior population distribution (PPD) and the dashed lines envelope the 90% credible region. The black and blue are GWTC-2 and GWTC-3 results, respectively. From [28].

$m_1 \approx 14M_\odot$, but at 96.1% credibility.

7.3.2 Spin Distribution

The spins of BBHs can imprint the evolutionary and formation history of the system, which, when constrained on a population level, can help us understand the astrophysical context in which BBH systems form [362, 322, 468, 218, 561, 510, 248, 402]. The magnitudes and directions of black hole spins may carry signatures of angular momentum transport of the stellar progenitors, prior merger history, and possibly the strength of supernova kicks. [242, 241, 89, 298, 299, 468, 402, 248, 581, 27] As noted previously, the spin directions $\cos\theta_{1,2}$ are of particular interest for distinguishing isolated from dynamical formation scenarios [362, 468].

In Figure 7.3, we present our constraints on the component spin magnitude and tilt distributions. Our constraints are considerably better than those from GWTC-2[1], finding χ is preferentially concentrated $\lessapprox 0.4$, with half of the support below $\lessapprox 0.25$. There is a tail extending toward high χ, but its contribution to the population is much less than lower values for χ are. In terms of the tilt angles, we rule out the case of perfect spin alignment ($\zeta = 1$ and $\sigma_t = 0$), finding stronger contribution of the isotropic spin tilt population. Under this flatter distribution, we find that 44^{+6}_{-11}% of black holes have anti-aligned spin components, $\cos\theta < 0$.

[1]However, this is likely due to consideration of precession in the selection effects.

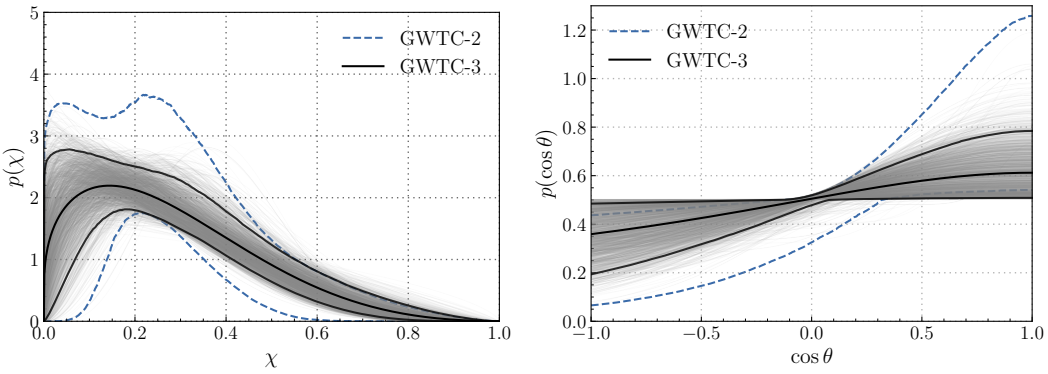

Figure 7.3: The distribution of spin magnitudes (χ) and tilt angles ($\cos\theta$) for the Default spin model for GWTC-3 (blue) and GWTC-2 (black). Thicker solid and dashed lines envelope the 90% credible regions, and light gray traces are individual draws from the posterior. From [28].

Searching for support for anti-aligned spins is of interest due to isolated, common evolution channels likely being unable to form such systems, unless natal spin kicks are stronger than expected. In adiditon to the constraints on $\cos\theta$ presented above, we infer the spin distribution under the $\chi_{\rm eff} - \chi_p$ parameterization [382], finding support for $\chi_{\rm eff} < 0$ in the population at $> 90\%$ credibility, with exact values depending on the model implementation details. Details of the spin distribution inferred under the $\chi_{\rm eff}$ and χ_p parameterization can be found in [28].

Finally, we find an apparent anti-correlation between the mass ratio and $\chi_{\rm eff}$ in the population [126]. We infer that unequal-mass BBH systems preferentially exhibit larger $\chi_{\rm eff}$ at 97.5% credibility.

7.3.3 Redshift Distribution

Observing BBH mergers across cosmic distances allows us to place constraints on the *evolution* of the BBH merger rate [228]. Motivated by the SFR evolution [351], we assume the merger rate follows a simple power law in redshift, $R(z) \propto (1+z)^{\kappa}$. While the true merger rate evolution may deviate from this trend, this should be sufficient to find the dominant evolutionary behavior at the redshifts at which we have observations.

The Power Law $R(z)$ model reduces to a merger rate that is uniform in comoving volume and source-frame time for $\kappa = 0$. We measure $\kappa = 2.9^{+1.7}_{-1.8}$, excluding $\kappa = 0$ at $> 99.6\%$ credibility, indicating a confident detection of an evolving merger rate with redshift. This is in contrast to GWTC-2, which found $R(z)$ was consistent with

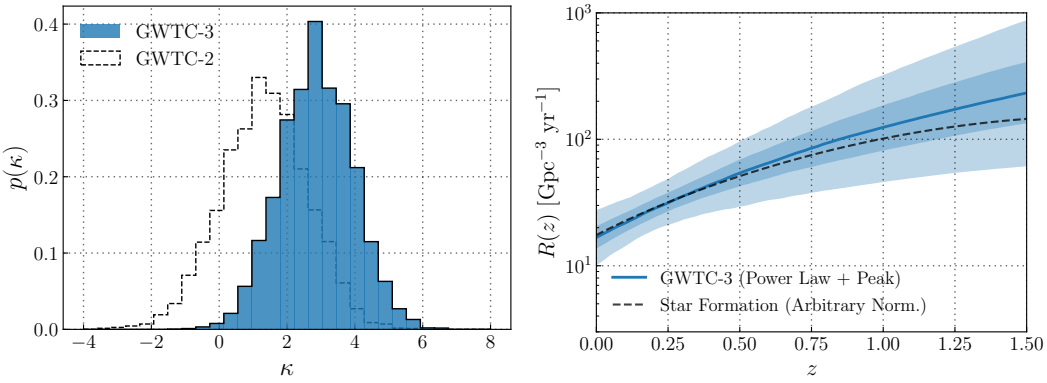

Figure 7.4: Posterior distribution on the slope of the $R(z)$ power law (top) and central 50% and 90% credible bounds of $R(z)$ (bottom). We constrain κ to be $2.9^{+1.7}_{-1.8}$, resulting in the rising merger rate seen in the bottom plot. For reference is the cosmic SFR from Madua-Dickinson. Note the best constraints come from $z \approx 0.2$.

being non-evolving. This difference, however, is due to an insufficient treatment of precession when estimating selection effects in GWTC-2, rather than the new data.

We present our measurements for $R(z)$ in Figure 7.4, comparing with the low-redshift SFR constraints from Madau-Dickinson, which correspond to $\kappa_{\rm SFR} = 2.7$. We infer an $R(z)$ consistent with this evolution. Given the delay times between star formation and merger, the star formation and BBH merger rates are not expected to be exactly the same.

We make our best measurements for $R(z)$ at $z = 0.2$, finding $R(z = 0.2) = 28^{+14}_{-8.9}$ Gpc^{-3}yr^{-1}. Given that our constraints come from these relatively low redshifts, we do not seek to constrain the behavior of $R(z \gtrapprox 1-1.5)$, which is expected to decay after reaching some peak. Given the SFR peaks at $z \approx 2$, it is likely that the peak of $R(z)$ for BBH mergers also lies beyond our detection horizon, but may be constrained with future observations and stochastic background searches.

7.3.4 Outliers in the BBH Population

For our constraints on the BBH population distribution, have excluded events whose massses make their classification as BBH systems ambiguous. In particular, we exclude GW190814 and GW190917, whose secondary masses extend to lower masses than the bulk of the events in the BBH population. In this section, we repeat the population analysis with these events included in turn, demonstrating the subsequent changes to our results.

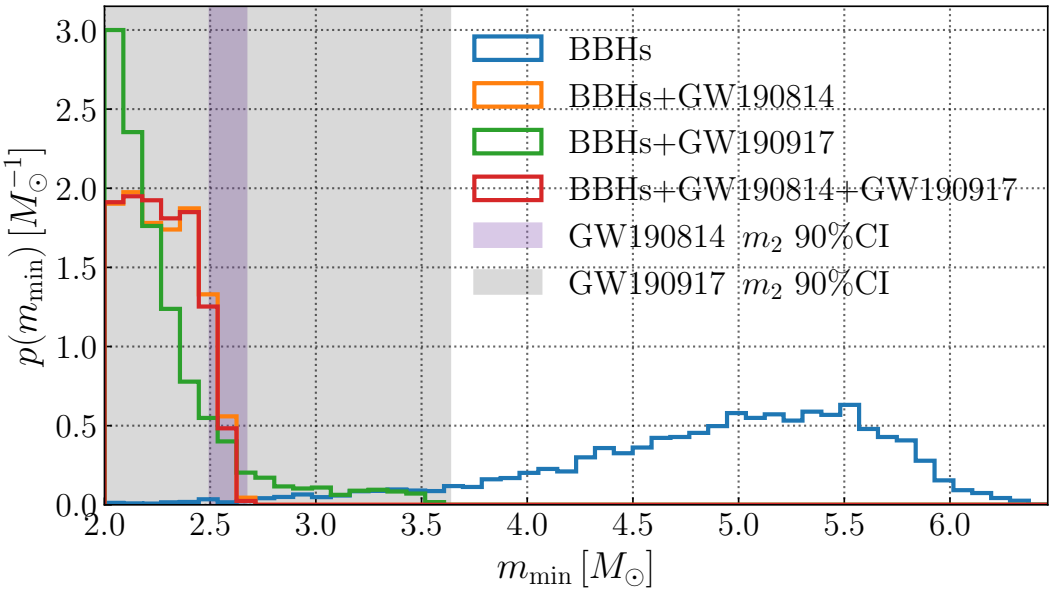

Figure 7.5: The posterior distribution on the minimum mass truncation $m_{\rm min}$ obtained with Power Law + Peak. Results are presented with different combinations of including and excluding GW190814 and GW190917. The cutoff at $m_{\rm min} = 2M_\odot$ corresponds to the lower bound of the prior. The inclusion of either GW190814 or GW190917 significantly effects the inferred $m_{\rm min}$. The shaded regions indicate the 90% credible interval on the m_2 posterior distribution for GW190814 (purple) and GW190917 (gray).

In Figure 7.5, we present the inferred lower mass truncation hyperparameter for the mass distribution $m_{\rm min}$ for different combinations of inclusion and exclusion of GW190814 and GW190917. With our default event list, which excludes these two events, we find $m_{\rm min} = 5.1^{+0.75}_{-1.5} M_\odot$ with a broad turn on $\delta_m = 5.0^{+3.3}_{-3.2} M_\odot$. When we include these two low-mass events, we find a much lower value of $m_{\rm min} = 2.3^{+0.27}_{-0.23} M_\odot$ and a sharp turn on of $\delta_m = 0.39^{+1.3}_{-00.36} M_\odot$. This difference is due to the fact that we force $m_{\rm min}$ for the primary and secondary mass distributions to be the same; therefore, the secondary masses of the low- and assymetric- mass BBH systems drive the changes at the low end of the mass spectrum.

The population inferred when including these events is in signficant significant tension with that which is inferred when they are excluded, meaning the population inferred without these events cannot even support the presence of them. We suggest it is likely that these systems are members of a separate subpopulation, disconnected from the bulk BBH population, and possibly connected to the newly-identified population of NSBH systems.

7.4 Rates

The merger rates of the different populations of compact binary systems are important for understanding the formation and evolution of these systems, as well as predicting the number of detectable events in future observing runs. We present the merger rates for the different populations across the mass spectrum in Table 7.3.4, and for just the BBH population in Table 7.3.4.

7.5 Conclusions

The observations in GWTC-3 have provided us with a more thorough census of the population of CBC sources in the universe. With the additional data collected through O3b, we infer the astrophysical distribution of BBH, BNS, and NSBH systems, constraining the merger rates of these systems as a function of mass, spin, and redshift.

We find that the neutron star mass distribution is broad and extends up to $m \sim 2M_\odot$, contrasting with the distinct narrow Gaussian structure of the Galactic BNS mass distribution [410]. While the limited data make the tension not significant, it is possible that the two populations disagree due to differing formation channels or poorly-understood observational and astrophysical selection effects [11, 481, 471].

We find structure in the BBH mass distribution, notably overdensities over a power law at $m_1 \approx 10M_\odot$ and $\approx 35M_\odot$. The peak at $10M_\odot$ needs further investigation as it is difficult to explain with mainstream astrophysical models. For example, significant contributions from globular clusters to the BBH population are expected to supress the merger rate at low masses, resulting in a peak at a mass $> 10M_\odot$. In a cluster environment, most black holes at these low masses are likely ejected by supernova kicks and therefore do not contribute significantly to dynamical formation [63, 487, 81, 80]. Models of isolated binary evolution predict peak locations consistent with $10M_\odot$, albeit with considerable uncertainties due to effects of mass transfer, wind, and supernova kicks [180, 90, 253, 571, 89]. It therefore may be the case that isolated formation channels are responsible for the bulk of the BBH population.

The location of the $35M_\odot$ peak and the lack of a subsequent truncation is inconsistent with a population of isolated first-generation BBH mergers according to current understanding of the PPISN process. This peak and the presence of higher-mass black holes therefore needs a different explanation, possibly involving contributions from cluster or galactic nucleus environments [589, 371, 236, 516, 516, 249].

Our finding that black holes spins tend to be low agrees with models of efficient angular momentum transport in stellar progenitors, forming black holes with litte spin [242, 241]. It is worth noting, however, that there is possible support in the tails at higher spins, which may be explained by hierarchical mergers or tidal spin up [404, 436, 360, 364, 361]. With spin tilts showing a slight preference for alignment, it may be the case that isolated binaries formed via common evolution make up most of the BBH population [362, 468, 248]. However, there is evidence for anti- and non-aligned spins, which may be explained by dynamical formation channels [371, 515].

Our novel finding of the increasing merger rate with redshift may have important implications for understanding the delay times of binary systems between star formation and merger. As constraints get better, we may be able to use this as a tool for understanding astrophysical formation scenarios and cosmology.

Acknowledgements

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 also gratefully acknowledge the support of the Science and Technology Facilities Council (STFC) of the United Kingdom, the Max-Planck-Society (MPS), and the State of Niedersachsen/Germany for support of the construction of Advanced LIGO and construction and operation of the GEO600 detector. Additional support for Advanced LIGO was provided by the Australian Research Council. The authors gratefully acknowledge the Italian Istituto Nazionale di Fisica Nucleare (INFN), the French Centre National de la Recherche Scientifique (CNRS) and the Netherlands Organization for Scientific Research (NWO), for the construction and operation of the Virgo detector and the creation and support of the EGO consortium. The authors also gratefully acknowledge research support from these agencies as well as by the Council of Scientific and Industrial Research of India, the Department of Science and Technology, India, the Science & Engineering Research Board (SERB), India, the Ministry of Human Resource Development, India, the Spanish Agencia Estatal de Investigación (AEI), the Spanish Ministerio de Ciencia e Innovación and Ministerio de Universidades, the Conselleria de Fons Europeus, Universitat i Cultura and the Direcció General de Política Universitaria i Recerca del Govern de les Illes Balears, the Conselleria d'Innovació, Universitats, Ciència i Societat Digital de la Generalitat Valenciana and the CERCA Programme Generalitat de Catalunya,

Spain, the National Science Centre of Poland and the European Union – European Regional Development Fund; Foundation for Polish Science (FNP), the Swiss National Science Foundation (SNSF), the Russian Foundation for Basic Research, the Russian Science Foundation, the European Commission, the European Social Funds (ESF), the European Regional Development Funds (ERDF), the Royal Society, the Scottish Funding Council, the Scottish Universities Physics Alliance, the Hungarian Scientific Research Fund (OTKA), the French Lyon Institute of Origins (LIO), the Belgian Fonds de la Recherche Scientifique (FRS-FNRS), Actions de Recherche Concertées (ARC) and Fonds Wetenschappelijk Onderzoek – Vlaanderen (FWO), Belgium, the Paris Île-de-France Region, the National Research, Development and Innovation Office Hungary (NKFIH), the National Research Foundation of Korea, the Natural Science and Engineering Research Council Canada, Canadian Foundation for Innovation (CFI), the Brazilian Ministry of Science, Technology, and Innovations, the International Center for Theoretical Physics South American Institute for Fundamental Research (ICTP-SAIFR), the Research Grants Council of Hong Kong, the National Natural Science Foundation of China (NSFC), the Leverhulme Trust, the Research Corporation, the Ministry of Science and Technology (MOST), Taiwan, the United States Department of Energy, and the Kavli Foundation. The authors gratefully acknowledge the support of the NSF, STFC, INFN and CNRS for provision of computational resources. This work was supported by MEXT, JSPS Leading-edge Research Infrastructure Program, JSPS Grant-in-Aid for Specially Promoted Research 26000005, JSPS Grant-in-Aid for Scientific Research on Innovative Areas 2905: JP17H06358, JP17H06361 and JP17H06364, JSPS Core-to-Core Program A. Advanced Research Networks, JSPS Grant-in-Aid for Scientific Research (S) 17H06133 and 20H05639 , JSPS Grant-in-Aid for Transformative Research Areas (A) 20A203: JP20H05854, the joint research program of the Institute for Cosmic Ray Research, University of Tokyo, National Research Foundation (NRF) and Computing Infrastructure Project of KISTI-GSDC in Korea, Academia Sinica (AS), AS Grid Center (ASGC) and the Ministry of Science and Technology (MoST) in Taiwan under grants including AS-CDA-105-M06, Advanced Technology Center (ATC) of NAOJ, Mechanical Engineering Center of KEK. *We would like to thank all of the essential workers who put their health at risk during the COVID-19 pandemic, without whom we would not have been able to complete this work.*

	BNS	NSBH	BBH	NS-Gap	BBH-gap	Full
	$m_1 \in [1, 2.5]M_\odot$	$m_1 \in [2.5, 50]M_\odot$	$m_1 \in [2.5, 100]M_\odot$	$m_1 \in [2.5, 5]M_\odot$	$m_1 \in [2.5, 100]M_\odot$	$m_1 \in [1, 100]M_\odot$
	$m_2 \in [1, 2.5]M_\odot$	$m_2 \in [1, 2.5]M_\odot$	$m_2 \in [2.5, 100]M_\odot$	$m_2 \in [1, 2.5]M_\odot$	$m_2 \in [2.5, 5]M_\odot$	$m_2 \in [1, 100]M_\odot$
PDB (pair)	170^{+270}_{-120}	27^{+31}_{-17}	$25^{+10}_{-7.0}$	19^{+28}_{-13}	$9.3^{+15.7}_{-7.2}$	240^{+270}_{-140}
PDB (ind)	44^{+96}_{-34}	73^{+67}_{-37}	$22^{+8.0}_{-6.0}$	$12^{+18}_{-9.0}$	$9.7^{+11.3}_{-7.0}$	150^{+170}_{-71}
MS	660^{+1040}_{-530}	49^{+91}_{-38}	37^{+24}_{-13}	$3.7^{+35.3}_{-3.4}$	$0.12^{+24.88}_{-0.12}$	770^{+1030}_{-530}
BGP	$98.0^{+260.0}_{-85.0}$	$32.0^{+62.0}_{-24.0}$	$33.0^{+16.0}_{-10.0}$	$1.7^{+30.0}_{-1.7}$	$5.2^{+12.0}_{-4.1}$	$180.0^{+270.0}_{-110.0}$
MERGED	$10 - 1700$	$7.8 - 140$	$16 - 61$	$0.02 - 39$	$9.4 \times 10^{-5} - 25$	$72 - 1800$

Table 7.1: Merger rates in $\mathrm{Gpc}^{-3}\mathrm{yr}^{-1}$ in various mass bins using fits to the mass distribution across $1 - 100\ M_\odot$. The BNS, NSBH, BBH, and mass-gap categories are defined by the component masses. We defined neutron star masses to be anything below $2.5M_\odot$, and the mass-gap to be from $2.5 - 5M_\odot$. The models here are Power Law + Dip + Break (PDB), MultiSource (MS), and Binned Gaussian Process (BGP). Merger rates are quoted with 90% credible intervals. The final row is the union of the 90% intervals from the previous three rows. The PDB (pair) model includes a pairing function between the primary and secondary mass distributions, which is not done in the other models. All models assume a merger rate that is uniform in comoving volume of the source, independent of redshift. From [28].

	$m_1 \in [5, 20] M_\odot$ $m_2 \in [5, 20] M_\odot$	$m_1 \in [20, 50] M_\odot$ $m_2 \in [5, 50] M_\odot$	$m_1 \in [50, 100] M_\odot$ $m_2 \in [5, 100] M_\odot$	All BBH
PP	$23.6^{+13.7}_{-9.0}$	$4.5^{+1.7}_{-1.3}$	$0.2^{+0.1}_{-0.1}$	$28.3^{+13.9}_{-9.1}$
BGP	$20.0^{+11.0}_{-8.0}$	$6.3^{+3.0}_{-2.2}$	$0.75^{+1.1}_{-0.46}$	$33.0^{+16.0}_{-10.0}$
FM	$21.1^{+11.6}_{-7.8}$	$4.3^{+2.0}_{-1.4}$	$0.2^{+0.2}_{-0.1}$	$26.5^{+11.7}_{-8.6}$
PS	$27^{+12}_{-8.8}$	$3.5^{+1.5}_{-1.1}$	$0.19^{+0.16}_{-0.09}$	$31^{+13}_{-9.2}$
MERGED	13.3–39	2.5–6.3	0.099–0.4	17.9–44

Table 7.2: Merger rates in $\mathrm{Gpc}^{-3}\mathrm{yr}^{-1}$ in various mass bins using fits to the mass distribution across $5 - 100\ M_\odot$. The models here are Power Law + Peak (PP), Binned Gaussian Process (BGP), Flexible Mixture (FM), and Power Law + Spline (PS). Merger rates are quoted with 90% credible intervals. The final row is the union of the 90% intervals from the previous rows. BGP assumes a merger rate uniform in comoving volume of the source, independent of redshift. The rates from the other models are cited at $z = 0.2$, approximately where we have the best constraints on $R(z)$. From [28].

Model	GWTC-2 $\log_{10}\mathcal{B}$	GWTC-3 $\log_{10}\mathcal{B}$
Power Law + Peak	0.0	0.0
Broken Powerlaw + Peak	−0.11	−0.46
Multi Peak	−0.3	−0.22
Broken Powerlaw	−0.92	−2.0

Table 7.3: Bayes factors for each of the previously used phenomenological mass models relative to the model with highest marginal likelihood, Power Law + Peak. The previous results from GWTC-2 are shown in the second column with the updated catalog results in the third column. From [28].

Appendix 7.A Other Mass Distribution Models

In Table 7.3, we present the Bayes factors for previously used phenomenological mass models relative to the model with highest evidence, Power Law + Peak. The previous results from GWTC-2 are shown in the second column with the updated catalog results in the third column. We find that the Power Law + Peak model is still favored over all other models, but the differences in evidence are mostly modest.

Appendix 7.B Impact of Sensitivity on Redshift Evolution Inference

One change in the sensitivity estimation procedure between this work and our previous study of GWTC-2 [29] is the use of injections that account for the effect of precession and as well as updates to our detection pipelines as detailed in [26]. Since precession was not included in the injections used in [29], the full spin distribution could not be reweighted to calculate the sensitivity via Eq. 7.3, and thus, for the purposes of sensitivity estimation, an approximation was made that $S_{x,y} \in (-0.5, 0.5)$. Since we now use precessing injections, we do the reweighting procedure including the full spin distribution as a function of Λ. To test if this difference in our sensitivity estimation procedure is responsible for the change in the inferred redshift evolution, we repeat the population analysis reported as the main BBH analysis, using our updated sensitivity model, but only including events analyzed in the GWTC-2 populations study [29]. From this analysis, we infer $\kappa > 0$ at 97.6% credibility, as opposed to the 85% credibility reported previously [29], indicating a much stronger preference for a merger rate increasing with redshift. We conclude that the differences between our current results for the evolution of the BBH merger rate and those reported previously [29] are due to improvements to our sensitivity model rather than the presence of the additional events in GWTC-3.

In Figure 7.6 we compare the redshift dependence of our current sensitivity model to that of the sensitivity model used previously [29]. To make this comparison, we reweight the injections used previously [29] to the same spin distribution assumed in that study, and assuming a fiducial Power Law + Peak and Power Law model for the mass and redshift distributions, respectively. We reweight the current injections to this same mass and redshift distribution, but reweight them to the median inferred spin distribution obtained previously [29], to mimic a astrophysically-realistic population. Both injection sets only cover the observing times of the O3a observing run. Taking the ratio of the corresponding sensitvities, we find our sensitivity has increased for low redshift events and decreased for high redshift events, relative to the sensitivity model used in our prior work [29]. We expect to see an increase in sensitivity between our previous analysis [29] and our current calculation due to updates to the detection pipelines. The relative decrease in sensitivity at higher redshifts indicates a bias in the previous sensitivity estimate, implying that the BBH merger rate at high redshift was underestimated in our earlier study [29]. Accounting for the shift in sensitivity as a function of redshift causes a relative decrease in local BBH merger rate and a relative increase in high-redshift BBH merger rate, leading to a higher inferred value for κ.

One possible explanation for the shift in sensitivity is that the use of precession in the injections for sensitivity estimation caused a non-trivial change in the inferred sensitive hypervolume, given that we do observe precession in the BBH population. Our current detection pipelines use template banks that include only aligned-spin components; this can result in up to tens of percent reduced sensitivity to a population of BBHs with spin precession, depending on the degree of precession possible [50, 48, 271]. The farthest precessing sources, which, due to their distances, correspond to FARs closest to the detection threshold, are therefore the most susceptible to dropping below the detection threshold with our current pipelines, causing us to see a decrease in sensitivity to a population of BBHs with precession relative to a strictly non-precessing population.

Additionally, both the use of population-informed reweighting of the spin distribution to calculate sensitivity to a population and the incorporation of additional detection pipelines may have contributed to a more accurate estimate of our sensitivity across parameter space.

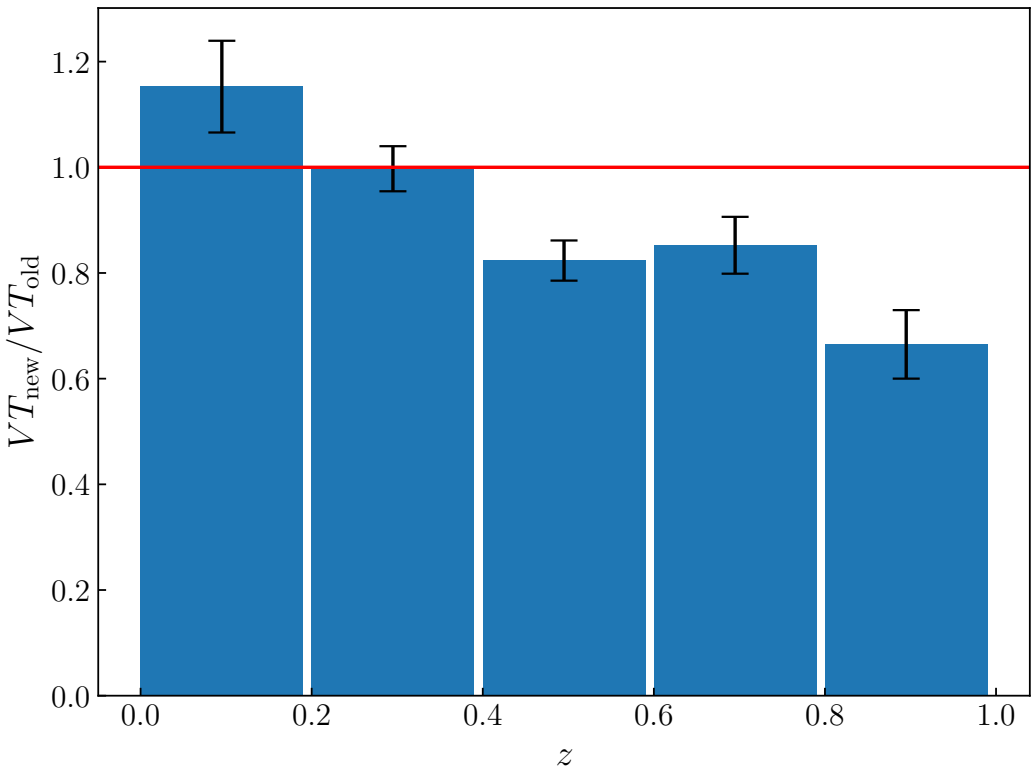

Figure 7.6: Comparison of our current BBH merger sensitivity estimate in the O3a observing run (VT_{new}) to that used in [29] (VT_{old}) as a function of redshift, for events with chirp masses between $20M_{\odot}$ and $50M_{\odot}$. Our current sensitivity model differs from what was used in [29] in two important ways: we use updated detection pipelines relative to those used in our GWTC-2 [29] and we use injections which include spin precession. There is a relative increase (decrease) in sensitivity at low (high) redshift. Computed by reweighting injections to a fiducial population for each of the two injection sets.

Chapter 8

BINARY BLACK HOLE POPULATION THROUGH GWTC-4

This chapter is intentionally redacted and will be embargoed until August 2025.

Chapter 9

GWPOPULATION: HARDWARE AGNOSTIC POPULATION INFERENCE FOR COMPACT BINARIES AND BEYOND

Colm Talbot, Amanda Farah, Shanika Galaudage, et al. "GWPopulation: Hardware agnostic population inference for compact binaries and beyond". In: *J. Open Source Softw.* (Sept. 2024). DOI: 10.21105/joss.07753. arXiv: 2409.14143 [astro-ph.IM].

9.1 Introduction

Since the first direct detection of gravitational waves by the LIGO–Virgo collaboration in 2015 [14], the size of the gravitational-wave transient catalog has grown to nearly 100 events [26], with more than as many observed during the ongoing fourth observing run. Extracting astrophysical/cosmological information from these observations is a hierarchical Bayesian inference problem. GWPopulation is designed to provide simple-to-use, robust, and extensible tools for hierarchical inference in gravitational-wave astronomy/cosmology. It has been widely adopted for gravitational-wave astronomy, including producing flagship results for the LIGO-Virgo-KAGRA collaborations (e.g., [28, 2]).[1] While designed to work with observations of compact binary coalescences, GWPopulation may be available to a wider range of hierarchical Bayesian inference problems.

Building on Bilby [67, 470], GWPopulation can easily be used with a range of stochastic samplers through a standard interface. By providing access to a range of array backends (numpy [270], JAX [111], and cupy [403] are currently supported) GWPopulation is hardware agnostic and can leverage hardware acceleration to meet the growing computational needs of these analyses. Included in the package are:

- Implementations of the most commonly used likelihood functions in the field.
- Commonly used models for describing the astrophysical population of merging compact binaries, including the "PowerLaw+Peak" and "PowerLaw+Spline"

[1] For a full listing of papers using GWPopulation, see the citations for the previous publication at https://ui.adsabs.harvard.edu/abs/2019PhRvD.100d3030T/citations.

mass models, “Default” spin model, and “PowerLaw” redshift models used in the latest LIGO-Virgo-KAGRA collaboration analysis of the compact binary population.[2]

- Functionality to simultaneously infer the astrophysical distribution of sources and cosmic expansion history using the “spectral siren” method [203].
- A standard specification allowing users to define additional models.

9.2 Statement of Need

Hierarchical Bayesian inference is the standard method for inferring parameters describing the astrophysical population of compact binaries and the cosmic expansion history (e.g., [536, 560]. The first step in the hierarchical inference process is drawing samples from the posterior distributions for the source parameters of each event under a fiducial prior distribution along with a set of simulated signals used to quantify the sensitivity of gravitational-wave searches. Next, these samples are used to estimate the population likelihood using Monte Carlo integration with a computational cost that grows quadratically with the size of the observed population. Since evaluating these Monte Carlo integrals is embarrassingly parallel, this is a prime candidate for hardware acceleration using graphics/tensor processing units. GWPopulation provides functionality needed to perform this second step and is extensively used by members of the gravitational-wave astronomy community including the LIGO-Virgo-KAGRA collaborations.

Maximizing the information we can extract from the gravitational-wave transient catalog requires a framework where potential population models can be quickly constrained with the observed data with minimal boilerplate code. Additionally, the availability of a standard open-source implementation improves the reliability and reproducibility of published results. GWPopulation addresses all of these points by providing a standard, open-source, implementation of the functionality needed to perform population analyses while enabling user-defined models to be provided by a Python function/class definition. The flexible backend system means hardware acceleration can be used with minimal coding effort. Using GWPopulation on Google Colab, it is possible to perform an exploratory analysis with a new population model in minutes and produce production-quality results without needing high-performance/throughput computing clusters. With access to high throughput

[2] See [28] for details of these models.

computing resources, a wide range of potential models can be easily explored using the associated GWPOPULATION_PIPE [517] package.

9.3 Related Packages

Several other packages are actively used and maintained in the community that can be used for population inference that operate in complementary ways to GWPOPULATION.

- GWINFERNO [187] is a package for hierarchical inference with gravitational-wave sources intended for use with NUMPYRO [424] targeting high-dimensional models. GWINFERNO includes many population models initially adapted from GWPOPULATION.
- There are a wide range of packages designed for joint astrophysical and cosmological inference with gravitational-wave transients including ICAROGW [370], GWCOSMO [260], MGCOSMOPOP [354], and CHIMERA [109]. ICAROGW supports some harware acceleration using CUPY but some cosmological calculations are limited to CPU support only. CHIMERA is JAX-compatible and supports flat Lambda-CDM cosmologies along with analysis using galaxy catalogs.
- VAMANA [539] models the compact binary distribution as a mixture of Gaussians and power-law distributions, POPMODELS [583] implements a range of parametric models for the compact binary distribution and supports sampling via EMCEE [237], neither supports hardware acceleration at the time of writing.

Acknowledgements

CT is supported by an Eric and Wendy Schmidt AI in Science Fellowship, a Schmidt Sciences program. SG is supported by the ANR COSMERGE project, grant ANR-20-CE31-001 of the French Agence Nationale de la Recherche. AF is supported by the NSF Research Traineeship program under grant No. DGE1735359, and by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1746045. HT is supported by Australian Research Council (ARC) Centre of Excellence CE230100016.

Chapter 10

SEARCHING FOR STRUCTURE IN THE BINARY BLACK HOLE SPIN DISTRIBUTION

Jacob Golomb and Colm Talbot. "Searching for structure in the binary black hole spin distribution". In: *Phys. Rev. D* 108.10 (2023), p. 103009. DOI: 10.1103/PhysRevD.108.103009. arXiv: 2210.12287 [astro-ph.HE].

Abstract

The spins of black holes in merging binaries can reveal information related to the formation and evolution of these systems. Combining events to infer the astrophysical distribution of black hole spins allows us to determine the relative contribution from different formation scenarios to the population. Many previous works have modelled spin population distributions using low-dimensional models with statistical or astrophysical motivations. While these are valuable approaches when the observed population is small, they make strong assumptions about the shape of the underlying distribution and are highly susceptible to biases due to mismodeling. The results obtained with such parametric models are only valid if the allowed shape of the distribution is well-motivated (i.e. for astrophysical reasons). Unless the allowed shape of the distribution is well-motivated (i.e., for astrophysical reasons), results obtained with such models thus may exhibit systematic biases with respect to the true underlying astrophysical distribution, along with resulting uncertainties not being reflective of our true uncertainty in the astrophysical distribution. In this work, we relax these prior assumptions and model the spin distributions using a more data-driven approach, modelling these distributions with flexible cubic spline interpolants in order to allow for capturing structures that the parametric models cannot. We find that adding this flexibility to the model substantially increases the uncertainty in the inferred distributions, but find a general trend for lower support at high spin magnitude and a spin tilt distribution consistent with isotropic orientations. We infer that 62-87% of black holes have spin magnitudes less than $a = 0.5$, and 27-50% (90% credible levels)

of black holes exhibit negative $\chi_{\rm eff}$. Using the inferred $\chi_{\rm eff}$ distribution, we place a conservative upper limit of 37% for the contribution of hierarchical mergers to the astrophysical BBH population. Additionally, we find that artifacts from unconverged Monte Carlo integrals in the likelihood can manifest as spurious peaks and structures in inferred distributions, mandating the use of a sufficient number of samples when using Monte Carlo integration for population inference.

10.1 Introduction

Gravitational waves offer a unique probe into the properties of merging black holes (BHs) and neutron stars. Since the first such detection in 2015, the LIGO-Virgo network [16, 1, 38] has reported the detection of $\sim$ 90 binary black hole (BBH) mergers, with each gravitational-wave (GW) signal encoding physical information about the BHs involved, such as their masses and angular momenta (spins) [26, 12, 12]. Extracting this information has has enabled the study of properties of BBH systems on both an individual and population-level basis. From an astrophysical perspective, combining GW detections to infer the mass, spin, and redshift distributions of BBH systems can help answer questions ranging from binary formation and stellar evolution [28, 29] to the expansion rate of the Universe and possible deviations from general relativity [21, 34].

The spin of the BHs in a BBH system offer insight into the history of the binary. For example, BH spins can help reveal whether the BHs in a BBH system formed directly from core collapse of a heavy star or from the previous merger of two lighter BHs [308, 230, 249, 227]. Although the processes governing the angular momentum transport out of a stellar core during collapse are not well-constrained, recent modeling work indicates that BHs resulting directly from core collapse supernovae should have negligible spin magnitudes [241, 437, 273]. While processes such as tidal interactions and mass transfer can induce higher spins on BHs in binary systems, it is uncertain how appreciable the resulting spin-ups can be [592, 404, 86, 85]. On the other hand, BHs formed from the merger of two non-spinning BHs are expected to form a final BH with a relatively high spin magnitude [279, 285, 227], motivating the possibility to use spin magnitude as a tracer of a BHs formation history.

The direction of the BH spin vectors also encode information related to the formation history of a BBH system. Models suggest that BBH systems formed from common evolution, in which the component BHs evolve together from a stellar binary in an isolated environment free from significant dynamical interactions, should have component spin vectors nearly aligned with the orbital angular momentum axis, with any tilt being efficiently brought into alignment by tidal interactions [218, 299]. On the other hand, BBH systems formed from dynamical encounters are not expected to have any correlated spins, such that the BH spin vectors are isotropic with respect to the orbital angular momentum [230, 362, 468].

While only a couple of events in the third gravitational-wave transient catalog

individually feature confidently high spin magnitudes or anti-alignment (i.e. a spin vector pointing opposite the angular momentum), hierarchically combining observations of GW events while folding in selection effects can reveal the degree to which these parts of spin parameter space contribute to the astrophysical distribution of BH spins. Previous work has used these inferred contributions to estimate the fraction of BBH systems in the local Universe which may have been formed hierarchically, dynamically, and by isolated evolution [179, 230, 308, 28]. However, recent publications have disagreeing estimates for the contributions of anti-aligned and non-spinning BBHs to the astrophysical population.

In [28, 29], the authors conclude that the BBH distribution must feature anti-aligned spins at $> 90\%$ credibility, in contrast to the conclusion drawn in [474] that such anti-alignment is not evident in the population. In addition, [243] finds evidence for a non-spinning subpopulation of BHs, a conclusion which was challenged by [127]. While technical differences exist between works, a major possible contribution to some of these differing conclusions is model misspecification (see, e.g. [420, 472]); that is, different assumptions being imposed on the functional form of the spin distribution.

The `Default` model in [28, 29] models the distribution of the magnitude of the BH spin vector and the tilt angle between the spin vector and the orbital angular momentum. They adopt a Beta distribution for the spin magnitude model [582, 28],

$$\pi(a_{1,2}|\alpha_\chi, \beta_\chi) = \text{Beta}(a_{1,2}|\alpha_\chi, \beta_\chi), \tag{10.1}$$

where a_1 (a_2) is the magnitude of the spin vector of the primary (secondary) BH, and α_χ and β_χ are population hyperparameters determining the structure of the Beta distribution. The model for the distribution of tilt angles, θ, is motivated by two subpopulations: one preferentially aligned ($\cos(\theta) \approx 1$) and one isotropic [521, 29, 28]. The model is parameterized as:

$$\pi(\cos\theta_{1,2}|\xi, \sigma_t) = \xi G_t(\cos\theta_1|\sigma_t) G_t(\cos\theta_2|\sigma_t) + \frac{1-\xi}{4}, \tag{10.2}$$

where G_t is a truncated Gaussian centered at $\cos\theta = 1$ with standard deviation σ_t and bounded in $[-1, 1]$, and ξ is the relative mixing fraction between the subpopulations. The second term corresponds to the contribution from the uniform (isotropic) distribution.

This population model has been extended in other work to allow for other astrophysically-motivated features to help draw conclusions related to the different formation scenarios present in the astrophysical distribution [243, 474, 127, 557]. Adopting an

astrophysically-motivated, strongly parametric model necessarily limits the possible features resolvable in the inferred distribution to what the chosen function can model. Accordingly, in this work, we consider a strongly parametric model to be one that has a specific, prior-determined shape as provided by the parameterization (e.g. a normal distribution), which is then constrained by the data. When using such a distribution to draw astrophysical conclusions from the inferred population, this is a reasonable and intended consequence, as the model is chosen to encode prior beliefs on the parameters that should govern the astrophysical distribution; however, if additional features exist in the true astrophysical distribution and a strongly parametric model cannot account for them, such features can be missed and a biased result may be obtained.

Previous work has shown that substructures in the BH mass distribution can be captured by cubic splines acting as a perturbation on top of a simpler parameteric model [186, 28]. In [186], the authors consider an exponentiated spline perturbation modulating an underlying power law in the mass distribution. In this work, we model the spin magnitude and tilt distributions using exponentiated cubic splines modulating a flat distribution to obtain a more data-driven result for the inferred population of BH spins. In doing so, we limit the potential bias caused by mismodeling the spin distribution and allow for the possibility of capturing features not accessible with a strongly parametric model.

The remainder of the paper is organized as follows. In Section 10.2, we detail the functional form and implementation of the cubic spline model. We provide the background of hierarchical Bayesian inference in Section 10.3, as it applies to population inference with GW sources. In Section 10.4 we present the resulting spin distributions we obtain for various spline models adopted in this work. Finally, we use these results to draw conclusions related to the astrophysical distribution of BBH spins and provide a relevant discussion in Section 10.5. We additionally supply three appendices; the first provides additional details about an efficient caching technique for the cubic spline model, the second explores the effect of uncertainty in our estimation of the selection function, and the third describes robustness of our results to different choices of prior distribution.

10.2 Models

Following the model for the black hole mass distribution considered in [186], we fit the distribution of spin magnitudes and cosine tilts using exponentiated cubic

splines

$$p(x) \propto e^{f(x)}. \tag{10.3}$$

A spline is a piecewise polynomial function defined by a set of node positions, the value of the function at those nodes, and boundary conditions at the end nodes. We use a cubic spline as it is the lowest order spline that enforces continuity of the function and its first derivative everywhere.

10.2.1 Node positions and amplitudes

In this work we consider models with 4, 6, 8, and 10 nodes spaced linearly in the domain of the parameter space. For the two distributions being modelled with splines, this gives 16 unique spin models (4 amplitude node placement models $\times$ 4 tilt node placement models). Our choice for the prior on the amplitude of each node is a unit Gaussian distribution. Comparisons with other node amplitude prior choices are detailed in Appendix 10.C.

In order to fully characterize a cubic spline, the first and second derivatives must be determined at each node. For all but the endpoints, these derivatives are specified by requiring continuity in the spline and its derivative. At the endpoints, there is no unique way to determine this and a range of boundary conditions are commonly used. For our implementation, we want the prior distribution of the derivatives at the endpoints to match that of the internal nodes. This requires providing two additional free parameters at each end of the spline. In practice, we add two additional nodes outside each boundary, with amplitudes that are free to vary according to the prior. Throughout this work, the number of nodes in a model refers to the number of nodes within the domain (i.e. not including these outside nodes). The spacing between these nodes is the same as that between nodes within the domain.

10.2.2 Modeling spins with splines

In this work, we use the spline model detailed above to model the population of spin magnitudes a and tilt angles $\cos\theta$. Consistent with [29, 28], we model these parameters as independent and identically distributed. The total spin population model is

$$\pi_{\rm spin}(\eta|\Lambda_s) = p_a(a_1)p_a(a_2)p_t(\cos\theta_1)p_t(\cos\theta_2), \tag{10.4}$$

where Λ_s is the set of population hyperparameters controlling the spline node location and amplitudes, and η is the set of single-event parameters. The functions

p are determined from Eq. 10.3. The domain of the spin magnitude distribution extends over $a \in [0, 1]$ and that of the spin tilt distribution covers $\cos\theta \in [-1, 1]$.

10.3 Methods

10.3.1 Hierarchical Bayesian Inference

In order to constrain the spin magnitude and tilt distribution, we carry out hierarchical Bayesian inference in which we calculate the likelihood of the entire observed dataset given a set of population hyperparameters Λ while marginalizing over the uncertainty in the physical parameters of each event. After analytically marginalizing over the total merger rate R with a prior $\pi(R) \propto R^{-1}$, we express the likelihood of the hyperparameters Λ parameterizing the population is expressed as (e.g. [536]):

$$\mathcal{L}(\{d\}|\Lambda) \propto p_{\rm det}(\Lambda)^{-N} \prod_i^N \int \mathcal{L}(d_i|\eta_i)\pi(\eta_i|\Lambda)d\eta_i. \tag{10.5}$$

Here, $\mathcal{L}(d_i|\eta_i)$ is the likelihood of observing the data d from the ith event, given physical (i.e. single-event) parameters η_i. In this work, η_i consists of masses, spins, and redshift of the ith event. The quantity $p_{\rm det}(\Lambda)$ encodes the sensitivity of the search algorithm that identified the signals and is described in more detail in Sec. 10.3.2.

Our population model $\pi(\eta|\Lambda)$ describes the astrophysical distribution of masses, redshifts, and spins. We model the primary mass distribution with the Powerlaw + Peak model [524], the mass ratio ($q = \frac{m_2}{m_1}$) distribution with a power law, and the redshift distribution also with a power law, with source-frame comoving merger rate density $R(z) \propto (1 + z)^3$ [228, 28, 29]. We choose to fix the redshift distribution because we use our own injection set to estimate sensitivity, thresholding on Signal-to-Noise Ratio (SNR) rather than False Alarm Rate (FAR) to determine "found" injections. Since this makes the threshold used to select real events (FAR $<$ 1 yr) slightly different from that used to threshold sensitivity injections, and the redshift distribution is particularly sensitive to the near-threshold events, we fix the redshift distribution in order to avoid biases (see [7] for an example of where a similar approximation was used). See Section 10.3.2 for details on sensitivity injections. We list the hyperparameters Λ and their corresponding priors in Table 10.1.

An initial choice that must be made when computing Eq. 10.5 is which events to include in the analysis. Typically this is done by establishing some detection threshold on the SNR or FAR and including all events that pass this threshold. We

Parameter	Description	Prior
α	Power Law index for m_1	(-4, 12)
β	Power Law index for q	(-2, 7)
$M_{\rm max}$	Maximum mass	(60, 100)
$M_{\rm min}$	Minimum mass	(2, 7)
λ	Fraction of sources in Gaussian peak	(0, 1)
M_{pp}	Location of Gaussian peak	(20, 50)
σ_{pp}	Standard deviation of Gaussian peak	(1, 10)
δ_m	Minimum mass turn-on length	(0, 10)

Table 10.1: Priors for mass distribution used in hierarchical inference, consistent with those used in [28]. Priors on the spin distribution are described in Section 10.2. Priors are uniform between the bounds listed in the third column.

choose to include the 59 events in the third observing run (O3) of the LIGO-Virgo network which have a False Alarm Rate of less than 1 year^{-1} and are included in the main BBH analysis of [28]. We limit ourselves to events in O3 for self-consistency, as the injections we perform to evaluate selection effects (see Section 10.3.2) use O3a and O3b detector sensitivities.

We compute Eq. 10.5 using the package `GWPopulation` [520], which constructs a Monte Carlo approximation of this integral by reweighting samples from the single-event posteriors into the population model. We use the nested sampling package `Dynesty` [507] to obtain hyperparameter samples from the posterior distribution.

10.3.2 Selection Effects

Since the sensitivity to an event is determined by the single-event parameters η, the observed population is biased toward events produced by the astrophysical population that are preferentially observable. To account for the bias arising from selection effects, we must compute the fraction of signals ($p_{\rm det}$) that will have a detection statistic ρ passing our detection threshold $\rho_{\rm th}$ by marginalizing over all possible signals and noise realizations n (e.g., [560, 6])

$$p_{\rm det}(\Lambda) = \int dn \int d\eta p(\eta|\Lambda) p(n) \Theta(\rho - \rho_{\rm th}). \tag{10.6}$$

Here Θ is the Heaviside step function. In practice, we use Monte Carlo importance sampling with respect to some simulated fiducial reference population Λ_0 to estimate Eq. 10.6. This method relies on injecting $N_{\rm inj}$ sources from this reference population into detector noise and determining which of these sources pass our detection

threshold [213]. This is computed as

$$p_{\rm det}(\Lambda) = \frac{1}{N_{\rm inj}} \sum_{\eta \sim \eta_{\rm found}} \frac{\pi(\eta|\Lambda)}{\pi(\eta|\Lambda_0)}, \tag{10.7}$$

where $\eta_{\rm found}$ corresponds to the single-event parameters of the events from the injection set that pass detection threshold.

For our sensitivity injection set, we simulate $O(5 \times 10^7)$ sources and inject them into Gaussian noise corresponding to O3 detector sensitivity specified by the representative Power Spectral Densities in [26, 24]. This results in $\sim 900,000$ injections passing our detection threshold of network optimal SNR greater than 10, where the square of the network optimal SNR is defined as the quadrature sum of the SNRs in each detector. We choose this threshold to be a surrogate for the 1 year^{-1} FAR threshold used for event selection. While this is not an exact mapping between the two detection statistics, the effects of spins on sensitivity is subdominant, so we expect that this approximation will not cause biases.

10.3.3 Uncertainties in the Likelihood

Since we approximate Eq. 10.5 and Eq. 10.7 using Monte Carlo summation, there exists a resulting statistical uncertainty associated with the use of finite samples to obtain estimates for the value of the log likelihood [213, 256, 369, 193]. For each sample of Λ, we compute this associated uncertainty in the log likelihood. Considering the computed approximation of $\ln \mathcal{L}(\{d\}|\Lambda)$ to be a realization from a distribution that asymptotically tends to a Gaussian distribution we approximate the covariance of this estimate using Equation 12 in [519]. For uncertainty propagation, we compute the variance associated with each $\ln \mathcal{L}(d|\Lambda)$ from Equation 7 in [519].

While we do not enforce any threshold directly on the variance $(\Delta \ln \mathcal{L})^2$, we retain this information for all points in the hyperposterior to investigate correlations between features in the population and uncertainty in the log likelihood (see Appendix 10.B).

10.3.4 Uncertainty in the Evidence

The evidence, or marginal likelihood, associated with a particular model is expressed simply as the expectation value of the likelihood conditioned on the population prior:

$$\mathcal{Z} = \int d\Lambda \mathcal{L}(\{d\}|\Lambda)\pi(\Lambda). \tag{10.8}$$

Comparing this quantity for two different models allows one to compute a Bayes factor, which is commonly used as a discriminator between models based on their relative strength at describing the observed data.

Because `Dynesty` computes the evidence by iteratively summing over a finite number of weights, there exists a statistical uncertainty associated with the estimated evidence. `Dynesty` reports this uncertainty along with the computed evidence.

Since there is also an uncertainty in the quantity $\ln \mathcal{L}(\{d\}|\Lambda)$ used in computing the evidence, and the evidence is the average of a set of $\ln \mathcal{L}$ values, we take the contribution of this uncertainty to the total evidence to be the average uncertainty in $\ln \mathcal{L}(\{d\}|\Lambda)$ over the draws from $\pi(\Lambda)$.

We take these two sources of uncertainty in the evidence to be independent and compute the total uncertainty in $\mathcal{Z}$ by propagation of errors. As a result, we obtain both the evidence and its uncertainty for each model of a fixed set of spline nodes. We note that all of the evidences for the spline models are consistent within their 1σ uncertainties, and the `Default` model has a natural log Bayes Factor of ≈ -1.5 with respect to the overlapping region of uncertainties in the evidences for the spline models.

10.4 Results

In this section, we present the results from analyzing the population of BBH spin magnitude and tilts using spline models. We use the standard Gaussian prior on node amplitudes as described in Section 10.2.1, with nodes placed linearly within the domain of parameter space. In Figure 10.1, we show the evidences and their uncertainties for the 16 node combinations we consider in this work. All models give similar evidences, with no significant preferences considering their associated uncertainties. The red shaded region shows where all of the evidence estimates overlap within 1σ. This indicates that adding more nodes does not tend to provide a better fit to the distribution and also does not over-fit it. We therefore cite the numbers in this section using the most flexible model, with 10 nodes for both the magnitude and tilt distributions. Unless otherwise noted, the plots of the spin magnitude (tilt) distributions assume 10 nodes in the tilt (magnitude) distribution.

As a general trend, we notice that the inferred 90% region of parameter space exhibit oscillating peaks at the location of the spline nodes. As shown in Figure 10.13, these oscillations appear for uninformative data. With the observations of BH spins being

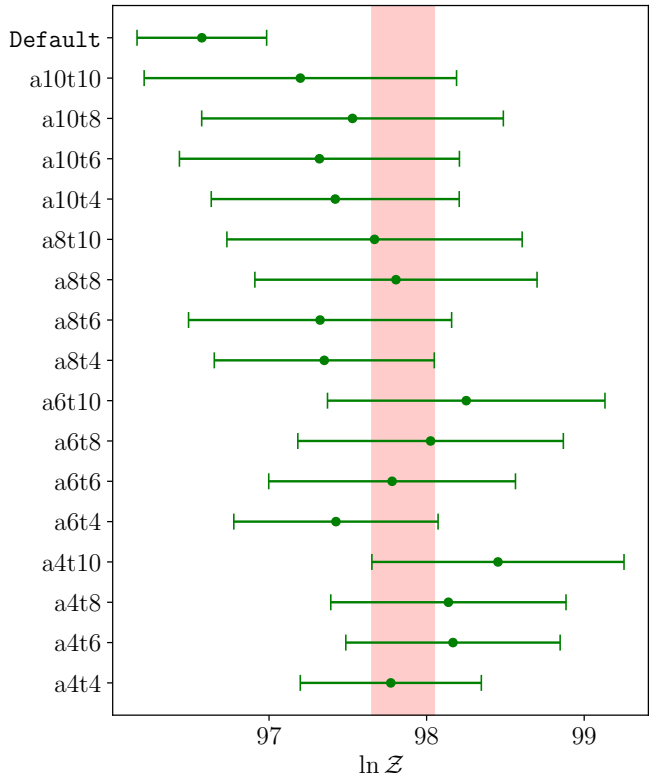

Figure 10.1: Comparison of evidences obtained from the different spline node combinations considered, as well as from the `Default` model. Uncertainties are computed by adding the average covariance in the log likelihood (see Eq. 12 in [519]) in quadrature with the uncertainty in the evidence as reported from `Dynesty`. The numbers after "a" and "t" are the number of nodes in the magnitude and tilt models, respectively. All evidences from the spline models are consistent with $\ln \mathcal{Z}$ in the red shaded region at 1σ.

weakly informative, we see this effect from the strong influence of the prior on the posterior distribution of the spline nodes.

10.4.1 The distribution of spin magnitudes

In Figure 10.2 we show the inferred distribution of spin magnitudes for our four choices of node numbers in spin magnitude, assuming 10 nodes in spin tilt. Although each model involves different positions and numbers of spline nodes, we note that the uncertainties (solid lines) and the average line (dot-dashed lines) in the distribution are comparable between models.

The 90% credible interval of the distribution is relatively broad, making it difficult to discern obvious trends in spin magnitude. However, we note the general pattern

of a preference for smaller spin magnitudes in the population and less support for higher spin magnitudes. Considering the model with 10 magnitude and 10 tilt nodes, we infer that $77.1\%^{+10.4\%}_{-14.8\%}$ of spin magnitudes are below $a = 0.5$, and 50% of spin magnitudes are below $a = 0.25^{+0.16}_{-0.10}$ (all uncertainties in this work are reported at the 90% symmetric credible levels unless otherwise stated).

While the models using fewer spline nodes tend to place increased support around $a = 0.2$, the significance is substantially reduced as we add more spline nodes. When more spline nodes are added, the model becomes more flexible and more data are necessary to constrain the distribution. While this feature may be real, it is not confident enough to remain present as the flexibility of the model increases.

Comparing to `Default` model (the green shaded region) used in [28], we observe substantially more uncertainty in the inferred spin magnitude distributions using our spline models.

As a point of comparison, in Appendix 10.C we show the distribution of spin magnitudes for different numbers of nodes, this time assuming 4 nodes in the tilt distribution. We find no significant differences from the distributions assuming 10 nodes in the tilt distribution which indicates that the number of nodes in the tilt distribution has a negligible effect on the inferred spin magnitude distribution.

10.4.2 The distribution of spin tilts

In Figure 10.3 we show the inferred distribution of spin tilts. Similar to the case with spin magnitudes, the uncertainties in the distribution are wide, but the average distributions for the different node combinations agree. In general, the distribution is consistent with being flat and featureless, but there is a slight trend for an increase in support for $-0.25 < \cos\theta < 0.75$.

As demonstrated by comparing Figure 10.12 with 10.3, the inferred distribution of spin tilts is very similar when we model the spin magnitude distribution with 4 nodes. We confirm this invariance for all sets of magnitude nodes tested, suggesting that the number of spin magnitude nodes does not meaningfully affect the recovered spin tilt distribution.

We infer that $38.6\%^{+17.3\%}_{-15.6\%}$ of spin tilts are below $\cos\theta = 0$, and 50% of spin tilts are below $\cos\theta = 0.15^{+0.22}_{-0.22}$. Notably there is no trend for an increase in support for $\cos\theta = 1$ as would be predicted by a preferentially aligned-spin population (see Section 10.5).

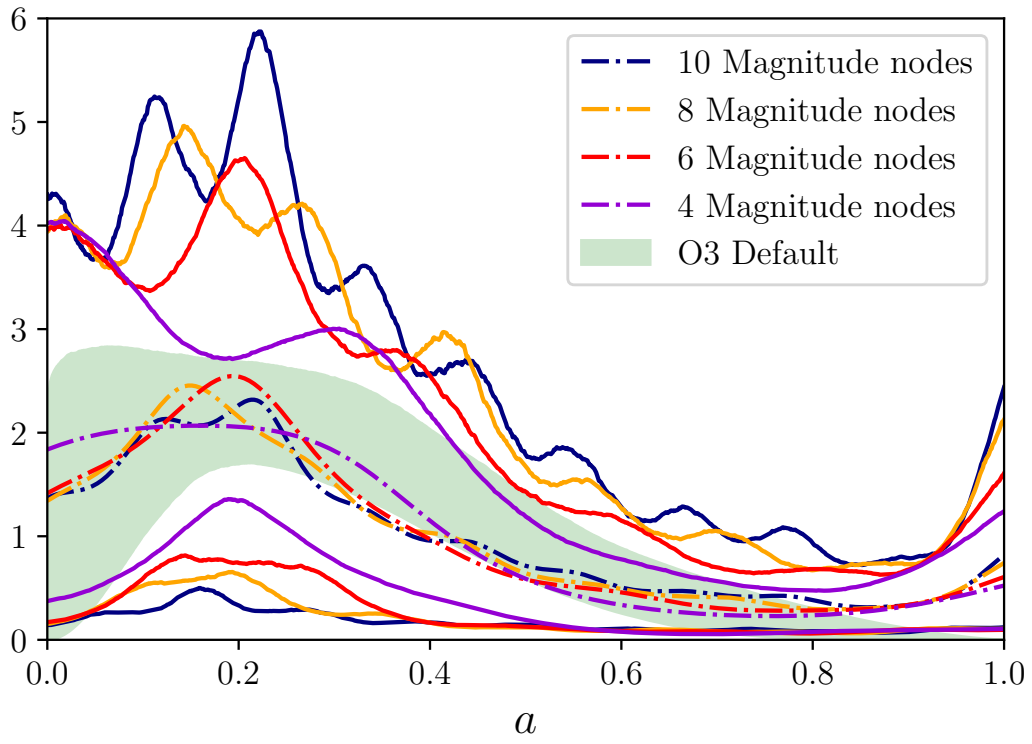

Figure 10.2: Distribution of spin magnitudes, with different numbers of nodes corresponding to different colors. All use 10 nodes in the tilt distribution.

10.4.3 The distribution of χ_{eff}

As an alternative to modeling the component spins, it is common to consider instead the total spin contribution aligned with the orbital angular momentum, the so-called "effective" aligned spin parameter. This term is parameterized as

$$\chi_{\text{eff}} = \frac{a_1 \cos\theta_1 + q a_2 \cos\theta_2}{1+q}. \tag{10.9}$$

While we do not directly model the distribution of χ_{eff} in this work, we can use the inferred distributions of a, $\cos(\theta)$, and q to reconstruct a distribution for χ_{eff} (c.f. [29, 243]).

Figure 10.4 shows this inferred distribution of χ_{eff} as we vary the number of tilt and magnitude nodes. As a point of comparison, we show the corresponding reconstruction of χ_{eff} when the distributions of component spin magnitudes and tilts are inferred using the `Default` model with same catalog of events. In addition, we also plot the χ_{eff} distribution recovered from uniformly sampling in a and $\cos(\theta)$ (solid black curve), assuming mass ratios drawn from a power law with an index of 2

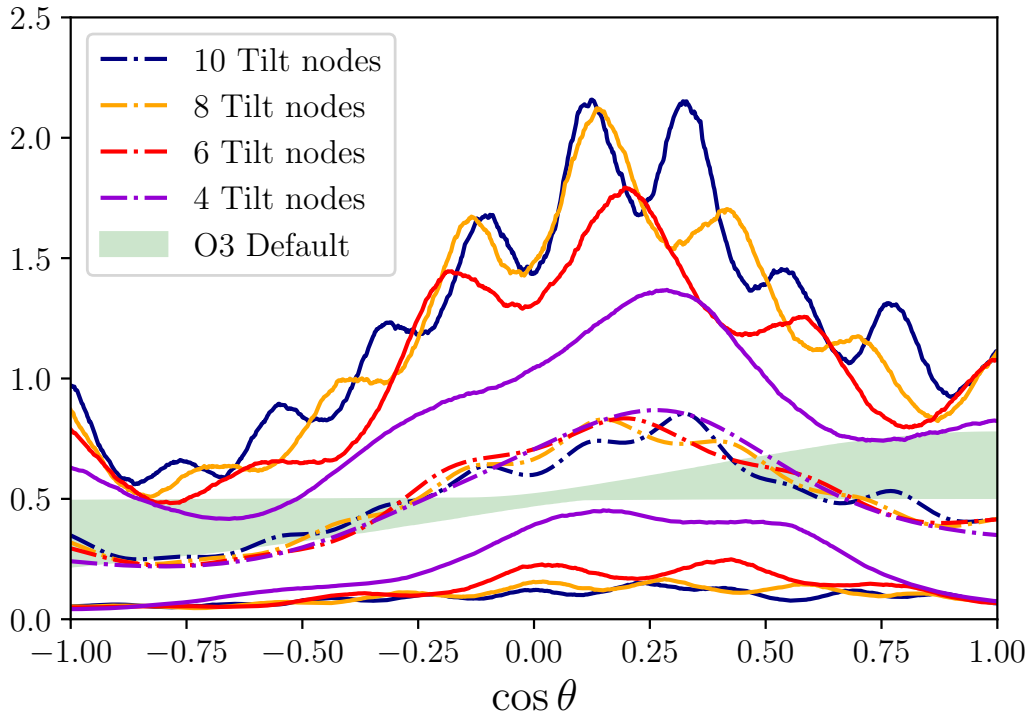

Figure 10.3: Distribution of spin tilts, with different numbers of nodes corresponding to different colors. All use 10 nodes in the spin magnitude distribution.

(q consistent with the results in [28]) . The $\chi_{\rm eff}$ distribution inferred from the spline model agrees well with the `Default` reconstruction, but is a noticeably narrower distribution than a uniform spin magnitude and tilt distribution would result in. Using the model with ten magnitude and tilt nodes each, we infer that $38.7\%^{+12.8\%}_{-11.5\%}$ of BBH systems have $\chi_{\rm eff} < 0$.

10.4.4 The distribution of χ_p

Another "effective" spin parameter commonly modeled in the gravitational wave literature is the effective precessing spin parameter, χ_p, which quantifies the amount of in-plane spin present in a BBH merger [490]. Here,

$$\chi_p = \max\left[a_1 \sin\theta_1, \left(\frac{3+4q}{4+3q}\right) q a_2 \sin\theta_2\right] . \tag{10.10}$$

Similar to the previous subsection, we can reconstruct the distribution of χ_p using the spline models of a, $\cos(\theta)$, as well as our inference on the population of q.

Figure 10.5 shows the inferred distribution of χ_p as a function of different magnitude and tilt nodes, respectively. We also show the inferred distribution recovered from

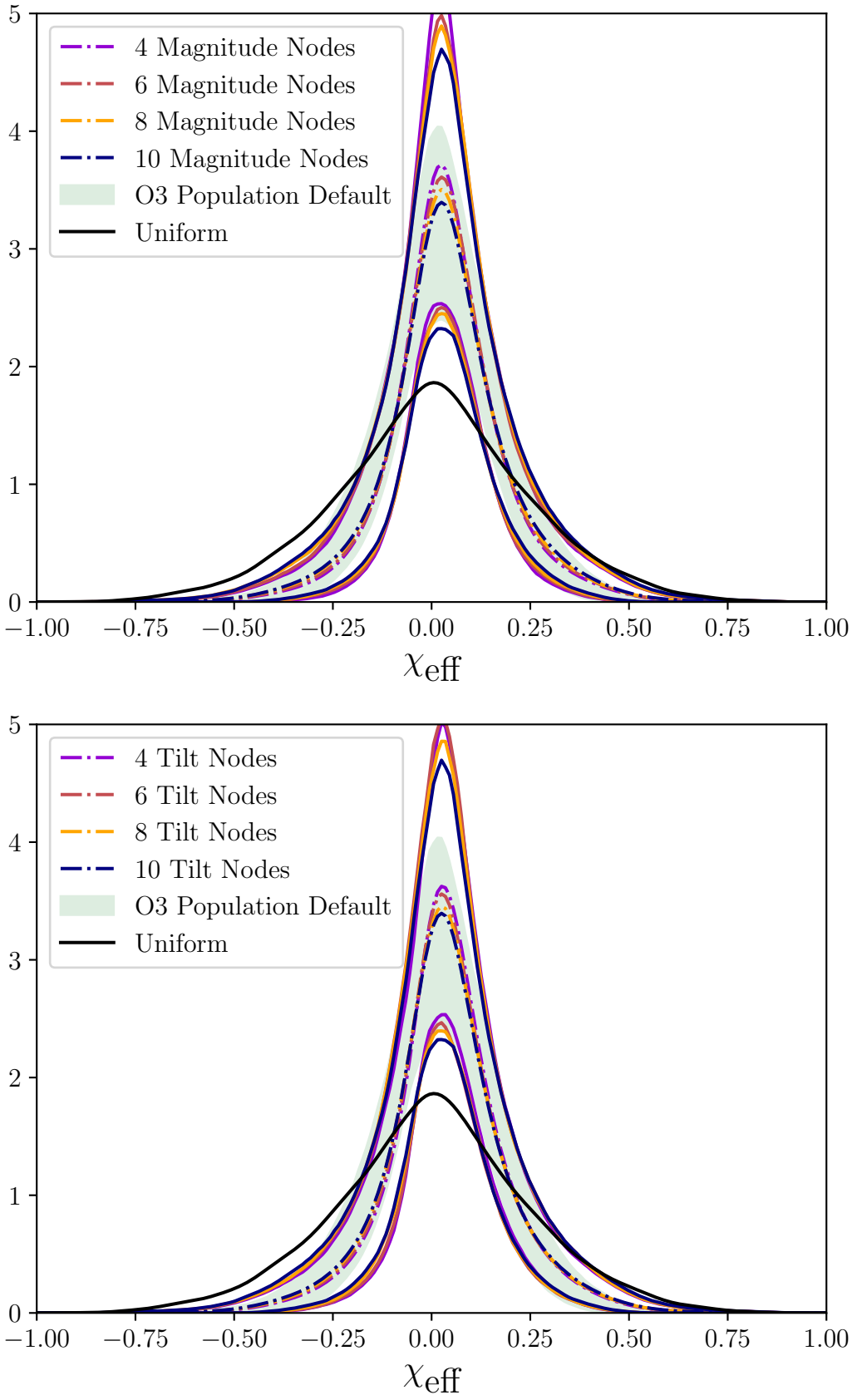

Figure 10.4: Distribution of effective inspiral spin parameter as recovered from the distributions in Figure 10.3 and Figure 10.2

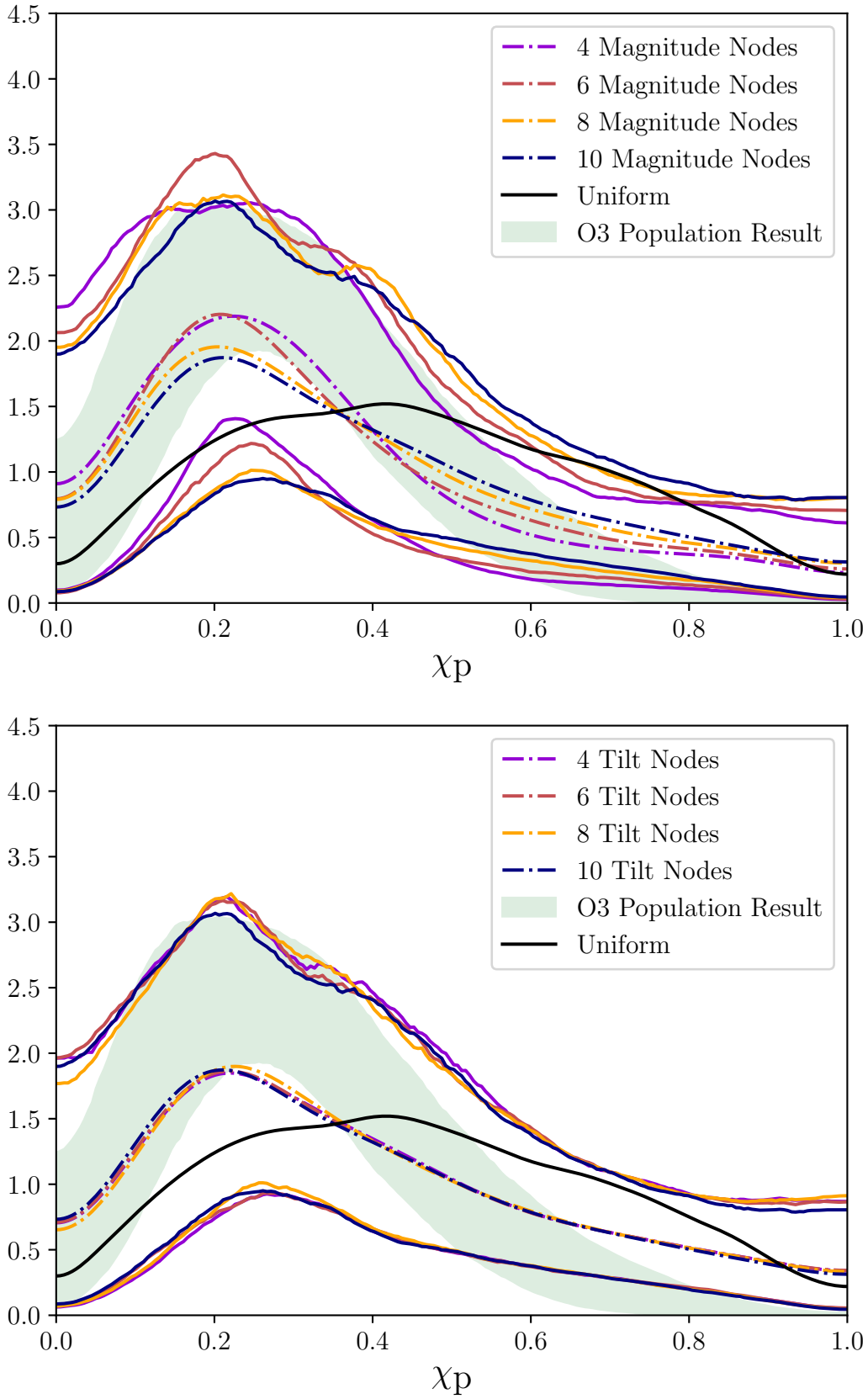

Figure 10.5: Distribution of precession spin parameter as recovered from the distributions in Figure 10.3.

the `Default` model [29, 28, 521, 582] analysis using the same event list, as well as the distribution corresponding to uniform distributions in a and $\cos\theta$. The inferred distribution of χ_p is consistent with an isotropic distribution, and shows agreement with the χ_p distribution reconstructed from the `Default` model. An exception to this agreement is the slightly increased support at high χ_p that is not present in the `Default` reconstruction. The higher support for large a and $\cos(\theta) \approx 0$ in the spline model relative to the `Default` model explains this increased support at high χ_p. Similarly, our result allows for more support at high χ_p relative to what is presented in [28], in which it is assumed that the $\chi_{\rm eff}$ and χ_p distributions follow a multivariate normal distribution [382]. In particular, Figure 16 in [28] shows vanishing support for $\chi_p > 0.4$ in the population, whereas we find some support in this region is included at the 90% credibility. We note that varying the number of nodes in spin magnitude has the largest impact on the averaged recovered χ_p distribution (dash-dotted curves) indicating that this measurement depends on our choice of prior for the spin magnitude distribution.

10.5 Discussion

Most previous analyses of the astrophysical distribution of merging binary black hole systems have focused on fitting parametric phenomenological models strongly constrained by the functional form of the model to the observed data (e.g., [28] and references therein) or directly compared with detailed simulations (e.g., [593, 575]). However, more data-driven methods have been employed to infer the binary black hole mass [358, 539, 463, 186] and spin distributions [539].

In this work, we use cubic splines (a model previously used to fit the black hole mass distribution [186]) to fit the astrophysical spin magnitude and spin tilt distributions of black holes as inferred from LIGO-Virgo observations. In doing so, we limit the influence of prior modelling assumptions on the inferred distribution and present a more data-driven result. While the uncertainties in the inferred distributions are large, we are able to interpret trends as they relate to astrophysical mechanisms of BBH formation.

Models of stellar physics suggest that angular momentum transport out of the core of a collapsing star is highly efficient, indicating that first generation stellar BHs should primarily have negligible spin [241]. Based on this, some models tend to favor nonspinning BHs when born in isolated environments and not susceptible to tidal spin-up.

Motivated by the work in [241, 89], the authors of [243, 127, 28, 542] search for contributions from a nonspinning subpopulation of BHs in the distributions of $\chi_{\rm eff}$ and spin magnitude. Some of this previous work has found support for $a = 0$ when using low-dimensional parametric spin distribution models that allow for support at that point [522, 564, 243, 309, 308]. The results of such inference are strongly model dependent, with the preference for the presence of a non-spinning component depending on the morphology of that component. We do not confidently recover such a feature as demonstrated in Figure 10.2. Given the width of the 90% credible interval at low spin magnitudes, we are unable to rule out the presence of this feature. This is consistent with what was found in [542, 127, 387], in which the authors find that there is insufficient data to resolve such a non-spinning subpopulation when employing a strongly parametric spin model with a spike at $a = 0$.

On the other hand, the merger of equal mass, nonspinning BBH systems are expected to result in a remnant BH with $a \sim 0.7$. As a result, population simulations predict that hierarchical mergers resulting from products of nonspinning first-generation mergers will leave a signature of a subpopulation of BH spins peaked around $a \sim 0.7$ with tails extending from $a \sim 0.5$ to ~ 0.9 [227]. Referring back to Figure 10.2, we do not see evidence of an obvious subpopulation in this high-spin region of interest, but rather some preference for low-spin magnitudes, possibly indicating that hierarchical mergers are not providing the dominant formation mechanism for the observed BBHs. The lack of support for a relatively high spin subpopulation is consistent with the conclusions drawn in [28, 127]. Assuming that BBHs from hierarchical mergers all have $a > 0.5$, we infer that no more than $23\%^{+14\%}_{-11\%}$ of the astrophysical population of merging BBHs form through a hierarchical merger channel.

A notable feature in our analysis is the increased uncertainty in the spin magnitude distribution as compared to that from the Beta distribution in spin magnitude. The motivation for using a Beta distribution in [582, 29, 28] is not physical but is statistical: the Beta distribution only has support in the interval [0, 1] and offers a flexible, parametric fit for the mean and variance of a distribution and has an analytic form. The spline model introduced in this work offers more flexibility than the Beta distribution, so lacking a physical motivation for the Beta distribution, we expect that the uncertainties in the spin magnitude distribution obtained in this work are more appropriate than those obtained from the `Default` model. Furthermore, the Beta distribution used in [29, 28] cannot model structures such as increased support

for nonspinning BHs or a secondary peak at high spin, making it a suboptimal model for the astrophysical spin magnitude distribution in the presence of a nonspinning subpopulation. Comparing to our data-driven approach, we therefore conclude that the resulting spin distributions presented both in in [28] as well as this work are partially model-driven.

When dynamical encounters take place within dense environments such as globular clusters, it is likely that some of the remnant BHs are retained in the cluster and merge in a subsequent dynamical encounter, contributing to the hierarchical merger population. The authors of [230] find that, for a broad range of populations considered, 16% of mergers in the hierarchical merger population have $\chi_{\rm eff} < -0.3$. Using our inferred $\chi_{\rm eff}$ distribution we infer $2.1\%^{+3.9\%}_{-1.5\%}$ of BBH mergers have $\chi_{\rm eff} < -0.3$. Using this interpretation from the $\chi_{\rm eff}$ distribution, we place a conservative upper limit on the contribution of hierarchical mergers to the BBH merger population of $13\%^{+24\%}_{-9\%}$ which agrees with the one obtained when using just the spin magnitude information. This limit broadly agrees with the upper limit of 26% for the fraction of hierarchical mergers presented in [230], in which the authors use low-dimensional parametric models to infer the $\chi_{\rm eff}$ distribution. This is also consistent with the results of [308] who found that depending on the escape velocity of the hierarchical merger environment up to $\approx$ 10% of merging black holes may come from hierarchical mergers.

Mergers of BBH systems which have spins that are isotropic in orientation, as is expected from dynamical formation scenarios, implies a distribution of $\chi_{\rm eff}$ symmetric about 0 (see the black line in Figure 10.4). This prediction comes from the idea that during a dynamical capture, there is no reason to expect that the two BHs should have correlated spin directions when they randomly encounter each other. In contrast, spins of BBH systems forming from common evolution are expected to remain primarily aligned with the orbital angular momentum, resulting in exclusively positive values for $\chi_{\rm eff}$ from this population. While the distribution of $\chi_{\rm eff}$ we recover appears symmetric, it is centered at $\chi_{\rm eff} = 0.033^{+0.034}_{-0.038}$, favoring a positive central location but consistent with being centered at $\chi_{\rm eff} = 0$ at the 90% level. This constraint is similar to that obtained by [187], using basis splines to model the component spin distributions. This result, coupled with the result that $38.7\%^{+12.8\%}_{-11.5\%}$ of events have $\chi_{\rm eff} < 0$, presents the possibility that dynamical encounters are a significant contribution to the formation mechanisms of BBH merger systems. This is in contrast to the results reported in [28] that the $\chi_{\rm eff}$ distribution is centered at

0.06 and rules out being centered at 0 at the 90% level; this result is obtained by modeling the $\chi_{\rm eff}$ distribution as a Gaussian with the mean and standard deviation as free parameters [382, 28]. While we use a largely identical event list in our analyses, the modeling assumptions for the $\chi_{\rm eff}$ distribution made in [28] probably explain some of the differences in our results. While [474, 127] also find a $\chi_{\rm eff}$ distribution consistent with being centered at zero, [474] does not find any support for $\chi_{\rm eff} < 0$; such a difference may also be due to different modeling choices for the population of $\chi_{\rm eff}$.

We see increased uncertainty in $\cos\theta$ with respect to those obtained from the `Default` model from [28]. While the tilt distribution of the `Default` model is astrophysically-motivated, it is incapable of capturing any possible substructure that may be present at locations other than $\cos\theta = 1$, as it is modeled by a monotonic function. Given the additional flexibility of the spline model, we notice a trend in the average line of the $\cos(\theta)$ distribution toward increased support for $-0.25 < \cos\theta < 0.5$. This trend is of low significance given the uncertainties surrounding it in the inferred $\cos\theta$ parameter space, but may indicate a nontrivial contribution from BBHs with in-plane spins to the astrophysical population. This trend is consistent with what is found in [557, 187], in which the authors use more flexible models to infer the $\cos\theta$ distribution. Our result that $38.6\%^{+17.3\%}_{-15.6\%}$ of BHs exhibit negative spin tilts is broadly consistent with previous studies that indicate the need for negative alignment in the astrophysical population. The presence of support for $\cos\theta < 0$ in the population was reported in [29] and confirmed in subsequent studies (e.g. [127, 28]).

It is generally considered unlikely for BBH systems formed under common/isolated evolution scenarios to exhibit spin-orbit misalignment, as any such misalignment in these systems is expected to be corrected by mass transfer and tidal effects [218, 299]. Lack of confidently-increased support for $\cos\theta = 1$ indicates that aligned-spin BBH systems do not contribute a statistically resolvable subpopulation of mergers. A possible explanation for this is a comparable or more significant contribution of BBH mergers from dynamical encounters in dense environments to the inferred astrophysical population of BBH mergers, as this would manifest as a more isotropic distribution in tilts.

Unlike other work which adopts data-driven models but enforce regularization or smoothing conditions across parameter space (e.g. [127, 187]), we set the correlation length explicitly by setting the number of spline nodes and their locations *a priori*.

While smoothing over the scale of perturbations in parameter space may avoid the biases noted in Section Appendix 10.B, it also limits the scale of the features that can be resolved by the model. By controlling for this scale limit on a model-by-model basis, we can evaluate if we need to resolve finer-scaled features in order to get a better fit to the data, for example by comparing the model evidences in Figure 10.1. We believe that the averaging or smoothing adopted in [127, 187] is what prevents the spurious features noted in Appendix 10.B from being recovered in those works, at the cost of not being able to resolve fine features in the population distribution. Nonetheless, as we see from the similar evidence values in Figure 10.1, it is not important that a model need to be able to resolve such narrow features in the spin distribution, as they do not seem to inform the posterior to a large extent. We anticipate that this will not be the case in analyses with future catalogs, as additional data will become more informative to the structures in the spin distribution, limiting the out-sized influence of the model prior on the posterior. In this case, data-driven models may become necessary to best describe the distribution and this will be reflected by a higher evidence for these models relative to that of the `Default` model.

The morphological differences in the recovered spin distributions between models are a natural outcome of using different models to infer a distribution given uninformative data, as the prior provides much of the information to the posterior. We note there are visible differences in the recovered distribution between spline models and their credible regions in Figures 10.2 and 10.3, such as the locations of large oscillations which appear depending on node placement. Because of the comparable evidence values in Figure 10.1, we conclude that none of the spline models do a significantly better job describing the data than any of the others. We anticipate that in future studies, additional observations will contribute to resolvable structure (if present) in the spin distributions; in this case, the spline models which best describe this structure will have the highest evidences. Similarly, the Bayes factor between the `Default` model and the spline model of the lowest evidence is only $\ln \mathrm{BF} \sim 1.5$ in favor of the spline model, indicating that the `Default` model does not provide a significantly worse fit to the data than the spline models. In other words, the additional flexibility of the spline models presented in this work may not be necessary to correctly describe the spin distribution, as the lower-dimensional model provides a fit with similar evidence. We caution against trusting the recovered features in the resulting distributions as different models with similar evidences can recover fairly different features in the spin distribution.

The work presented in this paper motivates the need for more data-driven models for inferring the BBH spin distribution from GW sources, as there may be features of astrophysical importance that cannot be captured by currently-used parametric models. While we cannot confidently discern many trends in the spin magnitude and tilt distributions, we can place constraints on the support in different parts of spin parameter space by substantially relaxing modeling assumptions. Using our more flexible model, we find that substantially increased uncertainties are a necessary cost to being able to model arbitrary features in the spin distribution, given current GW data. Data collected from events in future observing runs may help resolve such features which may exist in the spin distribution, as well as motivate a better choice of priors to use on these data-driven models.

Acknowledgements

We thank Ethan Payne, Salvatore Vitale, Derek Davis, and Sylvia Biscoveanu for helpful discussions regarding this work. We are grateful to Bruce Edelman and Ben Farr for fruitful discussions of spline modeling. We additionally thank Chase Kimball for reviewing this manuscript. JG acknowledges funding from NSF grants 2207758 and PHY-1764464. CT is supported by an MKI Kavli Fellowship and an Eric and Wendy Schmidt AI in Science Fellowship. This material is based upon work supported by NSF's LIGO Laboratory which is a major facility fully funded by the National Science Foundation. This work used computational resources provided by the Caltech LIGO Lab and supported by NSF grants PHY-0757058 and PHY-0823459.

Appendix 10.A Efficient evaluation of the spline model

Our model requires evaluating a different spline model at many values during each likelihood evaluation. This process can be divided into three stages: constructing the spline model, identifying where each of evaluation points lies in relation to the nodes, and evaluating the appropriate piece in the spline. The first stage must be performed at every iteration but does not depend on the number of points that the spline will be evaluated at. The second stage is independent of the value of the spline nodes but must be performed for each of the evaluation points. For a uniform spacing of spline nodes, this can be efficiently evaluated, however, for a generic spline this can be computationally intensive. At the third stage, we simply combine the results of the two previous stages. This can be trivially parallelized using a graphics processing unit (GPU).

For our use case, the locations at which the splines are evaluated and the node points are the same at every iteration. We can therefore cache the result of the second stage. We find that for our application the caching method accelerates the evaluation of the model by a factor of $\gtrsim 100$. Our implementation `cached_interpolate` is available via `pypi` and `conda-forge`.

Appendix 10.B Comparison between injection sets

Sensitivity estimates for Advanced LIGO and Virgo were released along with [26, 28] for the first three observing runs [529, 530, 26]. These sensitivity estimates consist of injections of simulated sources into detector noise, along with the SNRs and FARs of these injections as reported by the detection pipelines used in the LIGO-Virgo observing runs. In this section, we compare the use these injections to compute the $p_{\rm det}$ term as written in Eq. 10.7 to the use of our own injections to compute the same term.

Using a set of injections, we include those which pass a detection threshold in the summation over the found injections. For the injections provided in [26], we use a threshold of FAR < 1 yr^{-1}, consistent with the choice made in [28]. We do not run the detection pipelines to assign a FAR to each of the injections from our custom injection set in this paper, so we threshold these on an optimal SNR > 10. We visually check that the distribution of found injections is not biased by our approximation of the detection threshold in the selection function, such that the distribution of detection probabilities across parameter space are qualitatively

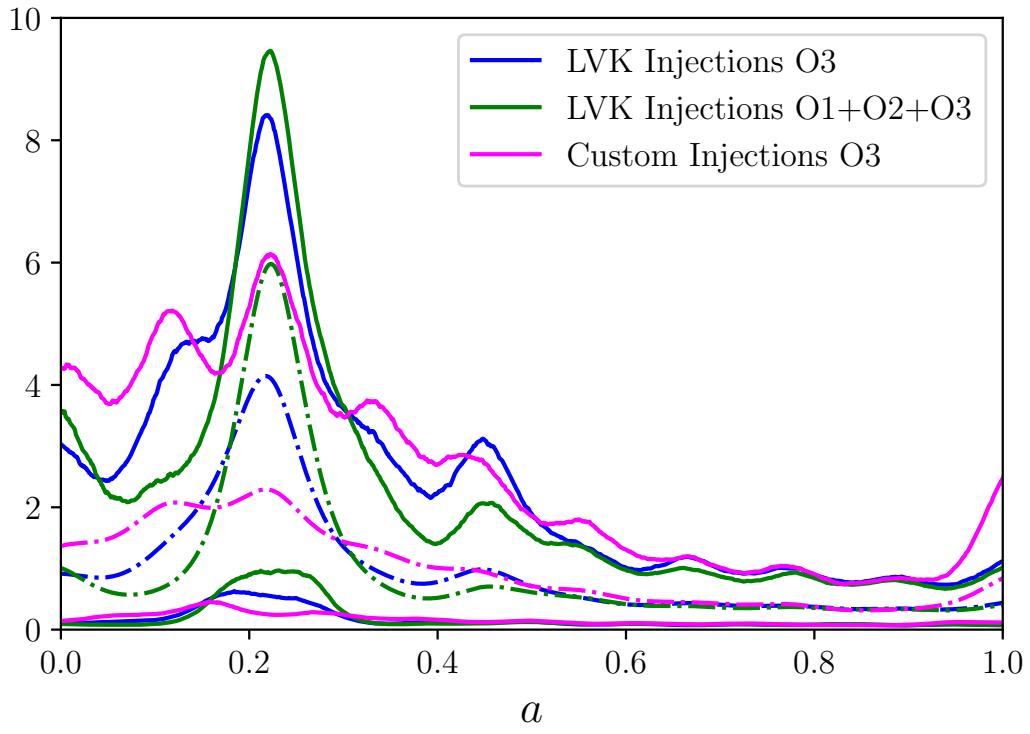

Figure 10.6: Distribution of spin magnitude using 10 nodes. Different colors correspond to different sensitivity injection sets. Injection sets use events from their corresponding observing runs. Note the peak at around $a = 0.2$ is most pronounced with using O1+O2+O3 set and least with our custom O3 set.

similar for the different injection sets. Since importance sampling Monte Carlo integration relies on drawing enough samples from the fiducial distribution that cover the support of the target distribution, using Eq. 10.7 as a reliable estimator for Eq. 10.6 requires a suitable number of "found" injections to get a well-converged estimate (i.e. see [213]). With too few samples being used to compute the Monte Carlo approximation, the variance of our estimator is large and the resulting estimate may be a poor approximation of the true log likelihood. The statistical uncertainty in the log likelihood estimates at each point in parameter space can cause a systematic bias to appear in the resulting posterior distribution for the population.

The injection sets provided in [26, 28] for the combined O1+O2+O3 sensitivity and O3 sensitivity respectively contain 41,972 and 81,117 simulated events which pass our detection threshold. As expected, the use of more samples reduces the uncertainty in the computed p_{det} estimate, and is reflected by the distribution of variances in $\ln \mathcal{L}$, as shown in Figure 10.10. We drastically reduce the variances in

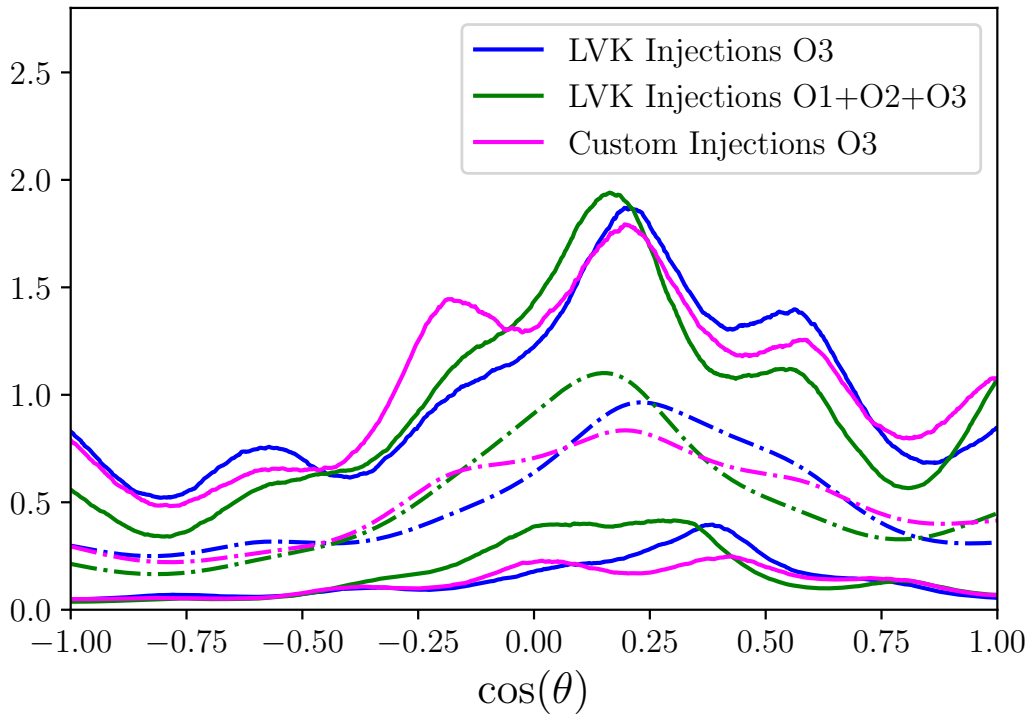

Figure 10.7: Distribution of spin tilt using 6 nodes. Different colors correspond to different sensitivity injection sets. Injection sets use events from their corresponding observing runs.

the log likelihood estimates by using our own injection set which contains 911,386 injections passing our detection threshold.

In order to validate that the number of samples used to compute p_{det} is a cause of systematic bias in the inferred population (as opposed to the difference in detection statistic used for the threshold), we repeat the above spin distribution inference but using injection sets that have been downsampled to have $\sim 40,000$ and $\sim 80,000$ found injections. In Figures 10.8 and 10.9 we note recovery of strongly peaked features in the spin magnitude and tilt distributions, respectively. The significance of these features becomes drastically reduced as the number of found injections increases, indicating that a lower number of effective samples used to compute Equation 10.7 can lead to biases that propagate into spurious features in the spin distribution. We therefore infer that a sufficient number of injections is necessary to recover an unbiased spin distribution using our spline model implementation, motivating our use of the custom injection set with substantially more found injections than what was released in [530].

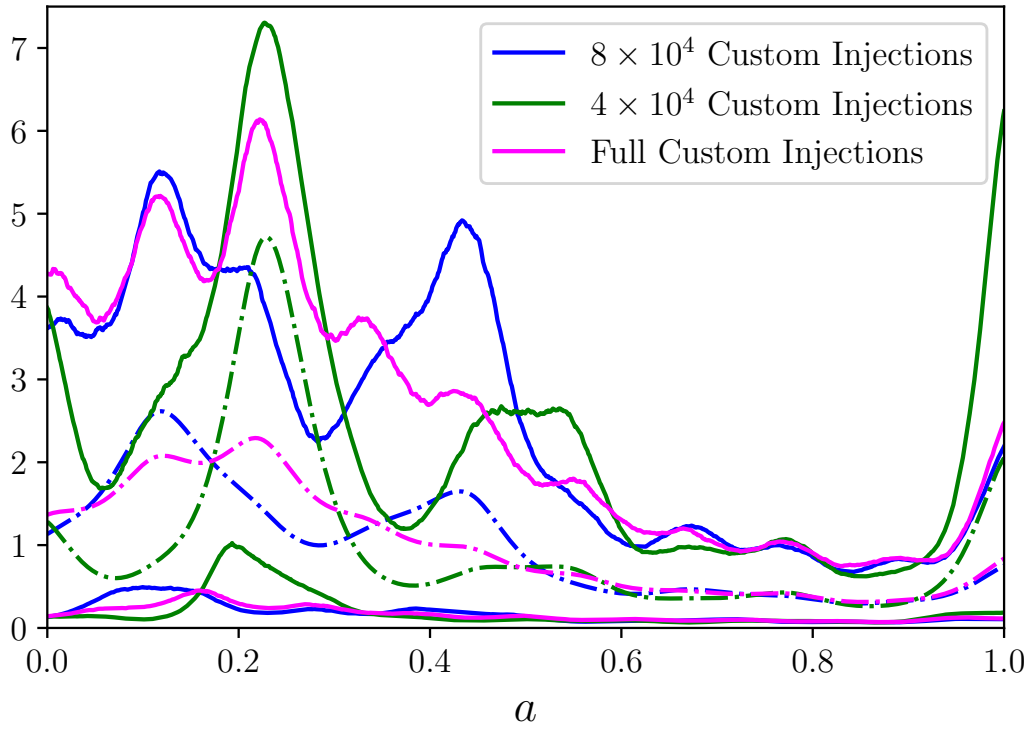

Figure 10.8: Distribution of spin tilt using 6 nodes. Different colors correspond to a different number of found injections: the 900,000 from the custom injection set, ~ 80,000, and ~ 40,000, where the latter two are close to the number of found injections in the LVK O3-only and O1+O2+O3 injection sets, respectively.

We quantify the uncertainty in the log likelihood across the parameter space by calculating the uncertainty in the difference in log likelihoods for all of Λ. In Fig 10.10, we quantify this uncertainty as the covariance in the log likelihood between each point Λ and Λ', the value of Λ that has the lowest variance.

We see in Figure 10.10 that the amplitude of the node at $a = 0.22$ is correlated with higher statistical variance; as the amplitude of this node increases, the uncertainty in the log likelihood increases as well, making the log likelihoods computed in this part of parameter space less trustworthy. As we decrease the variance by using injection sets with higher N_{inj}, the uncertainty in the log likelihood estimates decreases. With better estimates of the log likelihood, the support for the high amplitude of the node at $a = 0.22$ decreases, indicating that this feature in the spin magnitude distribution may be an artifact of poorly-converged Monte Carlo integrals. If this peak were a true feature in the astrophysical population, we would expect the inferred distribution computed with the custom injections would

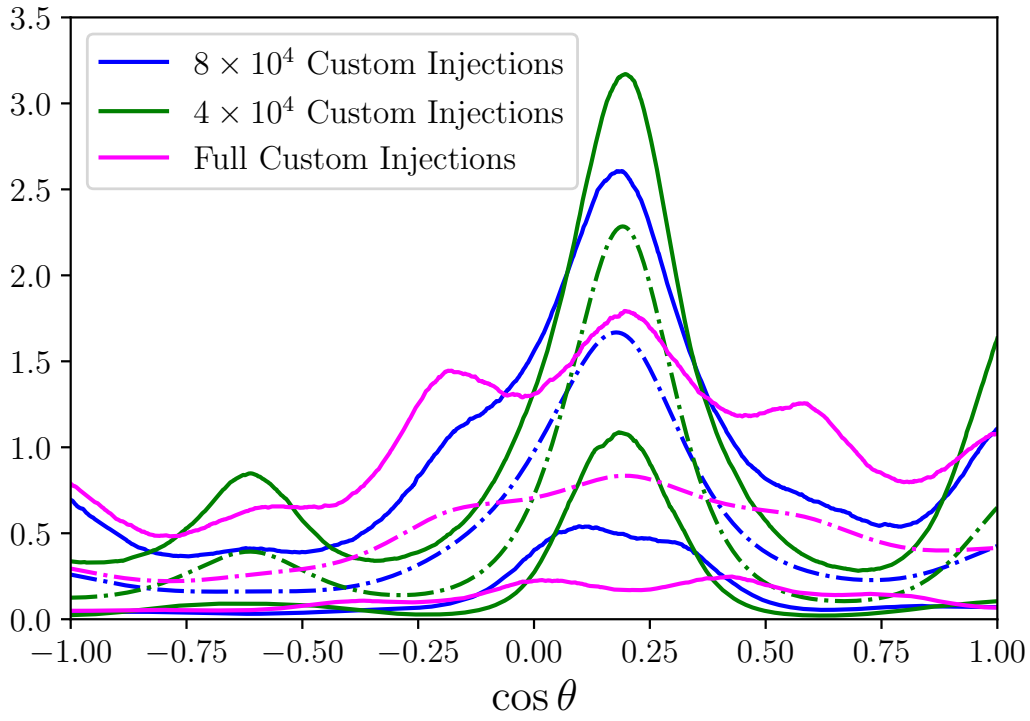

Figure 10.9: Distribution of spin tilt using 6 nodes. Different colors correspond to a different number of found injections: the 900,000 from the custom injection set, ~ 80,000, and ~ 40,000, where the latter two are close to the number of found injections in the LVK O3-only and O1+O2+O3 injection sets, respectively.

maintain support for high amplitude at this node, along with reduced uncertainty. We confirm that this uncertainty is associated with the selection function rather than associated with reweighting posterior samples in the population model by comparing the contributions of the uncertainties in both these Monte Carlo summations to the total propagated uncertainty in the log likelihood; for the models tested in this paper, we consistently find that the uncertainty associated with the contribution from Eq 10.7 dominates.

We note that we have noticed several other examples of similar behavior in our analyses, notably manifesting as spurious peaks in the spin distribution. This demonstrates the need for sufficient coverage of injections when using importance sampling to compute sensitivity estimates especially when evaluating a population distribution that can model narrow peaks.

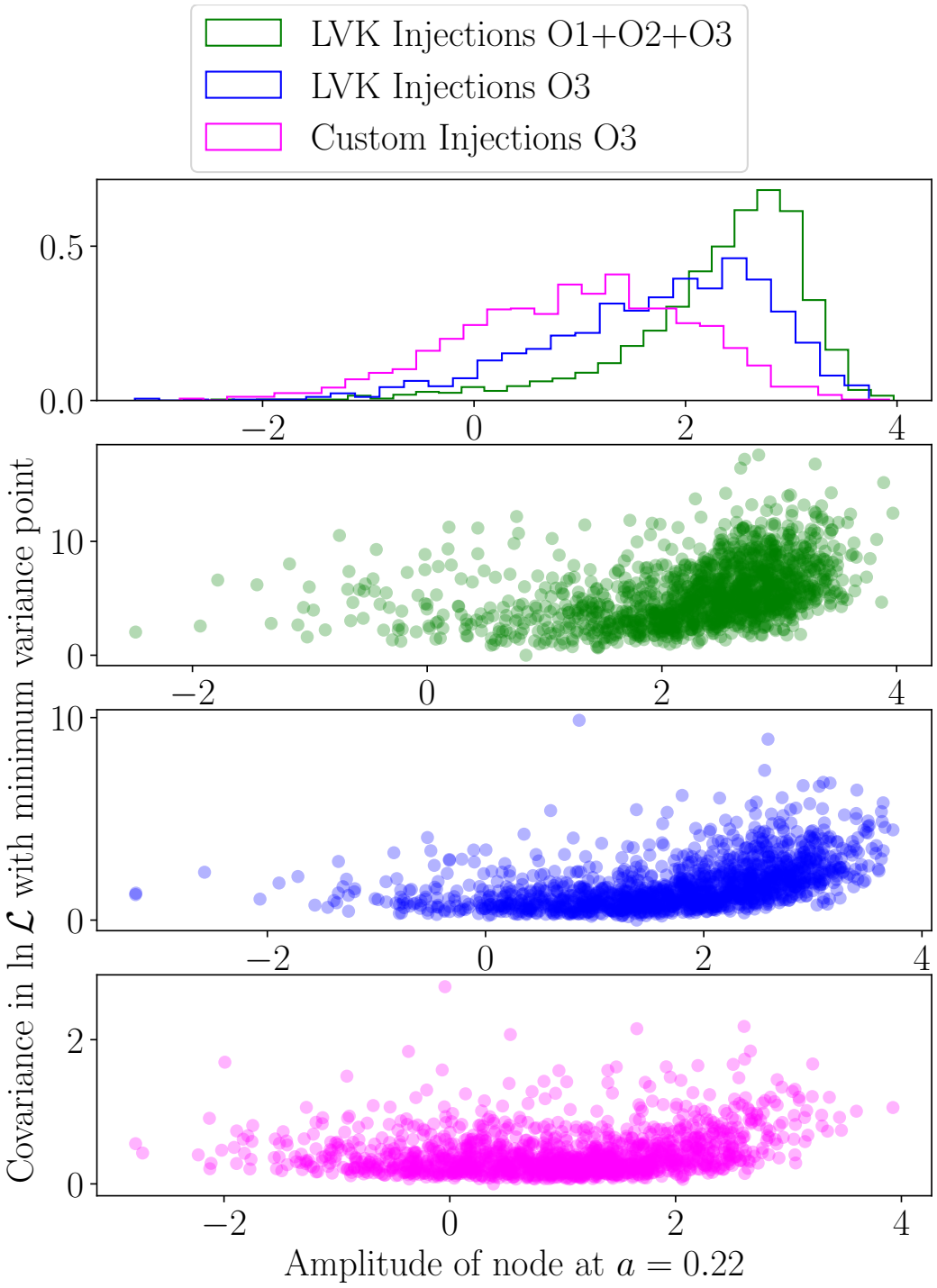

Figure 10.10: Posterior samples for the amplitude of the spline node at $a = 0.22$ and their associated variances in the log likelihood. Note the positive correlation between spline amplitude and variance. The results obtained using the injection set from all three observing runs, which has the least number of found injections, exhibits the highest variance in the log likelihood.

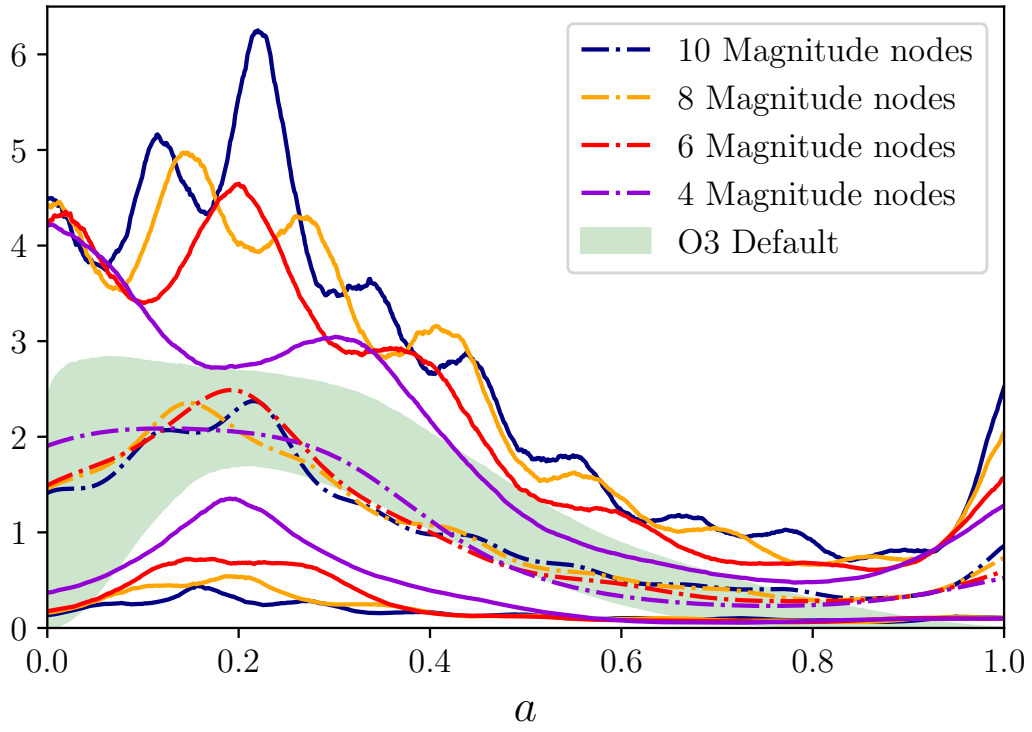

Figure 10.11: Distribution of spin magnitudes, with different numbers of nodes corresponding to different colors. All use 4 nodes in the tilt distribution.

Appendix 10.C Effect of priors on inferred distribution

In this work, we adopt a prior on the spline node amplitudes that is a standard normal distribution. In this appendix, we show the results obtained using different choices of prior on the node amplitudes and different positions.

In Figure 10.13, we show the distribution of spin magnitudes and tilts from prior draws only. We note that the average of the distribution is flat, reflecting the lack of any further structure imposed by the prior on the mean of the distribution. On the other hand, the upper limit of the 90% credible regions shows considerable oscillations. These are coincident with the node locations and thus correspond to where the distribution is informed directly by the spline amplitude sample. These oscillations are thus an expected feature of the spline model. The regions in between these oscillations correspond to where the spline provides an interpolation between node locations.

Figure 10.14 shows the inferred distribution for the 4 tilt node and 4 magnitude node model for three different choices of prior on the node amplitude and placement: a

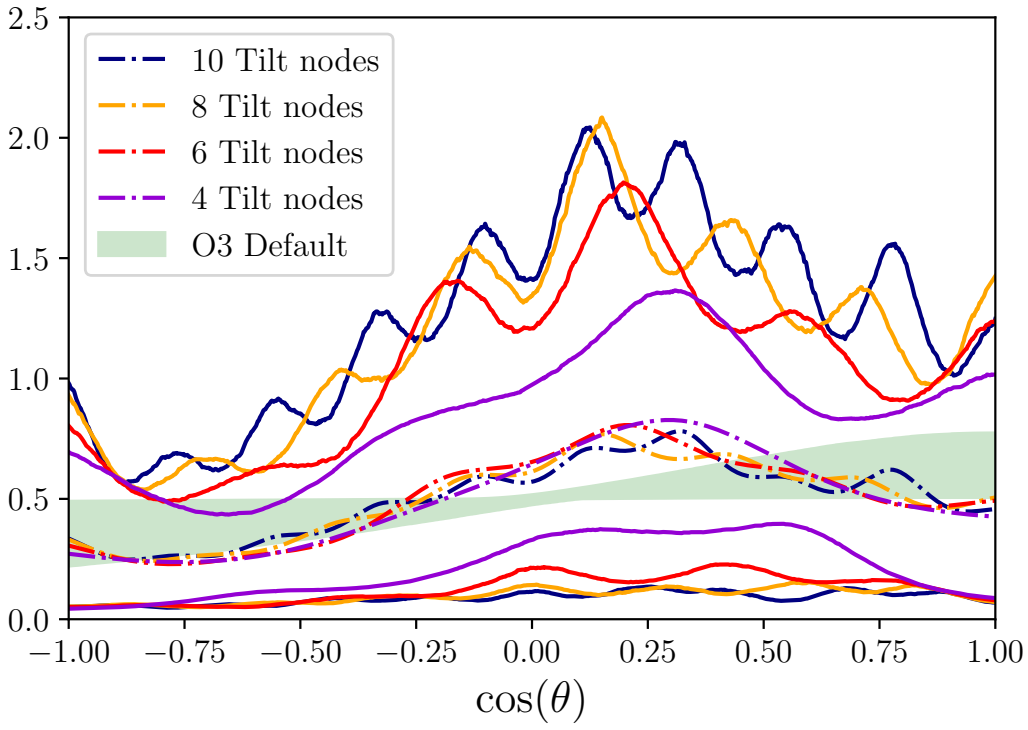

Figure 10.12: Distribution of spin tilts, with different numbers of nodes corresponding to different colors. All use 4 nodes in the spin magnitude distribution.

broader Gaussian (yellow), a narrower Gaussian (magenta), a uniform distribution in $[-3, 3]$ and the unit Gaussian without the additional end nodes (see, Sec. 10.2.1). They each result in comparable distributions within the statistical uncertainties, however, we note that the wider prior distributions leads to wider uncertainty bands, with the $\mathcal{N}(0, 0.5)$ prior giving the tightest constraints. We expect this to also hold for the other node configurations.

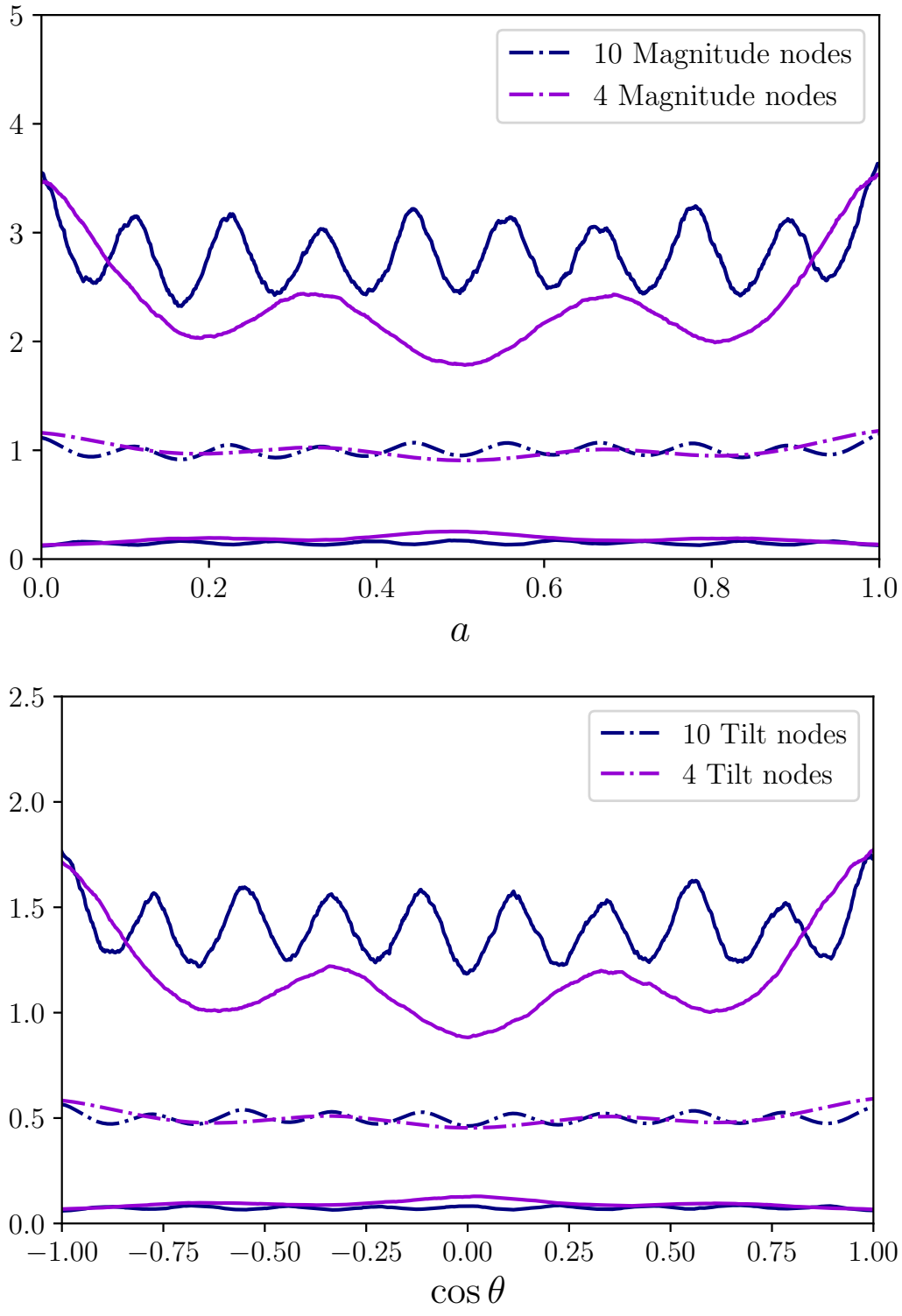

Figure 10.13: Comparison of the priors on the spin magnitude (top) and spin tilt (bottom) distributions. The envelopes (increased width in the distribution) are coincident with the node placement.

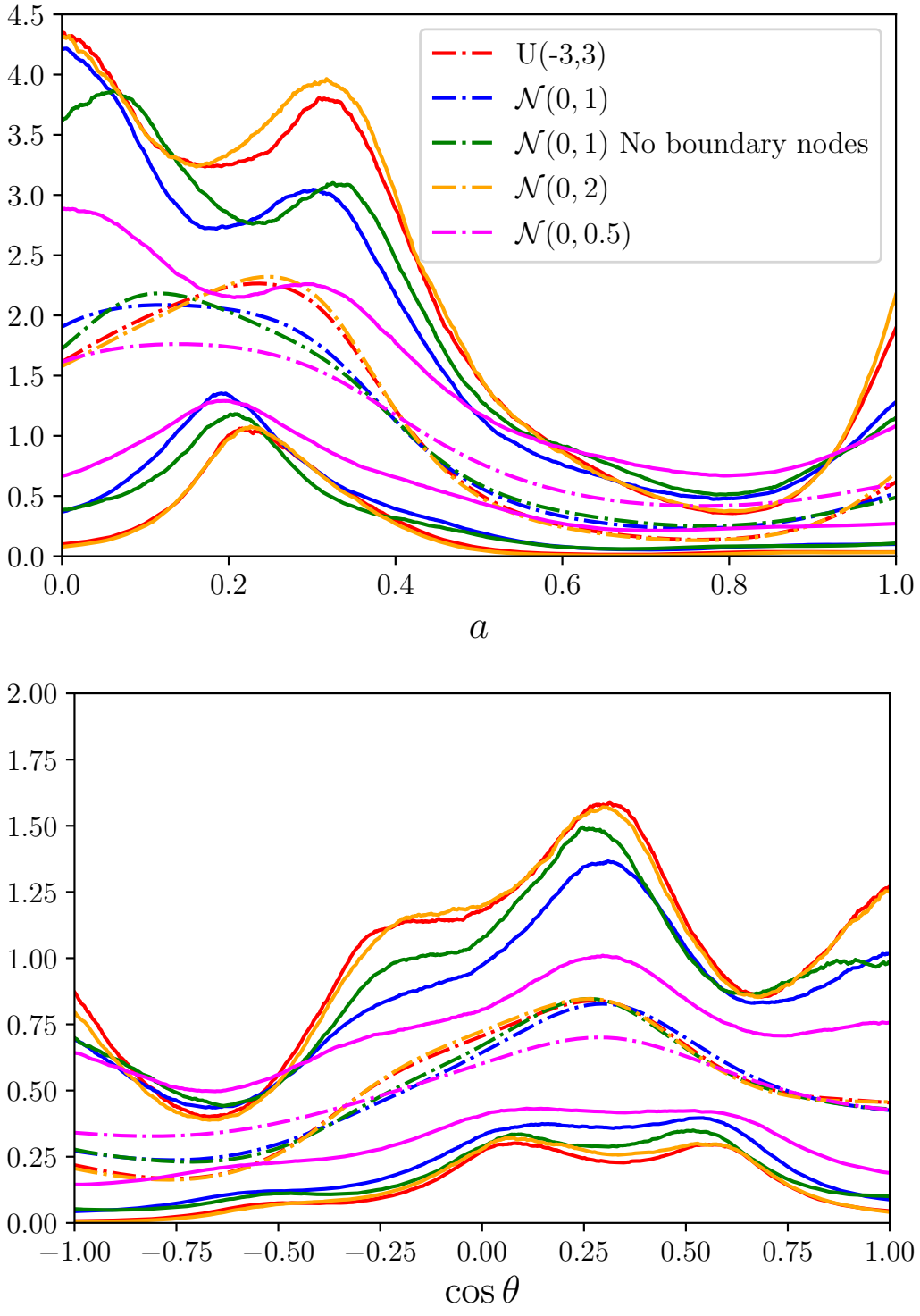

Figure 10.14: Comparison of spin reconstructions for different choices of priors on the spline nodes. The blue distribution corresponds to choice used in this work. The distribution "No Boundary Nodes" also places the boundaries linearly in the domain, but does not place any nodes on the domain boundary.

Chapter 11

PHYSICAL MODELS FOR THE ASTROPHYSICAL POPULATION OF BLACK HOLES: APPLICATION TO THE BUMP IN THE MASS DISTRIBUTION OF GRAVITATIONAL WAVE SOURCES

Jacob Golomb, Maximiliano Isi, and Will M. Farr. "Physical Models for the Astrophysical Population of Black Holes: Application to the Bump in the Mass Distribution of Gravitational-wave Sources". In: *Astrophys. J.* 976.1 (2024), p. 121. DOI: 10.3847/1538-4357/ad8572. arXiv: 2312.03973 [astro-ph.HE].

Abstract

Gravitational wave observations of binary black holes have revealed unexpected structure in the black hole mass distribution. Previous studies employ physically-motivated phenomenological models and infer the parameters that control the features of the mass distribution that are allowed in their model, associating the constraints on those parameters with their physical motivations *a posteriori*. In this work, we take an alternative approach in which we introduce a model parameterizing the underlying stellar and core-collapse physics and obtaining the remnant black hole distribution as a derived byproduct. In doing so, we constrain the stellar physics necessary to explain the astrophysical distribution of black hole properties under a given model. We apply this to the mapping between initial mass and remnant black hole mass, accounting for mass-dependent mass loss using a simple parameterized description. Allowing the parameters of the initial mass-remnant mass relationship to evolve with redshift permits correlated and physically reasonable changes to features in the mass function. We find that the current data are consistent with no redshift evolution in the core-remnant mass relationship, but place only weak constraints on the change of these parameters. This procedure can be applied to modeling any physical process underlying the astrophysical distribution. We illustrate this by applying our model to the pulsational pair instability supernova (PPISN)

process, previously proposed as an explanation for the observed excess of black holes at $\sim 35\,M_{\odot}$. Placing constraints on the reaction rates necessary to explain the PPISN parameters, we concur with previous results in the literature that the peak observed at $\sim 35\,M_{\odot}$ is unlikely to be a signature from the PPISN process as presently understood.

11.1 Introduction

Observations of gravitational waves from binary-black-hole (BBH) and binary-neutron-star (BNS) mergers with the LIGO-Virgo-KAGRA detector network have provided otherwise inaccessible information on the properties of those compact objects [16, 1, 38]. While individual events offer a glimpse into the details of a particular black hole (BH) or neutron star, studying observations collectively on a population-level allows us to draw inferences about stellar astrophysics, the formation channels of BHs and neutron stars, the overall rates at which BBH/BNS mergers occur in the universe, and cosmology [28, 21, 28, 29], as well as tests of general relativity [34, 417].

Models adopted for population inference of BBHs tend to take one of two major approaches. The first are so-called parametric methods, in which a phenomenological model is constructed using relatively few parameters, with these parameters directly controlling well-defined features encoded in the model. This commonly involves assuming a functional form for the global structure of the distribution (e.g., a truncated power law for the mass distribution), enhanced by features such as a bump, dip, or break, and jointly fitting for the properties of the global structure and additional features [524, 29, 28, 317, 226]. The other popular approach consists of data-driven methods (sometimes called "non-parametric" methods, in spite of their overabundance of parameters) in which a flexible model is allowed to fit nearly-arbitrary shapes to the distribution. Such fits have been achieved with tools such as splines, histograms, and Gaussian process regression [125, 257, 358, 447, 186, 540].

While the latter method is more general and can capture features not explicitly defined in a model, the former offers the ability to encode signatures from expected physical processes in parameters controlling the shape of the distribution, making it possible to interpret the constraints in terms of the underlying physical motivations. Nevertheless, interpreting the constraints on these parameters as constraints on the underlying physics can be difficult, as there are often unmodeled assumptions as to how the underlying physical processes translate into resulting distribution that is being modeled.

In this work, we provide an alternative approach to prescribing a parametric population model: instead of modeling the BBH distribution directly, we introduce parameters to describe the underlying astrophysics and the associated mapping to remnant BH properties and we derive the resulting population distribution as a re-

sult of these underlying parameters and its associated mapping. Subject to model assumptions, we more directly infer the physics as informed by BBH properties in a way that avoids strong phenomenological approximations in the BBH population model itself. This is related to the "backpropagation" approach in [574] and the method in [60], except we operate at the population level rather than individual events and impose a rigid physical mapping from progenitors to remnants.

Previous parametric approaches to modeling the BBH mass distribution have been particularly useful to place constraints with relatively little data when strong assumptions about the structure of the mass distribution are warranted. For example, parametric population analyses of the first catalogs of BBH events have revealed that the mass distribution is well-described by a truncated power law that peaks at $\sim 8\,M_\odot$, decaying to high masses, and featuring an overdensity (modeled as a Gaussian bump) at $\sim 35\,M_\odot$ [28, 29].

We apply our approach to the overdensity in the mass distribution at $\sim 35\,M_\odot$, which may or may not be accompanied by a subsequent dip [524, 186]. The original motivation for looking for this feature was the expected "pile-up" of BHs that resulted from progenitors that had undergone pair instability pulsations [524]. This pulsational-pair instability supernova (PPISN) process results in a nonlinear mapping from the masses of progenitors stars to their remnant BHs [576, 577, 578, 210]: for relatively low masses, the mapping is linear; however, it turns approximately quadratic for higher masses, with a turnover that caps the a maximum BH mass obtainable from the evolution of an individual star; in BH-mass space, the beginning of this BH "mass gap" is preceded by a pile-up from the quadratic turnover. This kind of relationship between progenitor and BH masses can be directly exploited to bridge parametric models of BBH populations with stellar physics for improved inference, as we do here.

In this work, we implement a simple model for the initial mass function of stellar progenitors and the associated map from progenitor mass to remnant mass, motivated by the type of relationship found from the simulations in [210, 211]. Rather than informing our model with individual massive sources, we construct a full population model for the mass distribution, including a subpopulation in the upper mass gap due to higher-generation mergers. Using data from gravitational wave events from the third Gravitational Wave Transient Catalog [GWTC-3, 28], we infer the shape of the initial mass function stellar progenitors, the associated mapping to the remnant BH distribution, and the relative contribution of sources formed through 1G mergers.

Baxter et al. [87] also explored the consequences of a similar initial mass-remnant mass relation to those in Farmer et al. [210] and Farmer et al. [211], but unlike our work here that directly parameterizes the relation, they used parameterized models that had been fitted to theoretical BH mass functions.

Having a physically-motivated model for the BH mass distribution facilitates extensions that incorporate richer physics. As an example, here we allow the underlying physics to evolve with redshift, as may be expected from cosmic history considerations. Such evolution in the underlying physical parameters captures the correlated changes in shape and height of the bump that must occur in the presence of changing progenitor to remnant mass relationships. It may be possible to use shape measurements to calibrate changes in the mass scale of the bump with redshift to reduce or eliminate systematic uncertianities in cosmological parameter inference from the BBH mass function, sometimes called the "spectral siren" method [215, 203].

While here we apply this approach to modeling the PPISN process underlying the astrophysical BBH mass distribution, it can more generally be used as a model to place constraints on any relationship between progenitor mass and remnant BH mass as informed by gravitational wave observations. The model introduced here can readily be applied to any process with accelerating mass loss as a function of progenitor mass, but this method can be useful for inferring the physics of any arbitrary relationship underlying an observable distribution associated with BBHs [cf. 231, for a related approach applied to inferring the delay times between binary formation and merger].

We begin with an overview of hierarchical Bayesian inference in Sec. 11.2. We then outline the models with and without evolution with redshift in Sec. 11.3. In Sec. 11.4, we present results for both model configurations, using data from the third oberseving run (O3) of LIGO-Virgo. We offer interpretations of our results in Sec. 11.5 and provide concluding remarks in Sec. 11.6. We find that the PPISN mechanism, as currently predicted by stellar evolution models, cannot predict the 35 $M_\odot$ feature in the BBH mass distribution.

11.2 Hierarchical Bayesian Inference

We conduct our inference on the population parameters Λ with a hierarchical Bayesian framework, in which we inform our population model with a catalog

$N_{\rm det}$ events, to compute the likelihood (see, e.g., [359, 536]):

$$\mathcal{L}(\{d\}|\Lambda) \propto \frac{K(\Lambda)^{N_d} e^{-K(\Lambda)}}{p_{\rm det}(\Lambda)^{N_d}} \prod_{i=1}^{N_{\rm det}} \int \mathcal{L}(d_i|\theta)\pi(\theta|\Lambda)d\theta \tag{11.1}$$

where $\mathcal{L}(d_i|\theta)$ is the likelihood of the data for the ith event, given physical parameters θ (i.e., masses, distances), and $\pi(\theta|\Lambda)$ is our population model with a predicted number of detections K. The $p_{\rm det}(\Lambda)$ prefactor accounts for the selection effects associated with observing a catalog biased toward sources with parameters that favor detectability (i.e., the Malmquist bias). See Appendix 11.A for details on this likelihood.

Following the approach in [213] and [538], we compute $p_{\rm det}(\Lambda)$ with injections of sources from a fiducial population in detector noise, and assigning weights to each of the sources that pass our detection threshold. These sensitivity injections are from the O3 injection set released in [337]. We compute the per-event population evidence (the integral in Eq. (11.1)) by reweighting samples from individual event posterior distributions and dividing by the event-specific sampling priors. Since our population model is written only in terms of masses and distances, we effectively adopt the prior from parameter estimation for the spin parameters (isotropic in direction and uniform in spin magnitude).

For our analyses involving third observing run (O3) data, we obtain posterior samples for each event from [338], using the same BBH events from O3 as in [28]. This results in 59 events meeting the False Alarm Rate (FAR) threshold of 1 per year. Throughout this work we assume the best-fit cosmological parameters from the Planck 2018 release [46].

We sample the population posterior using the No-U-Turn-Sampler (NUTS) in `Numpyro` [424, 95], and we write the functions for computing Eq. (11.1) in `jax` [111] to take advantage of automatic differentiation when sampling with Hamiltonian Monte Carlo [185].

We do not enforce the convergence conditions from [28] for the Monte Carlo integrals in our likelihood, but we confirm that all points in our posterior have a reasonable enough number of effective per event posterior samples and injections for good convergence.

11.3 Mass Distribution of Black Holes From Progenitor Mass Function

11.3.1 Mass Distribution Model

We begin to construct our mass distribution model by assuming a functional form for the initial mass function (IMF) of compact object progenitors. Surveys have shown that the stellar IMF on the main sequence can be well-modeled as a featureless power law at high masses, with a power law index of approximately -2.3 [see, e.g., 319, 485, 320]. Recent studies have shown through simulations and analytic approximations that the there may be an approximately linear relationship between a high-mass star's zero-age main-sequence (ZAMS) mass and the mass of its core before undergoing supernova, although this is uncertain [577, 90]. We therefore assume that the IMF of compact object progenitors can also be modeled with a power law to good approximation, but allow for a break at $20\,M_\odot$ for additional flexibility [see, e.g., 491]. Even if this relationship has nontrivial nonlinearities, modelling the shape of the broken power law should capture the dominant resolvable structure of the distribution.

We express the distribution of initial progenitor masses, M_I, as

$$\frac{dN}{dM_\mathrm{I}}(M_\mathrm{I}) \propto \begin{cases} \left(\frac{M_\mathrm{I}}{20M_\odot}\right)^{-a} & \text{if } M_\mathrm{I} < 20M_\odot\,, \\ \left(\frac{M_\mathrm{I}}{20M_\odot}\right)^{-b} & \text{if } M_\mathrm{I} > 20M_\odot\,. \end{cases} \tag{11.2}$$

In order to obtain the resulting BH mass distribution from the progenitor mass distribution, we require a mapping between M_I for a progenitor and the mass of its remnant after undergoing core collapse. Here we assume that the mean remnant mass follows the initial mass for small initial masses, before smoothly transitioning at black hole masses M_tr to a quadratic relationship that exhibits a maximum remnant mass M_BHmax, eventually decaying to zero remnant masses. We impose throughout this work a constraint that $M_\mathrm{BHmax} > M_\mathrm{tr}$ so that our mapping is well-defined. We express this piecewise mapping through a functional form $\bar{M}_\mathrm{BH}(M_\mathrm{I}|M_\mathrm{tr}, M_\mathrm{BHmax})$ given by

$$\bar{M}_\mathrm{BH}(M_\mathrm{I}|M_\mathrm{tr}, M_\mathrm{BHmax}) = \begin{cases} M_\mathrm{I} & \text{if } M_\mathrm{I} < M_\mathrm{tr}\,, \\ M_\mathrm{BHmax} + \frac{(M_\mathrm{I}-2M_\mathrm{BHmax}+M_\mathrm{tr})^2}{4(M_\mathrm{tr}-M_\mathrm{BHmax})} & \text{if } M_\mathrm{tr} < M_\mathrm{I} < 2M_\mathrm{BHmax} - M_\mathrm{tr}\,, \\ 0 & \text{otherwise.} \end{cases} \tag{11.3}$$

Such a simple model will inevitably miss some of the complexity of the initial-final mass relationship. We introduce scatter in the remnant mass at fixed initial mass to

account for this missing physics. We simulate such uncertainty in the $M_{\rm I} - M_{\rm BH}$ mapping by treating the natural logarithm of the remnant mass as a realization from a Gaussian distribution, with standard deviation σ:

$$p\left(\ln(M_{\rm BH}) \mid \bar{M}_{\rm BH}, \sigma\right) = \mathcal{N}\left[\ln\left(\bar{M}_{\rm BH}(M_{\rm I})\right), \sigma\right](\ln(M_{\rm BH}))\ , \tag{11.4}$$

where $\bar{M}_{\rm BH}(M_{\rm I})$ is given in Eq. (11.3). Since Eq. (11.4) specifies that the logarithm of $M_{\rm BH}$ values are normally distributed around with standard deviation σ, the uncertainty on the physical value of the mass $M_{\rm BH}$ will grow with $M_{\rm I}$ for fixed σ.

Any confident measurement of a nonzero value of σ would mean there is variation in the $M_{\rm I} - M_{\rm BH}$ mapping. This could originate from any number of factors, e.g. physical properties affecting stellar evolution manifesting differently between black holes in the catalog. For example, since metallicity is expected to have a slight effect on the remnant mass given an initial CO core mass in models of the pair instability [210], resolvable contributions from sources with differing birth metallicities in our dataset would result in a preferentially nonzero value for σ.

To obtain the mass distribution for stellar-origin ("first generation") BHs dN/dM_{1G}, we integrate over progenitor masses,

$$\frac{dN}{dM_{\rm 1G}}\left(M_{\rm BH} \mid a, b, M_{\rm tr}, M_{{\rm BH},max}, \sigma\right) = \int dM_{\rm I} \frac{dN}{dM_{\rm I}}\, p(M_{\rm BH} \mid M_{\rm I}). \tag{11.5}$$

(Note the implicit Jacobian from the logarithmic mass appearing in Eq. (11.4).) The turnover in the $M_{\rm I} - M_{\rm BH}$ relation above $M_{\rm tr}$ leads to a pile-up of black hole masses around $M_{{\rm BH},max}$. This pile-up is usually expressed as a relative overdensity through a Gaussian bump [524, 29, 28]; in our model, the location, width, height, and asymmetry of the bump are "naturally" derived from the parameters $M_{\rm tr}$, $M_{\rm BHmax}$, and σ. The black hole mass functions in our model (see Figure 11.1) are similar to those discussed in Baxter et al. [87] (they would agree in the limit of $\sigma \to 0$), but those authors did not attempt to describe the black hole mass function in terms of the underlying physical processes relating initial to final mass as we do here.

This simple, general parameterized model is deliberately reminiscent of the relationship expected between CO core mass and remnant BH mass induced by the pair instability [239, 443]. From simulations with the stellar evolution code `MESA` [414], [210] find that $M_{\rm I}$ prior to core collapse is in fact the dominant variable determining the remnant mass post core collapse. Figure 4 in [210] shows the resulting $M_{\rm BH}$ vs $M_{\rm I}$ relationships obtained for a range of choices of input physics and metallicity. The authors note that for a given choice of metallicity, this relationship is well-modeled

by a piecewise map: a linear relationship, turning over to a quadratic at the CO core mass at which pulsations begin to remove notable mass, followed by a decay to $M_{\rm BH} = 0\,M_\odot$, corresponding to the mass at which pulsational pair instability fully disrupts the star, leaving no remnant. This general trend has been confirmed by other simulation-based studies [e.g., 374, 576].

When $M_{\rm I}$ reaches $M_{\rm tr}$, the $M_{\rm I}$ to $M_{\rm BH}$ mapping transitions from its linear relationship to a nonlinear one. In terms of the pair instability, at this point the pulsation process causes mass loss whose efficiency increases with the star's mass [365, 210, 576, 578]. The form of the quadratic function in Eq. (11.3) puts the peak value at $M_{\rm BH,\mathit{max}}$ and enforces that the transition and its derivative be continuous at $M_{\rm tr}$. This results in a BH mass distribution in which remnants between roughly $M_{\rm tr}$ and $M_{\rm BH,\mathit{max}}$ can map back to a wider range of progenitor masses, and each BH mass bin $dM_{\rm BH}$ in this range contains more systems than it otherwise would had the $M_{\rm I} - M_{\rm BH}$ relationship continued to be linear.

As shown in [210], the map from $M_{\rm I}$ to $M_{\rm BH}$ is sensitive to unknown physics affecting the core collapse and stellar evolution process, even given a fixed $M_{\rm tr}$. We therefore do not know with certainty of a one-to-one map of $M_{\rm I}$ to $M_{\rm BH}$; this is captured by our σ parameter. Even if we knew global physical parameters for the core collapse process (e.g., reaction rates) with certainty, a given $M_{\rm I}$ will always have a range of possible associated remnant masses due to factors such as the unknown metallicity at formation (see Sec. 11.3.4).

In Figure 11.1, we show how the distribution of $M_{\rm BH}$ is derived from an initial distribution of progenitor masses $M_{\rm I}$ according to our model with some fiducial values. Each BH in the 1G population is assumed to come from a progenitor from the $dN/dM_{\rm I}$ distribution (bottom panel), which is mapped to a remnant BH mass through the $M_{\rm I}-M_{\rm BH}$ relationship (upper right panel). Finally, the resulting BH mass distribution $dN/dM_{\rm BH}$ is obtained by integrating this distribution in the upper right panel across $M_{\rm I}$, weighted by $dN/dM_{\rm I}$ (Eq. 11.5). This differs from the procedure in [87], as we directly infer the $dN/dM_{\rm I}$ and $p(M_{\rm BH} \mid M_{\rm I})$ distributions, which uniquely specify dN/dM_{1G} rather than inferring a phenomenological representation of a resulting dN/dM_{1G} distribution.

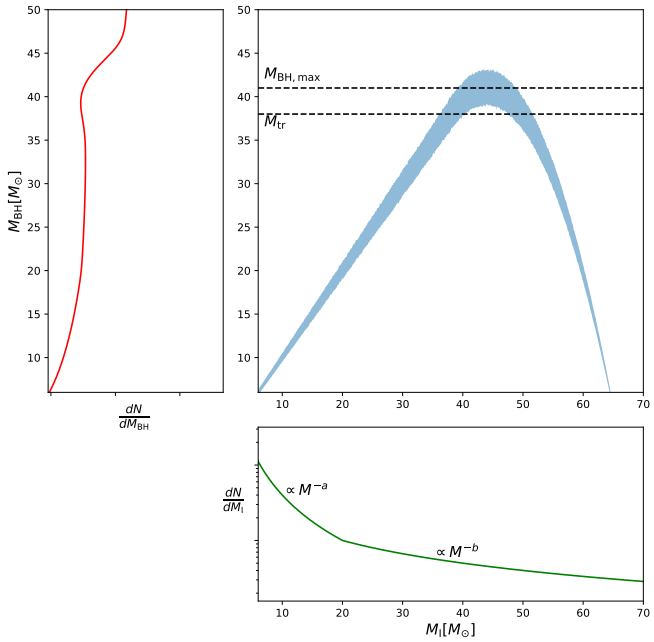

Figure 11.1: Obtaining a 1G BH mass distribution from the initial mass function. The progenitor IMF (bottom panel) gets transformed through the $M_{\rm I}$ to $M_{\rm BH}$ mapping (top right panel), resulting in a distribution for dN/dM_{1G} of BHs in 1G systems. We label parameters underlying the 1G BBH mass distribution: a and b are the low-mass and high-mass spectral indices of the progenitor IMF; $M_{\rm tr}$ and $M_{\rm BH,\mathit{max}}$ (dashed lines) control the onset of the nonlinearity and the maximum of the quadratic part of the mapping, respectively. The blue shaded region is the 90% credible region of the lognormal mapping for our choice of σ (see Eq 11.5 and preceding equations for the functional form). We vary these parameters in our fit to the LIGO-Virgo data, together with the parameters for the 2G population (see Section 11.3.2).

11.3.2 Full Mass Distribution

The model outlined in the previous section is only directly applicable to 1G mergers, in which $M_{\rm BH}$ is the remnant mass from the core collapse of one of the sources from $dN/dM_{\rm I}$. Realistically, a catalog of observed gravitational wave sources could also contain higher-generation merger events—namely, events that involve BHs who are themselves remnants of previous BH mergers, and therefore not of stellar origin—although it is commonly assumed that these systems will only subdominantly contribute to the inferred mass distribution [381, 467, 308, 249]. As the component masses of these events will be approximately (slightly less than) the sum of the masses of the BHs from its *previous* mergers, the masses in this population can exceed $M_{\rm BH,\mathit{max}}$ and will not follow the same distribution as the 1G BHs.

The details of the 2G distribution depend on unknown factors that make it difficult to prescribe a specific functional form [467, 308, 179]. In order to capture these events in a relatively agnostic manner, we enhance our model with a power law tail with a spectral index c that smoothly turns on just below $M_{\mathrm{BH},max}$, and has a height f_{pl} relative to dN/dM_{1G} at $M_{\mathrm{BH},max}$ (see the bottom right panel in Figure 11.2). We express the full mass distribution as:

$$\frac{dN}{dm} = \frac{dN}{dM_{1G}} + \delta(m \mid M_{\mathrm{BH},max})\, f_{pl}\, \frac{dN}{dM_{1G}}\bigg|_{M_{\mathrm{BH},max}} \left(\frac{m}{M_{\mathrm{BH},max}}\right)^{-c}, \tag{11.6}$$

where dN/dM_{1G} is given in Eq. (11.5) and $\delta(m)$ is an exponential tapering function that smoothly turns on to $M_{\mathrm{BH},max}$; the parameter f_{pl} controls the relative height between the peak of the 2G power law and the 1G mass distribution at $M_{\mathrm{BH},max}$. By adopting this two-component model, we can prevent 2G sources from biasing the inference of the parameters of the dN/dM_{1G} distribution, which has a sharp, log-normal falloff at masses $M_{\mathrm{BH}} > M_{\mathrm{BHmax}}$. This assumes that the 2G sources have a minimal contribution to the mass distribution below $\sim M_{\mathrm{BHmax}}$, consistent with the conclusions from, e.g., [230].

We model both component masses as coming from the same mass distribution dN/dM_{BH} and include a pairing function with power law slope β to get the full mass distribution:

$$\frac{dN}{dm_1 dm_2}(m_1, m_2) \propto (m_1 + m_2)^{\beta}\, \frac{dN}{dm_1}\frac{dN}{dm_2}, \tag{11.7}$$

where each $dN/dm_{1/2}$ factor corresponds to a density as in Eq. (11.6). The first factor in Eq. (11.7) constitutes the pairing function, by which the component masses do not only inform the mass distribution independently but also by how they pair together to form a total mass [225, 209]. The parameter β is the exponent on the total mass, such that positive (negative) values for β mean that masses pair up to preferentially form systems of higher (lower) total mass. We choose this form of the pairing function, first suggested in Fishbach and Holz [225], to permit the possibility of *breaking* factorization symmetry, so that when $\beta \neq 0$ the joint mass function is not the product of a function of m_1 and a function of m_2; many of the models highlighted in [28] are forced to be symmetric in this sense.

Figure 11.2 shows how the mass distribution dN/dm changes as a function of various population hyperparameters. The top left panel shows how M_{tr} predictably controls onset of the transition to a peak in the mass distribution by changing where the $M_{\mathrm{I}} - M_{\mathrm{BH}}$ mapping becomes nonlinear; additionally, as M_{tr} becomes closer to

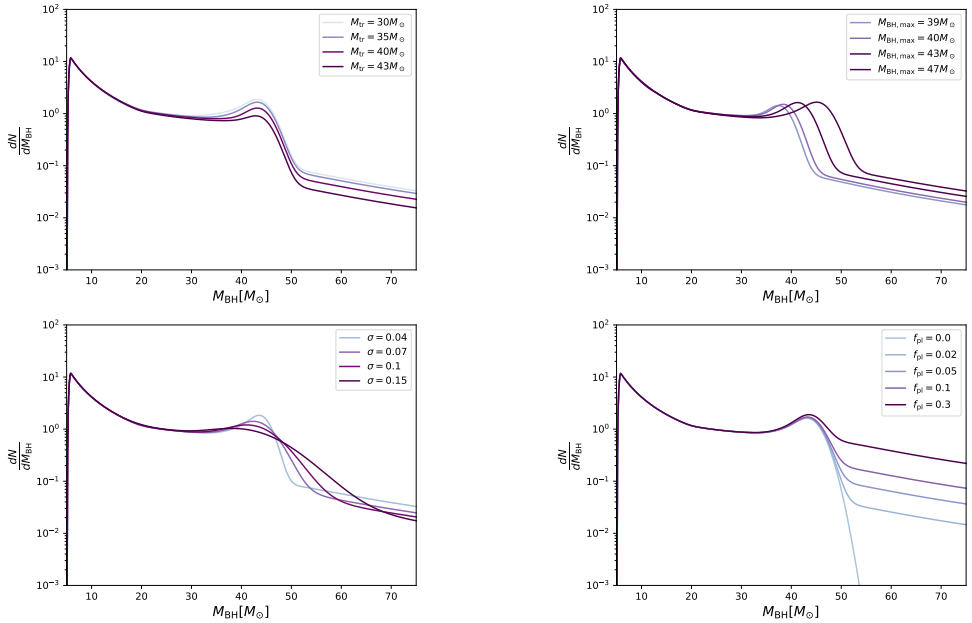

Figure 11.2: BH mass spectrum following our model in Eq. (11.7) for different choices of (clockwise from top left) $M_{\rm tr}$, $M_{{\rm BH},max}$, $f_{\rm pl}$, and σ. For the top left figure, we consider constant difference between $M_{\rm tr}$ and $M_{{\rm BH},max}$. Unless being varied, we assume the following fiducial parameters: $a = 2, b = 1, c = 2.5, M_{\rm tr} = 35M_\odot, M_{{\rm BH},max} = 45M_\odot, \sigma = 0.05$, and $f_{pl} = 0.04$.

$M_{\rm BHmax}$, the height of the peak decreases as a smaller range of the IMF is contributing to the peak region. The upper right panel shows how this peak moves to higher $M_{\rm BH}$ and gets wider as $M_{{\rm BH},max}$ increases for fixed $M_{\rm tr}$. The bottom left panel of Figure 11.2 shows the effect of varying σ on the resulting mass distribution. As σ increases, the remnant masses for a given core mass scatter more broadly, smoothing the remnant mass distribution and softening the peak, as well as making the cutoff above $M_{\rm BHmax}$ weaker. Together, these physical parameters govern the location, strength, and width of the peak in the BH mass distribution, as well as the strength of its cutoff. The final panel of Figure 11.2 demonstrates the increasing contribution of the high-mass power law tail when raising f_{pl}.

11.3.3 Redshift Model

Studies of cosmic star formation history with astronomical surveys show that the star formation rate increases to a redshift of $z \approx 2$, then smoothly decays at high redshifts, which is well-modeled by a smoothly broken power law [351, 547, 251]. When convolved with a reasonable delay-time distribution, this also gives rise to

a smoothly broken power law for the merger rate $R(z)$ [228], i.e., the number of mergers per comoving volume (V_c) per time interval in the source frame (t_s). Accordingly, we assume a redshift distribution such that

$$\frac{dN}{dV_c dt_s}(z) \equiv R(z) \propto \frac{(1+z)^\lambda}{1+\left(\frac{1+z}{1+z_{\rm peak}}\right)^\kappa}\,, \tag{11.8}$$

where, λ controls the low-redshift merger rate, estimated to be $\lambda \approx 3$ in current LIGO/Virgo studies [28]; meanwhile, the parameter $z_{\rm peak}$ controls the redshift at which the merger rate peaks and the slope becomes negative.

Both the peak redshift and the high-redshift merger rate are expected to be directly informed from detections beyond the horizon of current ground-based detectors or at redshifts where detections are scarce [125, 559, 28]. However, combining upper-limits from stochastic gravitational wave searches with population inference studies can place limits on these parameters with current detections [123]. Future observations with 3G detectors will allow us to significantly constrain the merger rate history across cosmic time using direct detections of BBH mergers at nearly all relevant redshifts [391, 355].

The merger rate as a function of redshift in the detector frame is expressed simply in terms of mergers per redshift z per detector-frame time $t_{\rm det}$:

$$\frac{dN}{dz dt_{\rm det}}(z) = \frac{dN}{dV_c\, dt_s}\frac{dV_c}{dz}\frac{dt_s}{dt_{\rm det}} = R(z)\frac{dV_c}{dz}\frac{1}{1+z}\,, \tag{11.9}$$

where $R(z)$ is as in Eq. (11.8), and dV_c/dz is the differential comoving volume per redshift bin as determined by cosmology.

11.3.4 Allowing the Mass Spectrum to Evolve with Redshift

Studies of stellar evolution predict that stars formed in lower-metallicity environments can reach higher remnant BH masses before hitting the PISN cutoff. This is generally attributed to the ability for metal-rich stellar winds to carry off significant mass, resulting in lower remnant BH masses after undergoing pulsations [504, 365, 210]. While we do not get direct information about the progenitor metallicities of gravitational-wave sources from the observed data, we can use known correlations between metallicity and observables in gravitational-wave data to look for this evolution. Redshift and metallicity are anticorrelated: stars formed earlier in the universe (i.e., at higher redshift) are metal-poor when compared to those formed more recently (at lower redshift), due to the need for the existence of pre-existing stars to

deposit metals into the interstellar medium in order to birth further generations of stars with higher metallicities [353, 90].

Several previous studies have used this trend as motivation to search for redshift-dependence in the observed BBH mass distribution.[1] These studies have typically adopted phenomenological approaches to modeling this effect, directly encoding redshift dependence in the location of features in the BBH mass distribution. Such features include the location of the Gaussian peak and the truncation point of the mass distribution, allowing these features to vary, for example, linearly with redshift or with some function of expected metallicity at a particular redshift [480, 228].

We can leverage the physical framework we introduced in Sec. 11.3.2 to model the redshift dependence in the mapping from $M_{\rm I}$ to $M_{\rm BH}$. This allows us to treat the redshift evolution in the observed BBH mass distribution as a derived byproduct from an astrophysical process expected to evolve with redshift, rather than encoding the redshift dependence in the BBH mass distribution directly.

We express this evolution in term of a linear expansion for the location of the $M_{\rm tr}$ turnover in the mass distribution:

$$M_{\rm tr}(z) = M_{\rm tr}(z=0) + \dot{M}_{\rm tr}\left(1 - \frac{1}{1+z}\right), \tag{11.10}$$

where $M_{\rm tr}$ and $\dot{M}_{\rm tr}$ are free parameters which we can interpret as the transition location at $z = 0$ and the change in this location over a Hubble time, respectively. In order to maintain the constraint that $M_{\rm BHmax} > M_{\rm tr}$, we apply an equivalent adjustment to $M_{\rm BHmax}$ to maintain

$$M_{\rm BHmax}(z) - M_{\rm tr}(z) = \text{const} \tag{11.11}$$

at all redshifts. This is an indirect model of the evolution of this feature from high metallicity (late universe, $z = 0$) to low-metallicity (early universe, $z = \infty$) environments. One could alternatively construct a more explicit model for $M_{\rm tr}$ and $M_{\rm BHmax}$ as a function of metallicity and then metallicity as a function of redshift.

Figure 11.3 shows how our model for dN/dM_{1G} appears for different values of redshift for two choices of $\dot{M}_{\rm tr}$. To be consistent with predictions from stellar models, as described above, we expect a positive value for $\dot{M}_{\rm tr}$ such that the turnover

[1] [504] proposes that differing delay time distributions between the high and low mass portions of the mass distribution may also result in an evolving mass distribution. Unlike evolution due to birth metalicity, this trend would not be tracked by the evolution in our model.

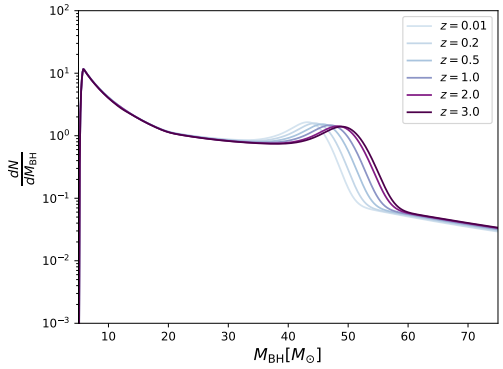

Figure 11.3: Model for a redshift-dependent mass distribution evaluated at selected redshifts assuming $\dot{M}_{\rm PISN} = 5\,M_\odot$ in Eq. (11.10). All other parameters are the same as the fiducial set in Figure 11.2.

to PPISN occurs at higher masses at higher redshifts; equivalently, this means we expect the bump in the mass distribution moves toward higher mass at higher redshifts.

11.4 Results

11.4.1 Non-Evolving Mass Distribution

Adopting the non-evolving mass and redshift models introduced in the previous section, we infer the corresponding hyperparameters using the hierarchical Bayesian inference approach described in Sec. 11.2 and the priors in Table 11.4.1.

We plot draws from the mass distribution posterior in Figure 11.4, which shows the inferred decaying power law shape of the mass distribution and the feature at $\sim 35 M_\odot$. Turning attention to the parameters that most directly control the location and strength of the peak in the mass distribution, Figure 11.5 shows the posterior distributions for σ, $M_{\rm tr}$, and $M_{{\rm BH},max}$. The recovered distribution for σ shows notable preference for low values, converging toward the lower bound of the prior ($\sigma = 0.05$). This means that the data are consistent with little to no scatter around the $M_{\rm I} - M_{\rm BH}$ mapping, while ruling out high values of σ that would over-smoothen the peak in the mass distribution (cf. Figure 11.2, bottom panel). By the same token, the strong support for low values of σ indicates that the data do allow for a relatively sharp cutoff in the peak; this is such that a suppression of the high end of the peak

Parameter	Prior	Description
a	$\mathcal{N}(2.35, 2)[-1.65, 6.35]$	Power law index of low-mass CO IMF
b	$\mathcal{N}(1.9, 2)[-2.1, 5.9]$	Power law index of high-mass CO IMF
c	$\mathcal{N}(4, 2)[0, 8]$	Power law index of 2G high-mass tail
$M_{\rm tr}$	$\mathcal{N}(35, 5)[20, 50]$	Linear-to-quadratic transition mass $[M_\odot]$
$M_{\rm BH,max} - M_{\rm tr}$	$\mathcal{N}(3, 2)[0.5, 7]$	Maximum remnant mass produced by the 1G channel, relative to $M_{\rm tr}$ $[M_\odot]$
σ	$\mathcal{N}(0.1, 0.1)[0.05,]$	Width of lognormal distribution for $M_{\rm I}$ to $M_{\rm BH}$ mapping
β	$\mathcal{N}(0, 2)$	Exponent on total mass pairing function
$\log(f_{pl})$	$\mathcal{U}[0.01, 0.5]$	Log of relative height between the start of the 2G powerlaw and end of dN/dM_{1G}
λ	$\mathcal{N}(2.7, 2)[-1.3, 6.7]$	Exponent controlling $R(z)$ at low redshift
$\kappa - \lambda$	$\mathcal{N}(2.9, 2)[1, 6.9]$	Exponent controlling $R(z)$ at high redshift
$z_{\rm peak}$	$\mathcal{N}(1.9, 1)[0, 3.9]$	Redshift at peak $R(z)$
$\dot{M}_{\rm tr}$	$\mathcal{U}[-2, 8]$	Difference in $M_{\rm tr}(z)$ over Hubble time $[M_\odot]$

Table 11.1: Priors used in this work. $\mathcal{U}$ is a uniform distribution and $\mathcal{N}$ is a Gaussian distribution with mean and standard deviation specified in the parentheses. Numbers in square brackets are upper and lower bounds of the prior.

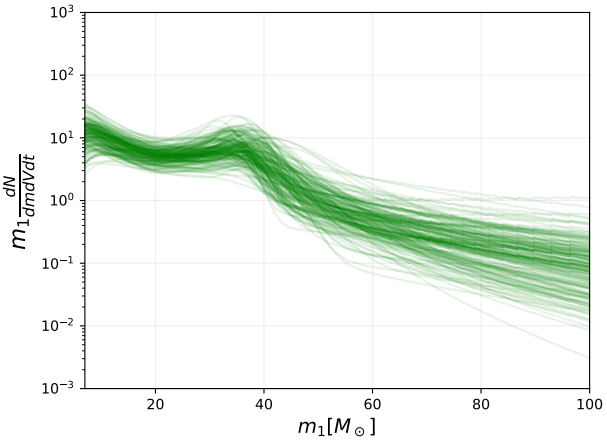

Figure 11.4: Draws from the non-evolving mass distribution posterior, evaluated at $z = 0$.

need not be compensated by a higher rate in the start of the 2G tail.

With the data supporting a peak at $\sim 35 M_\odot$, we measure a correlation between $M_{\rm tr}$ and $M_{{\rm BH},max}$, driven by the constraint that $M_{{\rm BH},max} > M_{\rm tr}$. We note that the posteriors for $M_{{\rm BH},max}$ and $M_{\rm tr}$ are different from the prior, indicating that the data are informing both the location and the width of the bump.

Taking the model at face value, we infer the location of $M_{{\rm BH},max}$ to be much lower than where stellar nucleosynthesis simulations generally predict the upper mass gap due to the PISN process to begin; a similar conclusion was reached in Farmer et al. [211]. For example, [210, 211] finds the lower edge of the PISN mass gap to range between $\sim 45-50\, M_\odot$, when varying the CO reaction rate within its 1σ uncertainty (with standard deviation σ_{C12}) with respect to the distribution of reaction rates given in `STARLIB` [484]. Using their fit to the start of the mass gap as a function of σ_{C12}, and extrapolating down to our inferred values of $M_{{\rm BH},max}$, we infer $\sigma_{C12} = 4.8^{+3.1}_{-2.8}$ at 90% credible levels.[2] In other words, to match the location of our observed peak would require a $\sim 5\sigma$ adjustment in the C12 reaction rate relative to its current nuclear-physics best-estimate and uncertainty. Further discussion of this point can be found in Section 11.5.3.

Although the simulation coverage is sparse at these masses and therefore these constraints are largely extrapolation-driven, the anomalous value inferred for this

[2]For reaction rates this high, the fraction of carbon in the core is too low ($X_{\rm C} \ll 10^{-3}$) to be considered a CO core. To be consistent with the assumptions stated in [210], $M_{\rm I}$ can instead be interpreted as the mass within the convective zone during helium burning.

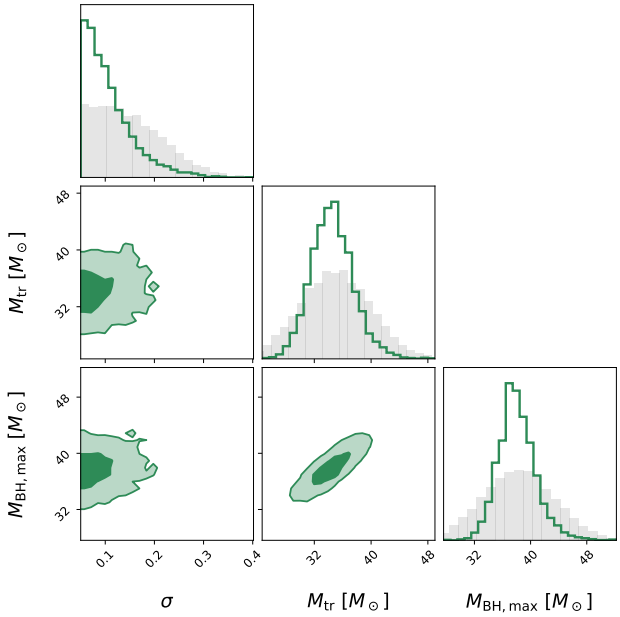

Figure 11.5: Posterior for selected mass distribution parameters using the model outlined in Sec. 11.3.2. Dark- and light-green shaded regions are the 1σ and 2σ contours, enclosing 39% and 86% of the probability respectively. Prior distribution is shaded grey for reference. We find that widening the prior does not meaningfully increase the posterior support for the PPISN feature at higher masses.

parameter casts doubt on this PPISN model as a an explanation for the ~35 $M_\odot$ peak in question. See Sec. 11.5 for further discussion.

In Figure 11.6 we plot the inferred inferred $M_{\rm I} - M_{\rm BH}$ mapping, the $M_{\rm I}$ IMF, and 1G $M_{\rm BH}$ mass distribution. We find that the $M_{\rm I}$ IMF steeply decreases before the break at $M_{\rm I} = 20M_\odot$ and then becomes shallower or flattens out; observations of massive stars in star forming regions suggest that the high-mass IMF could be shallower than at lower masses [491]. The $M_{\rm I} - M_{\rm BH}$ mapping is somewhat uncertain, but the turnover reliably creates a peak at $\sim 35M_\odot$.

11.4.2 Evolving Mass Distribution

By adopting the more general model from Sec. 11.3.4 we can relax some of the assumptions made in the previous sections and now infer the mass distribution in the presence of an $M_{\rm tr}$ that evolves with redshift. In Figure 11.7, we present the posterior probability density on several mass distribution parameters from this model. Most of the events in the O3 catalog lie at relatively low redshift, and therefore do not provide good coverage across redshift scales to inform $\dot{M}_{\rm tr}$ meaningfully. Due to

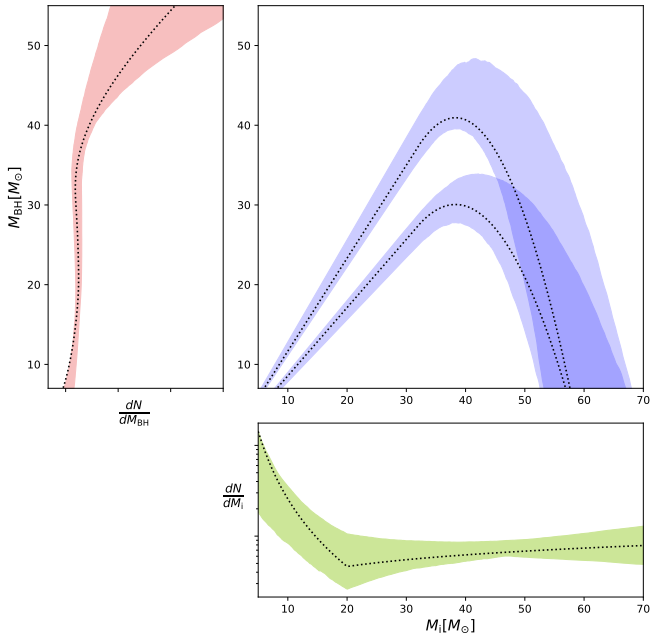

Figure 11.6: Representation of the $M_{\rm I} - M_{\rm BH}$ relationship of the non-evolving $\frac{dN}{dM_{1G}}$ model in Sec. 11.3.2. (Bottom panel) Inferred distribution for the initial mass function of $M_{\rm I}$ in merging binaries. (Top right panel) 1-σ credible region of the 95th percentile (higher blue shaded region) and 5th percentile (lower blue shaded region) for the inferred $M_{\rm I} - M_{\rm BH}$ mapping. Dotted line is a single representative draw from the posterior.

these poor constraints on $\dot{M}_{\rm tr}$, we find that the inferred distributions for $M_{\rm tr}$ and $M_{{\rm BH},max}$ are consistent with those obtained when using the non-evolving model. In other words, the feature at $M_{\rm tr}$ is being informed by structure in the data that does not appear to need to vary with redshift. The resulting mass distribution is consistent with that from the non-evolving model (plotted below in Figure 11.10).

When we extend the prior on $\dot{M}_{\rm tr}$ considerably, we find that we only rule out redshift evolution of the intrinsic mass function for very large values of $\dot{M}_{\rm tr}$. In Figure 11.8, we show the posterior distribution for relevant parameters: constraints on $\dot{M}_{\rm tr}$ are broad, encapsulating a 90%-credible region from $\dot{M}_{\rm tr}$ from $-20\,M_\odot$ to $36\,M_\odot$. This range is much broader than would be expected from the metallicity dependence of the onset of pair instability pulsations [210]. While this demonstrates that data rule out extreme values of $\dot{M}_{\rm tr}$, we cannot currently place constraints within a narrower, more physically-relevant prior range; future observations may change this. The strong anticorrelation between $\dot{M}_{\rm tr}$ and $M_{\rm tr}$ likely indicates that we are observing the peak from sources in a small range of redshifts, and we therefore cannot constrain

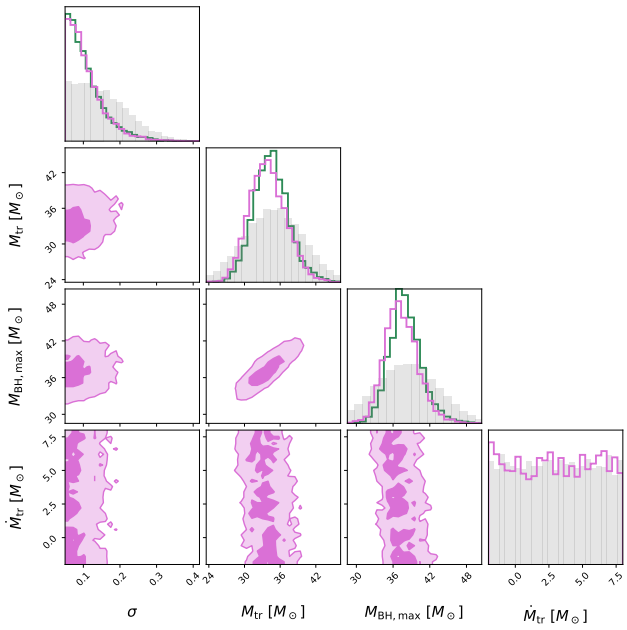

Figure 11.7: Posterior for selected mass distribution population parameters from the evolving mass distribution model in Sec. 11.3.4 (magenta). One-dimensional posteriors from the non-evolving mass model overplotted in green for reference.

both free parameters in Eq. (11.10). In order to break this degeneracy and get constraints on $\dot{M}_{\rm tr}$, we would need additional observations across redshifts. For the remainder of this section, we present results using the narrower prior for $\dot{M}_{\rm tr}$ (in Table 11.4.1).

Also of note is the similarity between the distribution of σ obtained with this model and that obtained with the non-evolving mass model (see the comparison in Figure 11.7). We discuss the implications of this in Sec. 11.5.

We present the distribution for redshift parameters, inferred jointly with the evolving mass distribution, in Figure 11.9. The parameter best constrained is λ, which controls the evolution of the low-redshift merger rate. We infer $\lambda = 4.3^{+1.6}_{-1.6}$, preferring a merger rate that evolves *steeper* than the low-redshift star formation rate ($\lambda \sim 2.7$). However, the evolution of the merger rate is still consistent with that implied by the star formation rate along with a short delay-time distribution. Narrower constraints on this parameter may reveal information on different formation channels contributing to the observed catalog of BBHs.

Additionally, the posterior distribution for $z_{\rm peak}$ is shifted slightly to the right of the prior, meaning that we are able to begin to place very conservative lower limits on

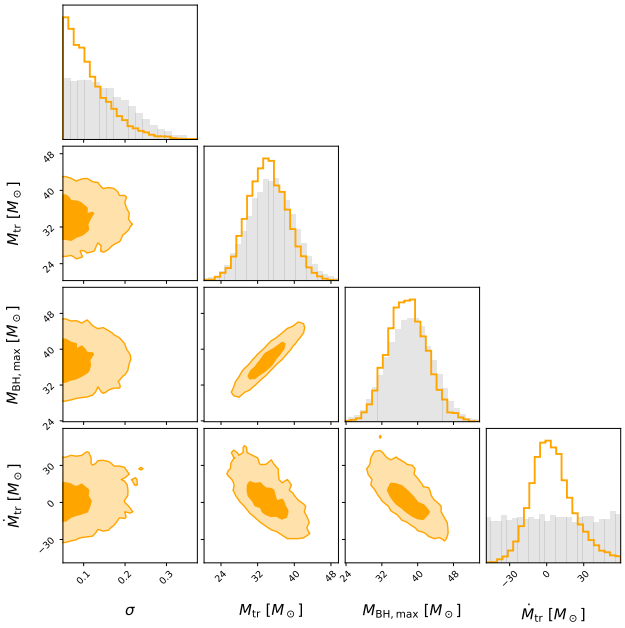

Figure 11.8: Inferred mass distribution parameters using the evolving mass model in Sec. 11.3.4, but adopting a wider prior on $\dot{M}_{\rm tr}$. Prior distribution is shaded grey for reference.

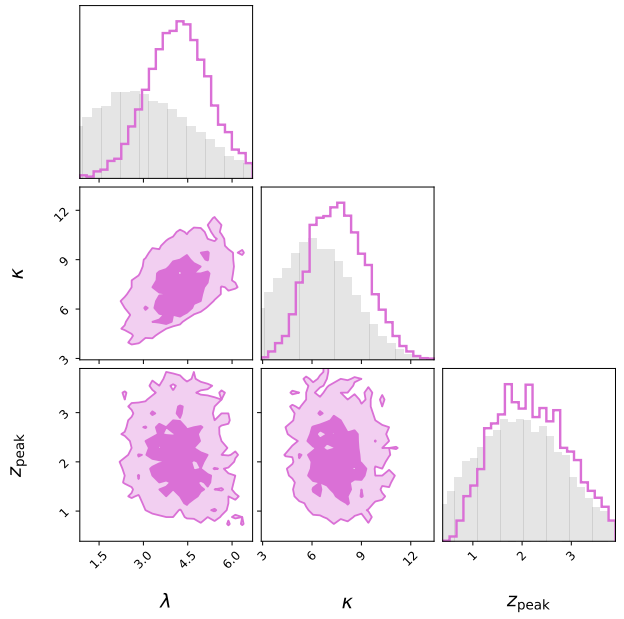

Figure 11.9: Redshift distribution parameters inferred with the evolving mass distribution model in Sec. 11.3.4. Prior distribution is shaded grey for reference.

$z_{\rm peak}$ due to the lack of a visible start of a turnover in the inferred $R(z)$ distribution. The lack of support at the tails towards higher $z_{\rm peak}$ is not due to information gained from the data, but rather from the prior (see Table 11.4.1). These constraints appear despite having only very little high-redshift information and are therefore very weak limits.

Given the similarities between the inferred distributions with and without redshift evolution in the $M_{\rm I}$ to $M_{{\rm BH},max}$ map, we do not currently obtain improved constraints on physical parameters of interest when modifying the model in this way. For example, we infer $\sigma_{C12} = 5.2^{+3.4}_{-2.9}$, which is similar to what we reported in Sec. 11.4.1. Future detections at higher redshift may further inform these aspects of the model.

11.5 Interpretation of Results

By adopting the model introduced in this work, we can draw conclusions from the inferred underlying physics represented in our models and explore how the population model compares to those reported in other works.

11.5.1 Global Shape of the Mass Distribution

For comparison, we obtain results using the same set of O3 events, adopting the Powerlaw + Peak mass distribution model, a flat spin magnitude and tilt model, and a broken power law redshift distribution as implemented in `gwpopulation` [520, 524, 228]. Qualitatively, we infer a mass distribution (marignalized over q) consistent with the Powerlaw + Peak model, with major features such as the slope at higher BH mass as well as the bump location showing good agreement in Figure 11.10. This indicates that this overdensity is a confident feature in the data whose location and prominence is not affected by systematic differences between these two models. This is reinforced by several other works, which find that models must include such a feature in order to faithfully capture the observed mass spectrum [28, 207, 186, 125].

A notable difference is in the merger rate relative to Powerlaw + Peak. While the 90% credible regions overlap in Figure 11.10, their relative heights show that our model tends to prefer a higher merger rate than what is predicted by Powerlaw + Peak, particularly at masses above $\sim 15 M_\odot$.

In our model, m_1 and m_2 both directly inform the physical mass distribution model (along with the pairing function, see Eq. 11.7). This is in contrast to the Power-

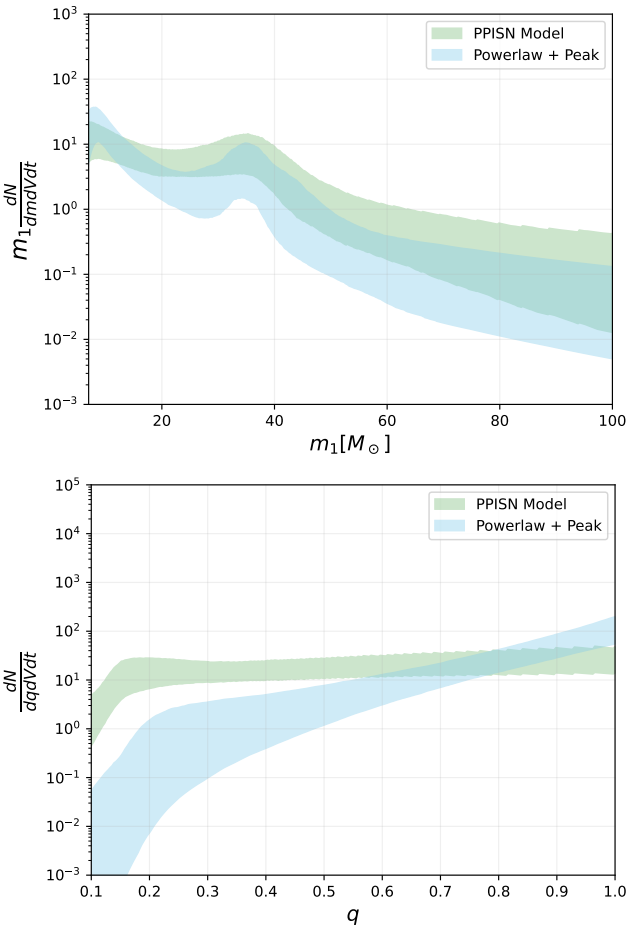

Figure 11.10: Mass spectrum derived from our model (green) with no evolution of the mass distribution with redshift compared to the POWERLAW + PEAK mass spectrum informed by the same events (blue). We do not include redshift evolution of the mass distribution in this comparison as the models in [28] do not include mass-redshift correlations. POWERLAW + PEAK results were obtained using `GWPopulation` [520]. (Top) Comparison of the primary mass distributions. (Bottom) Comparison of the mass ratio distributions.

LAW+PEAK model which has separate distributions for $p(m_1)$ and $p(q|m_1)$, such that m_2 does not directly inform dN/dm_1. As demonstrated in [209], this makes the mass distribution feature a peak in the joint $m_1 - m_2$ space, rather than in the marginal m_1 distribution. Furthermore, in the marginal mass ratio distribution (bottom panel of Figure 11.10), we see that our model prefers a much flatter distribution in q than what is preferred by POWERLAW + PEAK, which explicitly models $p(q|m_1)$ as a power law. This flat mass ratio distribution is consistent with what is found in [225] when adopting a pairing function that is a power law in q. Given that we infer β with a preference for negative values, we find that BHs tend to pair up in binaries that favor lower total masses; this may cause a relative lack of support at higher masses in the mass distribution which models both dN/dm_1 and dN/dm_2. We also note that the inferred local rate $R(z = 0)$ is consistent between models. The POWERLAW + PEAK model fits the underlying distribution (i.e., not including the bump) with a single power law, limiting the possible morphologies. We have checked that allowing the underlying power law in POWERLAW + PEAK to include a break does not resolve the discrepancy.

The distribution for $p(M_{\rm I})$ we infer (see Figure 11.6) disagrees with what one may expect from an IMF resulting from the ZAMS mass IMF assuming a linear relationship between ZAMS mass and $M_{\rm I}$. While the distribution is consistent with a decaying power law for low masses, the distribution appears to flatten out above our break point of $20\,M_\odot$. This trend is not strongly correlated with the $M_{\rm tr}$ and $M_{\rm BH,\mathit{max}}$ we infer.

Comparing our results to those obtained in [87], we find strong tension with the maximum BH mass in the 1G channel (the start of the upper mass gap). Motivated by stellar evolution simulations to model the 1G BH mass distribution with a phenomenological approximation to the shape and location of a overdensity due to PPISN pileup, [87] finds the PPISN feature and corresponding start of the upper mass gap to be at $\sim 46 M_\odot$, in very good agreement with predictions from typical values of the $^{12}\mathrm{C}(\alpha,\gamma)^{16}\mathrm{O}$ reaction rate. Notably, [87] does not find the feature at $\sim 35\,M_\odot$ we find and is consistently found in the literature.

11.5.2 Evolution of Mass Distribution with Redshift

Our finding that $\dot{M}_{\rm tr}$ is consistent with zero agrees with other studies that do not find strong preference for evolution of the BBH mass distribution with redshift. For example, [223] models the mass distribution as a broken power law where the mass

at which the power law breaks is allowed to vary with redshift. While this is a very different model, it should qualitatively reproduce some of the features of our model, particularly at the $\sim 35\,M_\odot$ feature (see Figure 11.2). We therefore expect that if [223] had found strong preferences for an evolving mass distribution, we would confidently find $\dot{M}_{\rm tr} > 0$. We also agree that the data are still consistent with a mass distribution that has some evolution with redshift, but again we do not have positive evidence that this is the preferred scenario.

[301] also looks for evolution of the mass distribution with redshift. The authors allow the Gaussian bump in a Powerlaw + Peak-like model to vary with redshift, where the placement of this peak at a given redshift is determined by the delay time distribution and a jointly-inferred model for the evolution of (birth) metallicity with redshift. The value they find for the lower edge of the upper mass gap of $\sim 44 M_\odot$ is nominally in better agreement with the prediction from stellar physics models. However, this value is cited at low metallicity, and they also find there must be a very strong evolution of this mass scale with metallicity. Extrapolating their results to the local universe, they find that the upper mass gap at $z = 0$ starts at $\sim 30\,M_\odot$, which is closer to the corresponding value we obtain for the start of the $M_{\rm I}$–$M_{\rm BH}$ turnover. This result seems in tension with theoretical predictions given how small of an effect metallicity is expected to have on $M_{\rm tr}$. There are unexplained differences in our results, however, as such a strong evolution of $M_{\rm tr}$ with metallicity should mean that we would infer a positive $\dot{M}_{\rm tr}$, assuming delays do not mix events from many different birth metallicities into similar merger times.

If metallicity evolution effects were causing some of the support for nonzero values of σ obtained in the non-evolving mass model (i.e., from scatter in the $M_{\rm I} - M_{\rm BH}$ relation), we would expect σ to be constrained closer to zero with the evolving model, as some of that scatter would have been absorbed by the redshift evolution. Given that this is not the case, we conclude that either (1) birth metallicity effects fundamentally have a subdominant impact on the $M_{\rm I}$–$M_{\rm BH}$ relationship compared to other physical parameters that vary between BBH systems, or that (2) the birth metallicities of the systems in our catalog are not strongly correlated with the redshifts at which they merge. The latter scenario could result from the delay time distribution between formation and merger having enough support in the long-delay tails such that we cannot yet discern a strong correlation between birth time and merger redshift for systems merging at redshifts of $z \lesssim 1$.

11.5.3 Physical Interpretations: PPISN Process

We can take advantage of the physical parameterization of our model to interpret the implied stellar physics, assuming $M_{\mathrm{BH},max}$ corresponds to the maximum 1G BH mass as determined by the PPISN process at a given redshift or metallicity.

Under the PPISN model, the pileup in BH masses around $\sim 35 M_{\odot}$ would result from the remnants of stars with zero-age main-sequence (ZAMS) masses between $\sim 60 M_{\odot}$ through $\sim 140 M_{\odot}$ [442, 578, 576], driven by nuclear processes in the core [239, 443] For stars with ZAMS masses above this range, similar processes completely disrupt the star, leaving behind no remnant. Since the PPISN process produces a small range of remnant BH masses from stars from a wide range of ZAMS masses, it is expected that the mass distribution will exhibit the bump due to this pileup (sometimes referred to as the "PPISN graveyard") followed by a suppression of sources, known as the upper mass gap [576, 577, 578, 210].

Simulations of stellar evolution [e.g., 210, 374, 206] have explored the relationship between initial stellar mass (in particular the mass of the Carbon-Oxygen (CO) core, M_{I}) and the final BH mass (M_{BH}) after core collapse. They have also quantified the dependence of the location of the lower edge of this mass gap and its associated mass range on other physical parameters such as nuclear reaction rates, metalicity, and details of neutrino physics. Previous studies have used this relationship to place constraints on the astrophysical properties of the pulsational pair instability process, assuming the most massive sources observed through LIGO are below the upper mass gap [211, 374, 206, 511]. [87] instead infers the population of BHs coming from the first-generation (1G) subpopulation below the upper mass gap along with the subpopulation of higher-generation (2G+) BHs (i.e., BHs that are themselves the product of past mergers) whose masses can lie within the upper mass gap.

Based on simulations [211, 374, 206], the $^{12}\mathrm{C}(\alpha,\gamma)^{16}\mathrm{O}$ reaction rate is likely to be the dominant physical factor controlling $M_{\mathrm{BH,max}}$. Under this assumption, Figure 11.11 shows the fit for $M_{\mathrm{BH},max}$ as a function of σ_{C12} (i.e., the number of standard deviations from the median reaction rate in [484], in turn, adopted from [321]), reproduced from the data release in Farmer et al. [211]. While their simulations only cover the range $-3 < \sigma_{C12} < 3$, there is a clear trend that σ_{C12} must rise very steeply to reach a maximum BH mass below $\sim 45\, M_{\odot}$. We offer details of this trend in Appendix 11.C.

After translating our inferred $M_{\mathrm{BH},max}$ into σ_{C12} via the top of Figure 11.11, we use

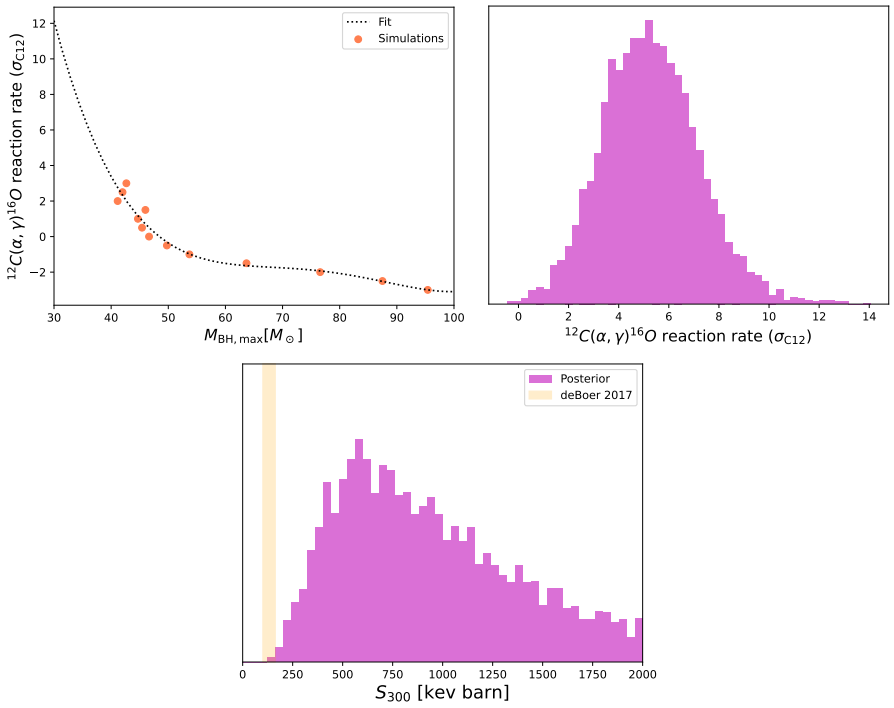

Figure 11.11: (Top) Relationship between σ_{C12} and lower edge of upper mass gap, reproduced from data release of [211]. (Center) Posterior distribution of $^{12}\mathrm{C}(\alpha,\gamma)^{16}\mathrm{O}$ reaction rate, in terms of standard deviations away from the median reaction rate in `STARLIB` [484]; computed by evaluating the fit in the top panel for the samples of $M_{\mathrm{BH},max}$ in the posterior in Fig 11.7. (Bottom) Inferred distribution of S_{300}, extrapolated from distribution of σ_{C12} as calculated from the fit in the top panel. Constraints on S_{300} from [172] plotted for comparison, showing tension with the values implied from our results.

the method in [211] to estimate corresponding S-factors.[3] We arrive at a value of the S-factor at 300 keV of $S_{300} = 932^{+1929}_{-581}$ keV $\cdot$ barn.

In Figure 11.11, we compare this estimate of S_{300} to the value predicted from nuclear experiments in [172]. That value is in strong tension with our estimate, which rules it out at $> 99.9\%$ credibility.[4] Given this, we conclude that at least one of our assumptions is invalid.

One assumption is that the $^{12}\mathrm{C}(\alpha,\gamma)^{16}\mathrm{O}$ reaction rate is the physical parameter behind $M_{\mathrm{BH},max}$. Multiple studies find that varying the $^{12}\mathrm{C}(\alpha,\gamma)^{16}\mathrm{O}$ reaction rate

[3]This "astrophysical S-factor" is the part of the cross section given by the matrix element for the nuclear reaction itself, ignoring Coulomb repulsion [see 311].

[4]This is in contrast with [206] and [211], which infer S_{300} to be consistent with [172]. However, those studies only consider the most massive observed BHs and assume they are 1G BBH mergers, without allowing for contamination from possible 2G mergers.

has a much stronger effect on the location of the start of the upper mass gap than other relevant reaction rates [211, 206]. As these simulations only go up to $\sigma_{C12} = 3$, we cannot confirm the effect that varying other reactions rates when σ_{C12} is high has on the location of the lower edge of the upper mass gap. While it is possible that one of these other reaction rates can be varied within their uncertainties to allow us to infer a lower σ_{C12}, it would have to change the location of the upper mass gap substantially to agree with the data [210, 206, 578].

Other assumptions inherent in present models of the PPISN process that could affect the maximum BH mass include the treatment of convection [458] and the hydrodynamic treatment of mass ejection from pair instability pulses [457].

Yet another, more fundamental, assumption in this interpretation is that the turnover in the $M_{\rm I} - M_{\rm BH}$ mapping can be associated with the pair-instability process at all.

In fact, our measurement in Figure 11.11 suggests that the observed bump in the mass distribution is not due to the PPISN turnover.[5] Previous studies have suggested that associating the observed peak in the mass distribution with the PPISN pileup would be in tension with known stellar physics and observed supernovae rates [274, 578]. Inferring the underlying $M_{\rm I} - M_{\rm BH}$ mapping that gives rise to the observed BH mass distribution, this work also provides evidence of such a tension in terms of the underlying physics that would be necessary to generate a turnover in the $M_{\rm I} - M_{\rm BH}$ map at the correct location. Our model allows us to directly infer this tension from the GW data.

The cause of the peak in the observed mass distribution at $\sim 35\,M_\odot$ may therefore requires alternative explanations [274]. Recent studies have proposed that this overdensity could be a signature from a subpopulation of binaries which had undergone stable mass transfer [113], BBH systems in globular clusters [64], and stars which have undergone significant wind-driven mass loss. Our model could be used to describe any mechanism that generates a peak in the high-mass tail of the 1G mass distribution via a transition to a nonlinear $M_{\rm I} - M_{\rm BH}$ relationship.

11.5.4 Model Limitations

Our model has a few additional caveats. For example, our model does not attempt to fit for the features beyond a power law that we know exist at lower masses [186,

[5] Mass-loss prescriptions and temporal resolutions of the simulations may introduce an unknown systematic bias [374, 206, 210]; however varying these settings has a subdominant impact on the location of the upper mass gap in simulations.

28, 207]. We have confirmed that neglecting this does not bias the inference in the higher-mass region that we care about here, and this will be explored further in future work. We also ignore the effect of spins in our population, but we demonstrate in Appendix 11.B that this does not cause a notable bias in our results of interest. Given that certain mass-spin correlations have been found in the BBH population [126], it may be insightful to use the spins to help distinguish the 1G and 2G subpopulations (see, e.g., [227, 247, 218].

11.6 Conclusions

Characterizing the population of BBH masses with direct phenomenological or nonparametric fits can provide insight into the shape of the mass distribution, but does not provide direct constraints on the underlying physics of BBH masses. With the method we propose here, we can infer the underlying physics by fitting the implied (derived) astrophysical BBH distribution to the observed data. We demonstrate the use of this method by evaluating the role of the PPISN process giving rise to the 1G BH mass distribution and its structure, including an excess (bump) in the mass distribution at the lower edge of the PPISN mass gap. Fitting this model to the observed data, we find that the necessary physical parameters to explain the excess of BHs at $\sim 35\,M_\odot$ are unrealistic from a nuclear physics perspective if we take this PPISN model at face value. We therefore conclude it is highly unlikely that the feature at $\sim 35\,M_\odot$ is associated with the PPISN process.

This framework motivates future investigations to better constrain the physics underlying astrophysical populations in general. Future work using additional observations and enhanced versions of our model may be able to constrain proposed astrophysical mechanisms underpinning the BBH mass, spin, and redshift distributions. This approach may offer fruitful applications such as calibration of "spectral siren" features for cosmology [215], investigating other proposed interpretations of the bumps in the mass distribution, and understanding progenitor populations by relating back to population synthesis configurations [574, 60, 594].

Acknowledgements

We thank Mathieu Renzo for helpful discussions about `MESA` simulations and the physics of stellar evolution and the PPISN process. We also acknowledge Alan Weinstein for useful comments and discussions, and Maya Fishbach for helpful comments on the manuscript. The Flatiron Institute is a division of the Simons

Foundation. The authors are grateful for computational resources provided by the LIGO Lab and supported by National Science Foundation Grants PHY-0757058 and PHY-0823459. JG is supported by NSF award No. 2207758. This research has made use of data or software obtained from the Gravitational Wave Open Science Center (gwosc.org), a service of the LIGO Scientific Collaboration, the Virgo Collaboration, and KAGRA.

This material is based upon work supported by NSF's LIGO Laboratory which is a major facility fully funded by the National Science Foundation, as well as the Science and Technology Facilities Council (STFC) of the United Kingdom, the Max-Planck-Society (MPS), and the State of Niedersachsen/Germany for support of the construction of Advanced LIGO and construction and operation of the GEO600 detector. Additional support for Advanced LIGO was provided by the Australian Research Council. Virgo is funded, through the European Gravitational Observatory (EGO), by the French Centre National de Recherche Scientifique (CNRS), the Italian Istituto Nazionale di Fisica Nucleare (INFN) and the Dutch Nikhef, with contributions by institutions from Belgium, Germany, Greece, Hungary, Ireland, Japan, Monaco, Poland, Portugal, Spain. KAGRA is supported by Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan Society for the Promotion of Science (JSPS) in Japan; National Research Foundation (NRF) and Ministry of Science and ICT (MSIT) in Korea; Academia Sinica (AS) and National Science and Technology Council (NSTC) in Taiwan.

This manuscript carries LIGO Document Number #P2300418.

Appendix 11.A Details of the Likelihood and Differential Rate Calculation

Setting θ to be the set of single-event parameters, we can write the contribution from ith-event to the population likelihood as [359]:

$$p(d_i|\Lambda) = \frac{\int d\theta_i \, p(d_i|\theta_i) \, p(\theta_i|\Lambda) \, p_{\rm det}(\theta_i, d_i)}{\int\int dd_i \, d\theta_i \, p(d_i|\theta_i) \, p(\theta_i|\Lambda) \, p_{\rm det}(\theta_i, d_i)}. \tag{11.12}$$

Recalling that the probability density should be normalized over the arguments on the left side of the bar, the denominator is included to explicitely normalize the numerator in terms of the data from the ith detection, and is commonly known as the "selection effects" term. We write the detection probability as $p_{\rm det}(\theta_i, d_i)$ in order to include the general possiblity of thresholding detection in terms of the event parameters, which may be implemented when considering, for example, a simulated catalog. For our purposes, the detection probability depends on the data, as this is the input to a detection pipeline when assigning a FAR. The normalization in the denominator also corresponds to the fraction of detectable events expected from the population given by Λ [213]. We make the following definition of the denominator:

$$\mu(\Lambda) \equiv \int\int dd_i \, d\theta_i \, p(d_i|\theta_i) \, p(\theta_i|\Lambda) \, p_{\rm det}(\theta_i, d_i). \tag{11.13}$$

The total likelihood comes from considering the probability of the entire dataset $\{d_i\}$ of N_d detections (where the ith event is detected if $p_{\rm det}$ is 1), given a population with parameters Λ that predicts N total events, $N\mu \equiv K(\Lambda)$ of which are expected to be detected. The total likelihood is just the product of the contributions from all the detected events, and the likelihood of detecting N_d events, considering the realization of N_d comes from a Poisson distribution with expected value K:

$$\begin{aligned} p(\{d\}|\Lambda, K) &= p(N_d|K(\Lambda)) \prod_i^{N_d} p(d_i|\Lambda) \\ &\propto K(\Lambda)^{N_d} e^{-K(\Lambda)} \mu(\Lambda)^{-N_d} \prod_i^{N_d} \int d\theta_i p(d_i|\theta_i) p(\theta_i|\Lambda). \end{aligned} \tag{11.14}$$

If we assume a prior of $\pi(K) \propto 1/K$, we can write the posterior over Λ and analytically integrate out the distribution over K:

$$
\begin{aligned}
p(\Lambda|\{d\}) &\propto \pi(\Lambda) \int dK \frac{K(\Lambda)^{N_d} e^{-K(\Lambda)}}{K} \mu(\Lambda)^{-N_d} \prod_i^{N_d} \int d\theta_i p(d_i|\theta_i) p(\theta_i|\Lambda) \\
&= \Gamma(N_d) \pi(\Lambda) \mu(\Lambda)^{-N_d} \prod_i^{N_d} \int d\theta_i p(d_i|\theta_i) p(\theta_i|\Lambda) \\
&\propto \pi(\Lambda) \mu(\Lambda)^{-N_d} \prod_i^{N_d} \int d\theta_i p(d_i|\theta_i) p(\theta_i|\Lambda)
\end{aligned}
\tag{11.15}
$$

where $\Gamma(N_d)$ does not depend on Λ, so marginalizing over $K(\Lambda)$ with this choice of $\pi(K)$ allows us to factorize the above equation, without explicitly considering its dependence on the Poisson term.

In practice, $p(\theta|\Lambda)$ does not need to be normalized, as any prefactors will divide out in Eq. 11.12. We therefore only need to calculate something proportional to $p(\theta|\Lambda)$. For reasons that will become apparent, we compute $p(\theta|\Lambda)$ in terms of something proportional to $\frac{dN}{d\theta}(\Lambda)$. We want to define a normalization factor for the population distribution such that:

$$
\frac{1}{\alpha(\Lambda)} \, m_1 \frac{dN}{dm_1 dq dV dt_s}\bigg|_{(m_{\rm ref}, q_{\rm ref}, z_{\rm ref})} = 1 \tag{11.16}
$$

where the differential rate is evaluated at a set of reference parameters.

With the distributions in Sec. 11.3 defined in terms of $\frac{dN}{dm}$ and $\frac{dN}{dV dt_s}$ (i.e. source frame merger rate density $\mathcal{R}(z)$), we can compute a normalization factor $\alpha(\Lambda)$:

$$
\begin{aligned}
\alpha(\Lambda) = m_1 \frac{dN}{dm_1 dq dV dt_s}(\Lambda)\bigg|_{(m_{\rm ref}, q_{\rm ref}, z_{\rm ref})} &= m_1 \frac{dN}{dm_1} \frac{dN}{dm_2} \frac{dm_2}{dq} \frac{dN}{dV dt_s}\bigg|_{(m_{\rm ref}, q_{\rm ref}, z_{\rm ref})} \\
&= m_1^2 \frac{dN}{dm} \frac{dN}{dm} \frac{dN}{dV dt_s}\bigg|_{(m_{\rm ref}, q_{\rm ref}, z_{\rm ref})}
\end{aligned}
\tag{11.17}
$$

Technically, we only know the dN distributions up to a constant. As we will see below, we will only be considering ratios of values that share the same unknown constant, so we are free to leave it out for now.

Instead of computing $p(\theta|\Lambda)$ as $p(\theta|\Lambda) = \frac{1}{N} \frac{dN}{d\theta}(\Lambda)$ exactly, we instead make the following transformation in Equations 11.12 and 11.15:

$$
p(\theta|\Lambda) \to \frac{1}{\alpha(\Lambda)} \frac{dN}{d\theta}(\Lambda) \tag{11.18}
$$

which is directly proportional to the differential rate and $p(\theta|\Lambda)$.

For each draw of Λ, we have the normalization factor $\alpha(\Lambda)$, related to the differential rate at our reference parameters, as defined in Eq. 11.17. We outline below how we use this re-expression to construct the rate independent of the likelihood. Note that this change in Eq. 11.18 does not affect the likelihood, as it only affects θ-independent prefactors, which factor out of both the numerator and denominator in Eq. 11.12.

Given the values of $\alpha(\Lambda)$ we have calculated, we wish to draw new samples of $\alpha(\Lambda)$, given that the number of detections is a Poisson-distributed realization. Recalling $K \equiv N\mu$, Eq. 11.18 means that when we compute the denominator of Eq. 11.12, we are actually calculating the ratio $\frac{K}{\alpha}$, and not μ. Noting that the K-dependent integrand of Eq. 11.15 is a Gamma-distribution for K with shape parameter N_d and a scale parameter of 1, we can make the identification that $\langle K \rangle = N_d$ under this distribution. With $K \equiv N\mu$, we can express the expectation value for α as $\langle \alpha \rangle = \frac{N_d}{K/\alpha}$.[6] As a final step in post-processing, we can construct the true underlying distribution for α by drawing samples $\alpha \sim \text{Gamma}(\frac{N_d}{K/\alpha}, 1)$. This gives us a distribution for the predicted merger rate at the reference coordinates, given the normalization factor α we computed during the hierarchical inference, assuming this is Poisson-distributed about the true value and assuming a $\frac{1}{K}$ prior. With the distribution of α, we can scale $\frac{dN}{d\theta}$ to get the differential merger rate at any set of coordinates θ.

Note that we have written everything in this section in terms of θ as if it is always the parameters in the population model, suppressing the fact that there will be Jacobians in Eq. 11.12 to transform from these coordinates to those in the detector-frame (or the priors from the single-event analyses).

Appendix 11.B Accounting for Spin Distribution

In the analysis presented in the body of this work, we assume the (uninformative) parameter estimation priors in the population reweighting. Based on population-level mass-spin and mass-redshift correlations presented in the literature (see, e.g., [126, 97, 28]), we may expect the assumed spin distribution can have an effect on our results. However, with the relatively poor spin constraints in the population, we empirically demonstrate that this is likely not the case.

[6] Since $\langle K \rangle = N_d$, it follows that $\sigma_\alpha^2 = \frac{N_d}{(K/\alpha)^2}$

In Figure 11.12, we compare posteriors obtained from our main analysis ignoring spins, with those obtained by reweighting the posterior samples from each event and the sensitivity injections to a fiducial spin distribution. For this fiducial spin distribution, the spin magnitudes are from a half-Gaussian centered at $a = 0$ with a standard deviation of 0.3, meant to model the preferentially-small spin mangitudes inferred in [28]. For the contribution aligned projection of the spin tilt angle ($\cos\theta$), we use the mixture model introduced in [521], with an aligned-spin fraction of $\xi = 0.8$ and an aligned-spin spread of $\sigma_t = 1.9$, consistent with the results reported in [28] (see references for definitions of these model parameters). We find that reweighting to this spin distribution has a negligible effect on our inferred population.

Appendix 11.C Details on PPISN Interpretation

The physical reason the the anticorrelation between $^{12}\mathrm{C}(\alpha,\gamma)^{16}\mathrm{O}$ and $M_{\mathrm{BH},max}$ is that during contraction of the stellar core, hydrostatic equilibrium can be maintained by convective carbon burning. Higher $^{12}\mathrm{C}(\alpha,\gamma)^{16}\mathrm{O}$ reaction rates lead to cores of lower carbon fractions, X_C. When the core gets hot enough to produce electron-positron pairs, the equation of state softens, leading to a contraction. With little carbon present to provide convective-driven pressure to stabilize the star, contraction can continue until it drives thermonuclear ignition of oxygen. This explosive process leads to an outward-moving shock, removing mass from the star when the shock reaches the surface with enough velocity. Once this shock breaks through the surface of the star, contraction begins again. This sequence of pulsations continues until oxygen in the core is depleted, core elements burn through the normal pre-SN process, and the star undergoes normal core collapse. If the $^{12}C(\alpha,\gamma)^{16}O$ reaction rate is lower, then relatively more carbon is present and able to burn convectively, counteracting the contractions in a stable manner. As the carbon fraction gets higher, the star is able to remain stable against pair-production contractions and stably burn through the core oxygen [211, 576, 577]

The location for the onset of Pair Instability Supernova (PISN) is highly sensitive to the $^{12}\mathrm{C}(\alpha,\gamma)^{16}\mathrm{O}$ reaction rate. With the core temperature strongly increasing with stellar mass, there exists a core mass at which the softening of the equation of state is too extreme to be resisted by available sources of outward pressure. Since stable outward pressure support at this stage is largely provided by shell carbon-burning, the lower carbon fraction, X_C, makes it now easier for a given contraction

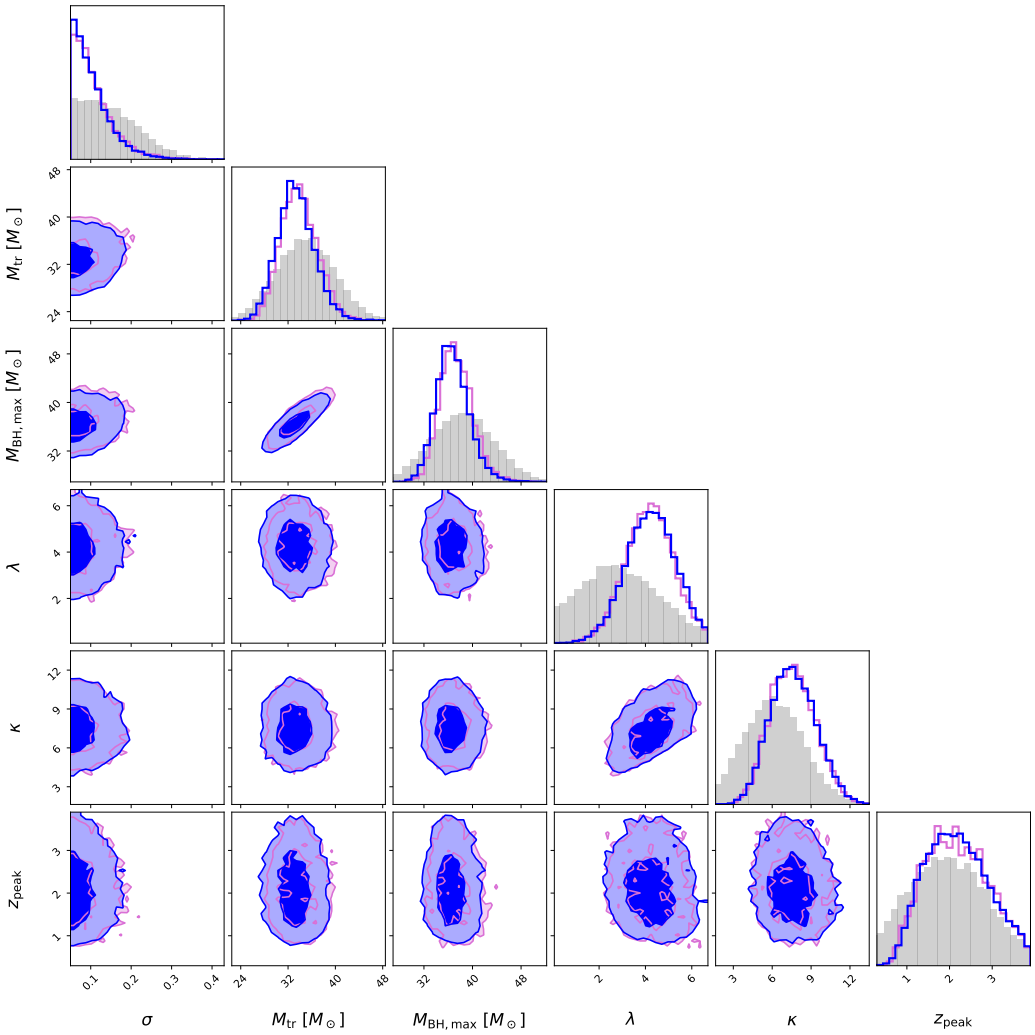

Figure 11.12: Selected population-level parameters from the evolving mass distribution analysis, without reweighting spins (pink, as presented in Section 11.4.2) and with reweighting spins to a fiducial population estimate (blue). The near-identical posteriors show that the spin population assumptions in this work do not cause a bias.

to compress and fully ignite the oxygen core, driving a subsequent pulsation so powerful that further pair production in the core cannot re-soften the equation of state fast enough to return it to a contraction phase. This is basically equivalent to a single pulsation during the pulsational pair-instability process blowing away the total mass of the star [576, 577, 578]. A lower carbon fraction results in this full disruption of the star (PISN) occurring at lower masses, controlling where the $M_{\rm I} - M_{\rm BH}$ map decays to zero after $M_{{\rm BH},max}$.

Chapter 12

PROJECTIONS OF THE UNCERTAINTY ON THE COMPACT BINARY POPULATION BACKGROUND USING POPSTOCK

Arianna I. Renzini and Jacob Golomb. "Projections of the uncertainty on the compact binary population background using popstock". In: *Astron. Astrophys.* 691 (2024), A238. DOI: 10.1051/0004-6361/202451374. arXiv: 2407.03742 [astro-ph.CO].

Abstract

The LIGO-Virgo-KAGRA collaboration has announced the detection of almost 100 binary black holes so far, which have been used in several studies to infer the features of the underlying binary black hole population. From these, it is possible to predict the overall gravitational-wave (GW) fractional energy density contributed by black holes throughout the Universe, and thus estimate the gravitational-wave background (GWB) spectrum emitted in the current GW detector band. These predictions are fundamental in our forecasts for background detection and characterisation, with both present and future instruments. The uncertainties in the inferred population strongly impact the predicted energy spectrum, and in this paper we present a new, flexible method to quickly calculate the energy spectrum for varying black hole population features such as the mass spectrum and redshift distribution. We implement this method in an open-access package, `popstock`, and extensively test its capabilities. Using `popstock`, we investigate how uncertainties in these distributions impact our detection capabilities and present several caveats for background estimation. In particular, we find that the standard assumption that the background signal follows a $2/3$ power-law at low frequencies is both waveform and mass-model dependent, and that the signal power-law is likely shallower than previously modelled, given the current waveform and population knowledge.

12.1 Introduction

The Laser Interferometer Gravitational-wave Observatory (LIGO) [1], Virgo [38], and KAGRA [53] detectors are progressively uncovering the features of the population of merging stellar-mass binary black holes in our Universe [29, 28]. As observing runs become more and more sensitive, the detection horizon increases and a higher number of gravitational-wave (GW) events are positively identified as binary mergers. The events observed so far lie at distances $z \lesssim 1$ [26], while the vast majority of binary mergers is expected to lie well beyond this horizon, as suggested by both theoretical expectations for the merger rate redshift evolution [181, 90, 466] and its inferred trend from GW data through the Third Gravitational-Wave Transient Catalogue (GWTC-3) [28]. This knowledge has motivated studies and inference of the sub-threshold, unresolved collection of binaries, treated as an overall gravitational-wave background (GWB) signal.

Works forecasting the GWB from compact binaries populate the literature: before the first GW detections, these were based on theoretical models of the binary population [425, 449, 450], while more recently GW data–informed projections [35] have become a benchmark for the GW community. An important distinction to make is between estimates of a specific realisation of the background, for example used in mock-data challenges [373], and estimates of the ensemble average of the background, which corresponds to the expectation value of the background amplitude targeted by GWB searches [598, 579]. In both cases, the calculation involves the (expected) distributions for the individual binary parameters, such as mass, distance, and event rate distributions. These population models are often taken to be simple parametric functions with a well-defined set of hyper-parameters, which are fixed to fiducial (or assumed) values for the background calculation. Re-calculating the background signal for varying population hyper-parameters can become computationally intensive, when employing large sample sets.

Several applications in the literature require marginalising the background signal over possible population configurations, including forecasting studies, such as those presented by the LIGO-Virgo-KAGRA (LVK) collaboration in [35, 28], and inference analyses, such as [123, 35, 544]. As the interest in this type of work grows, there is a need for efficient and flexible background estimation procedures. In this paper, we present a method to efficiently carry out these calculations, and make an open-access code base, `popstock`, available to the community. The novelty of our approach is in the design of a re-weighting technique which allows us to sample

the binary parameter probability distributions and evaluate a corresponding set of waveform approximants only once, enabling an extremely efficient re-estimation of the GWB when varying population hyper-parameters. An analogous re-weighting approach was previously implemented in [544]. Furthermore, we implement the use of waveform templates imported from the LIGO Scientific Collaboration Algorithm Library (LAL) [335]. To improve efficiency, previous codes have employed analytic waveform approximations directly embedded in the codebase; with our CPU–optimised spectral calculations and our GPU–optimised re-weighting technique, we are able to support a much broader range of waveforms through the commonly used python library for GW analysis, Bilby [67].

We use `popstock` to investigate the impact of the uncertainty on the binary population on the detection prospects of the GWB with ground-based interferometers. In particular, we compare with and extend work done in [28] to include more uncertainties on the redshift distribution of sources. We also employ the package to probe the effect of waveform choice on the estimation, and whether this is entirely degenerate with the population uncertainty. We find that the expected background amplitude can be significantly boosted, when admitting higher mass mergers and higher rates of mergers at low-redshift, within current population uncertainties. We further find that the choice of waveform can have noteworthy effects on background signal estimates, however these do not dominate current population uncertainties.

This paper is organised as follows: In Sec. 12.2 we introduce the theoretical aspects behind compact binary GWB calculations; in Sec. 12.3 we define the analytic models used here to describe the distributions of black hole masses and distances; in Sec. 12.4 we introduce `popstock`, our new package for GWB calculations, outlining its key functionalities; in Sec. 12.5 we present our background calculations and investigations; and finally we summarise our conclusions in Sec. 12.6.

12.2 The compact binary GW background

The amplitude of the GWB signal is parameterised by the fractional energy density spectrum emitted by GWs throughout the Universe, $\Omega_{GW}(f)$, [425]

$$\Omega_{GW}(f) = \frac{1}{\rho_c}\frac{d\rho_{\rm GW}}{d\ln f}\,, \tag{12.1}$$

which is normalised by the critical energy density of the Universe $\rho_c = 3c^2H_0^2/8\pi G$. This is in general the total energy density contributed by GWs throughout the Uni-

verse, and is not restricted to sub-threshold signals. Calculating a residual background requires the definition of a detection cutoff, which is detector–dependent[1], while here and elsewhere Ω_{GW} is considered to be an astrophysical property of the Universe.

While the GWB spectrum is detector–independent, it is useful to employ the detector frame to perform calculations. This allows us to immediately relate the intrinsic signal amplitude with the measured signal in GW detectors. As shown for example in [372, 451], the energy density spectrum $\Omega_{GW}(f)$ may be estimated as the average fractional GW energy density present in a detector during an observation time $T_{\rm obs}$,

$$\Omega_{GW}(f) = \frac{f^3}{T_{\rm obs}} \frac{4\pi^2}{3H_0^2} \sum_i^{N_i} \mathcal{P}_d(\Theta_i; f)\,, \tag{12.2}$$

assuming a finite number of GWs received at the detector, N_i. Here, $\mathcal{P}_d(\Theta_i, f)$ is the Fourier domain unpolarised power in the detector frame (hence the subscript d) associated to a GW with parameters Θ_i, defined as

$$\mathcal{P}_d(\Theta_i; f) = \tilde{h}_+^2(\Theta_i; f) + \tilde{h}_{\times,i}^2(\Theta_i; f)\,, \tag{12.3}$$

where $\tilde{h}_A(\Theta_i; f)$ is the Fourier transform of the GW waveform evaluated at Θ_i[2]. So defined, $\mathcal{P}_d$ has units s^{-2}.

In the limit of infinite observation time and infinite events, the Ω_{GW} spectrum approaches its ensemble average, $\bar{\Omega}_{GW}$. While the Ω_{GW} measured by an experiment depends on the specific realisation observed throughout the experiment observation time, $\bar{\Omega}_{GW}$ only depends on the distributions which describe the GW parameters Θ, which are in turn parameterised via population hyper-parameters Λ [425]. The $\bar{\Omega}_{GW}$ spectrum is thus a unique population signature, and is targeted in observations in practice by measuring Ω_{GW} for large $T_{\rm obs}$. The equivalence is easily seen in the limit of large GW numbers as

$$\frac{1}{T_{\rm obs}} \sum_i^{N_i} \mathcal{P}_d(\Theta_i; f) \underset{\substack{N_i\to\infty \\ T_{\rm obs}\to\infty}}{\approx} \frac{dN}{dt} \int d\Theta\, p_d(\Theta|\Lambda)\, \mathcal{P}_d(\Theta; f)\,, \tag{12.4}$$

[1] For discussions on residual backgrounds, in particular in the context of next generation ground-based interferometers, we refer the reader to [478, 596, 91], for example.

[2] In the case of a stochastic signal described as a superposition of plane waves, as is the case often for cosmological signals, the signal can be thought of as a wave where the Fourier amplitudes are stochastic fields, the parameters Θ describe the field, and the power is the second-order moment of the field [455].

where $p_d(\Theta|\Lambda)$ are the (normalised) detector frame probability distributions for the GW parameters Θ, such that

$$\bar{\Omega}_{GW}(\Lambda; f) = f^3 \frac{4\pi^2}{3H_0^2} R \int d\Theta \, p_d(\Theta|\Lambda) \, \mathcal{P}_d(\Theta; f) \, . \tag{12.5}$$

Here, $R \equiv \frac{dN}{dt}$ is the total rate of events per unit detector-frame time. It can be convenient to isolate the redshift integral in Eq. (12.5), assuming redshift is independent from other parameters, and incorporate the rate in the redshift evolution probability $p(z)$, defining the event rate per unit detector-frame time per redshift shell,

$$R(z) = R \, p(z) \, . \tag{12.6}$$

The $\bar{\Omega}_{GW}$ spectrum can then be interpreted as an integral over redshift shells of the average GW power present in each shell, analogously to Eqs. (4) and (5) of [123],

$$\bar{\Omega}_{GW}(\Lambda; f) = f^3 \frac{4\pi^2}{3H_0^2} \int dz \, R(z|\Lambda_z) \langle \mathcal{P}_d(z, \theta; f) \rangle \, , \tag{12.7}$$

where the $\langle . \rangle$ brackets imply the GW spectral power samples $\mathcal{P}_d$ are averaged over the ensemble described by the parameter probability distributions, in each redshift shell. To see how the GW spectral power is related to the energy spectrum, for each binary, see Appendix 12.A for a pedagogical derivation.

12.3 BBH population models

We illustrate the effect of the population model on Ω_{GW} by adopting two phenomenological mass distribution models used in [29, 28]. The Powerlaw+Peak model (PLPP), first introduced in [524], has been widely adopted in the literature as an astrophysically-motivated mass distribution model. The PLPP model consists simply of a truncated power-law, motivated by the shape of the stellar initial mass function, and a Gaussian bump (or peak), originally intended to account for a possible over-density of black holes around a certain mass, as motivated by, e.g., pulsational pair instability effects [524, 29, 28] [3]. We list the parameters of the PLPP model in Table 12.1. In addition to PLPP, we also consider a simpler mass distribution, consisting of a truncated power-law with a break at a particular mass. While [28] finds that this broken power-law (BPL) model is disfavoured with respect to the PLPP model, we also consider it for illustration purposes. Parameters used for the BPL model are described in Table 12.2.

Table 12.1: Power-law-plus-peak (PLPP) model parameters.

Parameter	Description
α	Slope of the primary mass power-law.
β	Slope of the mass ratio power-law.
$m_{\rm min}$	Minimum mass allowed in the system.
$m_{\rm max}$	Maximum mass allowed in the system.
$m_{\rm pp}$	Location of the Gaussian bump in the mass distribution.
$\sigma_{\rm pp}$	Width of the Gaussian bump in the mass distribution.
λ	Fraction of sources in the bump.

Table 12.2: Broken-power-law (BPL) model parameters.

Parameter	Description
α_1	Slope of the primary mass power-law before the break.
α_2	Slope of the primary mass power-law after the break.
$m_{\rm min}$	Minimum mass allowed in the system.
$m_{\rm max}$	Maximum mass allowed in the system.
$m_{\rm pp}$	Location of the Gaussian bump in the mass distribution.
b	Fractional distance between $m_{\rm min}$ and $m_{\rm max}$ of the break.

We also adopt redshift distribution models commonly used in the literature. For example, [228] introduced a broken power-law model to describe the merger rate as a function of redshift. This model is motivated by the observed star formation rate (SFR) across redshift: a rate rising to and peaking at some redshift ($z_{\rm peak}$, [361, 228]) then decaying down to high redshift, where the star formation rate was much lower. A SFR model commonly adopted in the literature is the broken power-law fit from [351], which is parameterised in terms of a low-redshift power-law index γ, high-redshift (negative) index κ, and a peak or turn-over redshift parameter $z_{\rm peak}$:

$$\mathcal{R}_{\rm MD}(z) \propto \frac{(1+z)^{\gamma}}{1+\left(\frac{1+z}{1+z_{\rm peak}}\right)^{\kappa}}. \tag{12.8}$$

In the following, as in [28] and [35], we use $\mathcal{R}_{\rm MD}$ to describe the number of mergers per unit comoving volume V_c per unit source-frame time t_s. For brevity, we refer to

[3]This feature is found in the data, but recent works have cast doubt on whether it can be attributed to the pulsational pair instability mechanism [255, 274]

this model as MD. The parameters employed throughout for the MD model and their default values are shown in Table 12.3. In reality, the merger rate of BBHs from stellar collapse is a function of the SFR and the delay time between star formation and merger of the remnant [229, 123, 544]; this may make the number of mergers per unit comoving volume and source-frame time $R(z)$ deviate from a SFR-like broken power-law.

Table 12.3: Broken-power-law Redshift model parameters.

Parameter	Description	Defaults
γ	Low-redshift power-law index.	2.7
κ	High-redshift power-law index.	5.6
$z_{\rm peak}$	Redshift of the peak rate.	1.9

Table 12.4: Default values drawn from [228, 123]. This model, including an overall local-merger-rate normalisation R_0, is referred to as MD throughout.

For illustration, we also adopt a less realistic model in which the merger rate is constant across cosmic time, $\mathcal{R}(z) = R_0$. This Uniform in Comoving Volume (UICV) model assumes that the merger rate as measured in the source frame of the emitter, is constant across redshift.

The redshift of individual compact binary merger events in the detector frame is drawn from a probability distribution $p(z)$ which takes into account the comoving volume per unit redshift gradient, dV_c/dz, and the redshifting of the rate from source-frame to detector-frame,

$$p(z) = \frac{1}{1+z}\frac{dV_c}{dz}\mathcal{R}(z)\,. \tag{12.9}$$

The $p(z)$ functions for the MD and UICM rate evolution are illustrated in the left panel of Figure 12.3.

In this paper, unless otherwise specified we draw uncertainties from the LVK collaboration GWTC-3 population posteriors, published in the data release [151] which accompanied the collaboration results [28]. The release includes samples from the posterior of population hyperparameters inferred through GWTC-3 (i.e., the population parameters governing the shapes of the mass, spin, and redshift distributions). We use these hyperparameter samples for the corresponding redshift and mass models described above in the analysis that follows. We use the PLPP mass model and MD redshift model as fiducial models for our studies. The PLPP model is considered a good parametric description of the mass spectrum of GWTC-3 [28], also

confirmed by non-parametric approaches [125], and has been widely used since in the context of GWB estimation (e.g., [596, 91, 483]), while the MD model is a well-motivated astrophysical model [228] and has been already employed in stochastic inference analyses [35].

12.4 `popstock`

We present `popstock`[4], a python–based open–source package for the rapid computation of background spectra such as Ω_{GW}, for a given realisation of events, and $\bar{\Omega}_{GW}$, for a given set of hyper-parameters Λ. Other than the standard python scientific libraries `numpy` [270] and `scipy` [556], the main dependencies of the `popstock` package are: `astropy` [433], a core python library used by astronomers; `bilby` [67], a most popular Bayesian inference library for GW astronomy; and `gwpopulation` [520], a hierarchical Bayesian inference package containing a collection of parametric binary black hole mass, redshift, and spin population models.

The `popstock` package relies on `multiprocessing` (included in most python distributions) to parallelise the computation of Ω_{GW} for large N_i. The GW waveforms required to compute $\mathcal{P}_d(\Theta_i; f)$ in Eq. (12.3) are evaluated at each Θ_i using the `Bilby` library, which in turn imports LAL [335]. This allows us to employ a vast array of modern waveforms in our computations.

To compute $\bar{\Omega}_{GW}$ for a given set of population hyper-parameters Λ and a given collection of population models, we directly sample the probability distributions $p_d(\Theta|\Lambda)$ and evaluate Eq. (12.5) via a Monte Carlo simulation. The accuracy of this evaluation depends on the number of samples employed, as discussed below. This approach is limited by the long sampling and evaluation times of the GW waveforms, and is not an optimal tool to perform in-depth studies of the impact of population uncertainties on the $\bar{\Omega}_{GW}$ spectrum. Hence, `popstock` includes a re-weighting technique to compute $\bar{\Omega}_{GW}$ for a new set of Λ parameters without re-evaluating Eq. (12.5).

The `popstock` repository includes tutorials with usage instructions and simple examples. The package's performance strongly depends on the computing setup employed (e.g., CPU and GPU availability). We refer the interested reader to the `popstock` documentation [5] for further details.

[4] https://github.com/a-renzini/popstock
[5] https://a-renzini.github.io/popstock/

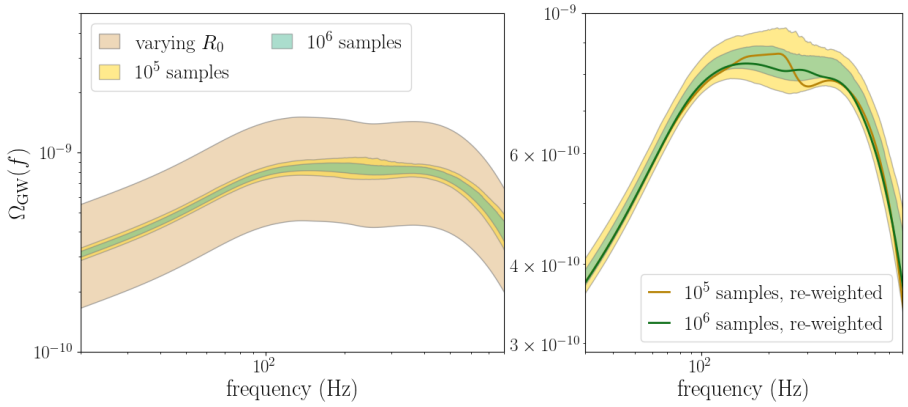

Figure 12.1: Impact of sample variance and re-weighting on the Ω_{GW} spectrum. Left: 95% confidence on the spectrum calculated using 10^5 (10^6) samples in yellow (green) drawn from a fixed hyper-parameter distribution, compared to the 95% confidence on the spectrum including uncertainty on the local merger rate parameter in brown. Right: sample variance from the left panel compared to re-weighted estimates of the Ω_{GW} spectrum. The re-weighted spectra lie neatly within the sample variance uncertainty bounds, implying that the re-weighted spectrum is indistinguishable from a regularly sampled spectrum, with these sample numbers.

In the rest of this section, we describe the `popstock` re-weighting technique and probe its efficiency and accuracy.

12.4.1 Re-weighting methodology

We lay out a simple method to efficiently calculate $\bar{\Omega}_{GW}$ for different sets of hyper-parameters Λ_i describing the (detector frame) target population distributions, $p_d(\Theta|\Lambda_i)$. The integral of (12.5) above allows the implementation of an importance sampling approach or re-weighting[6], whereby

$$\begin{aligned} \int d\Theta\, p_d(\Theta|\Lambda_1)\, \mathcal{P}_d(\Theta) &= \int d\Theta\, \frac{p_d(\Theta|\Lambda_1)}{p_d(\Theta|\Lambda_0)} p_d(\Theta|\Lambda_0)\, \mathcal{P}_d(\Theta) \\ &\equiv \int d\Theta\, w_{0\to1}(\Theta) p_d(\Theta|\Lambda_0)\, \mathcal{P}_d(\Theta)\,, \end{aligned} \tag{12.10}$$

where $p_d(\Theta|\Lambda_0)$ is chosen as the fiducial distribution, and $w_{0\to1}$ is the weight required to "transform" between the fiducial distribution and the target one, relative to Λ_1:

$$w_{0\to1} = \frac{p_d(\Theta|\Lambda_1)}{p_d(\Theta|\Lambda_0)}\,. \tag{12.11}$$

[6]Reweighting has become a popular tool for efficient Monte Carlo computations in GW astronomy. See, e.g., [520, 419, 287, 519] for a review of some applictions of this method in the GW field.

In practice, this reweighting approach is more efficient than direct Monte Carlo integration when $p_d(\Theta|\Lambda_1)$ is hard to sample from but easy to evaluate. Therefore, we first draw a large set of samples $\boldsymbol{\Theta}$ from the fiducial population Λ_0 and compute the probability of drawing those samples $p_d(\Theta|\Lambda_0)$. This is stored as the denominator in the weights w. Each time the integral of Eq. 12.10 is evaluated for some different population Λ_1, it is only necessary to evaluate the probability of those fiducial samples under the target population $p_d(\Theta|\Lambda_1)$, as w is the only term that depends on Λ_1. See also App. D of [544] for an analogous re-weighting approach to calculate the background spectrum; the authors' approach is technically equivalent, however they choose to re-weight the approximated event energy spectra calculated using [49], as opposed to the frequency domain $\mathcal{P}_d$ quantity calculated from the event waveform.

The re-weighting operation is directly implemented in the Monte Carlo evaluation described above, which is valid as long as a sufficient number of samples Θ_j are used,

$$\int d\Theta\, w_{0\to1}(\Theta) p_d(\Theta|\Lambda_0)\, \mathcal{P}_d(\Theta) \approx \sum_j w_{0\to1}(\Theta_j) \mathcal{P}_d(\Theta_j)\,. \tag{12.12}$$

This allows us to evaluate the (costly) $\mathcal{P}_d$ spectra only once, and re-weight the contribution of each wave according to a desired target distribution.

In practice, we rely on the source-frame population distributions to sample the GW parameters. To convert these to detector frame, we evaluate the relevant Jacobian matrix assuming a fixed cosmology,

$$p_d(\Theta|\Lambda) = p_s(\Theta|\Lambda)\, \frac{d\Theta_s}{d\Theta_d}\,. \tag{12.13}$$

12.4.2 Effective sample size and sample variance

As a check of the performance of our re-weighting approach, we estimate the effective sample size $N_{\rm eff}$ for different number of samples $N_{\rm s}$ and different Ω_{GW} spectra, and ensure $N_{\rm eff} \gg 1$. The effective sample size is calculated from the weights as $N_{\rm eff} = \Sigma w^2/(\Sigma w)^2$. We find that, for a fixed re-weighting set of Λ hyper-parameters, $N_{\rm eff} \approx 2\times10^4$ for $N_{\rm s} = 5\times10^4$, $N_{\rm eff} \approx 4\times10^4$ for $N_{\rm s} = 10^5$, and $N_{\rm eff} \approx 3\times10^5$ for $N_{\rm s} = 10^6$. In practice, these numbers will depend on the size of the parameter space probed by the re-weighting. In all implementations shown in this paper we use $N_{\rm s} = 1\times10^6$ unless otherwise stated, and we have checked the order of magnitude of $N_{\rm eff}$ reported here remains reliable for all results shown.

The Ω_{GW} spectrum is by definition a stochastic observable, and thus presents an intrinsic sample variance[7]. In particular, $\Omega_{GW}^{\rm BBH}$ is dominated by a Poisson process given the low merger rate of black hole binaries. We estimate the intrinsic variance of $\Omega_{GW}^{\rm BBH}$ associated to different number of samples (which can be directly converted to different observation times, assuming a total merger rate) by calculating the spectrum using different sample draws from a fixed set of population priors. These are shown in Figure 12.1 (left panel) using 10^5 and 10^6 samples, where the shading indicates the 95% interval over 1000 sets. In this case, increasing the number of samples by a factor of 10 decreases the width of the 95% interval by 52% on average, for frequencies between $20 - 500$ Hz. This specific example corresponds to the following set of hyper-parameters for the PLPP mass model: $\alpha = 3.5$, $\beta = 1$, $\delta_m = 4.5$, $\lambda = 0.04$, $m_{\rm max} = 100$, $m_{\rm min} = 4$, $m_{\rm pp} = 34$, and $\sigma_{\rm pp} = 4$. The redshift model is fixed to the default MD model defined above with a local merger rate of $R_0 = 15$ Gpc^{-3}yr^{-1}. In Figure 12.1 we further compare these sample variance uncertainty bands to the 95% confidence on the spectrum including uncertainty on the local merger rate R_0, drawn from [28]. Unsurprisingly, when using these large sample numbers the sample variance is subdominant compared to population parameter uncertainty.

We compare this intrinsic uncertainty to re-weighting: as may be seen in Figure 12.1 (right panel), re-weighted curves for $\Omega_{GW}^{\rm BBH}$ for the given set of Λ hyper-parameters lie within the 95% sample uncertainty on the spectrum, for different values of $N_{\rm s}$. This implies the re-weighted $\Omega_{GW}^{\rm BBH}$ spectrum for a given population model is within the intrinsic error on that spectrum, and is thus a fair approximation to make.

As we focus on BBHs in this paper, we drop the BBH subscript from $\Omega_{GW}^{\rm BBH}$ in what follows, and assume we refer to the BBH spectrum unless otherwise specified.

12.5 Background projections using `popstock`

We study the dependence of the amplitude, spectral shape, and uncertainty of Ω_{GW} on various models and data products. These studies will fundamentally inform compact binary population parameter estimation campaigns with upcoming GW datasets, for example in the style of [123, 35], which use constraints on (or, in the future, measurements of) the Ω_{GW} spectrum.

[7] See [324] for an analytic study of the massive black hole background spectral variance, in the case of a signal in the pulsar timing array detection band.

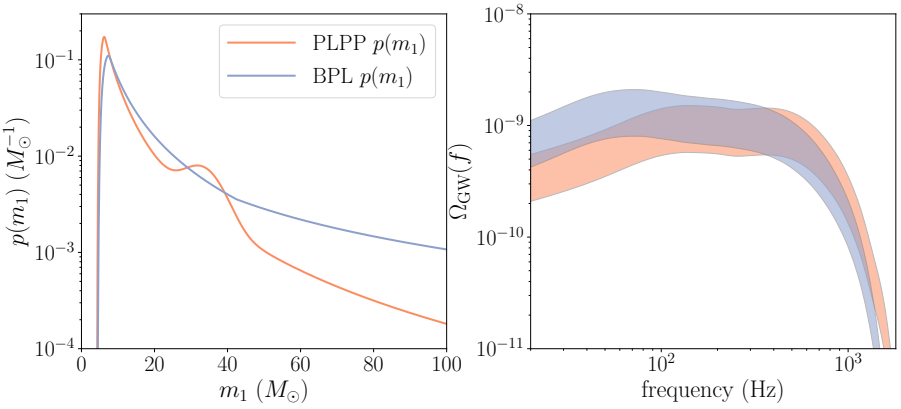

Figure 12.2: Impact of the primary mass distribution on the Ω_{GW} spectrum. On the left, the two primary mass model probability densities used throughout; on the right, 95% confidence intervals for Ω_{GW} using the two mass models, including uncertainty on the local merger rate from [28].

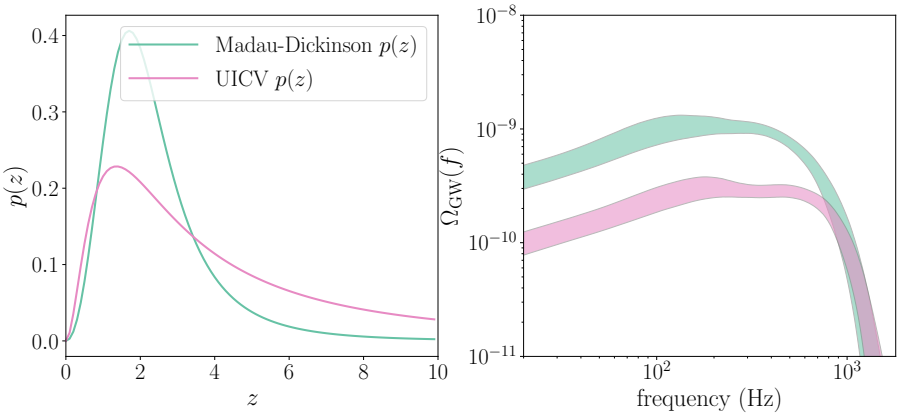

Figure 12.3: Impact of the merger rate redshift distribution model on the Ω_{GW} spectrum. On the left, the two redshift evolution model probability densities used throughout; on the right, 95% confidence intervals for Ω_{GW} using the two merger rate models, including uncertainty on the PLPP mass model from [28] as described in Sec. 12.3.

We consider here a frequency range of 10 – 2000 Hz as this corresponds to the sensitivity of second generation ground-based GW detectors such as the current configurations of the LIGO, Virgo, and KAGRA instruments as well as their near-future improvements.

12.5.1 Mass and redshift models

With `popstock` we can rapidly assess the impact of different mass and redshift models on the projected Ω_{GW}. Here we compare the PLPP and BPL mass models, and the UICV and MD redshift models, all introduced in Sec. 12.3. When comparing mass models, we fix the mass model hyper-parameters while including the uncertainty on the local merger rate R_0 from [28][8] assuming the MD redshift evolution with all other parameters fixed to the values discussed above. When comparing redshift models, we instead fix the redshift model hyper-parameters while including the uncertainty from [28] on the PLPP mass model. We deliberately choose values for certain Λ hyper-parameters which are unrealistic and not favoured by current data to showcase the effect different mass and redshift models may have on the Ω_{GW} spectral shape.

A comparison between the PLPP and BPL mass models is shown in Figure 12.2. The PLPP parameters are fixed to the same set used in Sec. 12.4.2, and BPL to $\alpha_1 = -2$, $\alpha_2 = -1.4$, $\beta = 1$, $b = 0.4$ (see Sec. 12.3 for details on the parameters). As PLPP is commonly used as a mass model when generating Ω_{GW} forecasts (as in [35, 28]), we take this as the fiducial model to produce Ω_{GW} spectra and compare those obtained with the BPL mass model against these. Note that in particular the choice of $\alpha_2 > \alpha_1$ for BPL here implies a larger amount of high-mass binaries in the distribution, as seen in the left panel of Figure 12.2, where primary mass probability distributions are shown. These more massive binaries merge at lower frequencies and their emission is further redshifted into the lower end of the frequency range considered here; remember that, for example, we detect an equal-mass binary with true component masses of $70M_\odot$ merging around ~ 60 Hz at $z = 0$ and ~ 30 Hz at $z = 1$ (see e.g. [454] for more considerations along this line). As seen in the right panel of Figure 12.2, this both boosts the amplitude of Ω_{GW} at all frequencies below a few hundred Hz, and changes the spectral shape at these frequencies, when compared to the PLPP mass spectrum. Specifically, the typical "turnover" in the spectrum corresponding to the frequency at which most binaries have merged is broken into two turnovers: one for the higher mass binaries (below 100 Hz) and one for the lower mass ones (around 300 Hz). This is effect is certainly fuelled by the unrealistic parameter choice made for BPL ($\alpha_2 > \alpha_1$). Comparatively, the PLPP model gives rise to a single turnover with a plateau between $\sim 100 - 400$ Hz, which

[8] Specifically: we employ samples from the posterior fit of the power-law redshift model, as in [28] no broken-power-law redshift model was used.

presents small features ("wiggles") which are related to the redshifting of the peak at $33M_{\odot}$. Futhermore, the spectral index at lower frequencies is more peaked than that of the PLPP model. A simple broken-power-law fit to the two sets of curves shown in the right panel of Figure 12.2 yields $\alpha = 0.61$ and $\alpha = 0.76$ for the lower frequency region of the Ω_{GW} spectrum corresponding to the PLPP and BPL mass models, respectively. A broader discussion on data-informed spectral indices is postponed to Sec. 12.5.3.

The uncertainty on the local rate R_0 implies that there is significant overlap between the 95% credible envelope of the spectrum from these two mass models, suggesting it would be challenging to distinguish mass spectrum features from redshift ones from a measurement of Ω_{GW} alone. The overlap would be even greater when including full uncertainty on the redshift model parameters. However, if a large amplitude and large spectral index (i.e., $\alpha > 2/3$) Ω_{GW} is observed at low frequencies, we expect a mass model which admits large mass binaries (such as the BPL one showed here) to be favoured.

In Figure 12.3 we compare the effect of the UICV and MD redshift models on Ω_{GW}. We fix the local merger rate to $R_0 = 15$ Gpc^{-3}yr^{-1}, and compare a UICV rate evolution to the default MD evolution (see Sec. 12.3) while we include the uncertainty on the PLPP mass model from [28]. Most notably, the UICV model impacts the overall amplitude of the Ω_{GW} spectrum across all frequencies. In this test case, the decrease in amplitude when assuming UICV is approximately constant (and equal to a factor of ~ 4) between $10 - 100$ Hz, and is due to the lower merger rate between $1 < z < 4$. This effect is much larger than the impact of the uncertainty on the mass model, confirming that a measurement of Ω_{GW} will have significant information on the merger rate redshift evolution (as also seen in [124, 123, 454]). The turn-over in the spectrum appears shifted to slightly higher frequencies in UICV, possibly due to the slightly higher rate fraction at low redshift compared to MD, which instead increases as a power-law[9]. Otherwise, the different redshift models do not appear to cause large variations in the overall shape of the spectrum, suggesting that the mass spectrum dominates these features.

A natural extension of this study is to consider BBH mass spectra that evolve with redshift: this will mix the effects seen here when considering independent contributions, and in principle will need to be appropriately included in parameter

[9]This is evident in the area on the left panel of Figure 12.3 where the UICV model $p(z)$ lies above the MD $p(z)$, at $z < 1$.

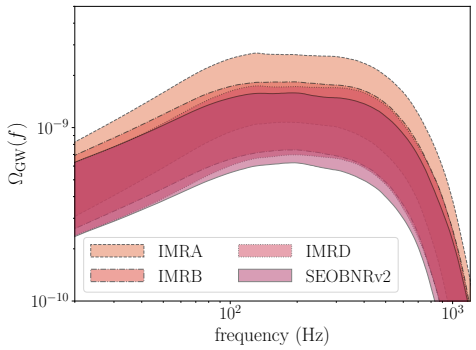

Figure 12.4: Impact of the waveform model on the Ω_{GW} spectrum. The shading indicates the 95% confidence on the spectrum including the uncertainty on the PLPP mass model and the local merger rate, assuming a fixed MD redshift evolution.

estimation studies to avoid biases.

12.5.2 Waveform approximants

The choice of waveform approximant, while central in certain studies of individual compact binary merger events, has been explored very little in the context of the compact binary background signal. In previous work (see e.g. approximations made in [372, 124, 388]), it was deemed sufficient to capture solely the evolution of the GW amplitude as a function of frequency, as the background Ω_{GW} contains no phase information; this evolution can be tracked analytically up to arbitrary Post-Newtonian (PN) order, considerably speeding up the calculation of Ω_{GW} compared to calculating full waveforms for large sets of events. Specifically, most works employ the amplitude component of frequency-domain inspiral-merger-ringdown (IMR) waveforms [47, 49] defined analytically by parts, where the transition of the GW from one phase to the next is set by the specific intrinsic parameters of the binary (mass and spin). As the background has remained a weak signal in the current LVK data, a precise quantification of the systematic differences between background estimates with different waveform approximants has not been necessary. However, as detector sensitivities improve and detection becomes a real possibility, all modelling systematics are important to quantify (see also [596], [505]). Here, we investigate the effect of the waveform approximant using `popstock` and confirm whether it is subdominant to the impact of population uncertainties.

We focus here on the IMRPhenom family of waveforms commonly used in the literature to compute Ω_{GW}, as well as an effective-one-body (EOB) numerical-

relativity (NR) waveform model. In all cases, we omit spin effects, setting both black hole spins to 0. The specific waveforms used are

- **IMRPhenomA** [47]: The first IMR waveform, developed for GW data analysis in the frequency domain for non-spinning binaries. Here the amplitude is expanded to leading post-Newtonian (0-PN) order, implying the inspiral phase presents the characteristic $f^{2/3}$ trend (in Ω_{GW} units).
- **IMRPhenomB** [49]: A direct successor of IMRPhenomA, this waveform includes higher order corrections in the amplitude term up to 1.5-PN and includes non-zero aligned spin. These correct the spectral shape of the waveform amplitude, as a function of both mass and spin.
- **IMRPhenomD** [305]: A recent waveform including corrections up to 3-PN order in the amplitude and a more sophisticated fit to numerical relativity compared to IMRPhenomA and B.
- **SEOBNRv2** [434]: An EOB NR waveform for spin-aligned BBHs, calculated numerically in the frequency domain.

A comparison between the Ω_{GW} spectra calculated for the same BBH population using the waveforms above is presented in Figure 12.4. We show the 95% confidence contours on the population Ω_{GW} including the uncertainty on the mass model (PLPP) and the local merger rate R_0 assuming a fixed MD redshift evolution (as defined above).

We find that when employing the 0-PN IMRPhenomA waveform, the background signal is overestimated at all frequencies by up to 50% in the range $10 < f < 1000$ Hz compared to IMRPhenomB; this is due to the missing amplitude corrections to the inspiral phase (see e.g. the differences in Eq. (4.13) of [47] and Eq. (1) of [49]). While the amplitude estimates for IMRPhenomA and B agree at $f \equiv 0$, these diverge for $f > 0$ as the amplitude evolution with frequency is $\Omega_A \propto f^{2/3}$ for IMRPhenomA and $\Omega_B \propto f^{\alpha<2/3}$ for IMRPhenomB. The trend of α will depend on the specific mass, redshift, and spin realisation as discussed in Sec. 12.5.1. This result shows that the somewhat basic assumption that the compact binary background at frequencies under $\sim$ 100 Hz is well-approximated by a $2/3$ power-law can be upgraded, and informed by likely population models to optimise background searches.

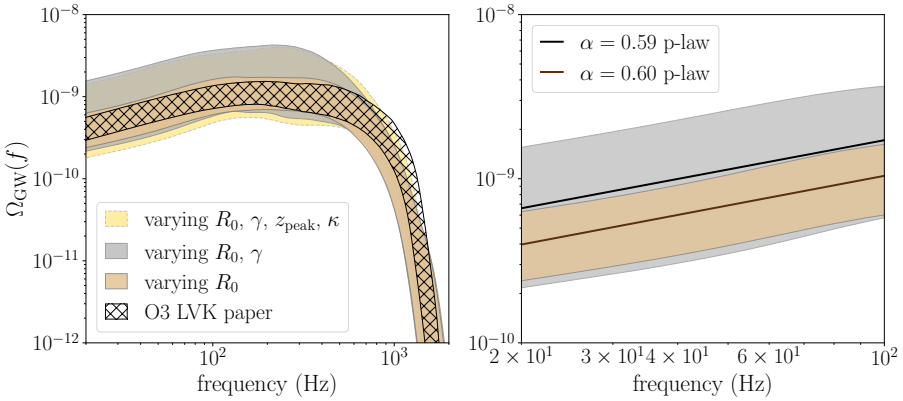

Figure 12.5: Uncertainty on the expected Ω_{GW} spectrum from BBHs due to uncertainty on the merger rate evolution parameters. Left: 95% confidence levels on the projected Ω_{GW} spectrum including uncertainty on the PLPP mass model and assuming a MD merger rate model with different levels of uncertainty. The hatched outline reports previous results published in [28]. Right: a zoomed-in comparison at low frequencies of the uncertainty on the Ω_{GW} spectrum when varying R_0 versus R_0 and γ, reporting average spectral indices referred to these two contours.

Differences due to the use of IMRphenomB/D and SEOBNRv2 approximants are comparable to each other and would be hard to distinguish from population uncertainties. Nevertheless, we comment that the different NR calibration used in IMRphenomD compared to B is evident in the impact due to the inspiral phase on the GWB signal, as the frequency evolution at low frequency is slightly modified, and SEOBNRv2 estimates an overall lower background than the IMRPhenom waveforms.

The impact of including higher-order modes in the waveform calculation on the background spectrum was found to be negligible; a comparison between spectra calculated with the IMRPhenomD and IMRPhenomXPHM waveforms is included for completeness in Appendix 12.B.

12.5.3 O3 population samples

We conclude our analyses by combining the uncertainties on the mass and redshift distributions drawn directly from the LVK GWTC-3 population analysis [28]. We limit our focus to the MD redshift model for BBHs as this is the only model we have viable posterior samples for: in the case of the low-redshift merger rate parameters (R_0, γ), we use samples from the power-law redshift inference results released in [151] for the power-law redshift model, while when including high-

redshift features ($z_{\max}$, κ) we use results obtained performing inference on the entire MD model as done in [35].

Results obtained progressively varying the redshift hyper-parameters are shown in Figure 12.5 (left panel). We include uncertainty over the entire PLPP mass hyper-parameter space, while varying:

(i) just the local merger rate R_0, assuming a fixed broken-power-law merger rate evolution with parameters fixed to those discussed in Sec. 12.3;

(ii) both R_0 and the local power-law index γ, keeping the high-redshift parameters fixed;

(iii) all parameters for the merger rate, including the turn-over redshift z_{peak} and high-redshift power-law index κ.

In cases (i) and (ii) samples are drawn from the power-law redshift model posteriors of [28]. In case (iii), samples are instead drawn from a full GWTC-3, O1–O2–O3 joint stochastic-population analysis (similar to [35], for details on how this analysis is carried out see [123]) as these also include posteriors on the higher redshift evolution of $R(z)$ [122].

We compare the 95% confidence levels on Ω_{GW} in case (i) with published results (shown in Figure 12.5 in hatched outline[10]) and find these to be consistent. Note that the corresponding LVK contours draw from the same PLPP mass posterior and local merger rate posterior, but assume a different redshift evolution (see the original paper discussion [28]), which explains the small differences between the curves at low frequency and the different turn-over trend at high frequency, which is dominated by low-redshift effects. Specifically, the redshift model used sampled a time-delay distribution between binary formation and merger, where binary formation was fixed to follow the star formation rate of [547], and the time-delay distribution was in the shape of $p(t_d) \propto t_d^{-1}$. This model can not be well-approximated with a simple broken power-law. The LVK contour was calculated assuming the IMRPhenomB analytic waveform model. Case (ii) and case (iii) give almost identical contours, which implies that the population analysis [28] and the stochastic constraints [35] carry little information about the high-redshift evolution of the merger rate. Furthermore, these show that the uncertainty on the local merger rate evolution alone could

[10]The hatched outline is exactly the green region highlighted in Figure 23 of [28], which is publicly available in [151].

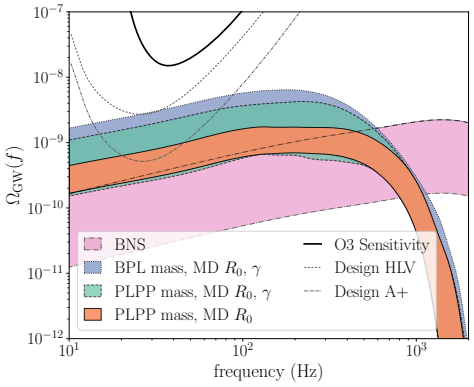

Figure 12.6: Projections of the background Ω_{GW} spectrum, given our knowledge on the compact binary population. Shaded regions outline 90% credible bands for the GWB from BBHs and BNS (in pink), including uncertainty on the mass and redshift models for these sources using samples released in [28, 151]. For BBHs we report uncertainty due to two mass models: the PLPP mass model, assuming a MD redshift model with uncertain local merger rate R_0 (orange), and also uncertain low-redshift power-law index γ (green); and the BPL mass model, assuming a MD redshift model with uncertain R_0 and γ. Current and projected sensitivity curves are included for reference.

account for an increase of up to a factor of ~ 5 in the Ω_{GW} spectrum. This could have considerable consequences on the detectability of the signal. In the right panel of Figure 12.5 we zoom into the $[20, 200]$ Hz portion of the (i) and (ii) spectra, and provide results of single power-law fits to the envelope of Ω_{GW} curves. The average α spectral indices found are consistent with each other, $\alpha_{(i)} = 0.59 \pm 0.02$ for (i), and $\alpha_{(ii)} = 0.60 \pm 0.03$ for (ii). This confirms γ has no impact on the low frequency spectral shape of Ω_{GW}, but only on its amplitude.

We repeat the Ω_{GW} calculation varying R_0 and γ using the BPL mass model, sampling over the joint mass and redshift posteriors obtained by performing population inference on the GWTC-3 catalogue (samples are publicly available as a part of the example sample sets in the `popstock` package repository). We compare these results with the PLPP results described above in Figure 12.6, overlaying the 2σ LVK power-law-integrated sensitivity curves [535] already shown in [28]. These track the present and future upgrades to the LIGO and Virgo facilities, where "O3 sensitivity" is given by the O3 measured spectra, "Design HLV" is produced assuming projections shown in [16] and is expected to approximate the sensitivity at the end of the O4 observing run (currently ongoing), and "Design A+" refers to the sensitivity projected for the next observing run O5 assuming 1 year of continuous data and all

planned improvements to the detectors are implemented and successful [16].

The BPL mass model predicts a systematically larger background, by a factor of 1.4 on average, which hints at the possibility of a louder signal than previously projected and thus the prospect of a detection of the stochastic signal before reaching the Design A+ LIGO–Virgo sensitivity. For completeness, we also include the expected background signal arising from binary neutron stars (BNS) in Figure 12.6. This signal is strongly dominated by uncertainties given the very few detections of BNS mergers [10, 11]. The projection is calculated employing the NRTidalv2 model discussed in [178]. We assume a uniform mass model between $1 - 2.5\ M_{\odot}$, as in [28], and a merger rate model corresponding to a time-delayed SFR as used for projections presented in [35][11]. We draw the local merger rate R_0 from the corresponding posterior samples presented in [28, 151]—for reference, we refer to the samples that set the $R_0 = 105.5^{+190.2}_{-83.9}\mathrm{Gpc}^{-3}\mathrm{yr}^{-1}$ upper limit.

12.6 Conclusions

We present a novel method and code-base to rapidly calculate the background spectrum for inspiralling and coalescing compact binaries starting from a given population model and hyper-parameter sets.

We quantify the joint uncertainty on the Ω_{GW} spectrum from both the mass and redshift distributions of the BBH population, given the most recent results from the LVK collaboration. Predictably, the uncertainty on the local merger rate and its evolution (together with the uncertainty on the mass model) dominate the expected amplitude of the spectrum, and can have significant implications on detectability.

Furthermore, we find that for the preferred mass model (power-law-plus-peak), the low-frequency spectral index of the stochastic background signal is $\alpha = 0.6$. Previous detection approaches assumed $\alpha = 2/3$ (for example [35, 30, 453]); this result was based on the waveform used to calculate the expected GWB and its PN order expansion, as first seen in [425] and then [493]. We find that, when employing 0-PN order waveforms, there is a tension between projections for Ω_{GW} from the presently-observed population which competes with population uncertainty itself. The mismatch between the treatment of the late inspiral phase in IMR waveforms is

[11] This model assumes the BNS progenitor formation rate is proportional to the SFR, and the distribution of time delays between binary formation and merger is inversely proportional to the time delays distributed between 20 Myr and 13.5 Gyr.

particularly significant, as it is present throughout the entire Ω_{GW} spectrum and in particular at lower frequencies, where current detector sensitivities peak.

Differences between specific background realisations and number of samples also produce different projections which may give rise to small tensions in the higher frequency range, where the spectra present a turn-over which is highly dependent on the binary mass distribution and local features of the merger rate. Current-generation detectors are not sensitive to this region of the spectrum, however this will have significant implications for next-generation interferometers.

In conclusion, the specific population properties of the coalescing compact binary population as well as the specific realisation during our observations will play a role in detection capabilities. In particular, within current binary black hole population uncertainties, a low-redshift amplification of the merger rate and a larger population of higher-mass binaries contribute to a significant boost in the background amplitude, in the LVK sensitivity band, which could lead to early detection. On the other hand, more astrophysically-motivated BBH rate evolution models relate the merger rate to binary progenitor features, and re-parameterise the merger rate density in terms of, for example, the time-to-merger delay distribution and the host galaxy metallicity [229, 149]. These models have recently been employed in joint analyses of the GWTC-3 catalogue and LVK stochastic upper limits [544], and may provide alternative forecasts of the uncertainty on the Ω_{GW} spectrum, as we gather more GW data. These models will progressively be included and updated in `popstock`. With `popstock`, we provide the GW community engaged in GW source modelling, data analysis, and astrophysical interpretation with a user-friendly tool for rapid background spectrum evaluation and easy integration of new models as our understanding of the GW universe expands.

Acknowledgements

We thank Patrick Meyers and Alan Weinstein for invaluable discussions and insight. We thank Thomas Callister for providing the population samples in [122], and for carefully reading our work. We thank Nicholas Loutrel for consulting on waveform models and their features. AIR is supported by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 101064542, and acknowledges support from the NSF award PHY-1912594. JG is supported by NSF award No. 2207758. The authors are grateful for computational resources provided by the LIGO Laboratory and supported by Na-

tional Science Foundation Grants PHY-0757058 and PHY-0823459. This material is based upon work supported by NSF's LIGO Laboratory which is a major facility fully funded by the National Science Foundation.

Appendix 12.A Deriving the Energy In GWs

We derive here the energy spectrum dE/df carried by gravitational waves, in vacuum. See, for example, [352, 482, 133]. We start by expanding the perturbed metric $g_{\mu\nu}$ to second order,

$$g_{\mu\nu} = \eta_{\mu\nu} + h^{(1)}_{\mu\nu} + h^{(2)}_{\mu\nu}, \tag{12.14}$$

where $\eta_{\mu\nu}$ is the Minkowski flat metric and it is assumed that the perturbation $h^{(i)}_{\mu\nu}$ is ith order in some small parameter controlling the scale of $h_{\mu\nu}$. Substituting the above into the Einstein field equation gives

$$G_{\mu\nu}\left[h^{(1)}\right] + G_{\mu\nu}\left[\left(h^{(1)}\right)^2\right] + G_{\mu\nu}\left[h^{(2)}\right] = 8\pi G\tau_{\mu\nu}\,, \tag{12.15}$$

Where the first term is the Einstein tensor linear in the first order perturbation, the second is the Einstein tensor terms quadratic in the first order perturbation, and the third term is linear in the second-order perturbation. In vacuum, $\tau_{\mu\nu} = 0$, and the solution for $h^{(1)}$ (i.e. the plane wave solution) reduces the first term to $G_{\mu\nu}\left[h^{(1)}\right] = \Box h^{(1)}_{\mu\nu} = 0$, in the Lorenz gauge. We can therefore rearrange the above into a form that resembles the Einstein field equation,

$$G_{\mu\nu}\left[h^{(2)}\right] = -G_{\mu\nu}\left[\left(h^{(1)}\right)^2\right], \tag{12.16}$$

where the first-order term $h^{(1)}_{\mu\nu}$ squared effectively forms a stress-energy (RHS) that sources the second order curvature (LHS). In this analogy to the RHS of the Einstein equation, we can define the effective stress energy (pseudo)-tensor of GWs [291]:

$$\tau_{\mu\nu} \equiv -\frac{1}{8\pi G}G_{\mu\nu}\left[\left(h^{(1)}\right)^2\right]. \tag{12.17}$$

A nice feature here is that the LHS of Eq. 12.16 satisfies the contracted Bianchi identities and therefore the RHS is divergence-free and can be interpreted as conserving energy according to some observer. Expanding out (12.17) gives

$$\tau_{\mu\nu} = \frac{c^4}{32\pi G}\left\langle \partial_\mu h_{\alpha\beta}\partial_\nu h^{\alpha\beta}\right\rangle\,. \tag{12.18}$$

The conservation law $\partial_\mu t^{\mu\nu} = 0$ implies

$$\partial_0\tau^{00} + \partial_i\tau^{i0} = 0\,, \tag{12.19}$$

where τ^{00} can be interpreted as a volumetric energy density, such that the energy E is defined as $E = \int d^3x\tau^{00}$ and therefore the associated power is

$$\frac{dE}{dt} = \partial_0 \int d^3x\tau^{00}\,. \tag{12.20}$$

Substituting into Eq. 12.19 yields

$$\frac{dE}{dt} + \int d^3x\partial_i\tau^{i0} = 0\,, \tag{12.21}$$

which simplifies to

$$\frac{dE}{dt} + \int d^3x\partial_z\tau^{z0} = 0 \tag{12.22}$$

for a wave moving along direction $\hat{z}$. We employ the divergence theorem to convert the volume integral into a surface integral,

$$\frac{dE}{dt} + \hat{z}\int dA\tau^{z0} = 0\,, \tag{12.23}$$

which gives an expression for the energy flux (energy per unit time per unit area):

$$\frac{dE}{dAdt} = -\tau^{z0}\hat{z} \equiv -\tau^{00}\hat{z}\,, \tag{12.24}$$

as $\partial_0 h_{ij} = -\partial_z h_{ij} = -\partial^0 h_{ij}$ holds for a wave solution. Considering our gauge, the only surviving terms are those for $\mu, \nu = 1$ or 2,

$$\begin{aligned}\tau^{00} = \frac{c^4}{32\pi G}\big\langle \partial_0 h_{11}\partial_0 h^{11} + \partial_0 h_{12}\partial_0 h^{12} + \partial_0 h_{21}\partial_0 h^{21} + \\ + \partial_0 h_{22}\partial_0 h^{22}\big\rangle\,,\end{aligned} \tag{12.25}$$

and substituting in the polarization components of the wave $h_{\mu\nu}$ in the TT gauge yields

$$\tau^{00} = \frac{c^2}{16\pi G}\left\langle |\dot{h}_+|^2 + |\dot{h}_\times|^2\right\rangle\,. \tag{12.26}$$

Solving for the energy flux of gravitational waves of Eq. (12.23) gives

$$\left|\frac{dE}{dAdt}\right| = -\frac{c^3}{16\pi G}\left\langle |\dot{h}_+|^2 + |\dot{h}_\times|^2\right\rangle\,, \tag{12.27}$$

where we have switched to the absolute value of this quantity with the understanding that GWs are removing energy from the system. The surface area energy density is then defined as

$$\frac{dE}{dA} = \int dt\frac{c^3}{16\pi G}\left\langle \dot{h}_+(t)^2 + \dot{h}_\times(t)^2\right\rangle\,. \tag{12.28}$$

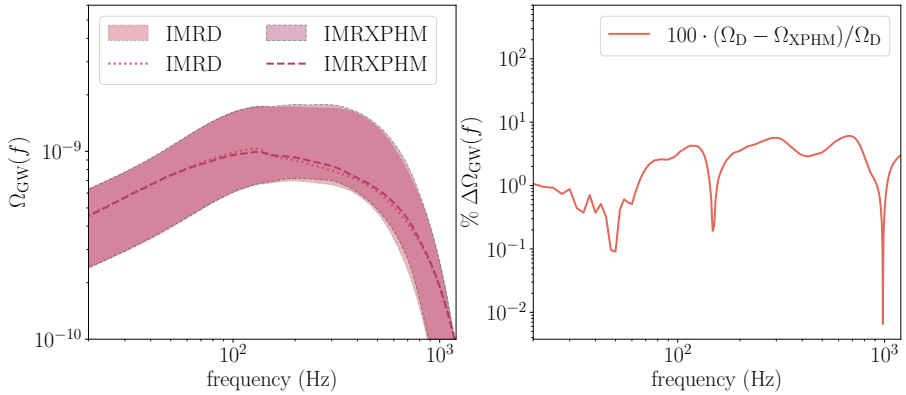

Figure 12.7: Impact of the inclusion of higher order modes in the waveform model employed to evaluate the Ω_{GW} spectrum. On the left: 95% confidence on the spectrum including the uncertainty on the PLPP mass model and the local merger rate, assuming a fixed MD redshift evolution. On the right: percent difference $\%\Delta\Omega_{GW}(f)$ between Ω_{GW} spectra calculated using the same event samples, shown as dashed and dotted curves on the left panel.

To expand the above we recall that, for a plane wave,

$$\dot{\tilde{h}}(f) = \int dt \dot{h}(t) e^{-i\omega t} = i\omega \int dt h(t) e^{-i\omega t} = i\omega \tilde{h}(f)\,, \tag{12.29}$$

where in the final equivalence we have directly employed the definitaion of a Fourier transform. Recalling Parseval's theorem, we can write the surface area energy density in terms of the Fourier transform $\tilde{h}(f)$,

$$\begin{aligned} \frac{dE}{dA} &= \int_{-\infty}^{\infty} df \frac{c^3\omega^2}{16\pi G} \left\langle \tilde{h}_+(f)^2 + \tilde{h}_\times(f)^2 \right\rangle \\ &= \frac{\pi c^3}{2G} \int_0^{\infty} df f^2 \left\langle \tilde{h}_+(f)^2 + \tilde{h}_\times(f)^2 \right\rangle, \end{aligned} \tag{12.30}$$

where the integral is taken over the sphere surrounding the source. Note that h_+ and $h_\times$ terms include dependence on inclination ι and reference phase ϕ_0 (or, equivalently, the observer's position along the azimuth) and therefore must be included in the integral over the area.

Appendix 12.B Impact of higher order modes on the GWB spectrum

For the sake of completeness, we append here findings on the impact on the Ω_{GW} spectrum due to the inclusion of higher order modes in the waveform model. Higher order modes are subdominant harmonics excited during GW emission, where the dominant harmonic is the $\ell = 2, m = 2$ mode [345]. We compare the 95% confidence

bands shown in 12.5.2 for the IMRPhenomD waveform model with bands obtained using the IMRPhenomXPHM approximant [428], which includes the $(\ell, |m|) = (2, 2), (2, 1), (3, 3), (3, 2), (4, 4)$ modes. As may be seen in Figure 12.7, for the population of binaries considered in 12.5.2 which in particular is non-spinning and non-precessing, there are negligible differences between the use of IMRPhenomD and IMRPhenomXPHM. This is particularly evident when comparing the Ω_{GW} spectrum calculated using the same event samples employing the two waveforms, shown as dashed and dotted curves in the left panel of Figure 12.7. The percent difference between these two curves is reported in the right panel of Figure 12.7, which remains consistently below 10% across the spectrum and is under 3% for frequencies below 100 Hz.

Chapter 13

THE ANGULAR POWER SPECTRUM OF GRAVITATIONAL-WAVE TRANSIENT SOURCES AS A PROBE OF THE LARGE-SCALE STRUCTURE

Yanyan Zheng, Nikolaos Kouvatsos, Jacob Golomb, et al. "Angular Power Spectrum of Gravitational-Wave Transient Sources as a Probe of the Large-Scale Structure". In: *Phys. Rev. Lett.* 131.17 (2023), p. 171403. DOI: 10.1103/PhysRevLett.131.171403. arXiv: 2305.02652 [astro-ph.CO].

KCL-PH-TH-2023-25

Abstract

We present a new, simulation-based inference method to compute the angular power spectrum of the distribution of foreground gravitational-wave transient events. As a first application of this method, we use the binary black hole mergers observed during the LIGO, Virgo, and KAGRA third observation run to test the spatial distribution of these sources. We find no evidence for anisotropy in their angular distribution. We discuss further applications of this method to investigate other gravitational-wave source populations and their correlations to the cosmological large-scale structure.

13.1 Introduction

Since the first detection of a gravitational wave (GW) signal from a binary black hole (BBH) coalescence in 2015 [14], LIGO, Virgo, and KAGRA (LVK) have detected dozens more such signals during the first three observation runs [26]. At the end of the next two observation runs, the number of detections is expected to reach the thousands [16]. This abundance of detected events will allow us to continuously refine our knowledge of the GW emitters.

In this context, an area of growing interest is the measurement of the spatial distribution of GW (SDGW) transient sources and its relation to the large-scale structure (LSS) of the universe [194, 389, 78, 415]. The SDGW provides a means to test the LSS that is complementary to electromagnetic measurements as well as dark siren analyses[5, 21], which rely on cross-referencing GW detections with galaxy catalogs and are prone to complications such as catalog incompleteness and selection bias. Developing a scheme to accurately measure the SDGW constitutes one of the critical milestones towards precision cosmology with GWs [76].

In this paper, we present a novel, simulation-based inference method to test the SDGW that borrows from techniques used in electromagnetic precision cosmology, in particular the study of the cosmic microwave background radiation (CMB). Specifically, we show how to calculate the observed angular power spectrum of *foreground* GW events and use it to probe the SDGW. This technique provides complementary information to analogous studies based on the astrophysical GW *background*, where the angular power spectrum is derived from the clustering statistics of the BBH host galaxies [297, 295, 296, 93, 92].

As a first application of our method, we test the isotropic source distribution hypothesis for the confident BBH mergers observed during the third LVK observing run (O3). However, it should be stressed that our approach is not limited to this specific instance. The technique that we present here can be easily generalized to various GW sources, future GW searches with additional detections, and different test hypotheses on the SDGW and its correlation with the LSS.

In the next two sections, we discuss the basics of our method, the selection of GW events, the generation of synthetic signals to test the isotropic hypothesis, and the production of sky localization maps via parameter estimation. In the last two sections, we present the main results and discuss future extensions of this work.

13.2 Methodology

Our method probes the spatial distribution of BBH merger events by computing their observed angular power spectrum [135] and comparing it to a fiducial distribution. In this work, we select the isotropic distribution, which corresponds to testing whether BBHs are isotropically distributed in the local universe. First, we compute the power spectrum of observed BBH events from the LVK GW catalogs. We choose a suitable subset of these events by imposing the selection cuts detailed in the next section. Then, we compute the power spectra of a number of mock sets obtained by injecting synthetic signals into real detector data. We sample their parameters from the latest LVK population analysis posterior distributions [28] and inject the signals isotropically on the sky. We then select a subset of events by imposing the same selection cuts used for the observed BBH mergers. The synthetic power spectra are combined to produce a fiducial distribution of an isotropically distributed angular power spectrum as would be measured by the LVK detectors. Finally, we perform a statistical consistency test of the observed BBH angular distribution with the fiducial isotropic distribution; for each multipole component of the power spectrum, we compute the p-value that the observed multipole belongs to the fiducial distribution.

We consider the subset of BBH events detected during the LVK O3 observing run with a false alarm rate (FAR) smaller than 1 yr^{-1} as reported in Ref. [28]. We further restrict our sample to three-detector events. This is required for the generation of a consistent fiducial angular distribution, as the accuracy of sky localizations depends on the number of detectors [499]. These conditions restrict the sample of O3 events to 34. These events constitute our catalog of observed signals. To generate the synthetic signals, we draw their source parameters from their inferred median population distributions [28], assuming the Power Law + Peak model for the primary mass [524] with a power law on mass ratio, the `Default` spin model [521, 582], and a power law model for redshift evolution [228]. The phase and orientation parameters are sampled from distributions with isotropic orientations. We inject the signals into real detector data with an isotropic distribution in the sky. The times of the injections are uniformly sampled during O3. We then downselect these times to periods that do not overlap with known non-astrophysical transient noise [171] and GWTC-3 confident detections [26]. The signals are simulated with the `IMRPhenomPv2` [267, 490] waveform model. Selecting the synthetic events based on their FAR is computationally expensive, as it requires doing PE for the full set

of events. To avoid this computational cost, we substitute the FAR selection cut with a threshold on the optimal network signal-to-noise ratio (SNR) ρ_N. We choose $\rho_N > 10$, following the approximate threshold used for the semianalytic sensitivity estimates in Ref. [28].

We generate a catalog of 3,400 synthetic events. This allows us to produce meaningful statistical results while limiting the computational cost required to perform PE and generate the sky localization maps. We use the synthetic signals to create 100 random mock sets of 34 events each. These sets provide independent realizations of what the detectors would observe under the hypothesis that the events are isotropically distributed in the sky. We use these sets to generate the fiducial distribution. We perform PE of all observed and synthetic events with `bilby pipe` [67]. We use the `IMRPhenomPv2` waveform for the signal model and draw the samples from the posterior distribution with the nested sampler `dynesty` [507].

We adopt the standard LVK uniform priors on the mass ratio and chirp mass from Ref. [26]. We restrict the chirp mass to a $\pm 12 M_\odot$ range around the injected values of the synthetic events and the median values of the O3 observed events. Additionally, we constrain the priors on the primary and secondary masses to be within the interval [1, 120] $M_\odot$. The prior on all other parameters is chosen according to the uninformative priors adopted in standard LVK analyses [26]. We then use the posterior samples for the declination and the right ascension to produce sky maps.

13.3 Angular Power Spectrum

Following Ref. [135], we treat the event sky localization error regions as probability density heat maps. We generate the combined sky localization map of the observed GW events, $M(\chi, \phi)$, by stacking the sky localization density maps of all events in the observed catalog. Here, χ and ϕ are the polar and azimuthal angles on the celestial sphere, respectively. Figure 13.1 shows the Mollweide representation of $M(\chi, \phi)$. We repeat this procedure to obtain a cumulative sky localization map for each set of synthetic events. Figure 13.2 shows the combined sky localization map obtained by stacking the 100 synthetic maps, each made from 34 events. The map shows that the synthetic events are isotropically distributed in the sky. It also depicts what the GW sky would look like with 3400 foreground BBH events, a not too unrealistic scenario in a few years.

We then compute the angular power spectra of the combined sky localization maps

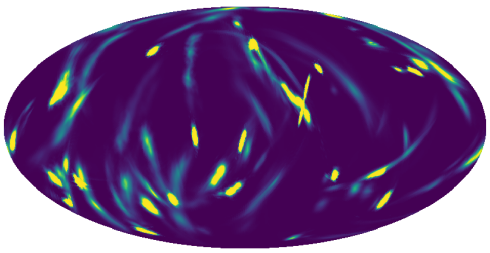

Figure 13.1: Combined sky localization map of the O3 BBH events considered in the analysis. The sky localization of each event is generated with Bayestar [499] from the PE posterior samples for the declination and the right ascension. The map is created with the `Healpy` package [600, 259].

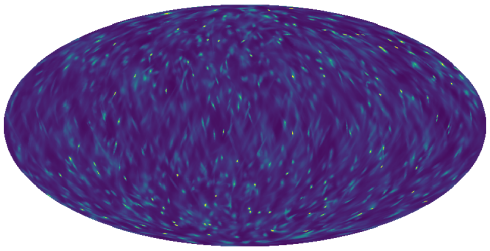

Figure 13.2: Combined sky localization map of the synthetic BBH events that are used to build the fiducial power spectrum. Their isotropic distribution in the sky is shown by the map.

by expanding each of them into spherical harmonics:

$$M(\chi, \phi) = \sum_{lm} \alpha_{lm} Y_{lm}(\chi, \phi) \,. \tag{13.1}$$

The multipole components of the angular power spectrum are obtained by summing the absolute square of the α_{lm} coefficients of the expansion over m:

$$C_l = \frac{1}{2l+1} \sum_m |\alpha_{lm}|^2 \,. \tag{13.2}$$

The physical information contained in the power spectrum can also be expressed in terms of the two-dimensional (angular) correlation function (CF). The CF describes the excess probability of finding two objects in the directions $\hat{n}_1$ and $\hat{n}_2$ and angular separation θ with respect to a uniform distribution. Given the cumulative sky

localization map $M(\chi, \phi)$, the CF is defined as $C(\theta) = \langle M(\hat{n}_1) \cdot M(\hat{n}_2) \rangle_{21}$, where the average is taken over the observed sky with angular separation held fixed [135]. The CF can be written in terms of the power spectrum as

$$C(\theta) = \frac{1}{4\pi} \sum_l (1 + 2l) C_l P_l(\cos\,\theta)\,, \tag{13.3}$$

where $P_l(\cos\theta)$ denotes the Legendre polynomial of order l and argument $\cos\theta$. Typically, the finite beam resolution of the detectors leads to a high-l cutoff $l_{\rm max}$ in Eq.(13.3). This effect can be modeled by introducing a window function $W_l \propto \exp\left[-l(l+1)\sigma^2_{\rm res}\right]$, where $\sigma_{\rm res}$ is the detector resolution [568].

The diffraction-limited angular resolution of the LIGO-Virgo network determines the high-l cutoff as $l_{\rm max} \sim \pi/\theta_{\rm res}$, where $\theta_{\rm res}$ is the angular resolution. We estimate $l_{\rm max}$ directly from the distributions of the skymaps. We fit the distribution of the observed skymap 90% contour regions as a proxy for the square angular resolution $\Delta\Omega_{\rm res} = 2\pi[1 - \cos(\theta_{\rm res}/2)]$ with a gamma distribution. We then perform a one-tailed test and choose $\Delta\Omega_{\rm res}$ such that 90% of the observed events have a larger localization area than that value. This provides an estimate for the angular resolution of $\theta_{{\rm res},o} \sim 6.95^\circ$, corresponding to $l_{{\rm max},o} \sim 26$. We then repeat the procedure for the whole set of synthetic events. This yields $\theta_{{\rm res},s} \sim 4.83^\circ$, corresponding to $l_{{\rm max},s} \sim 37$. The resolution of the simulated set is better than the resolution of the observed set. We expect this is due to the larger number of events in the simulated set compared to the observations. As a consistency check, we also estimate $\Delta\Omega_{\rm res}$ using the theoretical estimate of Ref. [566]. For a monochromatic GW at frequency f, the square angular resolution of a three-detector network is

$$\Delta\Omega_{\rm res} \approx 8\left(\frac{150{\rm Hz}}{f}\frac{10}{\rho_{\rm N}}\right)^2 \frac{10^{17}{\rm cm}^2}{A_{\rm N}} \frac{1/27}{\rho_1^2\rho_2^2\rho_3^2/\rho_{\rm N}^6} \frac{\sqrt{2}/2}{|\sin\,i_{\rm N}|}\,, \tag{13.4}$$

where A_N is the triangular area formed by the three detector sites, i_N is the angle between the wave direction and the three-detector plane, ρ_N is the network optimal SNR of the GW signal, and ρ_i (i =1,2,3) are the single-detector SNRs. We consider a triangular area $A_{\rm N} = 10^{17}{\rm cm}^2$ for the LIGO-Virgo network and a mean incidence angle of 45° with the detector plane. We use the posterior sample median values to estimate the SNRs and approximate f with the ISCO frequency obtained from the posterior median chirp mass and mass ratio. Using the means of the SNRs and f in Eq. (13.4), we obtain the angular resolution $\theta_{{\rm res},o} \sim 4.04^\circ$ for the observed events and $\theta_{{\rm res},s} \sim 4.44^\circ$ for the synthetic events, corresponding to $l_{{\rm max},o} \sim 45$ and $l_{{\rm max},s} \sim 41$, respectively. The theoretical estimate gives higher bounds than the data

sets. This is expected, as Eq. (13.4) is derived under optimal assumptions and a Fisher approximation. In the following, out of an excess of caution, we will use $l_{\rm max} = 26$ as a conservative upper bound.

13.4 Results

Figure 13.3 shows the power spectrum of the observed events (red curve) and the mean spectrum of the 100 synthetic sets (black curve) up to $l_{\rm max} = 26$. For each l, we fit the C_l distribution from the synthetic sets with a gamma distribution. The three gray-filled areas in Figure 13.3 (darker to lighter gray) denote the $1 - 3\sigma$ confidence level regions from the mean. All observed C_l values lie within the 2σ band. Therefore, we conclude that the observed angular distribution of observed BBH events shows no significant inconsistencies relative to an isotropic distribution. To quantify this statement, we performed two statistical tests. In the first test, we compute the cumulative distributions of p-values for the observed C_l under the hypothesis that the BBH are distributed isotropically in the sky.

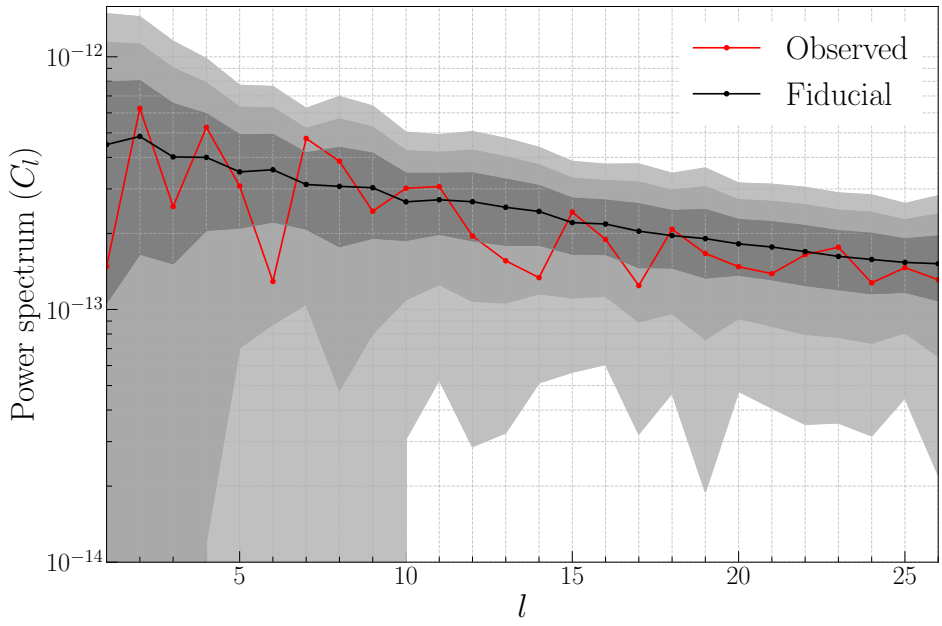

Figure 13.3: The observed power spectrum of the O3 BBH events considered in the analysis (red curve) and the fiducial power spectrum obtained from the 100 synthetic sets under the isotropic hypothesis (black curve). The gray-filled regions denote $1 - 3\sigma$ deviations from the mean.

Figure 13.4 shows the cumulative distributions of p-values (red dots). The expected distribution is represented by the black dashed line, with the gray-filled regions

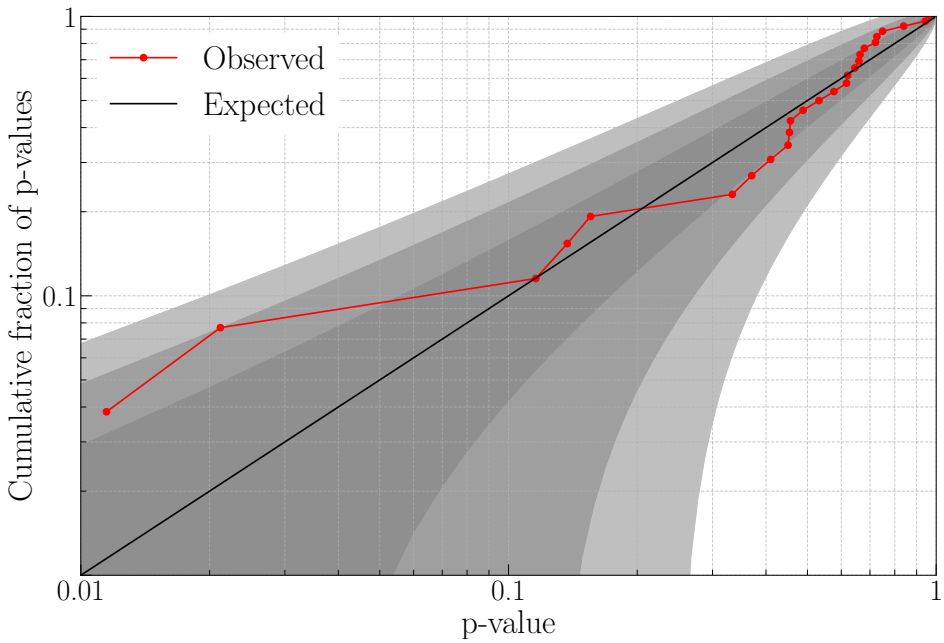

Figure 13.4: The cumulative distribution of observed p-values for the C_l. The black solid line indicates the expected distribution under the isotropic hypothesis. The gray-filled regions correspond to $1-3\sigma$ deviations from the expected distribution.

denoting the $1-3\sigma$ confidence levels. All p-values lie within the 2σ region, in agreement with the results of Figure 13.3. In the second test, we assess the goodness of fit of the observed power spectrum with the fiducial spectrum by performing a χ^2 test, which yields a p-value of 0.82, in agreement with the null isotropic hypothesis.

Finally, we test the isotropy hypothesis with the CF. Figure 13.5 shows the CF for the observed set and the fiducial correlation function obtained from the 100 synthetic sets under the isotropic hypothesis, where we have set the window function resolution to $\sigma_{\rm res} = l_{\rm max}$. Consistent with the power spectrum result, the observed CF is in agreement with the fiducial isotropic distribution within 2σ. The behavior of the CF at small scales, $C(\theta) = (\theta/\theta_0)^{1-\gamma}$, provides a test of isotropy [135]. We first compute the power-law slope γ of each synthetic CF at the minimum angular resolution $\theta_{{\rm res},s}$ with a log-log fit. Averaging the values, we obtain a fiducial value of $\gamma_{\rm s} = 2.05 \pm 0.35$, which is consistent with an isotropic distribution ($\gamma = 2$). We then compute the power-law slope for the observed set at the same angular scale. The observed power-slope is $\gamma_{\rm o} = 1.96$. This is in agreement with the null isotropic hypothesis with a p-value of 0.45.

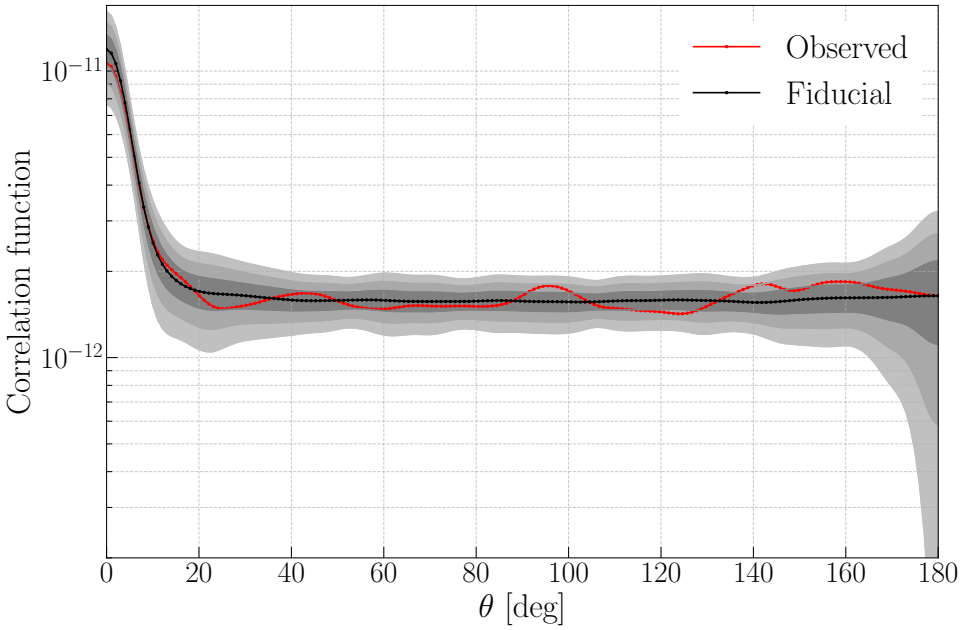

Figure 13.5: The observed correlation function of the O3 BBH events (red curve) and the fiducial correlation function under the isotropic hypothesis (black curve). The gray-filled regions denote $1-3\sigma$ deviations from the mean.

13.5 Conclusions

In this paper, we have developed a new, simulation-based inference framework to probe the spatial distribution of observed, foreground GW events. Our approach compares the power spectrum of observed GW signals to a fiducial power spectrum from a theoretical distribution. As an application of this method, we tested the isotropy hypothesis of the BBH mergers observed during the O3 LVK observing run. As foreseen [194, 389, 78, 415], we found no evidence of anisotropy at the 2σ confidence level.

Our method provides a powerful framework for testing the universe's LSS that complements current GW background searches [30, 20]. Due to the phase-coherence of matched-filter searches employed in GWTC-3 [26], we are able to access higher multipole moments than background searches [456]. Relying on resolved sources allows us to achieve astrometric resolution at the square degree level [76]. Although the two approaches essentially target the same signal in the limit of many detections, our method has a higher resolution and is more sensitive than background analyses.

A first, straightforward extension of this work is to refine the test of BBH isotropy as more GW events are discovered. Tests of specific theoretical models of anisotropic

distributions and cross-correlations with astrophysical populations in the EM domain are two additional applications. Our approach can also be directly extended to include information about the source distances. Statistical associations between the observed GW populations and other extragalactic populations may be within reach of current and next-generation GW detectors. This method will provide a means to rapidly detect and quantify any such associations.

Acknowledgements

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 acknowledge computational resources provided by the LIGO Laboratory and supported by NSF Grants PHY-0757058 and PHY-0823459. M.C. and Y.Z. are partially supported by the U.S. National Science Foundation (NSF) through award PHY-2011334. N.K. is supported by King's College London through an NMES Funded Studentship. A.I.R. is supported by the NSF award PHY-1912594. M.S. is supported in part by the Science and Technology Facility Council (STFC), United Kingdom, under the research grant ST/P000258/1. The authors thank Joe Romano for carefully reading the manuscript and providing useful comments. We also thank Sylvia Biscoveanu for discussions regarding PE using `bilby_pipe`, Derek Davis for discussion on detection statistics, Stuart Anderson for helping provide computing resources, and Reed Essick for useful comments on the manuscript. This manuscript was assigned LIGO-Document number P2300106.

Chapter 14

GROWING PAINS: UNDERSTANDING THE IMPACT OF LIKELIHOOD UNCERTAINTY ON HIERARCHICAL BAYESIAN INFERENCE FOR GRAVITATIONAL-WAVE ASTRONOMY

Colm Talbot and Jacob Golomb. "Growing pains: understanding the impact of likelihood uncertainty on hierarchical Bayesian inference for gravitational-wave astronomy". In: *Mon. Not. Roy. Astron. Soc.* 526.3 (2023), pp. 3495–3503. DOI: 10.1093/mnras/stad2968. arXiv: 2304.06138 [astro-ph.IM].

Abstract

Observations of gravitational waves emitted by merging compact binaries have provided tantalising hints about stellar astrophysics, cosmology, and fundamental physics. However, the physical parameters describing the systems, (mass, spin, distance) used to extract these inferences about the Universe are subject to large uncertainties. The most widely-used method of performing these analyses requires performing many Monte Carlo integrals to marginalise over the uncertainty in the properties of the individual binaries and the survey selection bias. These Monte Carlo integrals are subject to fundamental statistical uncertainties. Previous treatments of this statistical uncertainty has focused on ensuring the precision of the inferred inference is unaffected, however, these works have neglected the question of whether sufficient accuracy can also be achieved. In this work, we provide a practical exploration of the impact of uncertainty in our analyses and provide a suggested framework for verifying that astrophysical inferences made with the gravitational-wave transient catalogue are accurate. Applying our framework to models used by the LIGO-Virgo-KAGRA collaboration and in the wider literature, we find that Monte Carlo uncertainty in estimating the survey selection bias is the limiting factor in our ability to probe narrow population models and this will rapidly grow more problematic as the size of the observed population increases.

14.1 Introduction

Using data from the first three observing runs of Advanced LIGO [1] and Advanced Virgo [38] ≈ 70 signals from the merger of compact binary systems have been identified [26], along with a few tens of less-significant additional candidate events [398, 406]. While individual observations of compact binary mergers provide insights into astrophysics and cosmology, maximising the physical resolving power using the catalogue of gravitational-wave transients requires analysing the entire population as a hierarchical Bayesian inference problem. Due to computational constraints, these analyses are performed using a multi-stage process to calculate the population-level likelihood [see, e.g., 28, 536, 560, 359].

First, segments of data that are likely to contain gravitational-wave signals are identified by search pipelines [e.g., 56]. These pipelines are only sensitive to the loudest signals and so the observed sample is biased in favour of nearby high-mass binaries with black hole angular momenta ("spins") aligned with the orbital angular momentum [129]. This selection bias is typically accounted for by estimating the fraction of binaries that we expect to observe using simulated "injection" campaigns [193, 538, 213].

Next, the strain data from gravitational-wave detectors containing the observed transients is analysed with a fiducial reference model for the population (often referred to as the fiducial prior distribution) in order to obtain samples from the fiducial posterior probability distribution for the parameters (masses, spins, etc.) of each binary. While the fiducial prior distribution impacts the fiducial posterior, it is typically chosen to avoid imprinting astrophysical assumptions on the results. For example, binaries are assumed to be distributed homogeneously and isotropically throughout the Universe. The fiducial model for black hole masses is usually uniform in the mass of each black hole and uniform in spin magnitude and isotropic in direction.

In the final stage, these fiducial samples are importance sampled ("reweighted") using a parameterised model for the underlying population to compute the likelihood for the observed data given population-level parameters (e.g., the maximum allowed black hole mass) marginalised over the per-event parameters. For each model for the underlying population, the fraction of observable binaries is also estimated using importance sampling on the injected signals from the first stage [e.g., 222, 348, 216].

The importance sampling step is an example of using Monte Carlo summation to approximate an integral and as such comes with some intrinsic uncertainty that enters the analysis as a source of systematic error. Typically, this uncertainty is ignored when performing the analysis, however, in recent years several attempts have been made to quantify this uncertainty and theoretically motivated heuristics have been proposed to estimate and (hopefully) mitigate its impact [213, 193]. In this work, we perform a data-driven analysis of the potential systematic uncertainties from our use of Monte Carlo integration. We note that while we apply our formalism to the problem of population inference for gravitational-wave astronomy, it is widely applicable to any context in which an approximate estimator of the true likelihood is used in a Bayesian analysis.

The remainder of this paper is structured as follows. In the next Section, we describe how uncertainty appears in our estimate of the population likelihood through Monte Carlo integration and suggest a set of convergence criteria. In Section 14.3, we analyse a simple toy model to examine the impact of uncertainty on the accuracy of inference. Using this, we establish a threshold beyond which we expect our results to be significantly biased. Following this, we take a range of models previously considered for population analyses and quantify the uncertainty in these results in Section 14.4. Finally, we provide a closing discussion.

14.2 Uncertainty in the Population Likelihood Approximation

The likelihood function typically employed for an analysis of a population of N observed systems with source-dependent selection effects can be written [see, e.g., 536, 560, 359, for details]

$$\mathcal{L}(\{d_i\}|\Lambda) \propto \prod_i^N \frac{\mathcal{L}(d_i|\Lambda)}{P_{\rm det}(\Lambda)}. \tag{14.1}$$

Here, the $\{d_i\}$ are the data containing the observed signals (indexed by i). In the context of gravitational-wave astronomy, this is strain data recorded by gravitational-wave interferometers. The selection function $P_{\rm det}$ is the fraction of all signals that would be observed for a population described by population hyper-parameters Λ. We note that this likelihood has been marginalised over the overall rate of events (assuming a uniform-in-log rate prior) and the parameters describing each of the individual systems.

Each of the terms $\mathcal{L}(d_i|\Lambda)$ and $P_{\rm det}(\Lambda)$ are computed by marginalising over θ, the ≈ 15 parameters describing the individual binaries and many more describing the

noise properties of the interferometers

$$\mathcal{L}(d_i|\Lambda) = \int d\theta p(d_i, \theta|\Lambda) = \int d\theta \mathcal{L}(d_i|\theta) p(\theta|\Lambda) \tag{14.2}$$

$$P_{\rm det}(\Lambda) = \int dd \int d\theta p(d, \theta|\Lambda) \Theta(\rho(d) - \rho_*) \tag{14.3}$$

$$= \int dd \int d\theta \mathcal{L}(d|\theta) p(\theta|\Lambda) \Theta(\rho(d) - \rho_*). \tag{14.4}$$

In both expressions, we have expanded the joint distribution for the observed data and signal parameters into the population model $p(\theta|\Lambda)$ and the likelihood of observing data given single-event parameters $\mathcal{L}(d|\theta)$. The integral over d in the expression for $P_{\rm det}$ is over all of the data collected by the instrument while the d_i represents the data around the time of a specific observed signal. The final term is a Heaviside step function for the detection statistic (e.g., signal-to-noise ratio or false-alarm rate) ρ with threshold ρ_*. In order to minimise the cost of performing the analysis, these integrals are commonly computed using Monte Carlo estimators using some reference set of samples from the fiducial posterior distribution. We denote the estimator of quantity x as $\hat{x}$. As a specific example, the estimator of the log-likelihood (Eq. 14.1) is

$$\ln \hat{\mathcal{L}}(\{d_i\}|\Lambda) = \left(\sum_i^N \ln \hat{\mathcal{L}}(d_i|\Lambda) \right) - N \ln \hat{P}_{\rm det}(\Lambda). \tag{14.5}$$

In practice, these estimates are calculated using Monte Carlo integration:

$$I = \int dx f(x) p(x) \equiv \langle f \rangle_{p(x)} \tag{14.6}$$

$$\hat{I} = \frac{1}{M} \sum_{x_j \sim p(x)}^{j=M} f(x_j). \tag{14.7}$$

Here $\hat{I}$ is the estimator of the integral I and M is the number of samples in the Monte Carlo integral. We note that $p(x)$ is a normalised probability distribution and $f(x)$ is an arbitrary function of parameters x. Every Monte Carlo has an intrinsic statistical uncertainty

$$\sigma_I^2 = \frac{1}{M} \left[\langle f^2 \rangle_{p(x)} - \langle f \rangle_{p(x)}^2 \right] \equiv \frac{1}{M} \bar{\sigma}_I^2. \tag{14.8}$$

We define the quantity $\bar{\sigma}_I^2$ as the intrinsic variance between the proposal distribution $p(x)$ and the target distribution $f(x)p(x)$. In general, the uncertainty in a Monte Carlo integral will be minimised by choosing $p(x)$ and $f(x)$ to minimise $\bar{\sigma}_I$. For

example, for most gravitational-wave population analyses (including this work) we choose

$$f(\theta) \sim \frac{p(\theta|\Lambda)}{p(\theta|\varnothing)}, \quad p(\theta) \sim \mathcal{L}(d|\theta)p(\theta|\varnothing),$$

where $p(\theta|\varnothing)$ is the fiducial prior distribution. However, in some cases it is beneficial to define [e.g., 582, 256] $f(\theta) \sim \mathcal{L}(d|\theta)$, $p(\theta) \sim p(\theta|\Lambda)$. We also note that the variance scales inversely with the number of samples. A final quantity related to Monte Carlo integrals that we will need is the effective number of independent samples [312]

$$n_{\mathrm{eff}} = M \frac{\langle f \rangle^2_{p(x)}}{\langle f^2 \rangle_{p(x)}}. \tag{14.9}$$

In [213] the author demonstrates that for small values of n_{eff} a Gaussian approximation to the likelihood uncertainty breaks down. In previous works [e.g., 213, 28], n_{eff} has been used to assess the convergence of the likelihood estimator and to impose data-dependent cuts on the allowed parameter space. We prefer to work directly with the estimated variance and include n_{eff} here just to compare with previous work.

Since we assume that the reference samples used in each of the Monte Carlo integrals are independent, the variance in the estimate of the population (log-)likelihood is

$$\sigma^2_{\ln \hat{\mathcal{L}}}(\Lambda) = \sum_i^N \sigma^2_{\ln \hat{\mathcal{L}}_i}(\Lambda) + N^2 \sigma^2_{\mathrm{sel}}(\Lambda). \tag{14.10}$$

We note that the contribution to the total variance from the selection function grows quadratically with the population size, as $\mathrm{Var}(Nx) = N^2 \mathrm{Var}(x)$.

Assuming the individual observations are independent and identically distributed draws from the underlying population, we recast this expression in terms of an average per-observation uncertainty σ_{obs} to more clearly see the dependence of both terms with the population size

$$\sigma^2_{\ln \hat{\mathcal{L}}}(\Lambda) = N \sigma^2_{\mathrm{obs}}(\Lambda) + N^2 \sigma^2_{\mathrm{sel}}(\Lambda). \tag{14.11}$$

We have explicitly retained the dependence of this variance on the hyperparameters. We justify the assumption that σ_{obs} does not vary with time in Section 14.4.1.

Since we are predominantly interested in differences in log-likelihood for points with significant posterior support, we need to limit the error in the difference of log-likelihood estimators, $\Delta \ln \hat{\mathcal{L}}$. In general, the errors will not be independent, and so we calculate the variance in this quantity $\sigma^2_{\Delta \ln \hat{\mathcal{L}}}$ as defined in Eq. A11 in [193]. We

assume the error in the estimator of the log-likelihood is Gaussian distributed as the contribution to the population log-likelihood from the per-event terms is the sum of N independently and identically distributed estimators and so by the central limit theorem follows a normal distribution and in the high effective-sample size limit the selection function term also follows a normal distribution [213] We therefore write $\sigma^2_{\Delta \ln \mathcal{L}} = \sigma^2_{\Delta \ln \hat{\mathcal{L}}}$.

If the uncertainties in the estimators are uncorrelated with Λ, we will have $\sigma^2_{\Delta \ln \hat{\mathcal{L}}} = \sigma^2_{\ln \hat{\mathcal{L}}}$. In [193], the authors demonstrate that under certain conditions the variance in likelihood differences in "local neighbourhoods" avoids the worst-case scaling in Eq. 14.11 and rather find that

$$\sigma^2_{\Delta \ln \hat{\mathcal{L}}} = \sigma^2_{\rm obs}(\Lambda) + N\sigma^2_{\rm sel}(\Lambda) \tag{14.12}$$

for a simple example model due to correlation of the Monte Carlo errors between points with significant posterior support. It is unclear a priori when the local neighbourhood approximation is valid, in this work, we numerically test whether this scaling holds for the specific case of inferring the population properties of merging binary black hole systems.

14.2.1 Uncertainty as a draw from a Gaussian process

To build an understanding of the impact of uncertainty, we assert that the estimated difference in log-likelihood is a fair draw from the Gaussian process with mean function $\Delta \ln \mathcal{L}$ and (potentially non-stationary) kernel function $\Sigma(\Lambda, \Lambda')$

$$\Delta \ln \hat{\mathcal{L}}(\{d_i\}|\Lambda, \Lambda') \sim \mathcal{GP}(\Delta \ln \mathcal{L}(\{d_i\}|\Lambda, \Lambda'), \Sigma(\Lambda, \Lambda')). \tag{14.13}$$

Here $\Sigma(\Lambda, \Lambda') = \sigma^2_{\Delta \ln \mathcal{L}}$ is the $2D$-dimensional covariance matrix, where D is the dimensionality of the population model. In practice, we do not have access to the true kernel function, and so we approximate it using a numerical covariance matrix using the covariance between the likelihood estimator at each pair of points we consider. Specifically, we construct the approximate kernel by numerically calculating

$$\Sigma(\Lambda, \Lambda') = \sigma^2_{\Delta \ln \hat{\mathcal{L}}} \tag{14.14}$$

following Eq. A11 in [193]. We will use this quantity to estimate the average variance over the posterior for the hyperparameters

$$\left\langle \Delta \ln \hat{\mathcal{L}} \right\rangle \equiv \int d\Lambda \int d\Lambda' p(\Lambda|\{d_i\}) p(\Lambda'|\{d_i\}) \Sigma(\Lambda, \Lambda') \tag{14.15}$$

$$\approx \frac{1}{K^2} \sum_{k=1}^{k=K} \sum_{k'=1}^{k'=K} \Sigma(\Lambda_k, \Lambda_{k'}) \tag{14.16}$$

$$\Lambda_k \sim p(\Lambda|\{d_i\}). \tag{14.17}$$

We note that this is the average of the covariance matrix weighted by the posterior support.

This is a slightly different statistic than the one considered in Essick and Farr [193], where the authors replace the integral over Λ' with a fixed value at the mean of the hyper-posterior

$$\bar{\Lambda} = \int d\Lambda p(\Lambda|\{d_i\})\Lambda.$$

While the simpler expression used in Essick and Farr [193] likely produces comparable results for posterior distributions with Gaussian uncertainties, for posteriors with more complex shapes, e.g., multimodality or curving degeneracies, the mean of the posterior is not in general representative of points with significant posterior support. In contrast, the full integral over Λ, Λ' ensures that we accurately represent the variance between all pairs of points with posterior support.

14.3 How uncertain can we be?

Before turning to real examples, we first motivate an acceptable level of uncertainty in the log-likelihood estimator. Specifically, we want to know a threshold value of $\langle \Delta \ln \hat{\mathcal{L}} \rangle$ above which we expect to see significant biases. To test this, we consider a simple one-dimensional problem where the true posterior distribution is a unit normal distribution. To verify that the threshold is independent of the structure of the covariance matrix, we perform this experiment with four analytic kernel functions: a block-diagonal kernel where each block is fully correlated with a random number of blocks, a Matérn kernel with $\nu = 5/2$ and random correlation length, a completely uncorrelated kernel, and a completely correlated kernel. We find that the result is independent of the kernel choice.

For 4800 iterations, we choose a covariance matrix using one of our kernels with a random value of $\langle \Delta \ln \hat{\mathcal{L}} \rangle$ drawn logarithmically between $[10^{-2}, 20]$. For each

covariance matrix, we draw 100 realisations from the covariance matrix $\Sigma(\Lambda, \Lambda')$ to generate biased posterior probability distributions. For each of these realisations, we compute the fraction f of the posterior support below a random point drawn from the true posterior. If there is no bias, f should follow a uniform distribution in $[0, 1]$. We, therefore, compute a $p_{\rm value}$ comparing the 100 values of f to the uniform distribution. In Figure 14.1, we show a two-dimensional histogram of the result of this numerical experiment. We see that when $\langle \Delta \ln \hat{\mathcal{L}} \rangle \lesssim 1$, the $p_{\rm value}$ are uniformly distributed indicating unbiased recovery. However, as the magnitude of the uncertainty rises, the distribution of $p_{\rm value}$ skews heavily towards small $p_{\rm value}$. As a final quantitative test, we compare the distribution of $p_{\rm value}$ in each bin of $\langle \Delta \ln \hat{\mathcal{L}} \rangle$ to compute a combined $p_{\rm value}$. We see that the combined $p_{\rm value}$ is very small for $\langle \Delta \ln \hat{\mathcal{L}} \rangle \gtrsim 1$. We will therefore use $\langle \Delta \ln \hat{\mathcal{L}} \rangle \gtrsim 1$ as our heuristic threshold for significant bias.

14.4 How uncertain are we?

We now turn to a tangible example of uncertainty in the inference performed on the population of binary black hole mergers observed during the first three observing runs of Advanced LIGO and Advanced Virgo with a false alarm rate of less than one per year. The analyses performed in [28] imposed cuts on the convergence of the Monte Carlo integrals that implicitly limit the variance in the likelihood to avoid spurious features in the posteriors. All analyses in that work imposed a condition first proposed in [213] demanding that for the selection function $n_{\rm eff} > 4N$. Some models also enforced the condition that each marginalization over the single event posteriors had $n_{\rm eff} > N$. We consider one of the models that applied both convergence conditions.

We compute the uncertainty in the estimated likelihood for one of the models used in the latest LIGO-Virgo-KAGRA analysis. Specifically, we use the `PowerLaw + Peak` mass model [524], `Default` spin model [521, 582], and power-law redshift model [228]. For our default analysis configuration, we use the same 4278 per-event posterior samples [151] and injection set [530] used in the equivalent analysis in [28] and do not apply any constraints on the convergence of the Monte Carlo integrals.

For all of our analyses, we sample the population posterior using the `nestle` [82] nested sampling package as implemented in `Bilby` [67]. We use `GWPopulation` [520] to compute the likelihood function. We use the same prior distributions as in [28]. For each of the posterior samples, we evaluate the uncertainty in each of the 70 Monte

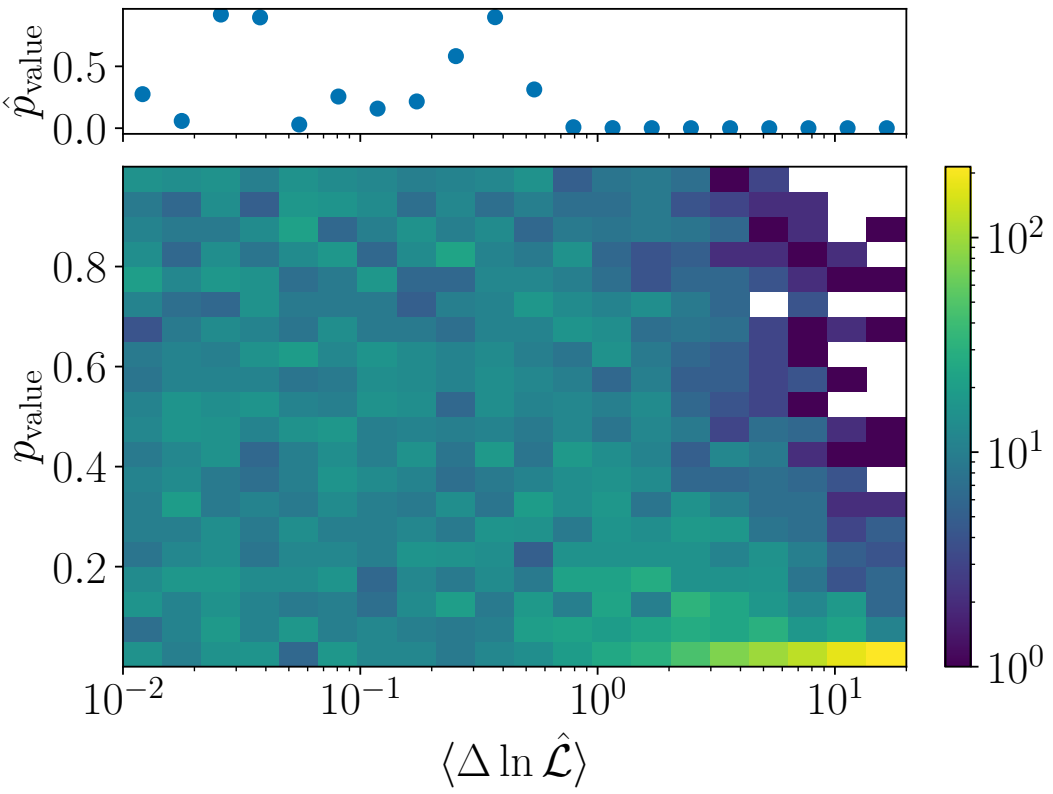

Figure 14.1: $p_{\rm value}$ vs uncertainty in difference in log-likelihood averaged over the posterior distribution ($\langle \Delta \ln \hat{\mathcal{L}} \rangle$). For unbiased analyses at a given value of $\langle \Delta \ln \hat{\mathcal{L}} \rangle$, we expect $p_{\rm value}$ to follow a uniform distribution in $[0, 1]$. The upper panel shows a combined $p_{\rm value}$ for all the points in the histogram falling within that range of $\langle \Delta \ln \hat{\mathcal{L}} \rangle$. We note that this is satisfied for $\langle \Delta \ln \hat{\mathcal{L}} \rangle \lesssim 1$, however, when the uncertainty is larger than that value, the analysis is biased on average.

Carlo integrals involved (one for each event and the selection function integral).

14.4.1 Evolution of $\sigma_{\rm obs}$

We begin by testing our assumption that rewriting the total variance in terms of the average per-event variance $\sigma_{\rm obs}$ is reliable. One method in which this could break down is if the average uncertainty changes as the sensitivities of the observatories improve. In Figure 14.2, we show the average contribution to the covariance over the posterior for the hyperparameters for each event ordered by observation date. The different colours correspond to events observed in different years. There is no obvious trend over time which validates our approximation of $\sigma^2_{\rm obs} = \langle \sigma^2_i \rangle$. We show this value with the dashed grey line. The event with the largest contribution to the uncertainty is GW190517, which has masses consistent with the excess at

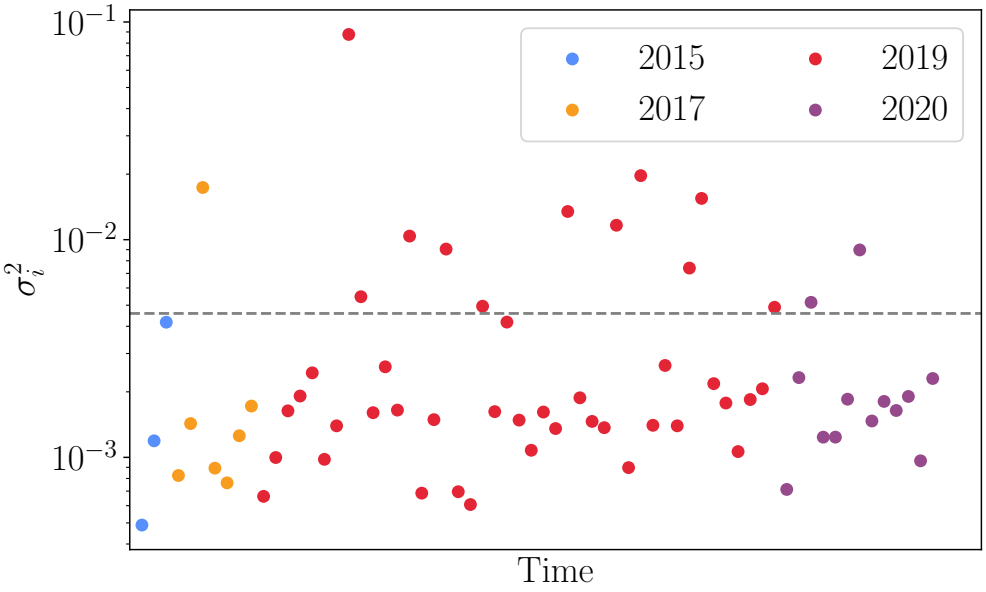

Figure 14.2: The per-event contribution to the likelihood covariance averaged over the posterior support for our population hyperparameters. We divide the events by the year of the observation, approximately corresponding to different observing runs of Advanced LIGO/Advanced Virgo. We note that there is no obvious trend with time, indicating that we can reliably consider the average uncertainty $\sigma^2_{\rm obs} = \langle \sigma_i^2 \rangle$ (shown by the dashed grey line).

$\sim 35 M_\odot$ and large inferred spins.

14.4.2 Scaling with the population size

In order to estimate the scaling of the uncertainty with the size of the catalogue, we randomly sample observations from the total catalogue to simulate smaller catalogues and scale the uncertainty on the selection function appropriately. Specifically, we consider catalogues with increments of 5 events from 5-65 and all 69 events. For each catalogue size, we sample from the hyper-posterior and compute the average variance in the estimated differences of log-likelihood values over the posterior samples $\langle \Delta \ln \hat{\mathcal{L}} \rangle$. We do not apply any of the ad-hoc restrictions on Monte Carlo integral convergence proposed in [213, 28] and described above in these analyses. We fit a simple model to the uncertainty coming from the per-event terms and the selection function to obtain fits for the contribution from the individual events and

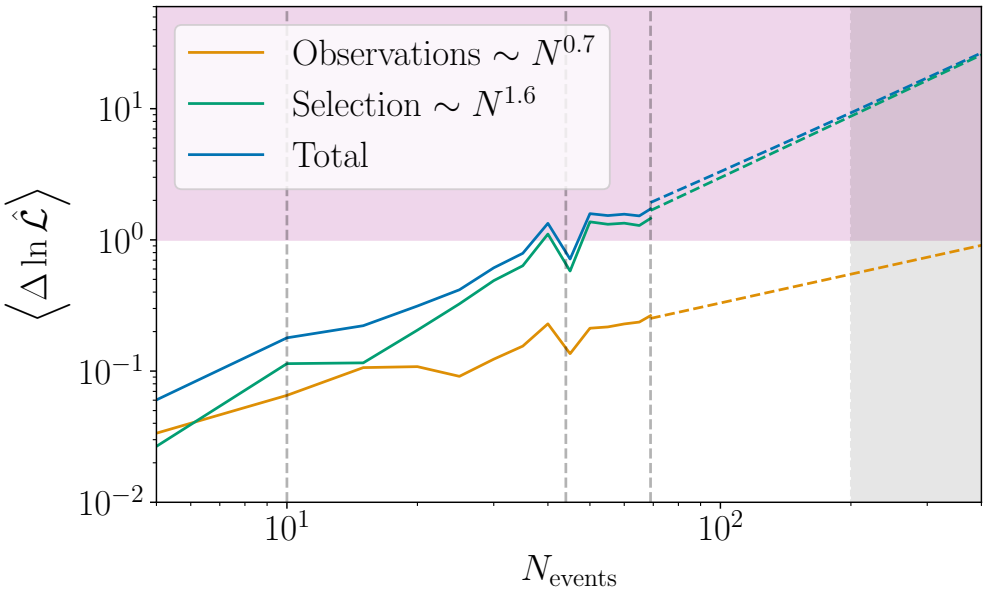

Figure 14.3: Scaling of the uncertainty in the log-likelihood averaged over the full posterior support with the population size for a simple parametric population model. The dashed vertical lines show the number of confident binary black hole events in the gravitational-wave transient catalogue at the time of publication of GWTC-1 [12], GWTC-2 [24], and GWTC-3 [26]. The gray filled region indicates the projected number of binary black hole observations during the next observing run of the international gravitational-wave detector network [423, 307]. The purple shaded region indicates heuristic values for when the uncertainty in the likelihood is likely to cause noticeable bias in the analysis. The solid curves show the empirically obtained uncertainties and the dashed curves are extrapolations based on the power-law fit to the per-event contribution (orange) and the contribution from the selection function (green). The total uncertainty is shown in blue.

the sensitivity. The model for the total variance is

$$\langle \Delta \ln \hat{\mathcal{L}} \rangle = \sigma^2_{\mathrm{obs}} N^a + \sigma^2_{\mathrm{sel}} N^b. \tag{14.18}$$

Here, we emphasise that $\Delta \ln \mathcal{L}$ is proportional to the variance in the estimator and not the standard deviation. We note that Eq. 14.11 implies $a = 1$, $b = 2$ while if the assumptions from [193] hold we will have $a = 0$, $b = 1$. We perform this calculation for both the mean variance and the mean covariance over the posterior support.

In Figure 14.3, we show the total uncertainty (blue) along with the contributions from the per-event terms (orange) and the selection function (green) as a function

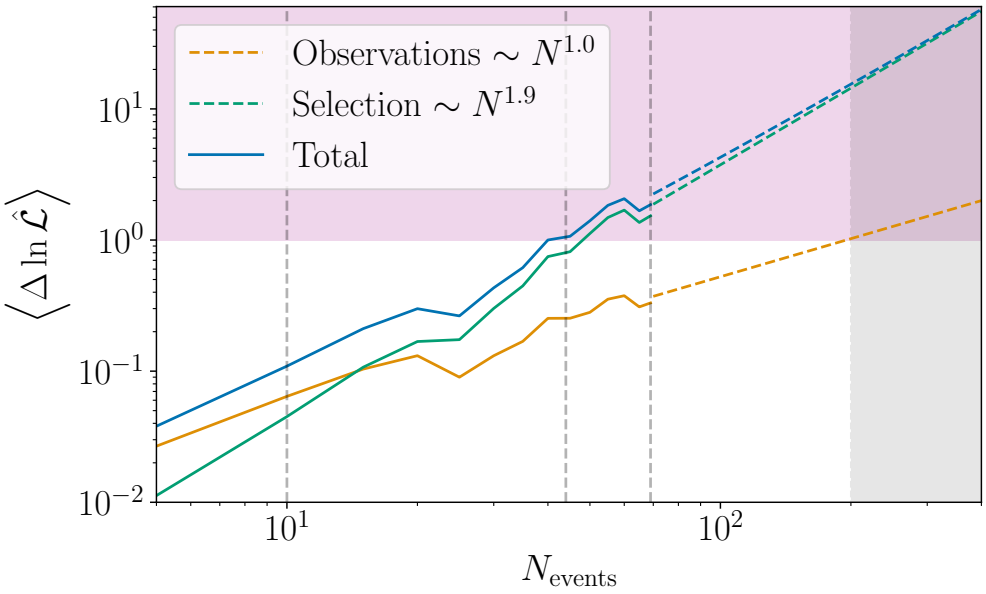

Figure 14.4: The same a Figure 14.3 but with a more flexible model. We note that the same general features are present, however, for this model, the uncertainty grows much more rapidly with population size.

of the number of events with the solid curves. The dashed-coloured curves show projections for larger populations based on the analytic fit. The dashed grey lines indicate the number of events in each of the first three gravitational-wave transient catalogues and the grey-shaded region shows a plausible range of observations we may expect after the upcoming fourth gravitational-wave observing run [423, 307]. The purple-shaded region shows where, heuristically, we may expect to see noticeable biases, following the criteria developed in Section 14.3.

We find that in practice, the scaling of the uncertainty lies between the best-case scenario from [213, 193] and the worst-case scenario in Eq. 14.11. Specifically, we find $a = 0.7$, $b = 1.6$. The dominant source of uncertainty is from estimating the selection function when the population is larger than ≈ 10 events. We note that for populations larger than ≈ 40 events, the uncertainty is consistently in the purple-shaded region. This is consistent with the fact that ad-hoc cuts on the prior space or Monte Carlo convergence were needed to avoid significant biases in [29].

To test if this scaling depends on the functional form used to fit the population, we repeat the above calculation with a more flexible model for the primary mass and spin parameters. Specifically, we take the exponential-spline-modulated power-law

mass distribution from [186] and the exponential-spline model for black hole spin magnitudes and tilt angles from [257]. For the mass distribution, we use ten spline nodes spaced logarithmically between $[2, 100]M_\odot$ and for the spin parameters we take six nodes equally spaced over the relevant domain. For all spline nodes, our prior on the amplitudes is a unit normal distribution, except for the endpoints for the mass distribution which are fixed to zero.

In Figure 14.4, we show the same as Figure 14.3 with this more flexible model. We see that the average covariance in both the per-event and selection function terms grows more rapidly in this case than for the simpler model (a=1.0, b=1.9). The more extreme scaling may be due to the greater flexibility of the spline model causing the "local neighbourhood" assumption of [193] to be less appropriate.

14.4.3 Scaling with the size of Monte Carlo integrals

Having established numerically how the size of the uncertainty in the likelihood estimates varies with the size of the population and configuration settings, we turn to how the number of samples per Monte Carlo integral impacts the uncertainty for this concrete example. To address this, we repeat the uncertainty calculation for the `PowerLaw + Peak` and `Default` configuration above ten times, once using all of the available samples, once with half of the samples, one third of the samples, etc., down to one tenth of the samples.

In Figure 14.5, we show the mean variance over the posterior distribution as a function of the number of samples per Monte Carlo integral in the upper panel. The solid blue, dashed orange, and dash-dotted green curves show the results using the full likelihood, selection function only, and per-observation terms respectively. In the lower panel, we show the variance scaled by the number of samples in the integral such that it will be constant if the uncertainty scales linearly with the number of samples. We observe that the variance is consistent with scaling inversely with the number of samples.

14.4.4 Impact on the inferred astrophysical distributions

To study the impact of the convergence-motivated prior cuts and bias in likelihood estimates we consider four analysis configurations:

- *LVK*. The first configuration is the same one used in the LIGO-Virgo-KAGRA analysis in [28]. This analysis used $\sim 4 \times 10^4$ found injections to estimate

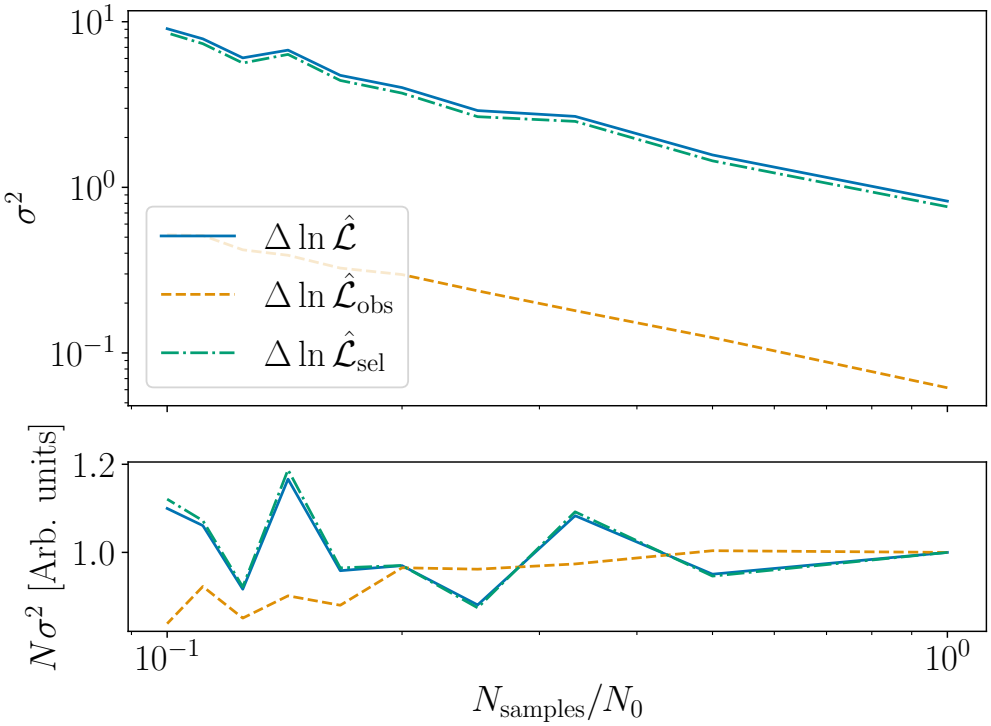

Figure 14.5: The scaling of the average variance in the log-likelihood with the number of events per Monte Carlo integral. The solid blue, dashed orange, and dash-dotted green curves show the results using the full likelihood, selection function only, and per-observation terms respectively. In the top panel, we show the variance. In the bottom panel, we show the normalized variance divided the number of samples per integral. As expected, these quantities scale inversely with the number of samples.

	$N_{\rm injections}$	α	$m_{\rm max}$	$m_{\rm min}$	δ_m	μ_m	σ_m	λ	$\langle \Delta \ln \hat{\mathcal{L}} \rangle$
LVK	4×10^4	2	100	2	0	-	-	0	0.63
No Convergence	4×10^4	2	100	2	0	-	-	0	5.06
Tailored	4×10^4	3.5	105	3	6	33	5	0.04	1.24
More Injections	8×10^5	1	100	2	0	-	-	0	0.50
No Injections	0	-	-	-	-	-	-	-	0.42

Table 14.1: Hyperparameters for the injection sets used in each of the analysis configurations we consider as described in Section 14.4. We additionally list the average variance in the difference between estimated likelihood values.

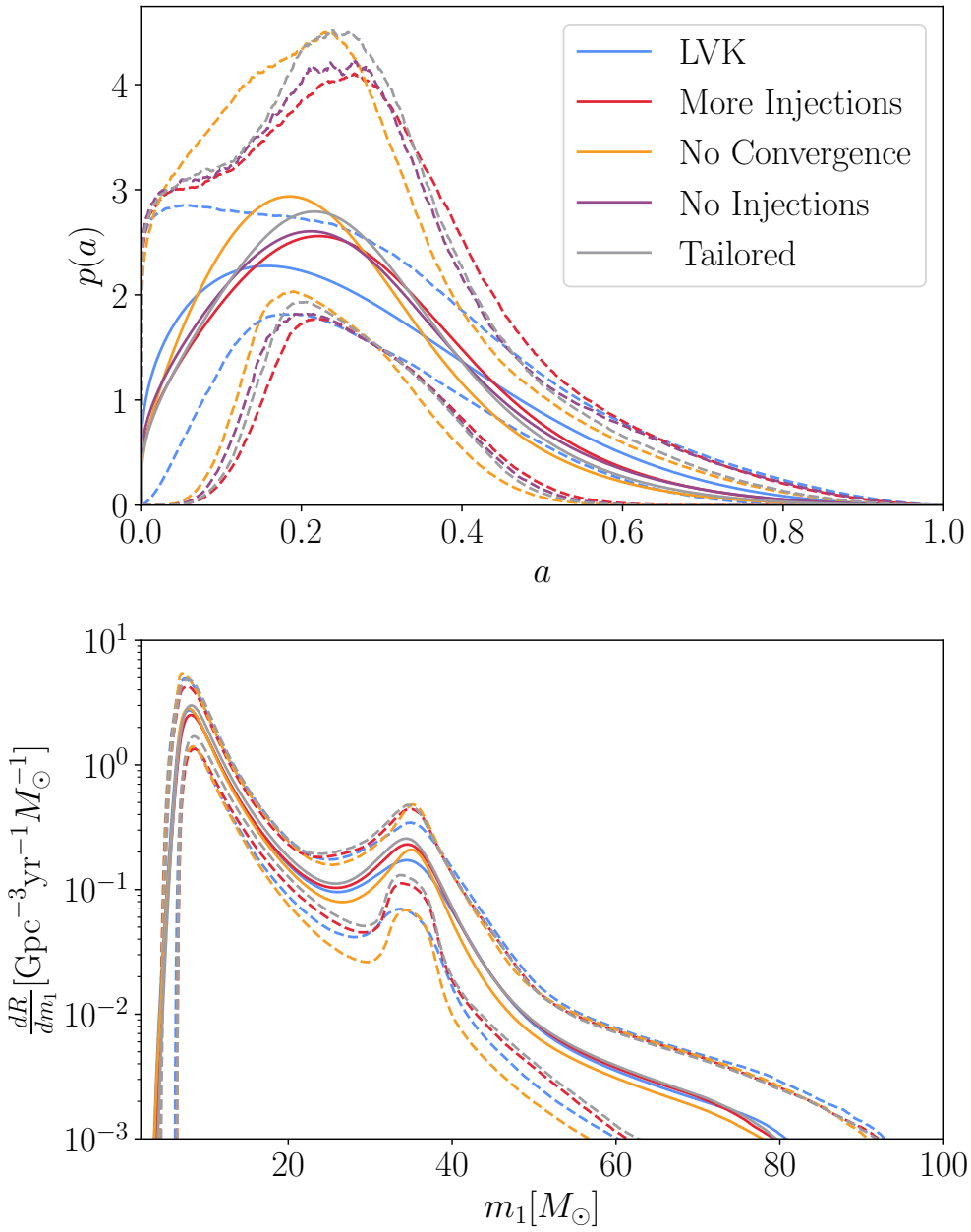

Figure 14.6: The inferred spin magnitude (top) and primary mass (bottom) distributions for a range of analysis configurations. The solid curves show the mean inferred distribution and the shaded regions show the 90% symmetric credible interval. The blue curves show the results presented in [28]. In orange, we show results obtained using the same input samples but without performing the ad-hoc constraints on the number of effective samples per Monte Carlo integral. In red, we show the results when using more found injections to compute the selection function. In purple, we show the results obtained when neglecting the selection function, we note that in this case, we do not show the inferred mass distribution as that is significantly biased by neglecting selection effects. In grey, we show the results obtained using our tailored injection set.

the selection function, and 4278 fiducial posterior samples were used for each event. Specifically, we use the posterior samples released in [151] and the set of sensitivity injections which combine injections covering the first three observing runs of Advanced LIGO/Advanced Virgo [530]. For this configuration, there is the prior cut on $n_{\rm eff}$ for each of the Monte Carlo integrals as described at the beginning of this section.

- *No Convergence.* The second configuration repeats the analysis from [28] but removes the prior constraints on $n_{\rm eff}$ in each Monte Carlo integral.

- *Tailored Injections.* We replace the injection set released by the LVK, we use synthetic injections drawn using a mass distribution that more closely matches the observed distribution. Specifically, we set the mass distribution using the `PowerLaw+Peak` model using the parameters in Table 14.1. Since the proposal distribution for our Monte Carlo integral more closely matches the target distribution, we expect this injection set to lead to smaller uncertainties with the same number of found injections.

- *More Injections.* Rather than using the $\sim 4\times10^4$ found injections used in [28], we use the $\sim 8\times10^5$ synthetic found injections used in [257] in order to reduce the uncertainty in the estimate of the selection function. While this uses many more injections, we note that the underlying distribution of signals is different than for the *LVK* configuration.

- *No Injections.* Rather than using the $\sim 4\times10^4$ found injections used in [28], we ignore the impact of selection effects completely. This will reduce the uncertainty in the estimated likelihoods at the cost of only estimating the observed distribution and not the underlying astrophysical distribution.

For both cases where we use synthetic injection sets, we do not repeat the full injection and recovery using a matched-filter search pipeline due to the large associated computational cost. Instead, we threshold the simulated signals on the optimal signal-to-noise ratio of the injected signal in Gaussian noise with PSDs matching the detector sensitivity during O3 rather than the false-alarm rate [26, 528]. We anticipate that this difference between the detection thresholds does not significantly bias the inferred mass and spin distributions [e.g., 7, 29, 28, 257, 192].

In Table 14.1 we summarise the population hyper-parameters describing the mass distribution used for each injection set. Additionally, we show $\langle \Delta \ln \hat{\mathcal{L}} \rangle$ computed

over the respective posterior distributions for the hyperparameters. We find that the *No Convergence* case clearly surpasses our threshold. The *Tailored* injection set reduces the variance by $\approx 4\times$ by more closely matching the true underlying distribution but is still in the regime where we expect to see some bias. For the other cases $\langle \Delta \ln \hat{\mathcal{L}} \rangle < 1$ and so we would expect the results to be unimpacted by Monte Carlo convergence.

In Figure 14.6 we show the inferred distribution for spin magnitude (top panel) and primary mass (lower panel) with our five analysis configurations. We note that the *No Injections* configuration is excluded for the primary mass distribution as that distribution is strongly biased by not accounting for selection effects. The solid lines indicate the mean inferred distributions and the dashed curves enclose the 90% uncertainty region. While the uncertainties of all of the results agree within their error bars there are visible differences between the inferred results. Specifically, we find that for both parameters, the width of the peak at $a \approx 0.2$ and $m_1 \approx 35 M_\odot$ are broadest for the result that imposes cuts on the prior based on Monte Carlo convergence (blue) and narrowest for the analysis that has the largest average uncertainties (orange) with the analyses with reduced uncertainty (grey, red, purple) lying in between. This indicates that for commonly used analysis configurations, the inferred shape of features in the distribution of black hole mass and spin are notably impacted by uncertainty in the estimate of the likelihood.

We note that the inferred spin magnitude distributions for the *More Injections* and *No Injections* configurations are the most consistent. This would be the expected outcome if the impact of the spin magnitude on the selection function is small and the uncertainty in the likelihood estimates is small. We thus infer that the larger injection set is sufficient to remove the bias present when using the found injections released by the LIGO-Virgo-KAGRA collaboration. While the cuts on the number of effective samples in each Monte Carlo integral in the *LVK* configuration control the average uncertainty in the likelihood estimates, the cuts have a visible impact on the inferred distributions.

14.4.5 Result differences are explainable due to Monte Carlo uncertainty

The posterior predictive distribution (PPD) for the binary parameters is defined as

$$p(\theta|\{d\}) = \int d\Lambda p(\theta|\Lambda) p(\Lambda|\{d\}) \approx \frac{1}{N} \sum_{\Lambda_i \sim p(\Lambda|\{d\})}^{N} p(\theta|\Lambda_i). \tag{14.19}$$

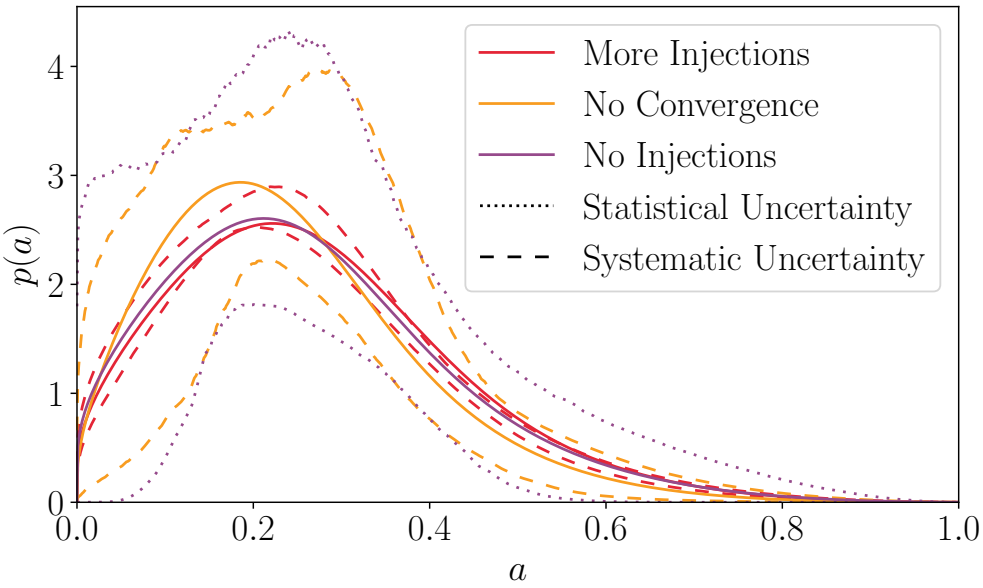

Figure 14.7: Comparison of statistical and systematic uncertainty in our inference of the distribution of black hole spin magnitudes a. The solid curves show the posterior predictive distribution for three of the analysis configurations described in Section 14.4. The dotted curves show the 5th and 95th percentiles of our statistical uncertainty for the lowest variance analysis (*No Injections*). The orange and green dashed curves show the 5th and 95th percentiles of the additional systematic uncertainty from estimating the selection function. We note that for the *More Injections* case the systematic uncertainty is much smaller than the statistical. However, for the *No Convergence* case the systematic uncertainty is comparable to the statistical.

In Figure 14.6 the solid curves show the PPD using our different analysis setups (solid curves). While the curves are visibly different, we wish to know whether the differences can be explained as the result of statistical fluctuations expected due to the uncertainty in our estimator of the likelihood.

Our aim is to estimate the range of different PPDs we might expect to measure given the PPD with no systematic uncertainty and a covariance $\Sigma(\Lambda, \Lambda')$. In the absence of a ground truth, we take the *No Injections* case as our reference analysis as it has the lowest uncertainty estimator of the likelihood and neglect the impact of the per-event integrals as all analyses use the same set of samples for each event.

We begin by taking the samples $\Lambda_i \sim p(\Lambda|\{d\})$ for the reference case. We then construct the covariance matrix by numerically calculating the covariance between the likelihood estimates for every pair of posterior samples. Using this covariance

matrix, we generate weights for each of the samples $\delta \sim \mathcal{N}(0, \Sigma(\Lambda, \Lambda'))$. Finally, we compute the PPD using these weights as

$$\hat{p}(\theta|\{d\}) = \frac{\langle \delta(\Lambda_i) p(\theta|\Lambda_i) \rangle_{\Lambda_i \sim p(\Lambda|\{d\})}}{\langle \delta(\Lambda_i) \rangle_{\Lambda_i \sim p(\Lambda|\{d\})}}. \tag{14.20}$$

By repeating this many times, we can construct the 90% credible interval for the systematic error.

In Figure 14.7 we show the same PPDs for the *No Injections*, *No Convergence*, and *More Injections* configurations and the statistical uncertainty for the *No injections* configuration (dotted curves) as in Figure 14.6. The estimated systematic uncertainty is shown by the dashed curves. We note that in both cases, the PPDs with our specific realisation are entirely consistent with the systematic uncertainty region indicating that the differences in the PPDs can be fully explained by Monte Carlo uncertainty. For the *No Convergence* case, the estimated systematic uncertainty is comparable to the statistical uncertainty. One limitation of our method is that the realisations cannot deviate outside the set of samples used for importance sampling and so cannot accurately resolve cases where the systematic uncertainty is larger than the statistical uncertainty in the posterior.

14.5 Conclusions

Often when performing Bayesian inference, we cannot calculate the true likelihood function, but rather a computationally tractable approximation. For example, the use of Monte Carlo integration to approximate marginal likelihoods is widespread in population inference in gravitational-wave astronomy and beyond. However, often, the uncertainty associated with these finite numerical integrals is neglected. We specifically examine the requirement of performing unbiased population inference on binary black holes with Monte Carlo integrals used to marginalise over the parameters of the individual sources. Previous work has claimed that as the size of the population increases, keeping the allowed uncertainty in each marginal likelihood constant (e.g., the number of samples used in each Monte Carlo integral doesn't have to increase with the population size) is sufficient for precise inference of the population parameters [193].

In this work, through a series of numerical experiments, we demonstrated that for models widely used to characterise the population of merging black hole binaries, this scaling is insufficient and the actual scaling depends on the functional form

chosen to fit the distribution. Failing to use a larger number of samples per Monte Carlo integral will result in an increasingly significant bias in the recovery of the population as the number of observations grows. We recommend that the calculations described in this work be routinely performed for any population analysis to identify cases where the inference may be impacted by Monte Carlo uncertainty. We provide scripts to evaluate this in the accompanying code release.

By considering a model routinely employed to characterise the distribution of masses and spins of merging compact binaries, we found that the uncertainty in the likelihoods estimated as part of population inference on the gravitational-wave transient catalogue is sufficient to lead to noticeable bias with the current size of the gravitational-wave transient catalogue. Additionally, by examining the impact of the specific choice of input samples and convergence requirements we observed changes in the width of features in the distribution of black hole masses and spin magnitudes. While the differences observed here are within the statistical uncertainties, more significant biases have been observed when using more flexible models, e.g., Appendix B of [257].

The results presented in this work are somewhat in conflict with the results from [193]. One difference between this work and theirs is that in [193] the authors only consider population models where the uncertainties on each measurement are smaller than the width of the population. By contrast, in many of the models considered here, including the models for the black hole mass and spin, the individual measurements are broader than the underlying population model. The spin magnitudes of individual black holes are very poorly measured, and so the individual posterior distributions are inevitably broader than the population for the majority of systems. For black hole masses, one might think that the total population model is broader than individual measurements; very few black holes are consistent with masses ranging from $5 - 80 M_{\odot}$. However, the relevant quantity is not the whole domain of the model, but rather than change in the population model over the individual event posterior support. For events intersecting the Gaussian peak at $\sim 35 M_{\odot}$ the uncertainty in the mass is almost always larger than the preferred width of $1 - 5 M_{\odot}$. We defer detailed investigations into whether this is a relevant difference to future work.

In the next observing runs, we can conservatively expect the size of the observed population to double or triple [423]. With a population of this size, we can expect that if we continue to use the same number of samples per Monte Carlo integral

the variance in the log-likelihood will reach $\sim 4 - 10$ and we will be in danger of making severely biased inferences. In order to avoid this, we will either need to use dramatically more samples in our Monte Carlo integrals or consider novel approaches.

There are a number of questions posed by our results that should be explored in future work. In Section 14.4, we found an approximate scaling for the growth of the uncertainty with the population size, developing a theoretical understanding of this scaling may prove instructive in developing improved methods to deal with large populations. Ensuring accurate estimation of the population likelihood is an increasingly complex task as the population size increases, and so we will require increasingly sophisticated methods.

As shown in Section 14.4, a simple method to reduce the uncertainty in Monte Carlo integrals is to reduce the divergence between the initial model and the target model. Fortunately, as the size of the population grows, we can use our existing knowledge to generate initial models that well approximate the true distribution, e.g., by drawing our injections to determine the survey sensitivity by our best estimate of the true population. Additionally, one can recast the Monte Carlo integral using continuous representations of the per-event likelihoods in order to minimise the uncertainty, e.g., [582, 256]. Finally, one can limit the analysis to only consider slowly varying source models, e.g., by imposing smoothing priors on the population model [187, 125]. However, this can lead to missing any sharp features in the distribution.

Each of these improvements is likely to fail eventually, and new methods will be needed. One possibility is to remove the Monte Carlo integral to determine the selection function and instead directly model the observed distribution of compact binaries. If desired, the astrophysical distribution can then be obtained as a post-processing stage using continuous estimates of the selection function such as those in, e.g., [554, 522]. Similar approaches have been proposed for analyses of online polling data [e.g., 189, 344]. Since the contribution of the uncertainty from estimating the selection function grows most rapidly with population size, this will significantly alleviate bias in the inferred distribution.

While we considered uncertainties in the likelihood function used for gravitational-wave population inference, our analysis holds for any problem where there are parameter-dependent biases in calculating likelihoods. For example, when characterising individual compact binary coalescences, there are a number of sources of bias in the likelihood function, including waveform systematics [435], detector

calibration uncertainty [418, 562], and likelihood acceleration methods [501, 384, 334]. While the specific results in Section 14.4 will not be relevant to these cases, the general expressions in Sections 14.2.1 and 14.3 are relevant.

Acknowledgements

We thank Reed Essick and Will Farr for multiple discussions. We thank Sylvia Biscoveanu, Tom Callister, Jack Heinzel, Eric Thrane, and Salvatore Vitale for useful conversations and comments. JG acknowledges funding from NSF grants 2207758 and PHY-1764464. CT is supported by an MKI Kavli Fellowship and an Eric and Wendy Schmidt Fellowship for AI in Science at the University of Chicago. This material is based upon work supported by NSF's LIGO Laboratory which is a major facility fully funded by the National Science Foundation. This work used computational resources provided by the Caltech LIGO Lab and supported by NSF grants PHY-0757058 and PHY-0823459. This work made use of the following software: `numpy` [270], `cupy` [403], `nestle` [82], `Bilby` [67], `GWPopulation` [520], `gwpopulation_pipe` [517].

Appendix 14.A Data Availability

This work used publicly available data produced by the LIGO-Virgo-KAGRA collaborations [151, 530]. We additionally used a larger synthetic injection set that we will make available on request. Scripts and `Jupyter` notebooks required to reproduce this analysis are available at github.com/ColmTalbot/monte-carlo-uncertainty-scaling.

Chapter 15

SYSTEMATIC BIASES FROM USING A HOPELESS SNR CUT WHEN ESTIMATING SENSITIVITY

This chapter is an unpublished manuscript based on work done in collaboration with Derek Davis.

15.1 Introduction

Observations of gravitational waves from Binary Black Hole (BBH) mergers have allowed astrophysicists to infer properties of not only individual black holes, but also about the population of black holes in the Universe [28, 3, 29]. Scientists use these population constraints to draw conclusions about BBH system formation scenarios, stellar astrophysics, and cosmology [21, 28, 3, 29].

A key ingredient when inferring the population properties of astronomical sources is an understanding of the Malmquist bias, as the catalog of LIGO-Virgo observations do not by themselves directly reflect the underlying astrophysical distribution of BBHs. As LIGO-Virgo-KAGRA observations compose a flux-limited survey, the catalog of GW observations from BBH events preferentially over-represent the masses, spins, and distances from the most sensitive parts of parameter space with respect to the true underlying population. For example, properties such as high masses, large aligned component spins, and small distances all generally increase the signal-to-noise ratio (SNR, denoted ρ) of a source and make it easier to detect.

It is common to quantify this observational bias by injecting a large number of simulated BBH sources into detector noise, and then determining how many of these injections pass the detection threshold chosen for the catalog of real events [538, 213]. Due to the computational expense of this procedure, this is quantified using a limited set of sources and several approximations are made in the process. This can lead to biases in the inferred astrophysical distribution if, for example, the model assumed when generating noise assumed when quanitifying detection probability is not faithful to the true noise distribution seen in the detector, or if there are insufficient numbers of injections across parameter space to resolve the Monte Carlo sum [213, 195, 193, 519]. Finally, if the threshold used to designate a “found" event in the injection set differs from the threshold for the inclusion of

real events, the detection probability of a given injected source may not be reflective of the true data selection and the resulting inferred population may be biased when taking sensitivity into account [192].

In the LVK population studies, the designation of an event as “detected" or “found" is made from its False Alarm Rate (FAR), as assigned by the various matched-filter pipelines searching the data. In short, the FAR is the rate at which a detection pipeline will falsely identify a candidate with comparable properties when in the absence of a signal. When creating the set of sensitivity injections, the vast majority of injected sources will be at low SNRs, as $\frac{dN}{d\rho} \sim \rho^{-4}$ in a Euclidean volume approximation. In order to avoid the burden of running these detection pipelines on an unnecessarily large number of injections sampled from a fiducial population, it is common to place a “hopeless cut" on the injection set, in which only injections exceeding some SNR are given to the pipelines to be assigned a FAR [21, 28, 192]. It is assumed to be not worth running the pipeline on an injection with an SNR below this hopeless cut, as it has a negligible chance of being assigned as detected with a reasonable detection statistic. Therefore, injections with SNRs below this cut are missing from the injection set released in [531], and are assumed to have an arbitrarily large FAR. For the GWTC-3 population analysis, this hopeless cut was placed based on the optimal SNR; that is, the SNR in the absence of added noise. While this cut was placed at an optimal SNR of $\rho_{\rm opt} = 6$, and the detection statistic used for designating a real event or an injection as detected was a FAR $< 1\text{yr}^{-1}$ (equivalently, Inverse FAR (IFAR) > 1yr). By placing a hopeless cut on the SNR, it is assumed that injections with SNRs less than the hopeless cut will not be designated as having an IFAR > 1yr by the detection pipelines.

In this work, we show how this hopeless cut effectively introduces a different threshold on injections and real events, and how artificially increasing this bias causes systematic biases to the inferred population. Additionally, we outline and implement a method to generate new FAR values for injections with SNRs below the hopeless cut, without having to run the detection pipelines. We validate this fitting and resampling method on existing injections as well as by generating new injections below the hopeless cut used in [28]. We then apply this method to the GWTC-3 Population analysis in order to test whether the hopeless cut used in that work biased any of the reported results.

15.2 Detection Probability in the Population Likelihood

The Malmquist bias correction is accounted for in the population likelihood in gravitational wave analyses as a population-dependent weight, quantifying the detectability of a particular population [538, 359, 213]. Assuming a prior on the rate of BBH mergers $\pi(R) \propto \frac{1}{R}$, it is standard to express the likelihood of the observed dataset $\{\mathrm{d}\}$ from N events given some parameters Λ describing the population as [536, 357, 228]:

$$p(\{d\}|\Lambda) = \prod_{i}^{N} \frac{\int d\theta_i \mathcal{L}(d_i|\theta_i)\pi(\theta_i|\Lambda)}{p_{\mathrm{det}}(\Lambda)}, \tag{15.1}$$

where θ is the set of parameters describing the individual events, such as their masses, spins, and distances.

The term $p_{\mathrm{det}}(\Lambda)$ is proportional to the sensitive four-volume and fraction of detectable sources of the population described by Λ. In order to calculate this, we must marginalize over realizations of the data d (which includes noise n) and parameter configurations from the population:

$$\begin{aligned} p_{\mathrm{det}}(\Lambda) &= \int d\theta \int_{\rho(d)>\rho_{\mathrm{th}}} dd p(d|\theta) p(\theta|\Lambda) \\ &= \int dn \int d\theta p(\theta|\Lambda) p(n) \Theta(Z(d,\theta) - Z_{\mathrm{th}}), \end{aligned} \tag{15.2}$$

where Z is some detection statistic, such as the False Alarm Rate (FAR), quantifying the significance of an event, and Z_{th} is the threshold on Z which is chosen to distinguish real events that will make the observed catalog of sources. As denoted by the Heaviside function in Equation 15.2, anything with a detection significance exceeding some threshold Z_{th} is taken to be “found" while any source with $Z < Z_{\mathrm{th}}$ is part of the background.

A straightforward way to approximate Equation 15.2 is to draw N_{inj} sources, each with $\theta \sim \pi(\theta|\Lambda)$, and inject these sources into noise realizations following the known noise distribution of the detectors $p(n)$. After assigning a detection statistic to each injection, one can calculate Equation 15.2 simply by N_{found} divided by N_{inj}.

While the above Monte Carlo procedure can be an accurate method to calculate p_{det}, it is computationally infeasible to do this for each iteration of computing Equation 15.1, which is commonly called by a stochastic sampler millions of times to evaluate the likelihood for different values of Λ. A common approach to efficiently

calculate Equation 15.2 for any Λ is by using importance sampling and reweighting a single set of injections. This is done by drawing from some fiducial distribution $\theta \sim \pi(\theta|\Lambda_0)$, injecting these sources into detector noise, and determining which sources are found. The injections with $Z(\theta) > Z_{\rm th}$ get their corresponding values of $\pi(\theta|\Lambda_0)$ attached as importance weights, treating $\pi(\theta|\Lambda_0)$ as a proposal distribution that can be simply reweighted to calculate the target $p_{\rm det}(\Lambda)$ [538, 213]. Only needing to do the injection-recovery procedure one time, we use the importance weights to approximate Equation 15.2 as

$$p_{\rm det}(\Lambda) \approx \frac{1}{N_{\rm inj}} \sum_{i}^{N_{\rm found}} \frac{\pi(\theta_i|\Lambda)}{\pi(\theta_i|\Lambda_0)}, \tag{15.3}$$

where θ_i are the parameters of the ith "found" injection. When generating the injection set from the fiducial population $\pi(\theta|\Lambda_0)$, it is important that the total number of injections drawn $N_{\rm inj}$ is recorded, as well as the probability of drawing each detectable injection (i.e. the value of $\pi(\theta_i|\Lambda_0)$), which is stored as a weight to be used in the above summation.

If $Z_{\rm th}$ is set by the FAR assigned by a detection pipeline, the set of "found" injections determined by the detection threshold necessarily can only include injections which actually went through the detection pipeline searches. As this process can be computationally difficult and the vast majority of injections are much too quiet to be detected, the set of injections chosen to go through the detection pipeline procedure are pre-filtered by their optimal SNRs ($\rho_{\rm opt}$). When placing this hopeless cut on $\rho_{\rm opt}$, it effectively alters Equation 15.2 to be:

$$p_{\rm det}(\Lambda) = \int dn \int d\theta p(\theta|\Lambda) p(n) \Theta(Z - Z_{\rm th}) \Theta(\rho_{\rm opt} - \rho_{\rm hopeless}), \tag{15.4}$$

where this additional Heaviside function mandates that an injection must pass the hopeless SNR cut in addition to the detection threshold set by Z in order to be counted as "found". Under an ideal choice for the hopeless cut, there would never be any cases where an injection has both $\rho < \rho_{\rm hopeless}$ and $Z > Z_{\rm th}$. In other words, assuming $Z_{\rm th}$ is the same threshold used to include the real events (i.e. from an event catalog) in the numerator of Equation 15.1, in order to get an unbiased value for $p_{\rm det}(\Lambda)$, the second Heaviside function in Equation 15.4 should never be zero when the first Heaviside function is one.

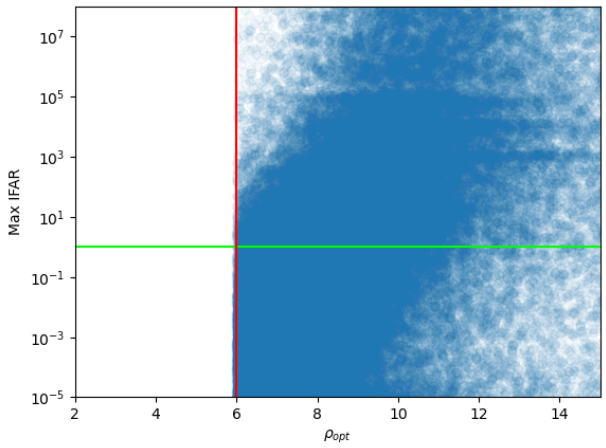

Figure 15.1: Maximum IFAR across pipelines vs Optimal netowrk SNR in Hanford and Livingston (as described in Section 15.3). Red line shows where the hopeless cut was placed and the green line shows the detection threshold used for the GWTC-3 Populations analysis (IFAR = 1 yr) [28].

15.3 Hopeless Cut in the GWTC-3 Injections

In [531], the LVK collaboration released a set of injections from a fiducial population for evaluating the sensitivity of the LIGO-Virgo detector network in the third observing run (O3) [171, 26]. Including only the injections that pass the hopeless cut, the data release lists the parameters of each of these injections as well as the detection statistics assigned to them. Not including injections from an Intermediate Mass Black Hole (IMBH) population, this injection set was created by simulating $\sim 7.4\times10^7$ sources in data from O3, with masses, spins, and distances sampled from a fiducial distribution assuming a power law on source-frame component masses, isotropic spin orientations with uniform spin magnitudes from $a = 0$ to 1, and a redshift distribution with a source frame merger rate $R(z) \propto (1 + z)$.

For each injection, $\rho_{\rm opt} = \sqrt{\rho^2_{\rm opt,\,H} + \rho^2_{\rm opt,\,L}}$ was calculated, where the subscripts “H" and “L" refer to the LIGO Hanford and Livingston detectors, respectively. The Virgo detector was not included in the calculation of $\rho_{\rm opt}$. This resulted in $\sim 2.84 \times 10^5$ injections with $\rho_{\rm opt} > \rho_{\rm hopeless} = 6$, which were injected into the detection pipelines to be assigned FAR values. In (pops paper), an event was considered “found" if any of the matched filter pipelines assigned it an IFAR > 1yr.

In Figure 15.1, we show the maximum IFAR across pipelines plotted against $\rho_{\rm opt}$ for the injections in [531]. Note that there are no points below the hopeless cut of $\rho_{\rm opt} = 6$ (the red line), as these injections are not included in the data release.

Injections lying above the green line (IFAR = 1yr) meet the detection threshold used in [28] and count as “found". Any points in the top left quadrant would correspond to sources which fall below the hopeless cut, but would have been designated as found if they had been injected into the detection pipelines. Thus, if ρ_{hopeless} were too high, this would mean that there should be points in the top left quadrant of Figure 15.1, had these injections been included in the released injection set. While we do not have access to the hopeless injections, by inspection we hypothesize that if one were to extrapolate the IFAR-SNR trend in Figure 15.1 below ρ_{hopeless}, several injections would lie in the top left quadrant.

15.3.1 Effects of Raising the Hopeless Cut

As outlined in Section 15.2, enforcing a hopeless cut on the injections before injecting them into the detection pipelines can effectively introduce a different detection threshold in the calculation of p_{det} than what is used to threshold real observations when forming a catalog. In order to illustrate the bias that can occur if ρ_{hopeless} is too high, we infer the population of masses, spins, and redshifts from the 59 BBH events detected in O3, using the same detection threshold as [28], fixing the observed catalog but raising the SNR of the hopeless cut.

In accordance with model choices from [28], we assume the `Power Law + Peak` model [524] for the mass distribution, the `Default` model [521] for spins, and a source frame rate density evolving as a power law in $(1+z)$ [228]. When calculating p_{det} for Equation 15.1, we enforce that a “found" injection must satisfy both IFAR $>$ 1yr and $\rho_{\text{opt}} > \rho_{\text{hopeless}}$.

We present selected results of the population inference in Figure 15.2. While the populations appear to be in agreement for small changes to the hopeless cut, we note a noticeably large deviation for the highest hopeless SNR cut in the posterior distribution for power law indices α, β, and κ, corresponding to the primary mass, mass ratio, and redshift distributions, respectively. We choose these parameters as examples important for controlling the shape of the population model. The redshift distribution seems to be most senesitive to changing ρ_{hopeless}. For illustration, we plot the evolution of this distribution as a function of hopeless SNR cut in Figure 15.3 in order to demonstrate the clear systematic trend of the inferred κ increasing with hopeless SNR cut. We therefore know that a higher hopeless cut would have introduced an additional bias.

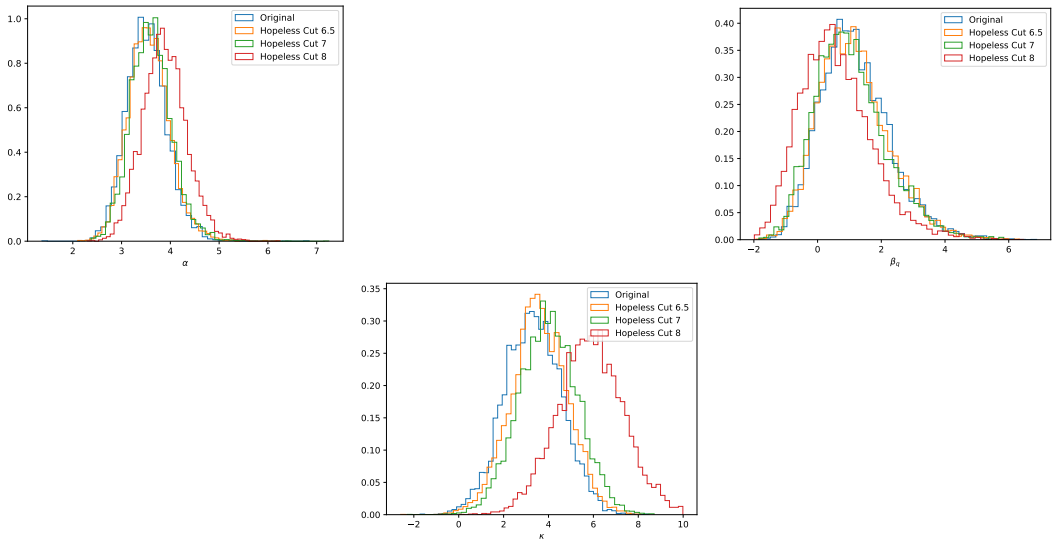

Figure 15.2: Selected results from the test described in Section 15.3.1, in which we perform population inference with iteratively raising the hopeless SNR cut. Shown are the posterior distributions for the power law index on the primary mass distribution (α, left), mass ratio distribution (β, middle), and redshift distribution (κ, right). We note the significant systematic bias is most present in the redshift distribution.

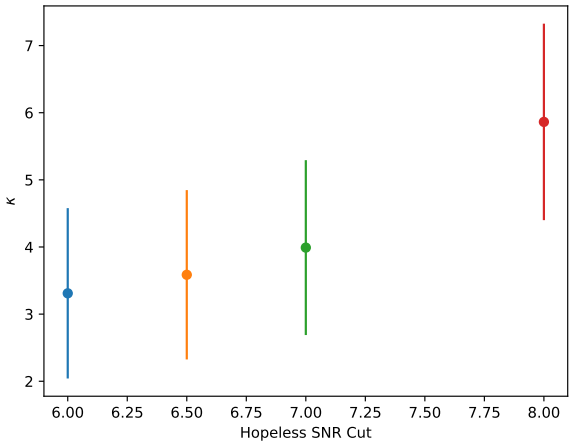

Figure 15.3: Average and standard deviation of the redshift power law index κ as a function of hopeless SNR cut. The increased hopeless SNR cut systematically biases κ to higher values.

15.4 Fitting Below the Hopeless Cut

The main obstacle for conducting a similar test as in the previous section, but lowering the hopeless cut, is that no information about the injections with $\rho_{\text{opt}} < 6$ are included in [531]. In order to evaluate p_{det} using the O3 injection set but with a lower hopeless cut, we introduce a method to add injections with $\rho_{\text{opt}} < \rho_{\text{hopeless}}$ along with corresponding IFAR values, without needing to run the detection pipelines. We do this by constructing a two-dimensional fit of $\frac{dN}{d\rho d\text{IFAR}}$ and extrapolating this fit to SNRs below the hopeless cut, where no data exist in the injection set. In this way, we can approximate $p_{\text{det}}(\Lambda)$ including injections with lower SNR.

As a first step, we consider a slice of fixed IFAR in the IFAR-SNR space of the injections and examine the distribution of SNRs. In Figure 15.4 we plot this two-dimension parameter space, populated by the injections from the injection set, as well as a histogram of SNRs in a single bin of IFAR for a few selected IFAR bins. For a source in Gaussian noise, a given ranking statistic such as IFAR should be normally distributed around the optimal (noiseless) value. In reality, a ranking statistic such as IFAR would tend to downrank some events of a given optimal SNR, due to effects such as template bank sparseness and data quality issues. We therefore expect that when we look instead at optimal SNRs in a given bin of IFAR, the distribution of optimal SNRs should behave more Gaussian for lower SNRs, whereas the high SNR tail contains more events downranked in significance due to the above reasons.

Following these assumptions, we model the bulk (80%) of optimal SNRs in a given IFAR bin as following a truncated Gaussian distribution, with bounds at the hopeless SNR cut ($\rho_{\text{opt}} = 6$ in this case) and the $80th$ percentile of the SNRs in the bin. Due to the above expectation that the high tail does not behave like the tail of a Gaussian, we simply make a Kernel Density Estimate (KDE) of the top 20% of SNRs in the IFAR bin. We then stitch the two density estimates together and renormalize by the total number of injections in the IFAR bin in order to get an approximation for $\frac{dN}{d\rho}$ for $\rho > 6$. We do this for a total of 99 bins of IFAR, logarithmically spaced in the range $[10^{-4}, 10^{15}]$ yr. In Figure 15.4, we show the histogram of optimal SNRs and corresponding fit for several choices of IFAR bins.

In order to extrapolate the model to SNRs below the hopeless SNR cut where we have no data, we enforce the above assumption that the distribution of lower-SNRs in an IFAR bin should closely follow a Gaussian. To encode this in the model, we simply extend the truncated Gaussian component of the fit to $\rho_{\text{opt}} = 0$, ensuring a normalization such that the fit integrated up from $\rho_{\text{opt}} = 6$ is consistent with the

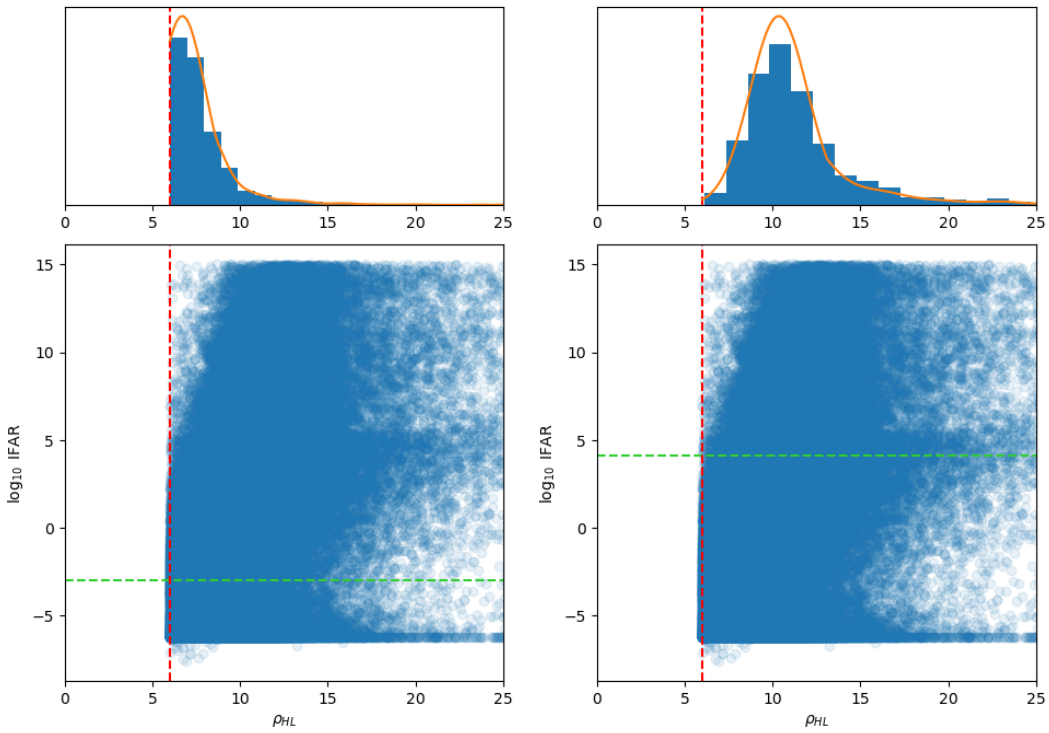

Figure 15.4: Two dimensional scatter plots of IFAR and SNR values for the injections in the injection set, along with histograms of the optimal SNR values along the IFAR bins corresponding to the green horizontal lines. The red dashed line marks the SNR 6 hopeless cut below which we have no data from the injections. The orange curve is the fit to the SNR distribution from the KDE and truncated Gaussian. We plot two example fits for demonstration, but this procedure is done for 100 IFAR bins.

number of injections in the IFAR bin used to make the fit.

In Figure 15.5 we show the same fits at Figure 15.4, including the extrapolation of the fit below $\rho_{\text{opt}} = 6$. Given these optimal SNR fits across IFAR space, we use a rectangular bivariate spline as implmented in SciPy [556] to construct a two-dimensional interpolant.

15.4.1 Creating New Injections

The two-dimensional fit to the full SNR-IFAR space provides a function that can be evaluated to get the differential number of injections at a given SNR and IFAR. Thus, for a trigger at a given SNR, we can evaluate the interpolant at a fixed SNR over a grid of IFARs. This is proportional to the probability distribution of IFARs for a trigger at a given SNR and therefore can be used as a generative model for an

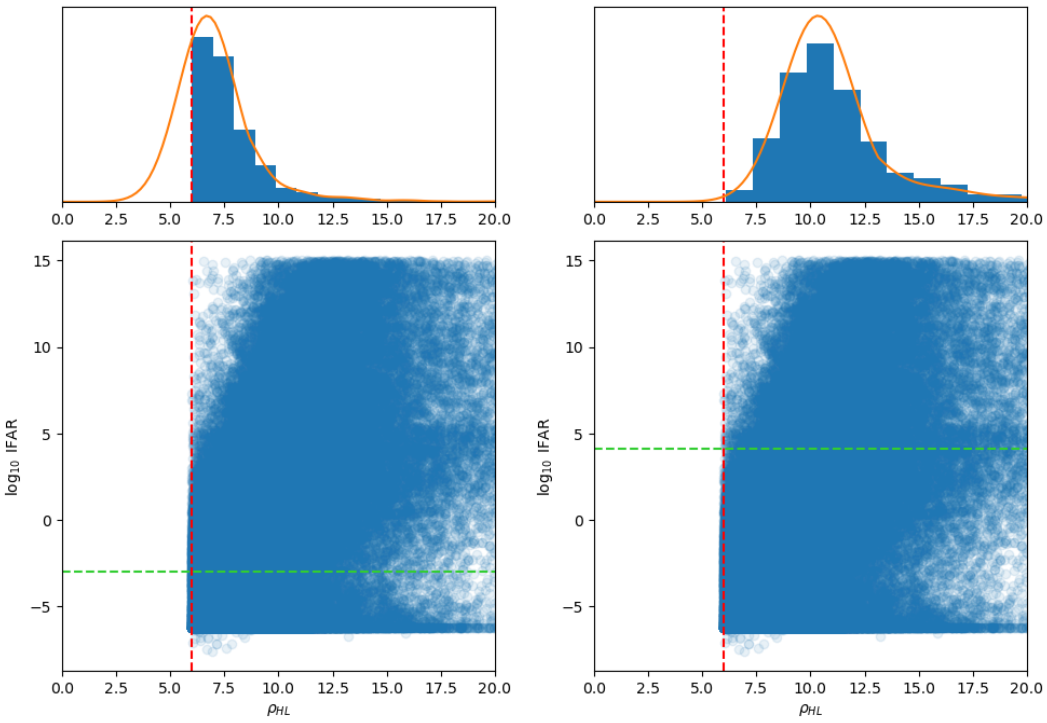

Figure 15.5: Same as Figure 15.4, but extrapolating the truncated Gaussian component to be bounded on the low end by $\rho = 0$ rather than $\rho = 6$.

IFAR value given an SNR.

Before assigning an IFAR to an injection of a given SNR, we must first determine if a detection pipeline triggered on the injection, as the probability of any pipeline triggering on a given source is never guaranteed.[1] In the injection set released in (cite), if a pipeline does not trigger on an injection, it is assigned an IFAR of 0 years, whereas the minimum IFAR a pipeline assigns to a trigger is $\sim 10^{-8}$ years. We can therefore assign the probability of the pipeline triggering:

$$p(\text{trigger}|\rho) = \frac{1}{N_{\text{total}}(\rho)} \int_{\text{IFAR}_{\text{min}}}^{\text{IFAR}_{\text{max}}} N(\rho, \text{IFAR}) d\text{IFAR}, \tag{15.5}$$

where $N_{\text{total}}(\rho)$ is the total number of injections at an SNR of ρ.

The first step is to draw samples for the source parameters for injections below the hopeless SNR cut, such that we get the distribution of SNRs in the region we do not yet have any data. This also allows us to calculate $N_{\text{total}}(\rho)$ in Equation 15.5, for

[1]Note that “triggered" is different than “found". When we discuss triggering on the data, we mean the pipeline matches the data to a template and assigns it a ranking statistic. After the pipeline triggers, it may be “found" if the ranking statistic exceeds the chosen threshold.

$\rho < \rho_{\text{hopeless}}$. For consistency, we adopt the same fiducial population model used in (cite pops). We then keep the injections within the SNR range of interest, using the same optimal SNR calculation used in (cite injections). If these injections are meant to supplement an existing injection set within a given SNR range, it is important to generate the appropriate number of injections within this SNR range to be the correct fraction with respect to the total number of injections.

After generating the new samples and computing their optimal SNR values, we assign each injection an associated IFAR. First, we use Equation 15.5 to decide whether each sample gets triggered by a pipeline, based on the condition that a trigger is defined as an injection with IFAR $> 10^{-8} yr$. For those that are assigned triggers, we sample an IFAR value from the two-dimensional IFAR-SNR fit, conditioned on the source's optimal SNR. If these are being used as a supplement to another injection set, they can be straightforwardly combined at this point. As the fit now extends to SNRs below the hopeless cut, along the low-SNR Gaussian tail, we can extrapolate to below $\rho_{\text{opt}} = 6$.

15.4.2 Fit Validation

Before testing the fit by extrapolating below the hopeless cut, we first test the method in the region of SNR space where injections already exist, in order to compare population results using the original injections with those obtained using the fit in a region of SNR space. This will test whether the method of assigning the IFARs of injections randomly based on the fit, conditioned on the injection's SNR, causes a bias in the recovered population.

We first take all of the injections from (cite) with $6 < \rho_{\text{opt}} < 8$, and remove their IFAR values. We then reassign the IFAR values for these injections by drawing samples as outlined in the procedure outlined above. All other injections in the injection set from (cite) remain unchanged.

We then run our population inference with the same models as in the previous section. Figure 15.6 shows a selected set of resulting recovered population hyperparameters, as compared to the population inference run with the original injection set. We note very good agreement; noteably the power law indices on the redshift and primary mass distributions, which are most susceptible to biases from the hopeless cut, are unchanged. Therefore, we conclude that our fit and IFAR resampling method, when implemented on injections with $6 < \rho_{\text{opt}} < 8$, does not cause a noticeable bias in the inferred population.

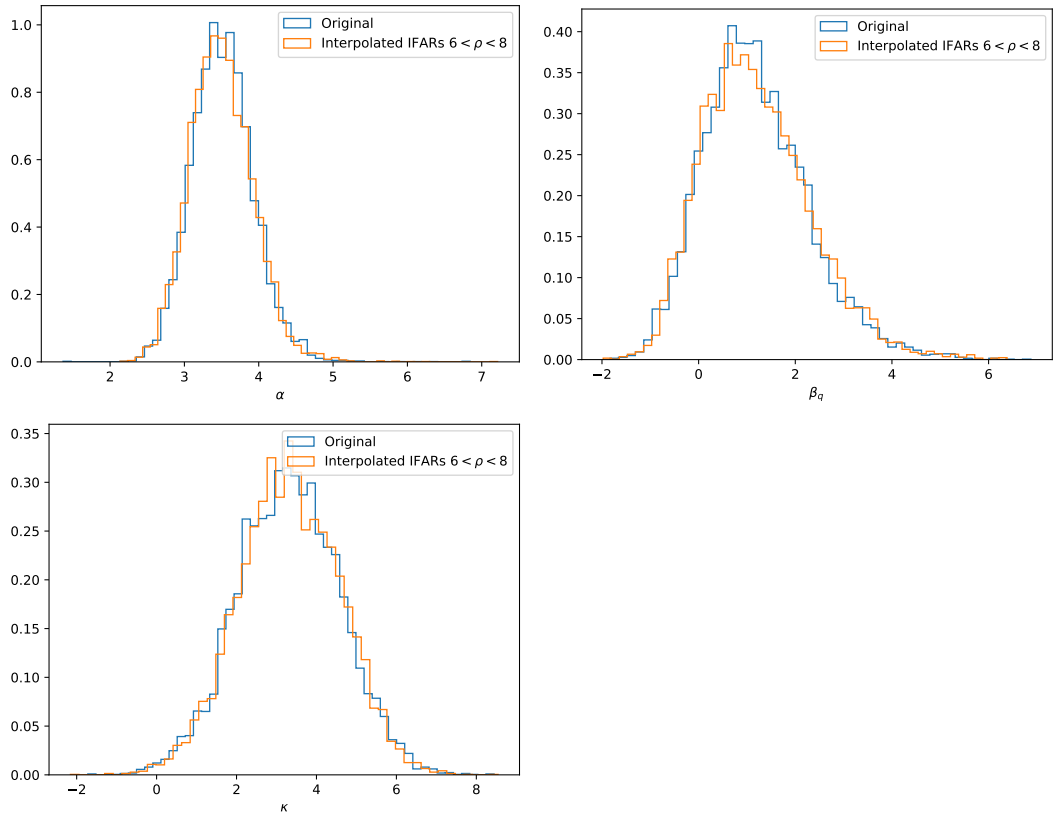

Figure 15.6: Selected results from the test described in Section 15.4.2, in which we perform population inference with new IFAR values for the injections with $\rho_{\text{opt}} < 8$ from the injection set in (cite). Shown are the posterior distributions for the power law index on the primary mass distribution (α, left), mass ratio distribution (β, middle), and redshift distribution (κ, right). We note no systematic bias in the parameter recovery.

15.5 Population Inference with Injections Below the Hopeless SNR Cut

We now move onto addressing the question: how does the inferred population change if the hopeless SNR cut used in the injection set is lower?

We begin by generating a new set of injections with $4 < \rho_{\text{opt}} < 6$, where the upper limit is set by the original hopeless cut in the injection set, and is the minimum SNR of those injections. To be consistent with the population model used to generate the injections in (cite), along with the total number of samples drawn to generate them ($N_{\text{total}} = 73,957,576$, as reported in (cite)), we calculate that we need to generate $\sim 478,949$ injections with $4 < \rho_{\text{opt}} < 6$.

For each of the new injections, we use the method outlined in the previous section to assign an IFAR value. Combining these with the original set of injections in (cite)

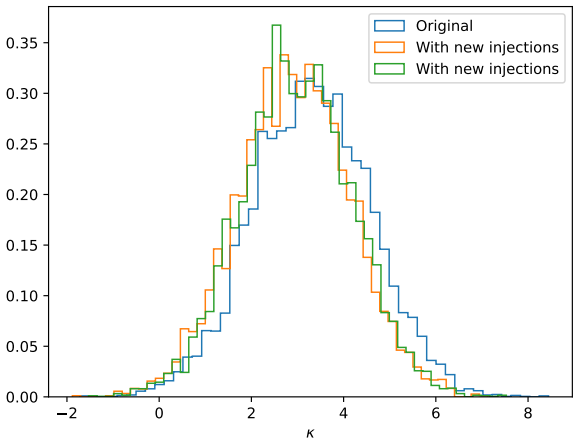

Figure 15.7: Posterior distribution on redshift power law index κ using different injection sets. In blue, we show the result using the original O3 injections from (cite). In green and orange we plot the equivalent distribution with including two different realizations of IFAR values of new injections below the hopeless SNR cut, as outlined in Section 15.5

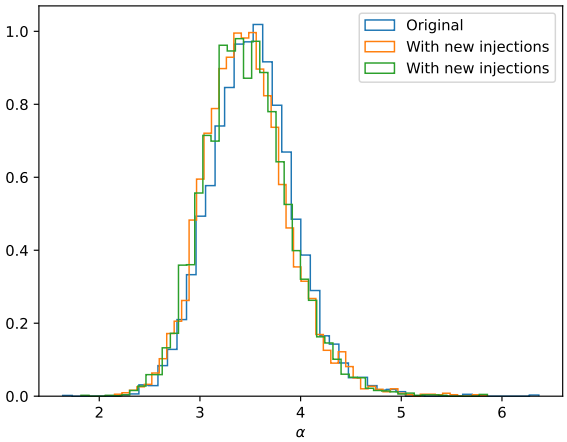

Figure 15.8: Posterior distribution on primary mass power law index α using different injection sets. In blue, we show the result using the original O3 injections from (cite). In green and orange we plot the equivalent distribution with including two different realizations of IFAR values of new injections below the hopeless SNR cut, as outlined in Section 15.5

gives us a larger injection set, which includes injections below the hopeless SNR cut of $\rho_{\text{opt}} = 6$. For demonstration, we plot the combined injections in IFAR-SNR space in Figure (ref), noting that there are several new injections with IFAR > 1.

In Figure 15.7, we plot the inferred distribution of the population hyperparameters using this new injection set, compared with the results from using the injection set

released in (cite). Comparing these two distributions, we notice a slight shift to increased kappa when including κ when including the new injections. We repeat the generate new IFAR values for the new injections by resampling the fit and re-run our population inference code and plot the distribution of κ on the same figure. We confirm that this shift in κ is stable to new realizations of IFAR values for these samples, indicating a systematic shift when including these new injections, indicating that the hopeless cut removed some number of detectable sources.

The direction of the shift is consistent with what would be expected from Figure 15.3: additional found injections tend to *decrease* the best fit values for κ. Including these additional injections with $4 < \rho_{\rm opt} < 6$ makes us infer slightly increased sensitivity at high redshift. For a fixed catalog of observed events, this leads to a decrease in the inferred high redshift merger rate, and therefore a lower value for κ. The fact that this shift in κ is noticeable and systematic indicates that the slope of the redshift evolution rate reported in (pops papers) may have been a slight overestimate; the placement of the hopeless cut may have biased the inferred sensitivity to high redshift events.

Nevertheless, we find that the confidence at which $\kappa > 0$ does not change significantly; with the LVK O3 injections (cite), we infer $\kappa > 0$ at 99.5% credibility, whereas we infer the condition at 99.1% credibility when using our new injection set. Hence, while the hopeless cut did introduce a bias, this bias was not large enough to change astrophysical conclusions.

15.6 Conclusions

In this work, we have examined the effects of the approximations made when implementing the hopeless SNR cut for evaluating sensitivity of the LIGO-Virgo network to GW sources. In doing so, we illustrate the need to choose this cutoff carefully and the changes to the inferred population that may result when it is poorly chosen.

We demonstrate this trend in two ways: by artificially raising the hopeless cut in order to see how leaving out detectable sensitivity injections biases the inferred population, and by creating a fit to the sensitivity injections and extrapolating the fit below the hopeless cut to check if the current placement of the hopeless cut was too high. The results from the first demonstration are in good agreement with [192], in that leaving out detectable sources when computing sensitivity causes us to overestimate the astrophysical rate in regions of parameter space we are least

sensitive to. The second demonstration allowed us to confirm that the placement of the hopeless cut only caused a minor bias in the inferred population.

Part IV

Matter Effects and EoS Implications

Chapter 16

HIERARCHICAL INFERENCE OF BINARY NEUTRON STAR MASS DISTRIBUTION AND EQUATION OF STATE WITH GRAVITATIONAL WAVES

Jacob Golomb and Colm Talbot. "Hierarchical Inference of Binary Neutron Star Mass Distribution and Equation of State with Gravitational Waves". In: *Astrophys. J.* 926.1 (2022), p. 79. DOI: 10.3847/1538-4357/ac43bc. arXiv: 2106.15745 [astro-ph.HE].

Abstract

Gravitational wave observations of binary neutron star mergers provide valuable information about neutron star structure and the equation of state of dense nuclear matter. Numerous methods have been proposed to analyze the population of observed neutron stars and previous work has demonstrated the necessity of jointly fitting the astrophysical distribution and the equation of state in order to accurately constrain the equation of state. In this work, we introduce a new framework to simultaneously infer the distribution of binary neutron star masses and the nuclear equation of state using Gaussian mixture model density estimates which mitigates some of the limitations previously-used methods suffer from. Using our method, we reproduce previous projections for the expected precision of our joint mass distribution and equation of state inference with tens of observations. We also show that mismodeling the equation of state can bias our inference of the neutron star mass distribution. While we focus on neutron star masses and matter effects, our method is widely applicable to population inference problems.

16.1 Introduction

Over the past six years, the LIGO-Virgo gravitational wave detectors [1, 38] have made > 50 observations of black hole and neutron star binary mergers [24], providing a new way to study some of the most energetic events in the universe. Growing catalogs of gravitational wave observations allow us to study the populations from which compact binary systems originate, offering further insights into the physical nature of these systems [7, 29, 590, 591, 553, 397].

Population inference from gravitational wave observations is performed by comparing catalogs of observed events to models of the astrophysical distribution. These astrophysical models include strongly physically motivated models (e.g., [593, 575]), phenomenological models inspired by theoretical predictions and prior observations (e.g., [219, 584]), or data-driven models (e.g., [540]).

Such population studies are an example of hierarchical Bayesian inference, combining a set of observed events, marginalizing over the single-event parameters for each event, and extracting global properties that govern the single-event parameters to probe the underlying distribution of events, putting observed properties of single events into a wider astrophysical context (see, e.g. Thrane and Talbot [536], Mandel, Farr, and Gair [359], and Vitale et al. [560] for recent reviews.) Using gravitational wave observations in a hierarchical framework provides a powerful method of constraining universal properties of merging binary neutron star (BNS) systems, such as the BNS mass distribution [217, 219] and equation of state (EOS) of dense nuclear matter [45, 543, 584, 323, 275, 252, 325, 141].

While terrestrial experiments have constrained the EOS of cold nuclear matter for densities approaching the nuclear saturation density, a complete picture of the microphysics of nuclear matter above these densities has yet to be confidently determined. With central densities reaching several times nuclear saturation density, neutron stars —observed via kilonovae spectra and light curves, X-ray pulsar measurements, and gravitational waves —provide a probe of nuclear physics at super-saturation densities [107, 377, 379, 497, 376, 157].

The first detection of a BNS merger via gravitational waves [10, 15] provided constraints on the EOS of dense nuclear matter, favoring more "soft" or compressible EOSs over stiffer EOSs [9]. The second observed binary neutron star merger [11] did not provide significant constraints on the EOS; however, the relatively high mass of this system suggests a tension with the galactic population of binary neutron star

systems [11, 244].

While only two confident detections of BNS mergers have been made by LIGO-Virgo, further detections in the near future will provide constraints on the EOS of high-density nuclear matter through the combination of observed events [366, 138]. Unlike gravitational wave signals from binary black hole mergers, signals from BNS mergers contain information about the neutron star EOS. This information is primarily encoded by the tidal deformabilities of the two bodies during late-stage inspiral, with the magnitude of this effect determined by the underlying EOS [283, 94, 595].

Specifically, the EOS directly governs the pressure-density relationship inside the star, a necessary ingredient for solving the Tolman-Oppenheimer-Volkoff equations for the mass-radius relationship of neutron stars [595, 341]. For a given EOS, the mass determines the magnitude of a neutron star's quadrupole moment induced from the external field during merger, imprinting a signature in the detected gravitational wave signal [283, 142, 138, 15, 9, 534]. This imprint is commonly expressed in terms of dimensionless tidal deformability (Λ), which is defined as [138, 9, 232]

$$\Lambda = \frac{\lambda}{m^5} = \frac{2}{3} k_2 \left(\frac{R}{m}\right)^5, \tag{16.1}$$

where k_2 is the quadrupole Love number, R is the radius of the neutron star, and m is its mass (we express this formula in units where $c = G = 1$). The EOS determines both k_2 and R for a given neutron star mass, resulting in a unique $\Lambda - m$ relationship for different (hadronic) EOSs [595, 138, 283, 563]. Under the assumption of a common EOS among neutron stars, we can infer this $\Lambda - m$ relationship by combining observations, constraining the underlying EOS.

Previous work has shown that hierarchical inference can be used to constrain the neutron star EOS by considering parameterized models of neutron star populations in conjunction with EOS relations [323, 325, 584]. In [584], the authors emphasize the importance of simultaneously inferring the mass distribution and EOS, due to bias that results from independent analyses. In this work, we introduce and implement a new method of performing a simultaneous hierarchical analysis to infer mass distribution and EOS. Specifically, we use a Gaussian mixture model (GMM) as an estimate of single-event posterior probability densities. Using this method, we demonstrate that mismodeling the EOS can lead to a biased inference of the neutron star mass distribution.

The paper is organized as follows. In Section 16.2, we detail the process of our density estimation procedure and how it can be implemented in general hierarchical inference problems. We then outline our choice of parameterized BNS mass population and EOS models in Section 16.3. We follow this in Section 16.4 with the details of the simulated data we use for the proof-of-concept and a description of how we apply this method to simultaneous EOS and mass distribution inference. We review the results of our simulated data study in Section 16.5 and conclude with takeaways and motivations for future work in Section 16.6.

16.2 Methods

We begin by reviewing Bayesian inference in the context of gravitational wave data analysis (see, e.g., [536, 560] for recent reviews). In Bayesian inference, one constructs the posterior distribution $p(\Theta|d)$ for a model with parameters Θ given some data d. Bayes' theorem is typically written as

$$p(\Theta|d) = \frac{\mathcal{L}(d|\Theta)\pi(\Theta)}{Z(d)} \tag{16.2}$$

where $\mathcal{L}(d|\Theta)$ is the likelihood of the data given the model parameters, $\pi(\Theta)$ is the prior probability distribution, characterizing our prior beliefs on the distribution of Θ, and $Z(d)$ is the Bayesian evidence, or the marginal likelihood for the model [1]. In gravitational wave analysis, $\mathcal{L}(d|\Theta)$ is typically taken to be a Gaussian likelihood distribution, whose mean is given by a (frequency domain) gravitational waveform characterized by Θ and variance given by the detector noise (e.g., [552]). The full set of Θ typically contains parameters intrinsic to the merger event (such as masses and spins) as well as the extrinsic parameters, such as position in the sky and luminosity distance.

Because the set of parameters Θ is typically > 10 dimensional, to recover the posterior distribution, Equation 16.2 is commonly sampled iteratively using a Markov Chain Monte Carlo (MCMC) sampler or nested sampler (e.g., [552, 67]). Once the sampler converges, we are left with a set of samples drawn from the posterior distribution. This posterior distribution represents our full knowledge about the physical parameters of the source of the gravitational wave event with our prior distribution.

Now, we combine multiple events hierarchically and sample the *hyper*posterior, in order to learn about the *hyper*parameters (or population parameters) Ω that describe

[1] While Equation 16.2 is technically conditioned on a model, we suppress the explicit dependence in our notation.

the global distribution of a subset of single-event parameters (e.g., the distribution of neutron star masses). We denote these single-event parameters of interest as θ, a vector of length D, which is a subset of Θ.

We do the combination by replacing the fixed model above with the set of hyperparameters describing the population model [536, 560].

$$p(\Omega|\mathbf{d}) = \frac{\pi(\Omega)\mathcal{L}(\mathbf{d}|\Omega)}{Z(\mathbf{d})}. \tag{16.3}$$

In the above notation, $\pi(\Omega)$ is the hyper-prior, and Z is the Bayesian evidence for all the observed data $\mathbf{d}$ marginalized over the hyper-prior. Assuming the observed events are N independent draws from the population, we express the (hyper-) likelihood as

$$\mathcal{L}(\mathbf{d}|\Omega) = \prod_{i=1}^{N} \int \mathcal{L}(d_i|\theta_i) p(\theta_i|\Omega) d\theta_i, \tag{16.4}$$

where for each observed event we marginalize over the single-event parameters θ conditioned on a population model $p(\theta|\Omega)$. The likelihood $\mathcal{L}(d_i|\theta)$ is implicitly already marginalized over all members of Θ not included in θ.

In order to account for selection effects, we augment Equation 16.4 by including a selection term:

$$P_{\rm det}(\Omega) = \frac{1}{N_{\rm inj}} \sum_{i=1}^{N_{\rm found}} \frac{p(\theta_i|\Omega)}{p(\theta_i|\Omega_0)}. \tag{16.5}$$

We compute this term by injecting into simulated detector noise $N_{\rm inj}$ simulated events from a fiducial source population Ω_0, and determining how many of those events pass our SNR detection threshold (see Section 16.4). In this equation, the term θ_i consists of the parameters of the ith "found" injection. Our new total likelihood is:

$$\mathcal{L}(\mathbf{d}|\Omega) = P_{\rm det}(\Omega)^{-N} \prod_{i=1}^{N} \int \mathcal{L}(d_i|\theta_i) p(\theta_i|\Omega) d\theta_i. \tag{16.6}$$

A key challenge in population inference is efficiently evaluating Equation 16.4 for a large catalog of events as Equation 16.3 is typically also constructed using stochastic sampling, requiring as many as several million evaluations.

Current techniques to compute this integral via Monte Carlo integration involve reweighting posterior samples (assuming the fiducial prior) by the corresponding population likelihood for each sample, for example in the analysis performed in [29].

This can be efficiently parallelized using graphics processing units (GPUs) [520] to control the run time. However, this method fails for very narrow distributions, due to having only limited samples from the posterior.

In this work, we consider the converse weighting for this marginalization step: we sample from the population model and compute the likelihood of the observed data for each event given these samples. This requires an efficient method for evaluating the likelihood at arbitrary points in parameter space. Previous work has used Kernel Density Estimates (KDEs, [473]) and Gaussian Processes (GPs, [446]) for density estimation [584, 325, 164]. While KDEs can be made quickly, distributions with sharp edges and high dimensions can cause the KDE to break down and the complexity of evaluating the KDE scales with the number of samples in the distribution. Similarly, GPs have been shown to provide good fits in small dimensions, but finite-binning effects from fitting the histogrammed samples can limit the accuracy of the density estimate. Another GP method for making density estimates is used in the parameter estimation code RIFT [328]. However, this requires fixed sets of intrinsic parameters and can be unsuitable for analyzing relationships between source-frame parameters, which is needed in this work (see Section 16.3).

While these methods provide estimates of single-event likelihoods, they each involve assumptions and/or computational complexities which may make them sub-optimal for any general given hierarchical inference problem [164, 522]. In the next section, we make density estimates of single-event likelihoods using GMMs for use in a hierarchical inference framework. The steps presented here provide a relatively computationally-inexpensive density estimation procedure that avoids the shortcomings of the methods outlined above.

16.2.1 Density Estimation

We begin with the goal of being able to evaluate the individual likelihoods $\mathcal{L}(d|\theta)$ at any arbitrary point in parameter space. To do this we must begin with our N sets of posterior samples (for N events) and create a functional form for each likelihood.

As an estimate of the D-dimensional marginalized likelihood for an observed event, we model the likelihood as a linear combination of several D-dimensional Gaussians, where D is the number of parameters of interest in each event's posterior. In such a Gaussian mixture model (GMM), a density estimate is made from the set of discrete samples from the single-event posterior, resulting in an analytic model for the likelihood, allowing for evaluation of $p(d|\theta)$ for any θ.

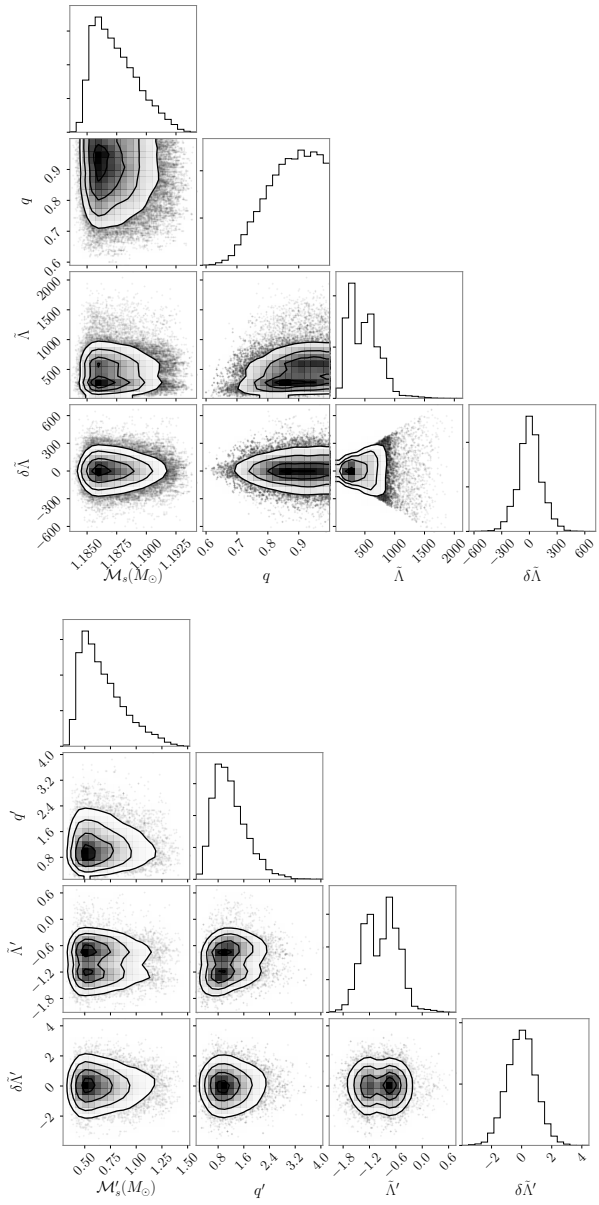

Figure 16.1: (Upper panel) Posterior distribution for the mass and tidal parameters of GW170817 in physical space (before transformation). (Lower panel) Posterior distribution for GW170817 after transforming samples into fitting space. Note that the domain of the transformed samples is much more uniform across parameter space and the sharp edges both in 1-D and 2-D histograms have been removed.

16.2.1.1 Pre-processing

Realistic posterior distributions for parameters in gravitational wave analysis are subject to sharp edge effects, widely differing domains between parameters, and other features which make the raw distribution unsuitable for reliable fitting with our density estimation method. In the top panel of Figure 16.1 we plot the posterior distribution for the chirp mass ($\mathcal{M}$), mass ratio (q), and two tidal parameters $\tilde{\Lambda}$, and $\delta\tilde{\Lambda}$ (see Section 16.4 for descriptions of these parameters) for GW170817 [9]. For this event, we note the marginal distribution for mass ratio q exhibits a hard cutoff at $q = 1$, and the correlation plot between $\tilde{\Lambda}$ and $\delta\tilde{\Lambda}$ has a sharp triangular shape.

We follow [522] to map the observed distribution to one that is smoother and on a better-behaved domain, and use this transformed distribution to train the density estimate.

First, we map each posterior sample in θ to the unit interval using the cumulative distribution function (CDF) F of the prior. Next, we transform the samples from the unit interval to the unit normal distribution. The transformed sample is

$$\theta' = \Phi^{-1}(F(\theta)), \tag{16.7}$$

where Φ^{-1} is the probit function, the inverse CDF of the unit normal distribution [105] [2].

The result of this transformation on the posterior distribution for GW170817 can be seen in the lower panel of Figure 16.1. The original posterior distribution (in physical space) can be compared to the transformed distribution, which has been made more suitable for fitting to a GMM.

16.2.1.2 Fitting the Distribution

After mapping the samples from the posterior distribution using the method in the previous section, the transformed samples follow a distribution more suitable to fitting with a GMM.

We train the model on the posterior samples and determine the maximum likelihood means, covariances, and weights assigned to each component of the GMM.

[2] Note that if an analytic form of $F(\theta)$ is not known, an interpolant may be necessary to compute $F(\theta)$ for arbitrary values of θ.

Mathematically, the density estimate of an observed posterior distribution is

$$p(\theta_i'|d_i) \approx \sum_k^K \phi_{ik} \mathcal{N}(\theta_i'|\mu_{ik}, \Sigma_{ik}), \tag{16.8}$$

where the k^{th} component of the mixture is a multivariate Gaussian of mean μ_k and covariance Σ_k, weighted by ϕ_k. Here, $\sum_k^K \phi_{ik} = 1$ for a K-component GMM, for each i. Note that the index k here runs over the *components* (individual Gaussians) in the mixture model, not the observed events, as this mixture model is unique to each event. We use the Gaussian mixture model as implemented in Scikit-Learn [421], which uses an expectation-maximization method to fit for the μ, Σ, and ϕ parameters in Equation 16.8 conditioned on a set of transformed samples from the posterior distribution of the ith event.

Since a GMM is a sum of individual weighted Gaussian components, we must determine how many such components to use to make the optimal fit characterizing the distribution without overfitting. To determine this, we take a set of posterior samples from an event and we randomly assign 80% of the transformed posterior samples to a training set and the other 20% to a testing set. We train the K-component GMM using the training data and evaluate the score (sample-wise average log-likelihood) of the testing samples for the GMM. To determine the optimal number of components to use, we vary K and repeat this process until the score noticeably flattens, indicating an increased value of K does not better characterize the distribution. When working with a catalog of observed events, it may be efficient to do this step using one selected event (possibly corresponding to the most complex posterior distribution), and use this optimal K for all GMM density estimates in the catalog. However, a more complete fitting method would consist of fitting the for the optimal K_i for each event i, rather than using the same K for all events. We note that one could also fit for the optimal K_i values for each event using a statistic such as the Bayesian information criterion (BIC) [302]. However, we find that the fits we obtain from the flattening of the score are sufficient for good recovery of our simulated distribution, as reported in Section 16.5.

In Figure 16.2, we show this curve using our GMM fits from GW170817. As the score flattens out by $K = 8 - 10$ components, this represents the optimal number of components to use in the GMM for this posterior distribution. As an illustration, we compare the GMM fit to the true posterior distribution using samples drawn from the GMM and the original transformed samples for GW170817 in Figure 16.3.

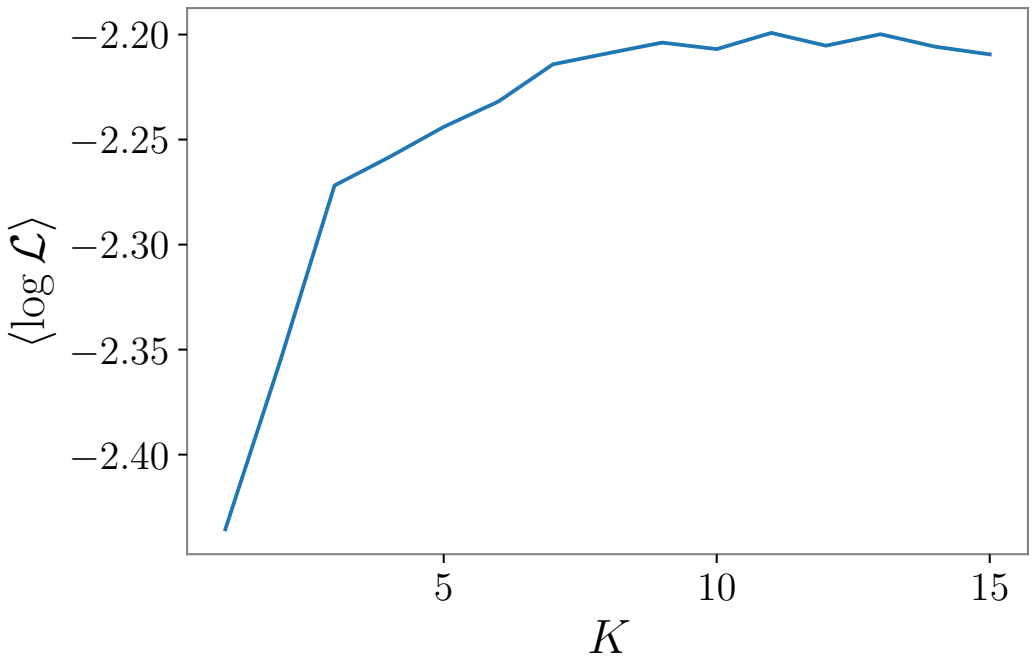

Figure 16.2: Average (natural) log likelihood of evaluation samples as a function of number of components (K) used to generate GMM, using posterior samples from GW170817. The score flattening out by $K \approx$ 8-10 indicates that the GMM does not provide a better density estimate for larger K.

Since the GMM is trained on transformed *posterior* samples, we convert the GMM density estimate $\mathcal{D}_i(\theta')$ from the ith event into a single-event *likelihood* via:

$$\mathcal{L}_i(\theta') = \frac{\mathcal{D}_i(\theta')}{\mathcal{N}(\theta', \mu = 0, \sigma = 1)}. \tag{16.9}$$

This results in the correct likelihood because we used the sampling priors for $F(\theta)$ in the original transformation $\theta \to \theta'$ (see Equation 16.7).

16.2.2 Hierarchical Likelihoods using Density Estimates

For a given population model $p(\theta|\Omega)$, we compute the likelihood of the event i (for $i \in (1, N)$) as:

$$\mathcal{L}_i(d_i|\Omega) = \frac{1}{M} \sum_{\theta \sim p(\theta|\Omega)} \mathcal{L}_i(\theta'), \tag{16.10}$$

where $\mathcal{L}_i$ is the likelihood of the ith event (Equation 16.9), and M samples θ are drawn from the population model $p(\theta|\Omega)$. This is a practical Monte Carlo integration scheme for the integral in Equation 16.4, dependent on the ability to

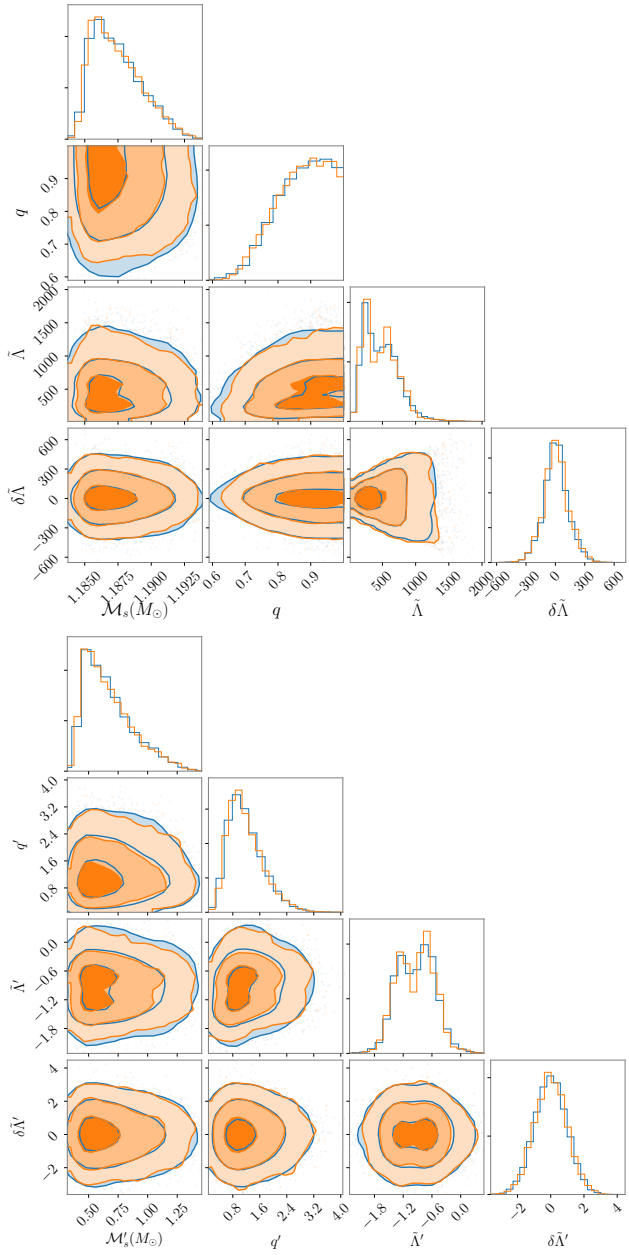

Figure 16.3: 4-dimensional posterior distribution for GW170817. Orange is the posterior samples and blue is from samples of the GMM fit. The overlap between the two distributions shows the GMM provides a good density estimate. Plotted in physical (transformed) space in the upper (lower) panel.

sample from the population model and evaluate the single-event likelihoods at the corresponding points in parameter space.

An implicit step in the above hierarchical likelihood equation is the mapping of the population samples θ into the corresponding transformed (fitting) space samples θ' for each density estimate, to match the space of the density estimates.

The total likelihood (i.e. Equation 16.4) for the data from N events therefore becomes:

$$\mathcal{L}(\mathbf{d}|\Omega) = \prod_{i=1}^{N} \frac{1}{M} \sum_{\theta \sim p(\theta|\Omega)} \mathcal{L}_i(\theta'). \tag{16.11}$$

Although Scikit-Learn provides a method of computing the log-likelihood of samples in a GMM fit, for a single evaluation of the total likelihood, each of N GMMs must be evaluated for M samples, which can become a computational burden for large N and M. When there are many observed events, it becomes more efficient to extract the best-fit means, covariances, and weights from the GMM fits and evaluate the likelihood matrices in Equation 16.8 directly on a GPU using CuPy [403]. This avoids explicitly looping over the N GMMs for each evaluation of the joint likelihood, while efficiently performing computations over the $O(N \times M \times K)$ array using array broadcasting and vectorization on a GPU.

16.3 Models

For the implementation of our GMM-based hierarchical inference method, we let Ω characterize the mass distribution as well as the EOS relating the Λ and m parameters. This requires choosing parameterized models for the mass distribution and the EOS model.

16.3.1 Mass Population Model

To model the distribution of neutron stars in merging binaries, we consider the observationally-motivated framework in [219]. In that work, the authors found evidence based on observations of galactic BNS systems for each companion of a BNS system being drawn from a separate population distribution. The first population distribution characterizes the member of the binary that forms first and spins up due to accretion, known as a recycled neutron star. The other member of the system is known as the slow neutron star, as it is born second and spins down quickly after formation. The authors found the model with the best support consists

of a two-component Gaussian model for the recycled neutron star mass distribution, and a uniform distribution for the slow neutron star mass.

Adopting this mass distribution model, the subset of Ω describing the mass population consists of 8 parameters. The lower-mass Gaussian of the recycled distribution is described by the parameters $(\mu_{R,a}, \sigma_{R,a})$ and is weighted by α, and the higher-mass component is described by $(\mu_{R,b}, \sigma_{R,b})$. The probability of observing a mass m from the recycled mass distribution is

$$\begin{aligned} p_R(m|\Omega) = \alpha \mathcal{N}(m|\mu_{R,a}, \sigma_{R,a}) \\ + (1-\alpha)\mathcal{N}(m|\mu_{R,b}, \sigma_{R,b}). \end{aligned} \tag{16.12}$$

For the slow mass distribution, we denote the low and high limits as $m_{S,l}$ and $m_{S,h}$, respectively. The maximum mass parameter, $M_{\rm max}$ represents an absolute cutoff of both distributions; we truncate the recycled (Equation 16.12) and slow mass distributions at $M_{\rm max}$ on the high end and 1 $M_\odot$ on the low end.

By considering the observed galactic BNS systems in the context of binary formation and evolution models, [597] found that modeling the slow companion as non-spinning was a robust approximation for gravitational wave data analysis. The recycled partner, while spun-up from the slow companion, has very little support for spins of $\chi > 0.05$ for population and EOS models they considered. We therefore do not consider spins at all in this work, and model all sources as nonspinning. Since we neglect spins, we have no way of concretely knowing which component mass (m_1, m_2) represents the slow or recycled mass, so each computation of the population probability must account for the possibility of either component being drawn from either distribution, with the constraint that each BNS system consists of exactly one recycled and one slow neutron star.

16.3.2 EOS Model: The Λ–m Relation

Several nonparametric and parametric models for EOS-sensitive observables exist in gravitational wave literature, based on the assumption that a neutron star of a given mass will have a corresponding Λ uniquely determined by its EOS [448, 342, 325]. Therefore, recovering parameters characterizing this mapping between the two observables, m and Λ, may provide a way to recover information about the underlying nuclear EOS.

To model the EOS-sensitive relationship, we follow the examples of [45, 173] and consider a simple expansion of $\lambda(m)$ about the canonical reference value of 1.4 $M_\odot$:

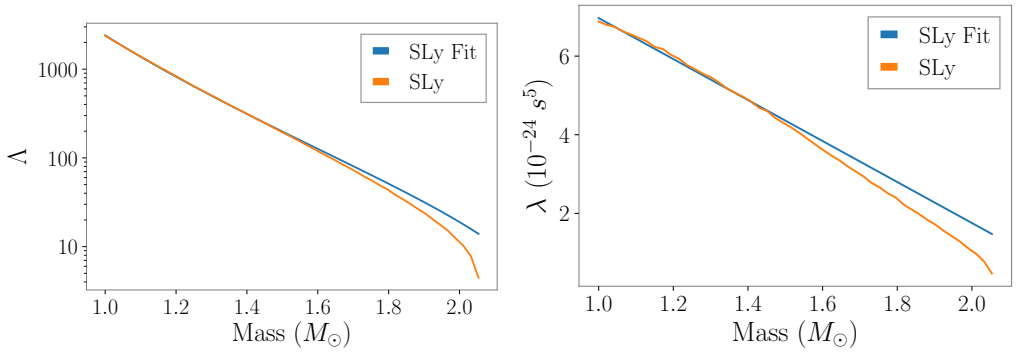

Figure 16.4: Dimensionless tidal deformability (left panel) and tidal deformability (right panel) as a function of mass. Units of λ are in seconds to the fifth power, following the convention in [45].

$$\lambda(m|c_0, c_1) = c_0 + c_1 \left(\frac{m - 1.4M_\odot}{M_\odot} \right). \tag{16.13}$$

The expansion coefficients in this model are our EOS-sensitive parameters, with different combinations approximating different EOSs. Previous work has shown that with this parameterization, LIGO observations will be unlikely to resolve terms c_j for $j \gtrsim 1$, so we only include these first two terms in our demonstration. With component masses of observed galactic BNS systems peaking around 1.4 $M_\odot$, this form for $\lambda(m)$ can provide meaningful constraints on the EOS, as the expansion is centered at this value.

For our fiducial choice of EOS to simulate, we use the values of c_0 and c_1 corresponding to the relatively-soft SLy EOS [183]. In Figure 16.4, we show the linear fit to the true EOS from Equation 16.13 for the $\Lambda - m$ (top) and $\lambda - m$ (bottom) relationships. The approximation breaks down for $m > 1.8M_\odot$ as previously noted in, e.g, [138], however, the majority of simulated events considered in this work are less massive than this. If observed BNS mergers contain high mass components, this linear parameterization will not be valid; however, it is sufficient for our proof-of-principle (see Section 16.6).

This model provides a weak connection between our EOS-sensitive model and the mass population model. Since c_0 and c_1 are allowed to vary in Equation 16.13, it is possible to arrive at a negative value for Λ, which is unphysical. Since $\Lambda(m)$ is a decreasing function with mass, we therefore constrain $M_{\rm max}$ such that $0 < \Lambda(M_{\rm max}) < 200$, consistent with limits on minimum $\Lambda(m)$ for commonly-used EOSs (see Figure 1 in [138]).

It is worth noting the limited significance of the $M_{\rm max}$ parameter in this work, as the constraint $0 < \Lambda(M_{\rm max}) < 200$ is simply a cutoff to keep values of Λ physical. This does not necessarily correspond to the maximum neutron star mass as determined by the EOS; this is known as the Tolman-Oppenheimer-Volkoff mass ($M_{\rm TOV}$), and is determined from the stability conditions set by a particular EOS [300]. By cutting off the population model at $M_{\rm max}$ in this work, we do not associate $M_{\rm max}$ with the most massive *possible* neutron star, but instead $M_{\rm max}$ is the upper limit on the mass of neutron stars in merging binaries. Thus, stellar binary evolution and population channel models play an additional role in the significance of $M_{\rm max}$ in the population. A more rigorous approach may be to calculate $M_{\rm TOV}$ for a given EOS and enforce the condition $M_{\rm max} < M_{\rm TOV}$.

Although this work is a proof-of-concept for this method, improvements to the EOS modeling and parameterization can make the results more realistic and applicable to real observations. The breakdown of our choice of EOS model at high masses serves to potentially bias the entire inference if a significant number of events are observed far from the reference mass used in the Taylor expansion of Equation 16.13. Additionally, previous work has shown that a simple Taylor expansion, such as the model used in this work, may not be robust to EOS models with phase transitions to non-hadronic constituents. Therefore, a more general model robust to arbitrary EOSs without deviations at high masses may be useful for realistic observations.

16.4 Data and Implementation

16.4.1 Data

In order to demonstrate our method, we simulate 100 BNS signals drawn from the mass distribution characterized by the maximum likelihood estimate in [219]. This corresponds to the combination of mass distribution parameters listed in Table 16.1. We specify injected tidal parameters using our linear fit to the SLy EOS, neglecting spins.

We draw the extrinsic parameters isotropically in position and orientation with distances uniform in source frame between 20 – 300 Mpc using the cosmology from the Planck 2015 data release [42].
For each simulated signal, we generate 128s of colored Gaussian noise corresponding to the three-detector Advanced LIGO-Virgo network operating at their projected design sensitivities [16]. We employ a GPU-implementation of the TaylorF2 wave-

form model [121, 520] and analyze data between 20 – 2048 Hz. We simplify our parameter estimation by marginalizing phase, merger time, and luminosity distance for sampling the posterior distribution for single-event analyses. We reconstruct the luminosity distance marginal posterior distribution in post-processing using the method outlined in [536].

We impose a detection threshold of SNR > 8 in at least one detector. For the 37 events passing our detection threshold, we infer the posterior distribution using the Bilby [67] implementation of PyMultiNest [116].
We employ a uniform prior on detector-frame chirp mass ($\mathcal{M}_d$) of $\pm 0.01 M_\odot$ around the injected value for each event. Our prior on q is uniform from 0.125 to 1. For $\tilde{\Lambda}$, defined as [563]

$$\begin{aligned}\tilde{\Lambda} = \frac{8}{13}\Big[&(1+7\eta-31\eta^2)(\Lambda_1+\Lambda_2) \\ &+\sqrt{1-4\eta}(1+9\eta-11\eta^2)(\Lambda_1-\Lambda_2)\Big],\end{aligned} \tag{16.14}$$

we use a uniform prior from 0 to 5000. Here, $\eta \equiv q(1+q)^{-2}$ is the symmetric mass ratio. We construct a conditional sampling prior on $\delta\tilde{\Lambda}$ as follows: for each sample θ_i, we analytically compute the maximum and minimum allowed values of $\delta\tilde{\Lambda}_i$ conditioned on q_i and Λ_i. The parameter $\delta\tilde{\Lambda}$ is defined as [563]

$$\begin{aligned}\delta\tilde{\Lambda}(q,\tilde{\Lambda}) = \frac{1}{2}\Big[&\sqrt{1-4\eta}\left(1-\frac{13272}{1319}\eta+\frac{8944}{1319}\eta^2\right)(\Lambda_1+\Lambda_2) \\ &+\left(1-\frac{15910}{1319}\eta+\frac{32850}{1319}\eta^2+\frac{3380}{1319}\eta^3\right)(\Lambda_1-\Lambda_2)\Big],\end{aligned} \tag{16.15}$$

such that $\delta\tilde{\Lambda}$ deviates from 0 as the differences in component tidal deformabilities increases. It therefore reaches a maximum (minimum) when Λ_1 (Λ_2) is 0. Thus, we calculate the bounds of the uniform prior on $\delta\tilde{\Lambda}$ conditioned on a sample of $(q,\tilde{\Lambda})$, where $\Lambda_{1(2)}$ is computed by setting $\Lambda_{2(1)} = 0$ for fixed q and $\tilde{\Lambda}$ in Equation 16.14, and using the resulting value of $\delta\tilde{\Lambda}$ as the upper (lower) bound. We then consider a uniform prior on $\delta\tilde{\Lambda}$ from these conditions.

While $\mathcal{M}$ is well-constrained in the detector frame, the hyperparameters we consider in this work are only relevant to source-frame masses. Therefore, if a given set of posterior samples only contain $\mathcal{M}$ in the detector frame, they must be converted to source frame via the relationship $\mathcal{M}_d = (1+z)\mathcal{M}$. We construct our corresponding prior on $\mathcal{M}_s$ as [536],

$$\pi(\mathcal{M}_s) = \int dz d\mathcal{M}_d \pi(z)\pi(\mathcal{M}_d)(1+z)\delta\left(\mathcal{M}_s - \frac{\mathcal{M}_d}{(1+z)}\right). \tag{16.16}$$

This is marginalized over both detector-frame mass and distance. This relation may be unique to each observed event if a different prior on $\mathcal{M}_d$ is used for each event. We therefore associate a unique prior on $\mathcal{M}$ with each GMM [3].

To make the density estimates of each posterior distribution, we follow the method outlined in the previous section, using our single-event sampling priors for F in Equation 16.7 for the transformation into fitting space for each event's posterior. As an example of GMM density estimation on BNS posterior distributions, we show the fit to GW170817 in Figure 16.1. We observe that BNS posteriors may include strong correlations between $\tilde{\Lambda}$ and $\delta\tilde{\Lambda}$ (i.e. the triangular shape in the $\tilde{\Lambda} - \delta\tilde{\Lambda}$ correlation plot in Figure 16.1), possibly impacting the quality of the density estimate. Using the CDF of our conditional prior as $F(\theta)$, the transformation decorrelates $\tilde{\Lambda}$ and $\delta\tilde{\Lambda}$. As can be seen in the bottom panel of Figure 16.1, the correlation between the tidal parameters no longer exists in transformed space when imposing this condition on the sampling prior.

To motivate our choice for the number of components in our GMMs, we implement the scoring method described in Section 16.2 on GMM density estimates of GW170817 and shown in Figure 16.2. Based on this example, we use 10 components for each GMM density estimate of our simulated BNS events.

16.4.2 Sampling the Hyper-posterior

For each calculation of the likelihood, we sample $M = 15,000$ masses from the population model and compute the corresponding tidal deformabilities conditioned on EOS-sensitive hyperparameters via the relationship in Equation 16.13. We then convert the M samples of $(m_r, m_s, \Lambda_r, \Lambda_s)$ to $(\mathcal{M}, q, \tilde{\Lambda}, \delta\tilde{\Lambda})$ and then into the fitting space for each of the GMMs. While we sample component masses in terms of recycled and slow mass, we convert (m_r, m_s) to (m_1, m_2), adopting the convention $m_1 > m_2$. Using these transformed samples, we can evaluate Equation 16.11.

To calculate $P_{\rm det}(\Omega)$ we draw 20,000 binaries from a mass distribution that is uniform in $[1, 2.2]M_\odot$. For each simulated binary, we compute the SNR in an independent noise realization and keep those that pass our threshold. We neglect the impact of tidal effects on sensitivity. Since our mass distribution model is in terms of slow and recycled components, but our analysis can only specify m_1 and m_2, for the purposes of computing $p(\theta|\Omega)$ we assume *a priori* each object is equally

[3] Note we express $\mathcal{M}_s = \mathcal{M}$ in this work, with detector-frame chirp mass written explicitly as $\mathcal{M}_d$

likely to be the recycled or slow companion, with the assumption each binary system contains exactly one slow and one recycled partner.

Each likelihood evaluation requires $O(M \times N \times K) \sim 6 \times 10^6$ computations. Running on an NVIDIA GeForce RTX 3080 GPU, each full likelihood evaluation took $\lesssim$ 50 ms for our 37 events which is comparable to the evaluation time for the method currently used to infer binary black hole mass distributions in LIGO-Virgo-KAGRA analyses [29, 520]. We note that the sampling, transformation, and selection function steps in the likelihood introduce subdominant effects to computation time relative to the computation of $M \times N \times K$ Gaussian likelihoods. We sample the hyper-posterior using the Bilby wrapper of the nested sampler PyMultinest [116], sampling with 250 live points.

16.5 Results

In Figure 16.5 we show the inferred mass distribution when the mass and EOS hyperparameters are sampled simultaneously. The solid line shows the posterior predictive distribution (PPD), the shaded regions show the symmetric 68% credible region, and the dashed lines show the true simulated distribution. With the priors on the mass distribution hyperparameters spanning a wide range, the posterior distribution is relatively well-constrained around the input hyperparameters (see Figure 16.11) for the full one- and two-dimensional posterior distributions.

Of note, we confidently recover the presence not only of the large peak in recycled mass distribution at $1.34 M_\odot$, but also the small and wide peak at higher masses, as shown by the μ_{Rb} and σ_{Rb} panels in Figure 16.11, as well as in the PPD in Figure 16.5. The inferred location of the large peak is $\mu_{Ra} = 1.32^{+0.03}_{-0.03} M_\odot$ (all ranges are 90% credible intervals), constrained to within ~4% of the input value. It is worth noting that the hyperparameters associated with the second peak of the recycled-mass distribution unsurprisingly show the poorest recovery. This is expected as the second peak in this distribution is very small (i.e. most masses from the recycled distribution are in the lower-mass peak) and thus very few events coming from these masses are expected. Nevertheless, we are able to recover evidence of this second peak around its input location. The presence of a secondary peak in the recycled mass distribution is favored over a single Gaussian, with a Savage-Dickey density ratio giving a Bayes factor of 2.6 in favor of a secondary peak ($\alpha \neq 0$ or $\alpha \neq 1$).

Additionally, the bounds of the slow mass distribution are well-constrained, with inferred values of $M_{Sh} = 1.45^{+0.08}_{-0.06} M_\odot$ and $M_{Sl} = 1.17^{+0.08}_{-0.04} M_\odot$. Both of these

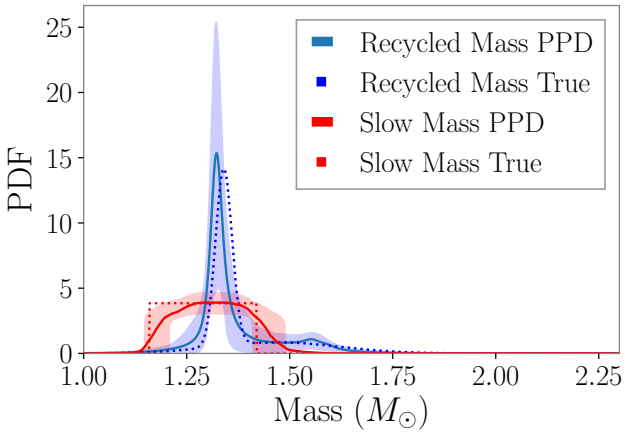

Figure 16.5: Inferred mass distribution from the full mass + EOS analysis of 37 simulated events. Solid lines represent the posterior predictive distribution (PPD). The recycled (slow) distribution is colored blue (red), with shading representing the $\pm 1\sigma$ (68%) credible region from the posterior. Dashed lines show the input distribution.

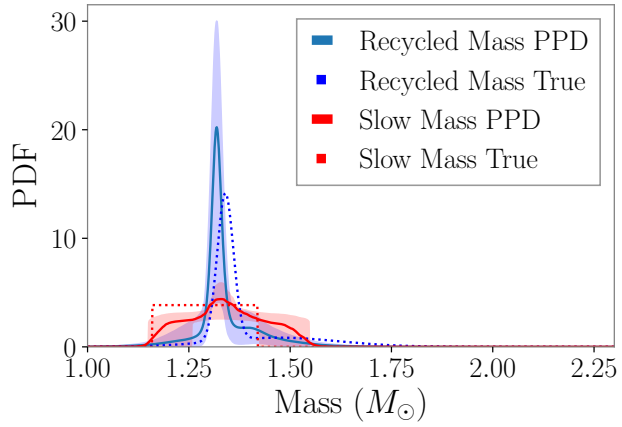

Figure 16.6: Inferred mass distribution from the mass-only analysis of 37 simulated events. Solid lines represent PPD. The recycled (slow) distribution is colored blue (red), with shading representing the $\pm 1\sigma$ (68%) credible region from the posterior. Dashed lines show the input distribution. Compare to Figure 16.5 to observe the bias in mass distribution recovery due to not including EOS inference.

parameters are therefore constrained to within $\sim 10\%$ of their input values of $1.42 M_\odot$ and $1.16 M_\odot$, respectively.

Similarly, good recovery is also seen in the EOS parameters (c_0 and c_1), around the values input for our SLy fit which predicts $\Lambda(1.4 M_\odot) = 313$. We infer $\Lambda(1.4 M_\odot) = 322^{+27}_{-25}$, which is constrained to within $\sim 17\%$ of the true value from the SLy fit. We

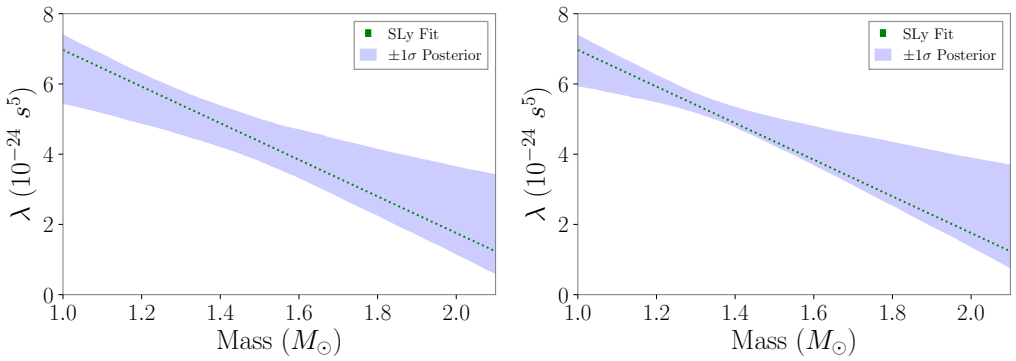

Figure 16.7: Inferred $\lambda - m$ parameter space from the analysis using only low SNR events (left) and high SNR events (right). Note the better recovery of EOS parameters from including few high SNR events compare to many low SNR events.

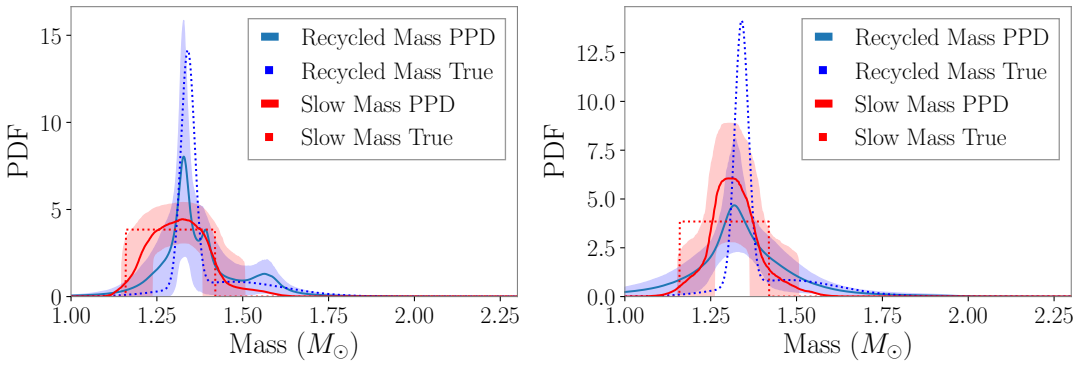

Figure 16.8: Inferred mass distribution from the full mass + EOS analysis of only low SNR events (left) and only high SNR events (right). Solid lines represent the posterior predictive distribution (PPD). The recycled (slow) distribution is colored blue (red), with shading representing the $\pm 1\sigma$ (68%) credible region from the posterior. Dashed lines show the input distribution. Compare to Figure 16.5 to note the worse recovery due to not including the full set of events.

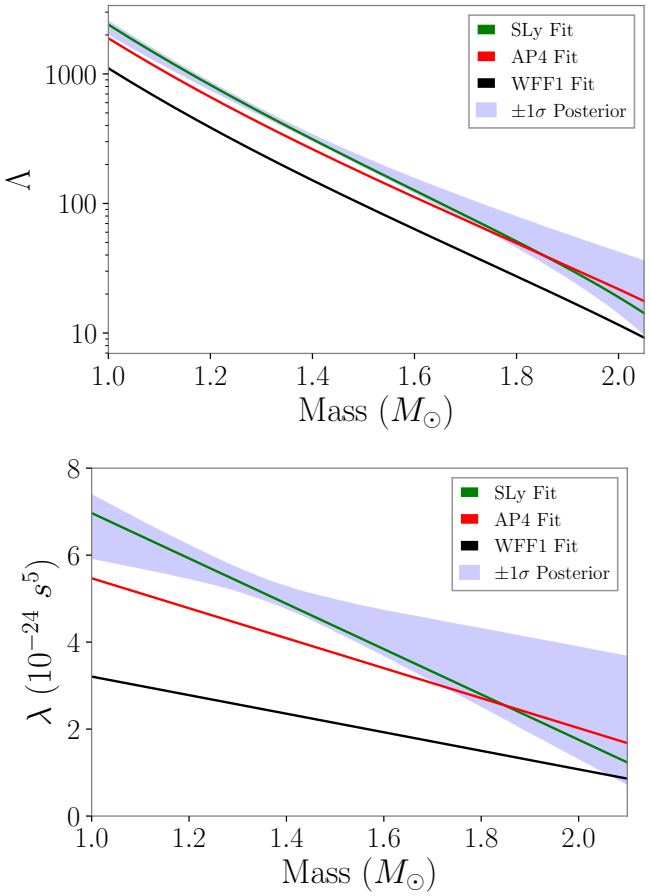

Figure 16.9: Inferred $\Lambda - m$ (top) and $\lambda - m$ (bottom) parameter space from the population + EOS analysis. Shaded region corresponds to $\pm 1\sigma$ (68%) region from c_0 and c_1 posterior samples. For reference, selected EOS curves overplotted.

also recover $\Lambda(1M_\odot) = 2282^{+435}_{-333}$ and $\Lambda(2M_\odot) = 29^{+18}_{-23}$, the inferred dimensionless tidal deformability of a 1 $M_\odot$ and 2 $M_\odot$ neutron star, respectively. The relatively wide credible intervals for these parameters can be understood as a result of the vast majority of our simulated neutron stars having masses closer to $1.4M_\odot$, with very little support in the mass distribution at $1M_\odot$ or $2M_\odot$. For reference, we also plot the $\Lambda - m$ relationship for two other soft EOSs: AP4 [52] and WFF1 [572] which are EOSs compatible with observations from GW170817 [15]. This way we show that the recovery can also favor the input EOS over some similar EOSs. Thus, we will be able to robustly distinguish between these EOSs after observing 37 events.

As noted in Section 16.3, the $M_{\rm max}$ parameter is fairly insignificant in this work, with its recovered posterior distribution relatively flat (see Figure 16.11). There is however a sharp, slanted cutoff in the correlation plot between c_1 and $M_{\rm max}$, owing

to the constraint we impose on $\Lambda(M_{\rm max})$ (see Section 16.3).

The recovery of the EOS parameters obtained in this work stands in contrast to what was found in [45]. Simulating component masses from a narrow Gaussian peaked at $1.35M_\odot$, but assuming a flat mass prior, the authors needed $O(> 100)$ events in their catalog to distinguish their candidate EOSs. In contrast to this work, those authors aimed to distinguish between soft, moderate, and stiff EOSs, whereas we only consider three similar soft EOSs, limiting the significance of a direct comparison between works. Despite comparing more similar EOSs in our work, we find a substantially lower threshold for distinguishing EOSs than in [45] by performing population inference simultaneously.

Motivated by [323, 275], we test how well this method can constrain the mass distribution and EOS hyperparameters from only lower-SNR or higher-SNR events by repeating the above analysis, but limiting ourselves to a portion of the events. The low SNR events are the 24 of our simulated events which had a network SNR $<$ 20 (but above our SNR threshold of 8), and the remaining 13 events are the events with SNR $>$ 20.

Consistent with those studies, we find that EOS-sensitive parameters show definitively worse recovery when only including the low SNR events. The inferred $\pm1\sigma$ parameter space in the left panel of Figure 16.7 in this case is much wider than when all the events are included, indicating that the SNR $>$ 20 events are providing a significant amount of the EOS information in this analysis. For instance, when only considering these low SNR events, we infer $\Lambda(1.4M_\odot) = 309^{+61}_{-63}$. On the other hand, only analyzing the 13 events with SNR $>$ 20 provides $\Lambda(1.4M_\odot) = 324^{+28}_{-29}$ (see right panel of Figure 16.7), constraints comparable to those from the full analysis with all SNR $>$ 8 events (see left panel of Figure 16.9).

As can be seen in Figure 16.8, the mass distribution recovery is poorer in the low-SNR case compared to including all events in the analysis. Specifically, we recover $\mu_{Ra} = 1.33^{+0.07}_{-0.11}M_\odot$; while this interval contains the true value of μ_{Ra}, it is three times larger than what we get from the analysis using the full set of events. The 13 high SNR events contribute less information to the mass distribution inference than the 24 low SNR events, giving an inferred value $\mu_{Ra} = 1.30^{+0.07}_{-0.21}M_\odot$; a credible interval ∼4.5 times larger than that from the analysis using the full set of events. Therefore, unlike the case for the EOS, the mass distribution is not preferentially informed by high SNR events but is most sensitive to the number of events in the population. Because the mass parameters (chirp mass in particular) tend to be

relatively well-constrained, several observations of lower-SNR BNS merger events can provide constraints on the mass distribution of merging neutron stars.

To estimate the bias from not inferring mass distribution and EOS hyperparameters simultaneously, we conduct the analysis from above but only sample the mass distribution hyperparameters and make GMM density estimates of $\mathcal{M}$ and q for our 37 observed events. By only considering the mass parameters in our analysis, we neglect the $\Lambda - m$ relationship from the EOS and implicitly (mis)model the tidal parameters as independent draws from the prior distribution used in the single-event sampling. As seen in Figure 16.6, the mass distribution becomes noticeably biased at the 68% credible level [4], with the PPD of the mass spectrum shifted from the input distribution. In this case, we recover $\mu_{Ra} = 1.32^{+0.02}_{-0.09} M_{\odot}$, a credible interval which almost does not contain the true value. This bias is due to ignoring the correlations between the mass parameters (particularly the mass ratio) and the tidal parameters, which can be seen in the 2d posterior panels between q and $\tilde{\Lambda}$ in single-event posteriors of our simulated events (see Figure 16.10). Therefore, inferring the $\Lambda - m$ relationship simultaneously with the mass distribution is necessary for an unbiased result.

16.6 Discussion

In this work, we demonstrate a new method of hierarchically combining posterior distributions from BNS merger events and inferring mass distribution and EOS parameters simultaneously. The initial step of using GMM density estimates in our transformed space reliably reflects the observed posterior distribution and allows for the evaluation of single-event likelihoods at arbitrary points in parameter space for arbitrary subsets of single-parameters efficiently.

We show that our method can recover underlying population model parameters when combining BNS events simulated with realistic observational parameters and noise realizations, while also constraining parameters of the neutron star EOS. Using our new method, we confirm the importance of inferring EOS and mass distribution parameters simultaneously to avoid potential bias in both the inferred mass distribution and EOS.

Additionally, we observe that both low-SNR (< 20) and high-SNR (> 20) observations contribute to mass population inference, with the few high-SNR observations

[4] The inferred distribution is consistent with the input distribution at the 90% credible level, however.

providing the bulk of the EOS constraints. This finding is generally consistent with the work in [323, 275]. Summarized in Figure 16.9, the EOS recovery from the 37 simulated observations constrains the $\Lambda - m$ parameter space around the input EOS.

The fourth observing run of the LIGO-Virgo-KAGRA network is expected to begin no earlier than Summer 2022 and last one year. At the targeted upgraded sensitivity, it is estimated that there will be 10^{+52}_{-10} BNS detections, significantly raising the prospects for providing constraints on neutron star EOS and population models [16].

The method presented in this work is generalizable to arbitrary population models and can incorporate parameterized models linking population observables to other single-event observables (i.e. $\Lambda - m$ relationship). Gravitational wave population analysis using mass and spin models (see [29]) could be similarly evaluated using this method by making GMM density estimates of mass and spin parameters, and sampling from the population models used in those studies.

We anticipate that the transformed Gaussian mixture model density estimation method employed here has additional potential applications, as it is robust to edge effects and has superior scaling with dimensionality compared to KDE and GP methods. In addition to being able to handle the delta-function population model for tidal parameters, this method can be applied to any situation where the uncertainty in individual measurements is much larger than the domain of the population model, e.g., spin distributions that are not probeable with the method currently employed by the LIGO/Virgo collaboration analyses [7, 29] (see also [584]). Further, Gaussian mixture models are a generative model and can therefore be used to generate $O(10^5)$ additional samples per second from the posterior distribution or as a proposal distribution for subsequent MCMC reanalyses building on the methods in [212, 68].

While this proof-of-concept study used a simple toy model for the neutron star $\Lambda - m$ relation, more sophisticated models can be folded into the method. Additionally, this method can be extended to include a model for the distribution of neutron star spins. Further, this study has focused on the situation when there are tens of measurements, the current population of binary neutron star systems is limited to two. In this small population regime the specific choice of population model/prior will significantly impact the inference. We leave these extensions to future work.

Acknowledgements

We would like to thank Katerina Chatziioannou and Alan Weinstein for useful discussions. We would also like to thank Stefano Rinaldi for useful comments on the manuscript. Finally, we thank the anonymous reviewer for helpful suggestions and critiques on this manuscript. JG and CT acknowledge the support of the National Science Foundation, and the LIGO Laboratory. 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-1764464. This paper carries LIGO Document Number LIGO-P2100215.

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 research has made use of data, software, and/or web tools obtained from the Gravitational Wave Open Science Center (https://www.gw-openscience.org/) [36], a service of LIGO Laboratory, the LIGO Scientific Collaboration and the Virgo Collaboration.

Parameter	Value	Prior	Units
μ_{Ra}	1.34	(1, 2)	$M_\odot$
σ_{Ra}	0.02	(0.005, 0.5)	$M_\odot$
μ_{Rb}	1.47	(μ_{Ra}, 2)	$M_\odot$
σ_{Rb}	0.15	(0.005, 0.5)	$M_\odot$
α	0.68	(0, 1)	N/A
M_{Sl}	1.16	(1, 1.7)	$M_\odot$
M_{Sh}	1.42	(M_{Sl}, $M_{\max}$)	$M_\odot$
$M_{\max}$	2.2	(1.9, 2.3)	$M_\odot$
$c_0/10^{-24}$	4.88	(10^{-1}, 10)	s^5
$c_1/10^{-24}$	-5.21	(-10, -1)	s^5

Table 16.1: Summary of hyperparameters used in demonstration. Value column indicates the number used to generate the samples, and the prior column is the bounds of the uniform prior used for sampling. The mass distribution parameters include the means, μ_R, and standard deviations, σ_R of the low-mass Gaussian, a, and higher-mass Gaussian, b, of the recycled mass distribution, along with a weight α (b weighted by $(1 - \alpha)$). The slow mass uniform distribution is characterized by its lower bound M_{Sl} and upper bound M_{Sh}. See the dashed line in Figure 16.5 for probability density function (PDF) of the input mass distribution. The c_0 and c_1 parameters are the EOS-sensitive parameters in Equation 16.13. All mass parameters in units of $M_\odot$.

Appendix 16.A Single-Event Posteriors

In this Appendix, we show the posterior distribution and samples drawn from the GMM fits for a range of our simulated events. By overplotting the samples from the posteriors and the GMM, we show that the GMM accurately characterizes the posteriors of individual events. We also note that the GMM is able to fit various features including peaks, correlations, and skews that appear in the transformed posterior distributions. This demonstrates the strength of using this density estimate as an analytic approximation of the likelihoods.

The posterior distributions also show the correlations between the tidal parameters and mass parameters (center panels in Figure 16.10). If the mass distribution is inferred independently from the tidal parameters, the information in the correlation (other than the $\mathcal{M} - q$ plot) is lost, impacting the recovery of the mass distribution.

Appendix 16.B Hyper-Posterior

In Figure 16.11 we plot the recovered posterior distribution for the mass distribution and EOS hyperparameters inferred from the analysis using the 37 simulated events.

Appendix 16.C Convergence of Monte Carlo Integrals

Population analyses such as those presented here rely on the use of Monte Carlo integrals to marginalize over the single-event parameters, either by summing over posterior samples for each event with some fiducial prior, e.g., [29], or by summing over samples from the population model as in this work. While such Monte Carlo integrals are asymptotically unbiased estimators, for a finite number of samples there is a finite uncertainty. This uncertainty is generally neglected in the literature, although it has been discussed for the integral estimating the sensitivity function, Equation 16.5 [213]. For a generic Monte Carlo integral of some function f over some set of K samples x_k the expectation is

$$\langle f \rangle = \frac{1}{K} \sum_{k}^{K} f(x_k) \tag{16.17}$$

and the fractional uncertainty is

$$\left(\frac{\Delta f}{\langle f \rangle} \right)^2 = \frac{1}{K} \frac{\left\langle f^2 \right\rangle - \langle f \rangle^2}{\langle f \rangle^2} = \frac{\sum_k^K f^2(x_k)}{\left(\sum_k^K f(x_k) \right)^2} - \frac{1}{K}. \tag{16.18}$$

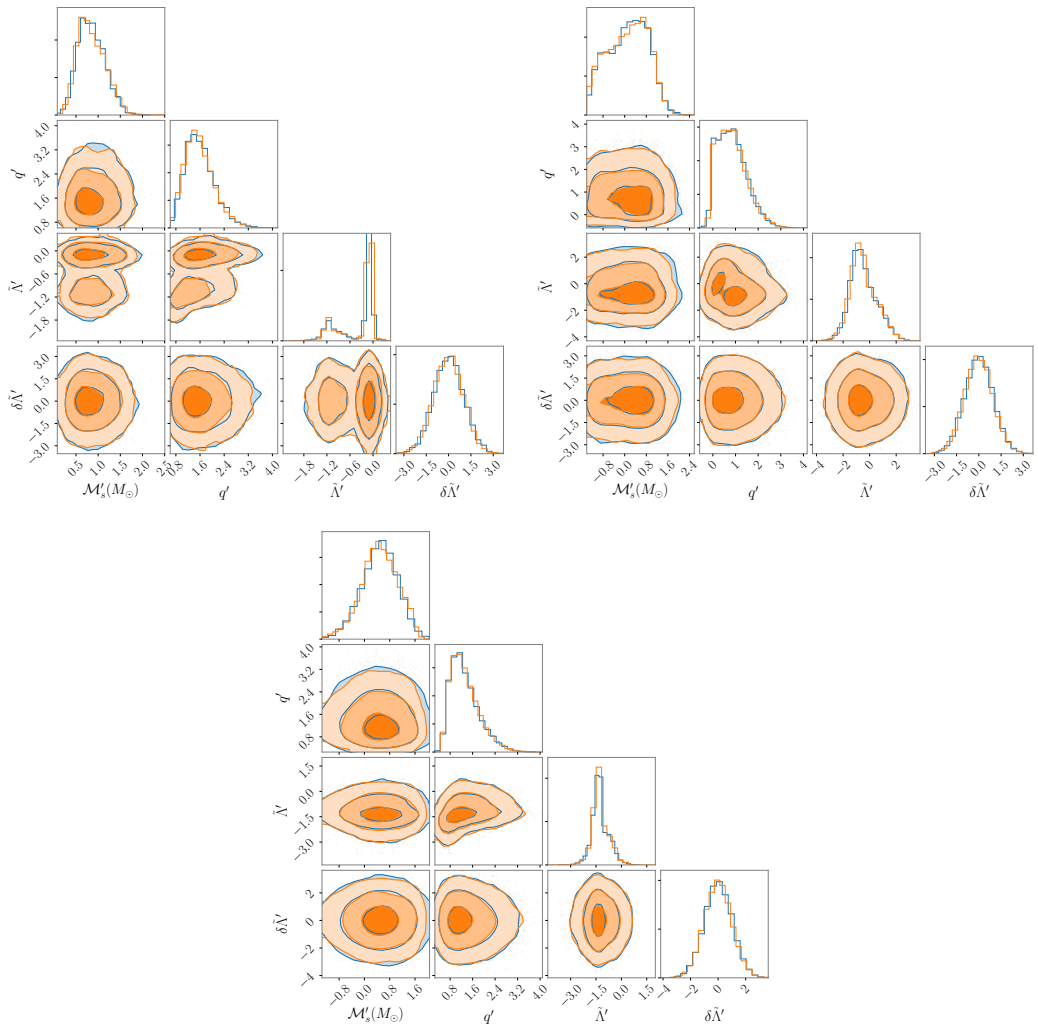

Figure 16.10: Posterior distributions of selected simulated events. Transformed samples are colored blue, and samples from the Gaussian mixture model density estimates are in orange. The overlap and consistency indicates that GMMs provide a good it in transformed space. Contours correspond to standard deviations in 2D space, such that 1-σ, 2-σ, 3-σ contours are 39%, 86%, and 99% confidence levels, respectively.

We include the two notational forms to highlight the asymptotic form (center) and the practical method for evaluating the quantity (right). From the asymptotic form, we note that if the moments of f can be analytically computed, we would have an expected variance that is exactly inversely proportional to the number of samples being averaged over.

Neglecting the uncertainty in the estimate of the selection function, the uncertainty in our total likelihood is the logarithm of the product of many Monte Carlo integrals,

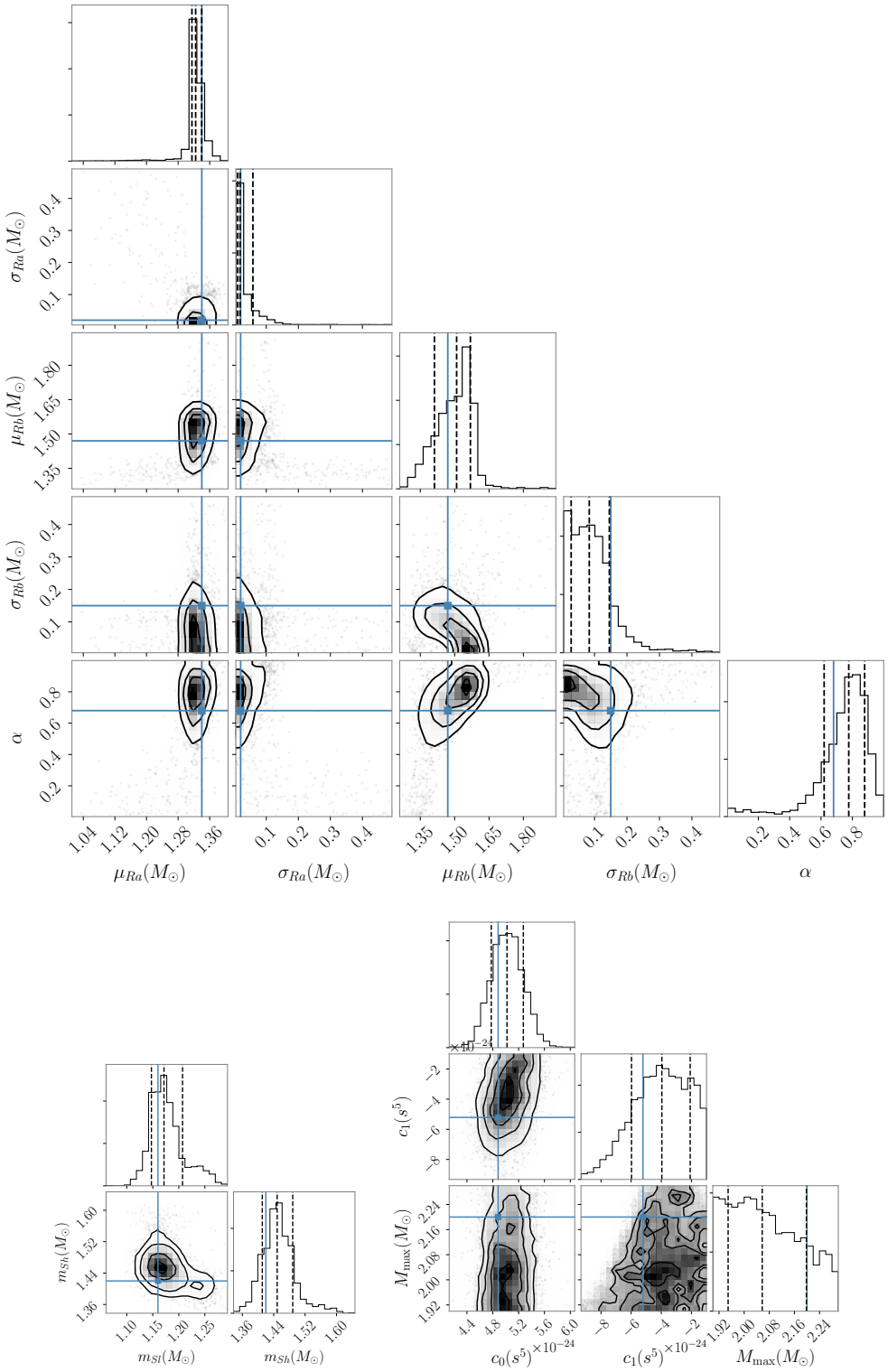

Figure 16.11: Inferred hyperparameter distributions for the recycled mass distribution hyperparameters (top), slow mass distribution hyperparameters (bottom left), and EOS hyperparameters (bottom right). Contours correspond to standard deviations in 2D space, such that 1-σ, 2-σ, 3-σ contours are 39%, 86%, and 99% confidence levels, respectively. See Table 16.1 for priors on each parameter.

the standard rules of propagating uncertainties yield

$$\Delta \ln \mathcal{L}(\{d_i\}|\Omega) = \sqrt{\sum_i^N \left(\frac{\Delta \mathcal{L}_i}{\langle \mathcal{L}_i \rangle}\right)^2}. \tag{16.19}$$

Here, the quantity in the sum is the equivalent of Equation 16.18. We emphasize that we are interested in the absolute uncertainty in the log-likelihood.This uncertainty will increase with the number of events for fixed per event variance. In order to maintain a constant uncertainty, the required number of samples per hyperparameter is proportional to the number of events. Thus the total number of samples required for constant uncertainty scales like N^2.

Appendix 16.C.1 Population sample weighting

In this work, we draw samples from the population model and evaluate Equation 16.9, i.e., $f(x_k) \rightarrow \mathcal{L}_i(\theta')$. In addition to estimating the statistical uncertainty, we note that we can directly sample the distribution of $\ln \mathcal{L}$ by repeatedly evaluating the likelihood with different realizations of samples from the population model.

In Figure 16.12 we plot the average log likelihood and 1σ uncertainty for 100 trials as a function of increasing number of population samples M (blue). This provides a test of convergence of the Monte Carlo integration, as a converged integral should be invariant under changes to the number of samples M. We note that using $M = 15,000$ samples, with an associated $\Delta \ln \mathcal{L} = 0.23$, is sufficient for convergence of this likelihood integration. In order to confirm that this integral is well-behaved for our choice of M, we perform a check by computing the likelihood again for each hyperparameter sample in our posterior distribution and reweight our original posterior distribution by this new distribution. We do this ten times to get ten new mock realizations of our posterior distribution. Differences between each simulated realization should therefore be due to statistical uncertainty when computing the Monte Carlo integral over random samples from the population distribution. We confirm qualitatively that these additional realizations are nearly identical to the original posterior distribution, indicating the Monte Carlo integral for the likelihood is stable for this choice of M.

[584] further reduces the uncertainty in their implementation of the population sample reweighting method by sampling only from regions in the population model which have non-vanishing support for mass parameters in the single-event likeli-

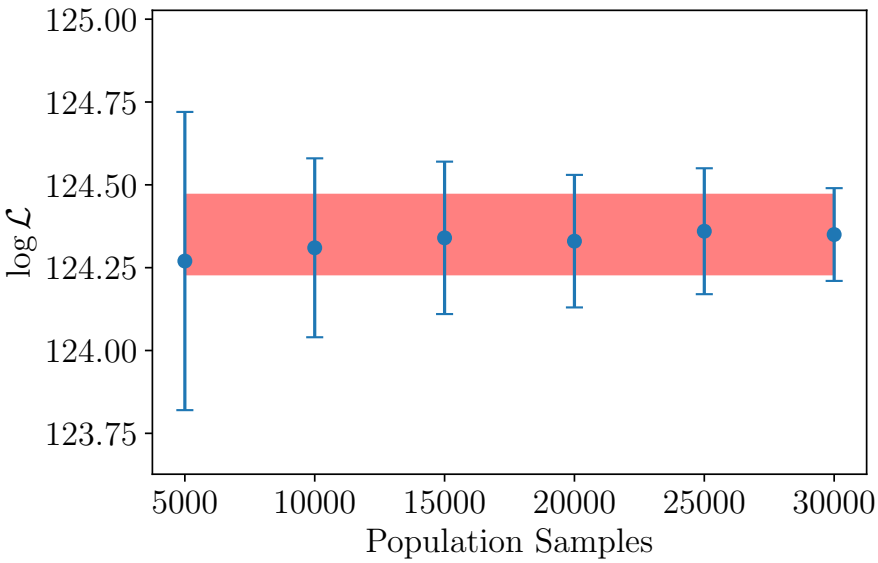

Figure 16.12: Log likelihood (Equation 16.6) for the true population hyperparameter values as a function of the number of samples M from the population. Each data point is an average over 100 likelihood iterations and error bars are $\pm 1\sigma$ (68%) uncertainties. By $M = 15,000$, the Monte Carlo integration is stable and increased values of M only increase the precision marginally. The red shaded region corresponds to the 68% net uncertainty resulting from using Monte Carlo integration via reweighting the single-event posterior samples in the mass population model (see Equation 16.18)

.

hoods. This increases the number of effective samples per Monte Carlo computation, presumably resulting in a reduced $\Delta \ln \mathcal{L}$ computed via Equation 16.19.

Appendix 16.C.2 Posterior sampling reweighting

As a comparison, we compute the uncertainty in the calculated likelihood when reweighting single-event posterior samples in the likelihood. In this case the quantity inside the sum is the ratio of the population model to the fiducial prior distribution $f(x_k) \rightarrow \pi(\theta|\Lambda)/\pi(\theta|)$. For this method, we are restricted to a single set of samples from the fiducial posterior distribution and so we must rely on the statistical uncertainty. We are also limited to only using posterior samples from the masses (no samples from the tidal parameters), as we cannot reweight posterior samples in a population model which accounts for a $\Lambda(m)$ relationship. In Figure 16.12, we show the uncertainty in an equivalent calculation in the red shaded region. We center the region at the mean estimator using the population model sampling method with

30,000 population samples for ease of comparison. To calculate this uncertainty we use 4,480 samples per event. We reiterate here that, while this reweighting method may give less uncertainty in $\ln \mathcal{L}$ in this application (i.e., the single-event posteriors are much narrower than the population model), it cannot account for tidal effects in the population model, as the distribution for Λ is a delta function for a given mass (i.e., the single-event posteriors are much wider than the population model).

Chapter 17

THE INTERPLAY OF ASTROPHYSICS AND NUCLEAR PHYSICS IN DETERMINING THE PROPERTIES OF NEUTRON STARS

Jacob Golomb, Isaac Legred, Katerina Chatziioannou, et al. "Interplay of astrophysics and nuclear physics in determining the properties of neutron stars". In: *Phys. Rev. D* 111.2 (2025), p. 023029. DOI: 10.1103/PhysRevD.111.023029. arXiv: 2410.14597 [astro-ph.HE].

Abstract

Neutron star properties depend on both nuclear physics and astrophysical processes, and thus observations of neutron stars offer constraints on both large-scale astrophysics and the behavior of cold, dense matter. In this study, we use astronomical data to jointly infer the universal equation of state of dense matter along with two distinct astrophysical populations: Galactic neutron stars observed electromagnetically and merging neutron stars in binaries observed with gravitational waves. We place constraints on neutron star properties and quantify the extent to which they are attributable to macrophysics or microphysics. We confirm previous results indicating that the Galactic and merging neutron stars have distinct mass distributions. The inferred maximum mass of both Galactic neutron stars, $M_{\mathrm{pop,\,EM}} = 2.05^{+0.11}_{-0.06}\,M_\odot$ (median and 90% symmetric credible interval), and merging neutron star binaries, $M_{\mathrm{pop,\,GW}} = 1.85^{+0.39}_{-0.16}\,M_\odot$, are consistent with the maximum mass of nonrotating neutron stars set by nuclear physics, $M_{\mathrm{TOV}} = 2.28^{+0.41}_{-0.21}\,M_\odot$. The radius of a $1.4\,M_\odot$ neutron star is $12.2^{+0.8}_{-0.9}$ km, consistent with, though $\sim 20\%$ tighter than, previous results using an identical equation of state model. Even though observed Galactic and merging neutron stars originate from populations with distinct properties, there is currently no evidence that astrophysical processes cannot produce neutron stars up to the maximum value imposed by nuclear physics.

17.1 Introduction

The properties of neutron stars (NSs) depend on both the dense-matter physics that governs their interiors and the astrophysical context in which they form, evolve, and are observed [330, 410, 138, 140]. This interplay is driven by an apparent coincidence: the mass scale of maximally-compact matter in its ground state is comparable to the Chandrasekhar mass. The NS characteristic compactness (defined as M/R with M the mass and R its radius) is just below the black-hole (BH) limit of 1/2[1]. This implies $2M/R \sim c_s^2 \sim 1$ [479], where c_s^2 is the characteristic speed of sound squared in the body. In the standard model, cold matter can only achieve such sound-speeds at densities greater than an atomic nucleus, $\rho_{\rm nuc} \sim 2.8 \times 10^{14} {\rm g/cm}^3$ at high neutron-to-proton ratio. This requirement fixes both the compactness *and* density of such a near-maximally compact object, and therefore its mass and radius scales to $M \sim 1\,M_\odot$ and $R \sim 10\,{\rm km}$ respectively. The former is remarkably close to the Chandrasekhar mass, $\sim 1.4\,M_\odot$, the maximum mass that can be supported by electron degeneracy. As NSs form from cores that are too massive to be supported by electron degeneracy, this sets another characteristic mass scale for NSs [156].

Substantial uncertainties in the details of NS formation and dense-matter physics mean it is not immediately clear which of the two drives the distribution of NS masses. For example, general relativity and the dense-matter equation of state (EoS) set a maximum mass for nonrotating NSs, the Tolman-Oppenheimer-Volkoff (TOV) limit $M_{\rm TOV}$ [541, 407]. Originally speculated to be near $0.7\,M_\odot$, $M_{\rm TOV}$ is now understood to be $\sim 2{-}3\,M_\odot$ [460, 300, 329, 413, 439, 332, 378], but it is unknown whether astrophysical formation mechanisms can produce NSs up to this mass. Moreover, NSs form in a variety of ways, including in core-collapse supernovae and binary mergers, each of which likely results in different natal mass and spin distributions. Even after formation, NSs are modified via binary interactions: for instance, "spider" pulsars [464] may achieve large masses and spins via accretion.

Galactic observations have constrained the masses of dozens of NSs in binaries via pulsar timing [104]. The mass distribution of Galactic NSs with a mass measurement includes a primary peak at $\sim 1.35\,M_\odot$ preferred at 3:1 over a secondary peak at $\sim 1.8\,M_\odot$ [62, 58, 213]. The observed cutoff in the distribution above $\sim 2\,M_\odot$ [58, 213] may correspond to the TOV mass, or to a different maximum mass imposed by astrophysical processes; the most general interpretation of the cutoff identifies it as an astrophysical maximum mass that may differ from $M_{\rm TOV}$. The Galactic

[1] In units where $G = c = 1$, which we use unless otherwise stated.

NS population is broadly consistent with the masses of NS-like compact objects in wide-period binaries revealed by Gaia astrometry [74, 72]. However, this inferred mass distribution does not account for selection effects in the various surveys, and lumps together NSs in different astrophysical systems, e.g., NS–NS binaries (BNS) and NS–WD binaries, that may have different inherent distributions. Indeed, the known Galactic BNSs are all contained within the $\sim 1.35\,M_\odot$ component of the bimodal distribution [219].

A subset of Galactic millisecond pulsars [356] show persistent pulsed X-ray emission originating from surface hotspots. Detailed modeling of the hotspot emission has placed constraints on the mass and radius of three pulsars using NICER and XMM-Newton [377, 461, 378, 462, 148], two of which are in binaries and thus have radio-based mass constraints. Since two of the NICER targets are known radio pulsars, they are commonly treated as part of the Galactic NS population. For example, the properties of PSR J0740+6620, one of the most massive known pulsars [161, 235], have been inferred simultaneously with the Galactic population [214]. Requiring PSR J0740+6620 to hail from the bimodal Galactic NS mass distribution revises its mass downward to $2.03^{+0.14}_{-0.11}\,M_\odot$ [214].

A different population consists of NSs in merging compact binaries with NSs or black holes (BHs) observed with gravitational waves (GWs) [28]. Among BNSs, GW170817 [10] is consistent with the Galactic BNS population with a total mass of $\sim 2.7\,M_\odot$. GW190425, at a total mass of $\sim 3.4\,M_\odot$ [11], is however an outlier. Attempts to explain this discrepancy include selection effects [471, 481] and non-BNS interpretations [264, 234]. Regardless, this discrepancy suggests that the Galactic and merging BNS distributions should be treated separately. The distribution of all NSs observed in merging binaries to date, including both BNSs and likely NSBH systems [27, 2], is relatively flat with no prominent peak at $\sim 1.35\,M_\odot$ [327, 28]. The population of NSs in BNSs and NSBHs might, however, be different owing to different formation and evolutionary histories [28, 100]. NS spins are ignored from these constraints due to large measurement uncertainties [101]; it is therefore unknown how merging NS spins relate to the well-measured spins of Galactic NSs. GW-based NS observations (primarily GW170817) also drive constraints on the EoS through mass and tidal deformability constraints [10, 9, 15].

The picture is much simpler when it comes to the nuclear physics and the EoS of NSs. Even when originating from different formation mechanisms, cold NSs are expected to be described by the same universal EoS. This expectation has been

widely utilized to combine mass, radius, and tidal deformability measurements from various observations to place constraints on the EoS, e.g. [9, 380, 438, 177, 326, 261, 378, 439, 413, 332, 205, 102]. Even then, assumptions about NS masses have to be made.

Such assumptions typically include a uniform mass distribution, and whether astrophysical mechanisms create NSs up to the TOV mass or up to a different predetermined value [326, 332].

In this paper, we study the properties of NSs in binaries with a focus on separating the impact of nuclear physics and astrophysics. We use radio, X-ray, and GW data to jointly infer the dense-matter EoS and the NS mass distribution, each with their own maximum mass. We go beyond considering a single mass distribution for all NSs that terminates at the TOV mass [102, 205] and separately infer the populations of Galactic NSs and merging BNSs. Moreover, rather than the TOV mass, we allow the possibility of the astrophysical mass distribution terminating at a different "astrophysical maximum mass" that is lower than the TOV mass. Our model and inference set up allow us to begin to answer whether the maximum mass of NSs in various subpopulations is limited by the EoS or by astrophysical processes. Beyond access to such questions, simultaneous inference mitigates biases that can arise with as few as $O(10)$ GW detections when inferring either the EoS or the mass distribution alone while making improper assumptions about the other [584, 256]. We also account for GW selection effects, which cause the detected population to be biased towards higher masses; as the selection effects in the electromagnetic surveys are unknown, we do not consider them.

The subpopulations, datasets, and models are described in Sec. 17.2. The EoS is modeled with a mixture of Gaussian processes (GPs) [325, 198], which allows for a wide range of EoS morphologies including phase transitions [199] and imposes minimal intra-density correlations that hamper the flexibility of parametric models [333]. We consider two subpopulations:

1. The Galactic NS population is modeled with a bimodal distribution with a maximum mass cutoff [58, 62]. The relevant datasets include radio, optical, and X-ray observations of pulsars in binaries [58] and X-ray pulse-profile observations of pulsars J0030+0451 [377, 461], J0740+6620 [378, 462], and J0437-4715 [148].

2. The merging BNS population observed with GWs is modeled with a power-law with a maximum mass cutoff. The dataset consists of GW170817 [10] and GW190425 [11], both assumed to be BNSs.

Joint inference on the EoS and mass subpopulations is performed with a reweighting scheme that is described in Sec. 17.3 and Appendix 17.A.

Our results are presented in Sec. 17.4. We find no evidence that the maximum mass of the two subpopulations is different than the TOV mass. The difference between the maximum Galactic NS (merging BNS) mass and the TOV mass is less than $0.53\,M_\odot$ ($0.73\,M_\odot$) at 90% credibility, with zero difference consistent with the posteriors. Even though the maximum masses are consistent, we confirm previous results that the mass distributions of Galactic NSs and merging BNSs are different. The latter possesses no prominent peak at $1.35\,M_\odot$, indicating that the two distributions should be modeled separately in an inference framework and have the freedom to differ from one another.

For the NS EoS, we infer a sound-speed profile that exceeds the conformal bound of $1/\sqrt{3}$ for weakly interacting nucleonic matter [88], in line with a previous study using the same EoS model [332]: the 90% lower bound on the maximum speed of sound squared anywhere inside the NS is 0.59. We constrain the radius of a canonical NS, a proxy for the stiffness of the EoS, to $R_{1.4} = 12.2^{+0.8}_{-0.9}\,\mathrm{km}$, and the TOV mass to $M_{\mathrm{TOV}} = 2.28^{+0.41}_{-0.21}\,M_\odot$. Uncertainties are lower than Legred et al. [332] due to the recent NICER observation of PSR J0437-4715 and the impact of the ensemble of Galactic NS mass measurements via the updated treatment of the maximum mass.

We conclude in Sec. 17.5.

17.2 Modeling the Equation of State and the mass distribution

In this section we describe the data, as well as the EoS and astrophysical populations we model them with.

17.2.1 Data

The observations that inform the joint inference of the NS EoS and astrophysical population come from three sources: radio/optical pulsar mass measurements

(PSR), X-ray pulse profile modeling for pulsar masses and radii (NICER), and GW constraints on BNS masses and tidal deformabilities (GW).

The PSR dataset includes the 74 Galactic pulsars with a mass measurement from Ref. [58], minus PSR J0437–4715, which is counted as part of the NICER dataset.[2] The PSR observations are heterogeneous, including NSs in various types of binaries and several different mass determination methods.

The NICER dataset consists of the observations of PSR J0030+0451 [380, 461], PSR J0740+6620 [378, 462], and PSR J0437–4715 [148]. The constraints on the masses and radii of these pulsars are sensitive to the details of the X-ray pulse profile modeling, such as the assumed hotspot geometry and the stochastic sampling of the multidimensional parameter posterior; thus different interpretations of the NICER data exist. Here we use results from the three-hotspot model of Ref. [380] for J0030+0451, the combined NICER-XMM Newton analysis with the two-hotspot model from Ref. [378] for J0740+6620, and the `CST+PDT` model from Ref. [148] for J0437–4715. As the NICER analyses for J0740+6620 and J0437–4715 incorporate pre-existing radio-based mass estimates, we exclude them from the PSR dataset to avoid double-counting. In Appendix 17.D we quantify the sensitivity of our inference to alternative data selection choices for the NICER observations.

For the GW dataset, we consider compact binary coalescences from the third Gravitational Wave Transient Catalog [26] of the LIGO-Virgo-KAGRA network [1, 38, 54] with source-frame chirp mass $\mathcal{M} \lesssim 2.176\,M_\odot$, corresponding to equal-mass component masses below $2.5\,M_\odot$. This leaves us with GW170817 [10] and GW190425 [11] as the only events consistent with BNS mergers. We do not consider the recent observation of GW230529_181500 [2], which is potentially a BNS merger according to this criterion, as sensitivity estimates for the fourth observing run do not exist. For GW170817, we generate new posterior samples with the waveform approximant `IMRPhenomPv2_NRTidal`, which includes spin-precession and tidal effects [175], using the parameter estimation package `bilby` [67, 470] and the nested sampler `dynesty` [507]. We fix the source location to the host galaxy NGC4993 and adopt spin priors that are isotropic in orientation and uniform in dimensionless magnitude up to 0.05, motivated by the spin distribution of pulsars in binary systems expected to merge within a Hubble time [599]. For GW190425, we use the publicly released parameter estimation samples [336] for the `IMRPhenomPv2_NRTidal` waveform.

[2] While J0437–4715 is in the NICER dataset, we use its radio mass measurement to inform the mass distribution.

Since GW190425's total mass is inconsistent with those of Galactic BNSs, we allow for dimensionless spin magnitudes up to 0.4, roughly corresponding to a 1 ms spin period [276]. Appendix 17.F investigates the impact of a spin-magnitude upper limit of 0.05 for both GW170817 and GW190425.

17.2.2 EoS model

The dense-matter EoS, i.e., the pressure-density relation, is described with a model-agnostic Gaussian process [325, 198], which builds a prior EoS process via a mixture of GP hyperparameters probing a large range of correlation scales and strengths. This procedure produces an EoS distribution that is relatively insensitive to the nuclear models it is conditioned on [198] and imposes minimal model-dependent correlations between the low- and high-density EoS [333]. The GP flexibility is particularly important for our goal of disentangling the maximum TOV mass $M_{\rm TOV}$ and the maximum astrophysical mass. Less flexible parametric EoS models implicitly correlate the radius or tidal deformability and $M_{\rm TOV}$ [333] which in turn translate to model-dependent correlations between $M_{\rm TOV}$ and the astrophysical parameters. All NSs are assumed to be described by the same EoS. For efficiency, we restrict the prior to EoSs with $M_{\rm TOV} > 1.8\,M_\odot$.

17.2.3 Astrophysical population models

For the astrophysical mass distribution we use parametric distributions with hyperparameters η. We consider two classes of observations modeled with separate distributions: Galactic NSs observed via electromagnetic (EM) radiation as part of the PSR and NICER datasets, and NSs in merging BNSs observed via GWs constituting the GW dataset.

We restrict to the NS masses while ignoring spins and assume that all objects are NSs.

17.2.3.1 Galactic neutron stars with radio and X-rays

Motivated by Refs. [62, 58, 214], we model the Galactic NS masses m as a mixture of two Gaussians:

$$\pi(m|\eta_{\rm EM}) = f\mathcal{N}(\mu_1, \sigma_1) + (1-f)\mathcal{N}(\mu_2, \sigma_2)\,, \tag{17.1}$$

for $m \in [M_{\rm min}, M_{\rm pop,\,EM}]$, and where $\mathcal{N}(\mu, \sigma)$ is a truncated normal distribution with mean μ and standard deviation σ, and f is the mixture weight. Following Ref. [58], we fix $M_{\rm min} = 1\,M_\odot$ and infer the hyperparameters $\eta_{\rm EM} = \{\mu_1, \mu_2, \sigma_1, \sigma_2, f, M_{\rm pop,\,EM}\}$ with flat priors: $M_{\rm pop,\,EM} \in (1.8, 3.0)\,M_\odot$, $\mu_1 \in (1, 2)\,M_\odot$, $\mu_2 \in (\mu_1, 2.5)\,M_\odot$, $f \in (0, 1)$, and $\sigma_{1,2} \in (0.05, 1)\,M_\odot$. Since all analyzed objects are NSs, we impose $M_{\rm pop,EM} < M_{\rm TOV}$.[3] This prior restriction leads to a marginal priors on $M_{\rm pop,\,EM}$ and the EoSs that are not uniform, although the full multidimensional prior is flat within its domain of support.

Although the PSR and NICER datasets include NSs in different astrophysical settings, i.e. in binaries with various companions, or in isolation in the case of J0030+0451, and could in principle hail from different subpopulations, we model these NSs as a single population for consistency with previous results and due to the lack of selection effect estimates. (We are not aware of any established methods to account for selection effects in radio surveys or for NICER's target selection procedure [106].) Given the lack of selection effect estimates for the PSR and NICER datasets, we simply assume the observed mass distribution to be equivalent to the astrophysical distribution.[4] We quantify the impact of this assumption in Appendix 17.E, where we present results with a fixed uniform mass distribution in place of Eq. (17.1).

17.2.3.2 Merging neutron stars with gravitational waves

We model BNS masses with a truncated power-law for both binary components m_1 and m_2:

$$\pi(m_1, m_2|\eta_{GW}) \propto m_1^\alpha m_2^\alpha\,, \tag{17.2}$$

for $m \in [m_{\rm min}, M_{\rm pop,\,GW}]$ and random pairing between m_1 and m_2 in the two-dimensional space. We again fix $m_{\rm min} = 1\,M_\odot$ and infer the hyperparameters $\eta_{GW} = \{\alpha, M_{\rm pop,\,GW}\}$ with flat priors $\alpha \in (-5, 5)$, $M_{\rm pop,\,GW} \in (1.6, 2.5)\,M_\odot$. Since we assume that both GW170817 and GW190425 are BNSs, we again impose $M_{\rm pop,GW} < M_{\rm TOV}$.

[3] We ignore the impact of pulsar spin on the maximum mass. Using approximate relations to fourth order in spin magnitude [386, 112], we estimate that the maximum allowed mass will differ from $M_{\rm TOV}$ by $\lesssim 1\%$ compared to statistical uncertainties $\sim 20 - 30\%$ for the range of pulsar periods in our dataset, $P \gtrsim 2$ ms.

[4] This procedure can result in a bias even for the detected population [195]. Such a bias however is expected to be small. For example, Figure 4 of [195] shows the bias for $\sim$800 simulated GW observations.

GW selection effects are well understood, and we incorporate them in our inference. Because the GW data selection procedure involves identifying events as BNSs based on a component mass cut at 2.5 $M_\odot$, our analysis only places constraints on the mass distribution below 2.5 $M_\odot$. The GW selection modeling is described in Sec. 17.3.1.

17.3 Joint inference via reweighting

The joint mass-EoS model is a combination of EoS draws from the GP prior process and the parametric mass models of Eqs. (17.1) and (17.2). While the joint posterior could be sampled with standard stochastic sampling methods with pre-computed GP draws, we instead use a multi-stage reweighting scheme and the GP draws from Ref. [332].

The reweighting scheme includes the following steps, with technical details relegated to the Appendices:

1. Use standard stochastic sampling to infer the mass population and the EoS using Eqs. (17.1) and (17.2) for the mass distribution and a simplified, low-dimensional EoS model. Details about the EoS model are given in Appendix 17.B. The EoS model is included here to mitigate potential biases of a mass-only inference [256].

2. Treat the inferred mass distribution as a proposal distribution. For each sample from the distribution of $\eta = \{\eta_{\rm EM}, \eta_{GW}\}$, calculate the likelihood for each pre-computed GP draw. The likelihood form depends on the dataset considered [327] and is described in Secs. 17.3.1 and 17.3.2 for the GW and EM data respectively.

3. With these likelihoods, calculate weights from the proposal mass distribution to the target joint mass-GP EoS distribution as described in Appendix 17.A.

4. Combine the new posterior distributions for each dataset. This procedure allows us to obtain weighted samples from the joint posterior of the mass distribution and the GP EoS. We validate the reweighting scheme in Appendix 17.C with simulated GW observations.

Each of the datasets considered (GW, NICER, and PSR) results in unique constraints and thus requires a unique formulation of the likelihood [327, 138]. Below we

discuss each dataset likelihood noting that the full likelihood is the product over the individual datasets.

17.3.1 GW likelihood

Given $N_{\rm GW}$ independent events, the likelihood for the EoS ε and population hyper-parameters η_{GW} is[5] [560, 536, 359]

$$\begin{aligned}\mathcal{L}_{GW}(d|\varepsilon,\eta_{GW}) \propto\ & p_{\rm\ det}(\eta_{\rm\ GW})^{-N_{GW}}\times \\ & \prod_i^{N_{GW}} \int \mathcal{L}(d_i|m_1,m_2,\varepsilon)\pi(m_1,m_2|\eta_{GW})dm_1dm_2\,,\end{aligned} \tag{17.3}$$

where $\pi(m_1,m_2|\eta_{GW})$ is the model of Eq. (17.2) and

$$\mathcal{L}(d_i|m_1,m_2,\varepsilon) = \mathcal{L}(d_i|m_1,m_2,\Lambda(m_1,\varepsilon)\Lambda(m_2,\varepsilon))\,, \tag{17.4}$$

is the ith individual-event GW likelihood (e.g., [221, 552]) marginalized over all binary parameters other than the component masses m_1, m_2 and tidal deformabilities Λ_1, Λ_2. Consistency with the EoS is ensured by calculating the likelihood for $\Lambda_1 = \Lambda(m_1,\varepsilon), \Lambda_2 = \Lambda(m_2,\varepsilon)$, i.e., the EoS prediction for the tidal deformability given the mass. We estimate the individual-event likelihood from the posterior samples for the source-frame masses and tidal deformabilities using a Gaussian mixture model [256], and the integral in Eq. (17.3) is computed as a Monte Carlo sum.

The term $p_{\rm det}(\eta_{\rm GW})$ encodes the selection effect which characterizes how parts of the parameter space are over-represented in a catalog of GW events, as determined by the sensitivity of the detectors. Defining $p_{\rm det}(d)$ as the probability that search algorithms detect a significant signal in data d results in

$$\begin{aligned}p_{\rm det}(\eta_{GW}) &\equiv \int \mathcal{D}d \int d\theta\, p(d|\theta)\pi(\theta|\eta_{GW})p_{\rm det}(d) \\ &= \int d\theta\, \pi(\theta|\eta_{GW})p_{\rm det}(\theta)\,,\end{aligned} \tag{17.5}$$

where we identify $p_{\rm det}(\theta) \equiv \int \mathcal{D}d\, p(d|\theta)p_{\rm det}(d)$ as the probability of detecting an event with parameters θ, marginalized over possible realizations of data d. For example, neglecting the specifics of the noise-generating process, the sensitivity to an event increases with its chirp mass $\sim\mathcal{M}_c^{5/6}$ and decreases inversely with its

[5] This expression assumes a $1/R$ prior on the event rate R and marginalizes over it [359, 228].

distance. We then further marginalize over possible realizations from the population $\theta \sim \pi(\theta|\eta_{GW})$. The presence of $p_{\rm det}(\eta_{\rm GW})$ in Eq. (17.3) ensures that the final result reflects the true astrophysical population rather than the observed population. In practice, $p_{\rm det}(\eta_{GW})$ might also depend on the EoS, but Ref. [162] showed that the effect is negligible except for very stiff EoSs and low-mass NSs: there is a $\lesssim 2\%$ change in the match between a template that sets $\Lambda = 0$ and the true waveform.

We compute $p_{\rm det}(\eta_{GW})$ by reweighting recovered simulated signals in data from the first three observing runs, using standard techniques [28, 213, 538].

17.3.2 NICER likelihood

Given $N_{\rm NICER}$ observations, the likelihood for the EoS ε and population hyperparameters $\eta_{\rm EM}$ is obtained by marginalizing over the pulsar mass

$$p_{\rm NICER}(d|\varepsilon,\eta_{\rm EM}) = \prod_i^{N_{\rm NICER}} \int \mathcal{L}(d_i|m,\varepsilon)\pi(m|\eta_{\rm EM})dm\,, \tag{17.6}$$

where i indexes the NICER observations, $\pi(m|\eta_{\rm EM})$ is the mass distribution of Eq. (17.1), and

$$\mathcal{L}(d_i|m,\varepsilon) = \mathcal{L}(d_i|m,C(m,\varepsilon))\,, \tag{17.7}$$

is the individual-pulsar likelihood marginalized over all NICER parameters other than the mass m and compactness C, which is again evaluated on the EoS prediction. The likelihoods are described in the publications associated with each observation [148, 377, 378]. We use a Gaussian mixture model [256] to evaluate Eq. (17.7), and a Monte Carlo sum for the integral in Eq. (17.6).

The NICER analysis of PSR J0437-4715 in Ref. [148] uses a prior that is flat in radius, rather than flat in compactness (or inverse compactness) like the analyses of PSR J0030-0451- [377] and PSR J0770+6620 [378]. We correct for this with the appropriate Jacobian term to obtain a likelihood function in mass and compactness. Unlike Eq. (17.3) for the GW observations, the NICER likelihood ignores selection effects per the discussion in Sec. 17.2.1.

17.3.3 PSR likelihood

Finally, the likelihood for $N_{\rm PSR}$ pulsar mass measurements is

$$p_{\rm PSR}(d|\varepsilon,\eta_{\rm EM}) = \prod_i^{N_{\rm PSR}} \int \mathcal{L}(d_i|m)\pi(m|\eta_{\rm EM})dm\,, \tag{17.8}$$

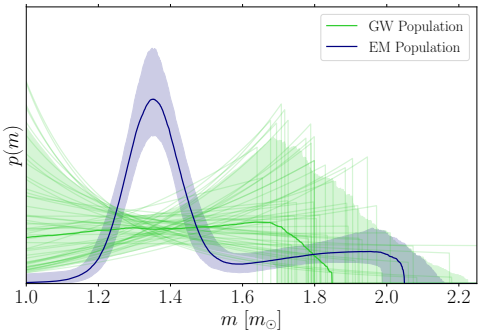

Figure 17.1: Posterior on the mass distribution of the GW BNS (orange) and the Galactic NS (blue) population. We plot the median and 90% highest-probability credible regions. The EM population is constrained to much better precision than the GW one due to the low number of GW BNS detections. With the caveat that they correspond to the astrophysical BNS and observed Galactic NS distributions respectively, we find that the two distribution are inconsistent, in agreement with Ref. [28]. Faint lines are random draws from the GW mass distribution, illustrating the bimodal uncertainties in the mass distribution.

where i indexes the pulsars and $\pi(m|\eta_{\rm EM})$ is the mass distribution of Eq. (17.1). The form of $\mathcal{L}(d_i|m)$ for each observation is prescribed analytically in Refs. [58, 214], depending on whether the measurement constrains the pulsar mass, the binary mass function and the total mass, or the binary mass function and the mass ratio. Like the NICER likelihoods, the PSR likelihoods do not account for selection effects, and we evaluate the integral in Eq. (17.8) via Monte Carlo.

17.4 Implications of joint mass-EoS inference

In this section, we present results from the joint inference over the EoS and the mass distribution of two NS populations. We begin with mass-specific and EoS-specific results in Secs. 17.4.1 and 17.4.2 respectively, before contrasting their impact on NS properties in Sec. 17.4.3.

17.4.1 Constraints on astrophysical populations

Figure 17.1 shows the inferred mass distribution of merging BNSs observed with GWs (modeled with a truncated power-law) and the observed distribution of Galactic NSs observed with EM (modeled with a truncated Gaussian mixture). The BNS population is consistent with being flat and has large uncertainties due to the now number of events (a total of 4 NSs). The smallest uncertainty is at $\sim 1.4\,M_\odot$, corresponding to the relatively well-measured masses on GW170817, while there

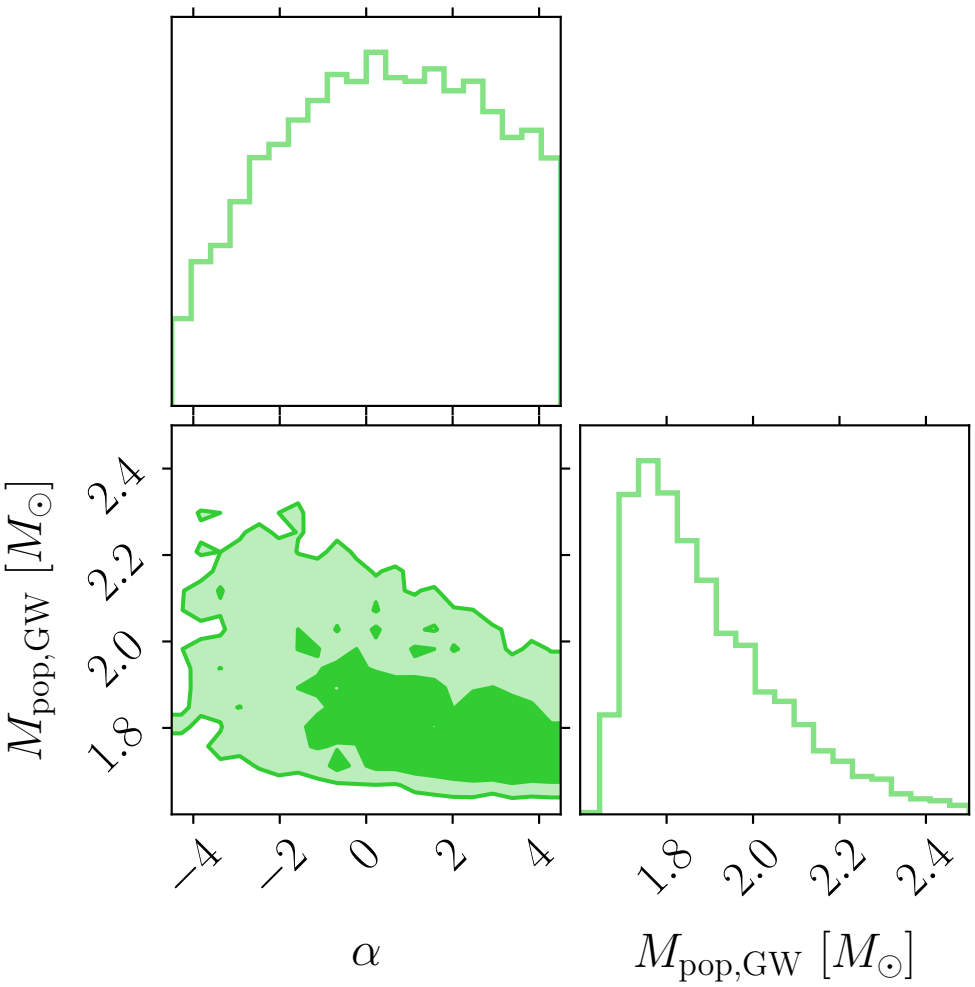

Figure 17.2: Marginalized posterior for the power-law slope α and maximum mass $M_{\rm pop,GW}$ of the GW population. The slope α is poorly constrained and thus its posterior rails against the upper prior bound, in turn affecting the $M_{\rm pop,GW}$ posterior.

is vanishing support for masses above $\sim 2.2\,M_\odot$ with $M_{\rm pop,\,GW} = 1.85^{+0.39}_{-0.16}\,M_\odot$. This shape is broadly consistent with the results of Refs. [28, 327] that additionally considered the two NSs in the NSBH binaries GW200105 and GW200115 and did not model the EoS. The seemingly "bimodal" shape with peaks at high and low masses at the 90% level is model-dependent: it is an outcome of the fact that the distribution is well-measured at $\sim 1.4\,M_\odot$ and we model it with a truncated power-law. Figure 17.2 indeed shows that the power-law index α and the maximum mass, $M_{\rm pop,\,GW}$, are correlated and the upper limit on $M_{\rm pop,\,GW}$ depends on the α prior. In particular, while the one-dimensional posterior peaks at $\alpha \approx 0$, $\alpha \gtrsim 4$ cannot be ruled out but is only consistent with $M_{\rm pop,\,GW} \lesssim 2.0 M_\odot$.

The observed EM population is comparatively better constrained as it is based on a total of 74 pulsar mass measurements. We find consistent results with Refs. [58, 214] that used the same pulsar mass data but did not infer the EoS with $\mu_1 = 1.35^{+0.02}_{-0.02}\,M_\odot$ and $\mu_2 = 2.01^{+0.43}_{-0.27}\,M_\odot$, $f = 0.65^{+0.11}_{-0.13}$, and $\sigma_1 = 0.07^{+0.02}_{-0.02}\,M_\odot$ and $\sigma_2 = 0.39^{+0.37}_{-0.22}\,M_\odot$. The maximum mass is $M_{\rm pop,\,EM} = 2.05^{+0.11}_{-0.06}\,M_\odot$, compared to $2.12^{+0.12}_{-0.17}\,M_\odot$ in [58] and $2.25^{+0.82}_{-0.26}\,M_\odot$ in [214]. Our estimate is lower due to the fact that we simultaneously infer the EoS and impose $M_{\rm pop,\,EM} < M_{\rm TOV}$.

Assuming that the three NICER pulsars are part of the general Galactic NS population leads to updated mass inference. The original mass estimates quoted in Refs. [377, 378, 148] refer to flat mass priors, while our analysis effectively updates the prior to be the population distribution [214].[6] The mass for each NICER target under a population-informed (flat) prior is $1.37^{+0.22}_{-0.11}$ ($1.44^{+0.25}_{-0.23}$) $M_\odot$ for J0030+0451, $1.39^{+0.08}_{-0.05}$ ($1.42^{+0.06}_{-0.06}$) $M_\odot$ for J0437-4715, and $2.01^{+0.08}_{-0.09}$ ($2.07^{+0.11}_{-0.12}$) $M_\odot$ for J0740+6620. The J0740+6620 result is somewhat larger than the value in Farr and Chatziioannou [214], $2.03^{+0.17}_{-0.14}\,M_\odot$. The effect is most stark for J0030+0451 whose mass is poorly measured from the X-ray data alone, but now resides in the dominant peak of the mass distribution.

17.4.2 Constraints on EoS quantities

Figure 17.3 shows the prior and posterior for various macroscopic and microscopic EoS properties: the TOV mass, $M_{\rm TOV}$, the radius and tidal deformability of a canonical $1.4\,M_\odot$ NS, $R_{1.4}$ and $\Lambda_{1.4}$ respectively, the radius of a $1.8\,M_\odot$ NS, $\Lambda_{1.8}$, and the pressure at twice and 6 times nuclear saturation, $p_{2.0}$ and $p_{6.0}$ respectively. We infer $\Lambda_{1.4} = 438^{+224}_{-166}$ and $R_{1.4} = 12.2^{+0.8}_{-0.9}\,$km. For comparison, we also plot the corresponding analysis from Legred et al. [332] that fixes all mass distributions to uniform. To isolate the effect of the mass distribution inference, we repeat the analysis of Ref. [332] while adding the X-ray mass-radius measurement of J0437-4715 such that the two analyses use the same NICER and GW data. We obtain largely consistent results: mass-marginalization leads to mild changes in $R_{1.4}$ and $\Lambda_{1.4}$, while including spider pulsars in the analysis and introducing an EoS-limited astrophysical maximum mass leads to a mild increase in the inferred value of $M_{\rm TOV}$.

These results are consistent with previous estimates. Legred et al. [332] used the GP EoS model with the same GW dataset, the first two NICER objects, J0030+0451

[6] The same is true for the two GW events, but the effect is minimal as the mass distribution uncertainty is wide and consistent with flat which was the inference prior to begin with.

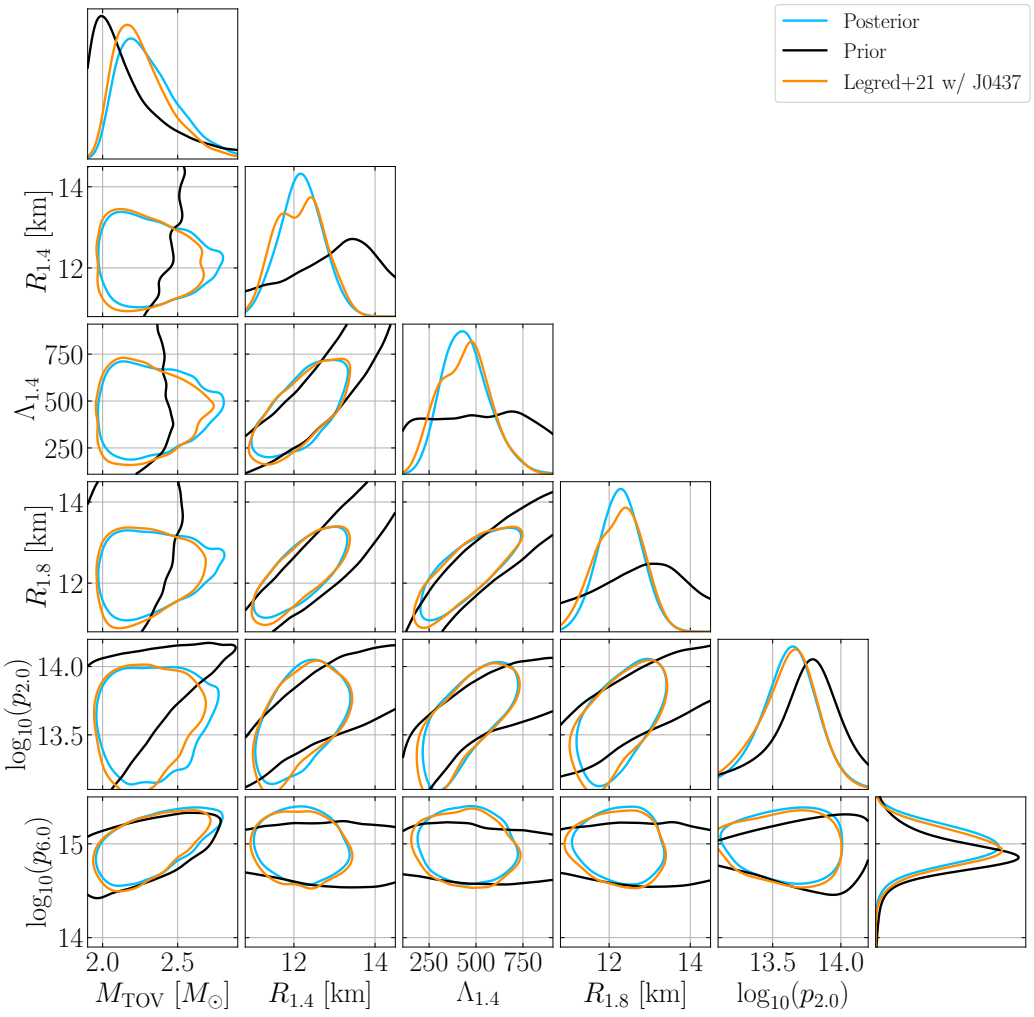

Figure 17.3: One- and two-dimensional posteriors for select EoS macroscopic and microscopic parameters: the TOV mass, $M_{\rm TOV}$, the radius and tidal deformability of a canonical 1.4 $M_\odot$ NS, $R_{1.4}$ and $\Lambda_{1.4}$ respectively, the radius of a 1.8 $M_\odot$ NS, $R_{1.8}$, and the log-base-10 pressure (divided by the speed of light squared) at twice and 6 times nuclear saturation, $p_{2.0}$ and $p_{6.0}$ respectively, when measured in g/cm^3. Two-dimensional contours denote the boundaries of the 90% credible regions. We show the prior (black), the posterior from the main analysis that marginalizes over the mass distribution (blue), and the analogous posterior that arises from additionally including the mass-radius measurement of J0437-4715 in the analysis of Ref. [332].

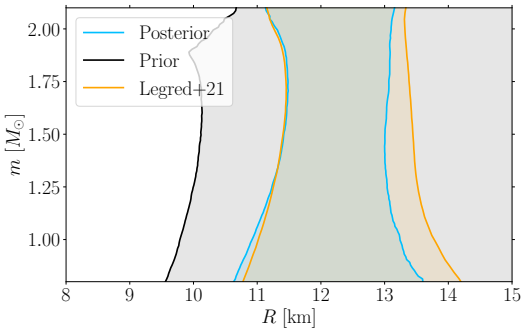

Figure 17.4: Mass-radius inference, we show the 90% symmetric credible region for the radius at each mass. We plot the prior (black), posterior from the main analysis that marginalizes over the mass distribution (blue), and posterior from Ref. [332] that fixes the mass distribution to flat and does not include J0437-4715. The upper limit on the radius decreases by ~ 0.5 km for all masses.

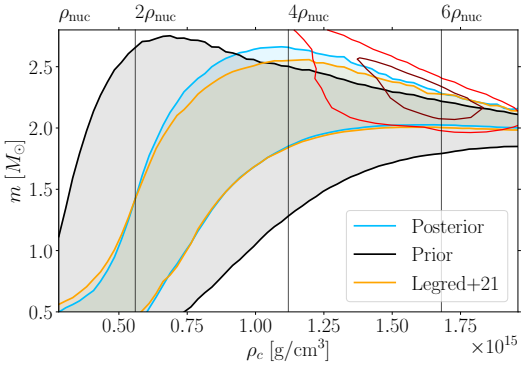

Figure 17.5: Mass-central density inference, we show the 90% symmetric credible region for the NS mass at each value of the central density ρ_c. We plot the prior (black), posterior from the main analysis that marginalizes over the mass distribution (blue), and posterior from Ref. [332] that fixes the mass distribution to flat and does not include J0437-4715. Vertical lines denote multiples of the nuclear saturation density. Maroon and red contours mark 1 and 2-σ credible regions, respectively, for the joint posterior on ρ_c-$M_{\rm TOV}$.

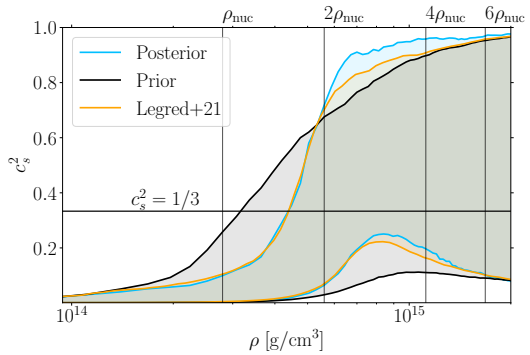

Figure 17.6: Speed of sound-density inference, we show the 90% symmetric credible region for the speed of sound squared, c_s^2 at each rest-mass density ρ. We plot the prior (black), posterior from the main analysis that marginalizes over the mass distribution (blue), and posterior from Ref. [332] that fixes the mass distribution to flat and does not include J0437-4715. Vertical lines denote multiples of the nuclear saturation density. The speed of sound increases by $\sim$ 5% around densities 2 – 3 times saturation density.

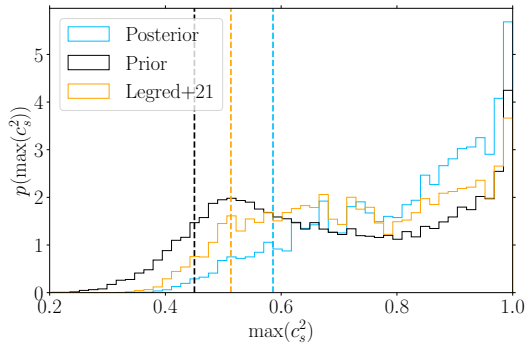

Figure 17.7: Marginalized posterior for the maximum speed of sound squared inside a stable NS. We plot the prior (black), posterior from the main analysis that marginalizes over the mass distribution (blue), and posterior from Ref. [332] that fixes the mass distribution to flat and does not include J0437-4715. The 90% lower limit on the maximum speed of sound, marked by dashed vertical lines, increases from ~0.51 to ~0.59.

and J0740+6620, and the mass of J0348+0432 (all with a fixed flat mass prior) to find $R_{1.4} = 12.6^{+1.0}_{-1.1}$ km and $M_{\rm TOV} = 2.21^{+0.31}_{-0.21}\,M_\odot$. Our updated radius estimate has a ∼0.4 km lower median due to the new J0437-4715 data that favor softer EoSs and a ∼20% smaller uncertainty due to the fact that we use more NICER and massive pulsar data. Our updated $M_{\rm TOV}$ estimate of $2.28^{+0.41}_{-0.21}\,M_\odot$ is marginally larger than the value found in Legred et al. [332], which can be attributed to the spider pulsars, and the removal of the EoS Occam penalty for massive pulsar measurements; see the Appendix of Ref. [332].

The full mass-radius inferred relation is shown in Figure 17.4 which plots the 90% symmetric credible region for the radius at each mass. We include the prior, the posterior from our analysis, and compare against the posterior from Legred et al. [332], i.e., without J0437-4715. While the radius lower limit is broadly consistent with Ref. [332], we obtain a lower radius upper limit for all masses by ∼500 m, which we attribute to the new data for the J0437-4715 radius. We additionally plot credible regions for the relation between the NS mass m and its central density ρ_c in Figure 17.5. The upper limit on the mass of a NS with central density four times the nuclear saturation density ($\rho_{\rm nuc}$) increases from $\sim 2.55\,M_\odot$ to $\sim 2.69\,M_\odot$, primarily due to the removal of the Occam penalty and the inclusion of spider pulsars. The central density of the maximum mass star is inferred to be $5.53^{+1.07}_{-1.24}\ \rho_{\rm nuc}$ (red contours).

We examine the EoS microscopic properties and specifically the speed of sound as a function of density in Figure 17.6 and the maximum speed of sound inside NSs in Figure 17.7.

Compared to Legred et al. [332], our analysis favors a larger speed of sound around $2 - 4\rho_{\rm nuc}$ and a larger maximum speed of sound throughout. The 90% lower limit on the maximum speed of sound, increases from $\sim$ 0.51 in Ref. [332] to $\sim$ 0.59 for our analysis. This higher maximum speed of sound is necessary to explain the high mass of certain Galactic pulsars which, though poorly measured, can have exceptionally large median values, e.g., J01748-2021B with an estimated mass of $2.74^{+0.21}_{-0.21}\,M_\odot$ [240] at 68% credibility. The addition of the NICER radius measurement J0437-4715 also marginally impacts the inferred maximum sound speed; removing the radius measurement of J0437-4715, Appendix 17.D, leads to a maximum c_s^2 value of $0.8^{+0.19}_{-0.31}$.

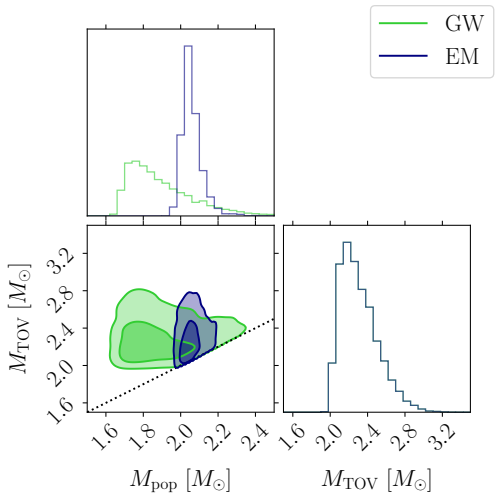

Figure 17.8: One-and two-dimensional posteriors for $M_{\rm TOV}$ and the maximum astrophysical mass $M_{\rm pop}$ for the Galactic NSs (blue) and the merging BNSs (orange). The black dashed line represents $M_{\rm pop} = M_{\rm TOV}$, which is imposed in our analyses as we assume that all objects are NSs. The TOV mass is consistent with the astrophysical maximum mass for both populations. Contours are drawn at 50% and 90% levels.

17.4.3 Joint constraints on the population and EoS

The joint EoS-mass inference allows us to separate the TOV mass, $M_{\rm TOV}$, from the maximum astrophysical mass in the two subpopulations, $M_{\rm pop,\,EM}$ and $M_{\rm pop,\,GW}$. Figure 17.8 shows the joint posterior for $M_{\rm TOV}$ and the two population maximum masses, denoted collectively as $M_{\rm pop}$. The limit $M_{\rm TOV} = M_{\rm pop}$ is marked with a dashed line; points near the line correspond to maximum population masses that are equal to the TOV mass. As also evident in Figure 17.1, the two population maximum masses are consistent with each other within their statistical uncertainties. The difference between the maximum mass in the EM (GW) population and $M_{\rm TOV}$ is less than 0.53 $M_\odot$ (0.73 $M_\odot$) at 90% credibility.

We therefore have no evidence that the maximum mass of neutron stars formed astrophysically is different than the maximum mass possible from nuclear physics.

17.5 Conclusions

As a first step toward untangling the properties of NSs that depend on nuclear physics versus astrophysics, in this study we presented a joint inference of the dense matter EoS and the NS mass distribution. We considered two subpopulations

of NSs corresponding to merging BNSs observed with GWs and Galactic NSs observed with EM. All NSs share the same universal EoS modeled with a flexible GP mixture. Our results are consistent with existing EoS-only or mass-only inference where applicable [332, 214, 28]. However, the joint inference scheme allows us to begin addressing the interplay between nuclear physics and astrophysics in determining NS observational properties. Focusing on NS masses, we find no evidence that the maximum mass of NSs observed with either EM or GWs is different than the maximum mass allowed by nuclear physics. Moreover, we updated the estimates of the canonical NS radius and the TOV mass to $R_{1.4} = 12.2^{+0.8}_{-0.9}\,\mathrm{km}$ and $M_{\mathrm{TOV}} = 2.28^{+0.41}_{-0.21} M_\odot$, respectively.

17.5.1 Past work

Our results are broadly consistent with comparable studies. Whereas we model the EoS phenomenologically as a GP, Rutherford et al. [476] used a piecewise-polytropic EoS model and the same data as Legred et al. [332] plus the radius measurement of J0437-4715; they found $R_{1.4} = 12.3^{+0.5}_{-0.8}\,\mathrm{km}$. Our result has a ∼30% larger uncertainty likely due to the more flexible EoS model.

Fan et al. [205] simultaneously inferred the mass distribution and the EoS, though they assumed the same mass distribution for all NSs, and that the upper truncation mass for the NS population is M_{TOV}. They used the same data as our study except the radius measurement of J0437-4715, and included ∼50 additional pulsar mass measurements. They used a variety of parameteric and nonparametric EoS models, but recovered similar values of $R_{1.4}$ and M_{TOV} for all models, indicating their non-parametric models may have limited flexibility (analogous to the "model-informed prior" of [325, 198]). They further incorporated information from perturbative quantum chromodynamics (pQCD) at high densities, and chiral perturbation theory at low densities, both of which strongly informed the estimate of M_{TOV} due to the choice of modeling of correlations. They found $M_{\mathrm{TOV}} = 2.25^{+0.08}_{-0.07}\,M_\odot$.

Biswas and Rosswog [102] also simultaneously inferred the population and the EoS, similarly requiring the NSs to form a single population which is truncated by M_{TOV}. For the EoS they used a piecewise-polytropic parameterization, hybridized with a low-density prescription constrained by chiral effective field theory. They analyzed the same data as Fan et al. [205], and additionally the PREX-II [43] and CREX [44] measurements of the neutron skin thickness of ^{208}Pb and ^{48}Ca respectively.

They found $R_{1.4} = 12.5^{+0.3}_{-0.3}\,\mathrm{km}$, and $M_{\mathrm{TOV}} = 2.27^{+0.08}_{-0.09}\,M_\odot$. These uncertainties

are substantially lower than our results. The radius constraint can at least in part be attributed to information from chiral perturbation theory, while the EoS parameterization also results in tighter inference throughout due to less modeling flexibility [200, 333]. Moreover, the use of a single mass distribution places a very strong prior on the masses of the GW events, with the mass of GW170817 for example likely tightly constrained to be within the primary peak of the bimodal mass distribution. Such improved mass measurement will translate to tighter tidal and hence EoS constraints. The impact of pQCD information [315] remains unclear [503, 316], though the prescription used in that analysis is likely informative of $M_{\rm TOV}$.

Other studies have obtained multimessenger constraints on the EOS by combining GW, gamma-ray burst, and kilonova observations surrounding GW170817 with fits to the EM emission from BNS simulations [441, 157, 159, 412]. While there are systematic and statistical uncertainties in the models and observations, these studies infer $R_{1.4}$ and $\Lambda_{1.4}$ broadly consistent with our results.

17.5.2 Caveats

Our findings depend on several analysis choices and assumptions. In the appendices, we examine their impact, and here we summarize our conclusions.

In our main analysis, we assume that selection biases in the radio and X-ray surveys are negligible. In Appendix 17.E we consider the impact that modeling all Galactic NSs with the same bimodal distribution without taking selection effects into account has. Compared to an analysis that fixes the pulsar mass distribution to uniform up to $M_{\rm TOV}$ [332], inference of the mass distribution leads to an EoS that is marginally softer at low densities and marginally stiffer at high densities. As a consequence, the evidence for a violation of the conformal limit $c_s^2 = 1/3$ increases and the lower limit on the maximum speed of sound increases by $\sim$10%.

Data selection further influences our results. In particular, different interpretations of the NICER observations exist in the literature. Given systematic studies on the impact of analysis assumptions on NICER measurements [191, 555] we present results without J0030+0451 and/or J0437-4715 in Appendix 17.D. Excluding J0437-4715 leads to a stiffer inferred EoS with $R_{1.4} = 12.5^{+1.0}_{-0.9}$ km and consistent results with Ref. [332]. Excluding J0030+0451 results in a substantially reduced value of $R_{1.4} = 11.6^{+1.3}_{-0.9}$ km. However, all results are consistent with each other at 90% credibility, see Figure 17.11 in Appendix 17.D.

Additionally, our main results assume a fixed spin distribution, extending in magnitude up to 0.05 for GW170817 and 0.4 for GW190425. Assumptions about the spin affect mass inference through the mass-spin correlation [163] and hence mass population inference. We explore the impact of restricting the spin of GW190425 further in Appendix 17.F. Imposing an upper limit of 0.05 results in a tighter constraint on its mass ratio and a lower primary mass, which correspondingly reduces the value of $M_{\mathrm{pop,GW}}$. Consistency between $M_{\mathrm{pop,\,GW}}$ and M_{TOV} is reduced with their difference less than $0.77\,M_{\odot}$ at 90% credibility. Therefore we still find no strong evidence that the TOV and the maximum astrophysical mass are different. Simultaneous inference of the spin distribution [101], along with the EoS and mass distribution, is reserved for future work.

Finally, in this study, we restricted to two subpopulations of NSs: GW observations of BNSs and Galactic NSs from radio or X-ray surveys. As a consequence, our mass distribution inference is only predictive below $2.5\,M_{\odot}$, which we took to be the (fixed) demarcation between NSs and BHs. Extending to higher masses would require simultaneously classifying GW events as BNSs, NSBHs, or BBHs within the analysis framework [197, 141], while introducing a third NS subpopulation associated with the NSBH mergers. This would allow us to treat other GW discoveries, such as GW230529_181500 [2] and GW190814 [23], whose nature is ambiguous. These and further extensions to the joint inference methodology presented here will become necessary to fully explore the interplay between nuclear physics and astrophysics on the properties of NSs as our catalog of informative NS observations increases in size.

Acknowledgements

We thank Will Farr for helpful discussions on hierarchical inference of subpopulations. We are also grateful to Reed Essick for useful discussions on population modeling with our dataset. We also thank Sylvia Biscoveanu for helpful comments on the manuscript. J.G. would like to gratefully acknowledge the support from the National Science Foundation through the Grant NSF PHY-2207758. I.L. and K.C. acknowledge support from the Department of Energy under award number DE-SC0023101, the Sloan Foundation, and by a grant from the Simons Foundation (MP-SCMPS-00001470). P.L. is supported by the Natural Sciences & Engineering Research Council of Canada (NSERC). Research at Perimeter Institute is supported in part by the Government of Canada through the Department of Innovation, Science

and Economic Development and by the Province of Ontario through the Ministry of Colleges and Universities. 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 material is based upon work supported by NSF's LIGO Laboratory which is a major facility fully funded by the National Science Foundation. Software: `bilby` [67, 470], `dynesty` [507], `scipy` [556], `numpy` [569], `matplotlib` [289], `lwp` [196].

Appendix 17.A Reweighting scheme for the joint posterior

The joint posterior for the GP EoS ε and the population hyperparameters $\eta = \{\eta_{GW}, \eta_{\rm EM}\}$ is [326, 138]

$$p(\varepsilon, \eta|d) = \frac{\mathcal{L}(d|\varepsilon, \eta)\pi(\varepsilon, \eta)}{p(d)}\,, \tag{17.9}$$

where d is the data, $\mathcal{L}(d|\varepsilon, \eta)$ is the likelihood, $\pi(\varepsilon, \eta)$ is the prior, and $p(d)$ is the evidence. We choose a prior of $\pi(\varepsilon, \eta) = \pi(\varepsilon)\pi(\eta)\Theta(M_{\rm TOV} - M_{\rm pop,\,EM})\Theta(M_{\rm TOV} - M_{\rm pop,\,GW})$, where $\pi(\epsilon)$, is the model agnostic prior defined in Refs. [325, 198] (uniform over GP draws), and $\pi(\eta)$ is the prior on the population hyperparameters, as described in the main text (uniform over all parameters). Since the GW and EM datasets are independent, the total likelihood factors into individual likelihoods

$$\mathcal{L}(d|\varepsilon, \eta) = \mathcal{L}_{GW}(d|\varepsilon, \eta_{GW})\mathcal{L}_{\rm NICER}(d|\varepsilon, \eta_{\rm EM})\mathcal{L}_{\rm PSR}(d|\varepsilon, \eta_{\rm EM})\,,$$

given in Eqs. (17.3), (17.6), and (17.8) respectively.

We evaluate the likelihood $\mathcal{L}(d|\varepsilon, \eta)$ with a reweighting scheme based on a simpler lower-dimensional EoS model ε_0, details about which are given in Appendix 17.B. We first obtain samples from the joint posterior for ε_0 and η using standard stochastic sampling [67].

$$p_0(\varepsilon_0, \eta|d) = \frac{\mathcal{L}_0(d|\varepsilon_0, \eta)\pi_0(\varepsilon_0, \eta)}{p_0(d)}\,. \tag{17.10}$$

We then use the marginal mass distribution posterior

$$p_0(\eta|d) = \int p_0(\varepsilon_0, \eta|d)d\varepsilon_0\,, \tag{17.11}$$

as a proposal distribution to rewrite Eq. (17.9) as

$$p(\varepsilon, \eta|d) \propto \mathcal{L}(d|\varepsilon, \eta)\frac{\pi(\eta)\Theta(M_{\rm TOV} - M_{\rm pop})}{p_0(\eta|d)}p_0(\eta|d)\pi(\varepsilon)\,, \tag{17.12}$$

where we have dropped the normalization $p(d)$ and defined $\Theta(M_{\rm TOV} - M_{\rm pop}) \equiv \Theta(M_{\rm TOV} - M_{{\rm pop},EM})\Theta(M_{\rm TOV} - M_{{\rm pop},GW})$. Reweighting includes

1. Compute a Kernel Density Estimate (KDE) of $p_0(\eta|d)$ so that we can directly evaluate the density for each value of η.

2. Draw samples $\varepsilon \sim \pi(\varepsilon)$ and $\eta \sim p_0(\eta|d)$. If $M_{\rm TOV} < M_{\rm pop,\,EM}$ or $M_{\rm TOV} < M_{\rm pop,\,GW}$, reject the sample.

3. For accepted (ε, η) samples compute the weight

$$w = \mathcal{L}(d|\varepsilon, \eta)\frac{\pi(\eta)}{p_0(\eta|d)}\,. \tag{17.13}$$

   The term $p_0(\eta|d)$ is computed with the KDE from step #1 and the likelihood $\mathcal{L}(d|\varepsilon, \eta)$ is computed with a Monte Carlo sum over individual-event posterior samples.

4. Each sample (ε, η) is a weighted draw from the joint posterior $p(\varepsilon, \eta|d)$ with weight w.

In practice, we consider the EM likelihood for the two EM datasets

$$\mathcal{L}(d_{\rm EM}|\varepsilon, \eta_{\rm EM}) = \mathcal{L}(d_{\rm NICER}|\varepsilon, \eta_{\rm EM}) \times \mathcal{L}(d_{\rm PSR}|\varepsilon, \eta_{\rm EM})\,, \tag{17.14}$$

and the combined likelihood

$$\mathcal{L}(d|\varepsilon, \eta) = \mathcal{L}(d_{\rm EM}|\varepsilon, \eta_{\rm EM}) \times \mathcal{L}(d_{GW}|\varepsilon, \eta_{\rm GW})\,, \tag{17.15}$$

from Eq. (17.12). In order to calculate the likelihood for the GW population parameters η_{GW}, we approximate

$$\begin{aligned}&\mathcal{L}(d|\eta_{GW}) = \\ &\int \mathcal{L}(d_{\rm EM}|\eta_{\rm EM}, \varepsilon)\mathcal{L}(d_{GW}|\eta_{GW}, \varepsilon)\pi(\eta_{\rm EM}, \varepsilon)d\varepsilon\, d\eta_{\rm EM}\end{aligned} \tag{17.16}$$

with the Monte Carlo sum:

$$\begin{aligned}\mathcal{L}(d|\eta_{GW}) \approx &\sum_{\varepsilon\sim\pi(\varepsilon)} \mathcal{L}(d_{GW}|\eta_{GW}, \varepsilon)\times \\ &\Big[\sum_{\eta_{\rm EM}\sim p_0(\eta_{\rm EM})} \frac{\mathcal{L}(d_{\rm EM}|\eta_{\rm EM}, \varepsilon)}{p_0(\eta_{\rm EM}|d)}\pi(\eta_{\rm EM}|\varepsilon)\Big]\,.\end{aligned} \tag{17.17}$$

The likelihood for the EM population parameters is obtained by by swapping $GW \leftrightarrow$ EM in Eq. (17.17).

Similarly, we compute the likelihood for the EoS ε as

$$\mathcal{L}(d|\varepsilon) \approx \sum_{\eta_{GW}\sim p_0(\eta_{GW})} \frac{\mathcal{L}(d_{GW}|\eta_{GW},\varepsilon)}{p_0(\eta_{GW}|d)}\pi(\eta_{GW}|\varepsilon)\times \sum_{\eta_{\rm EM}\sim p_0(\eta_{\rm EM})} \frac{\mathcal{L}(d_{\rm EM}|\eta_{\rm EM},\varepsilon)}{p_0(\eta_{\rm EM}|d)}\pi(\eta_{\rm EM}|\varepsilon)\,. \tag{17.18}$$

Appendix 17.B Approximate lower-dimensional EoS model

The reweighting scheme of Appendix 17.A utilizes a lower-dimensional EoS model ε_0 that gets marginalized away in Eq. (17.11), solely for constructing an efficient proposal distribution for the hyperparameters η. The goal of including ε_0 in the first place is to avoid potential systematic biases in $p_0(\eta|d)$ if inferred without any reference to an EoS [256]. Such biases would make it an ineffective proposal distribution for the reweighting of Eq. (17.12). Our requirement for ε_0 is therefore that it can be evaluated efficiently and that it roughly captures typical EoS behaviors. Existing parametric models such as the piecewise-polytropic [448], spectral [342], or speed-of-sound [527, 261] models could play this role. However, we find that something even simpler suffices.

We take advantage of the simple relation between the NS moment-of-inertia I and mass m [587, 199] for hadronic EoSs. For EoSs without rapid changes in the speed of sound [199],

$$\frac{d\ln I}{d\ln m} \sim 1.6 \pm O(10^{-2})\,. \tag{17.19}$$

We therefore define ε_0 with a linear relationship between $\ln I$ and $\ln m$:

$$\ln I = a\ln m + b\,, \tag{17.20}$$

where the free parameters a and b define a specific EoS. From the $I(m)$ relation we can obtain $\Lambda(m)$ (used for analyzing GW data) and $R(m)$ (used for analyzing X-ray data) with the I-Love [586] and C-Love [585, 134] universal relations respectively. Since the model does not have a miscrophysics interpretation, it does not self-consistently lead to a maximum-mass solution. Instead we define its TOV mass as $\Lambda(M_{\rm TOV}) = \Lambda_{\rm thresh} = \exp(1.89)$ which empirically produces reasonable values for $M_{\rm TOV}$,

We find that this model is inexpensive to sample and accurate enough that that it leads to an improved reweighting efficiency. However, it would not be a reliable model for EoS inference due to its simplistic nature.

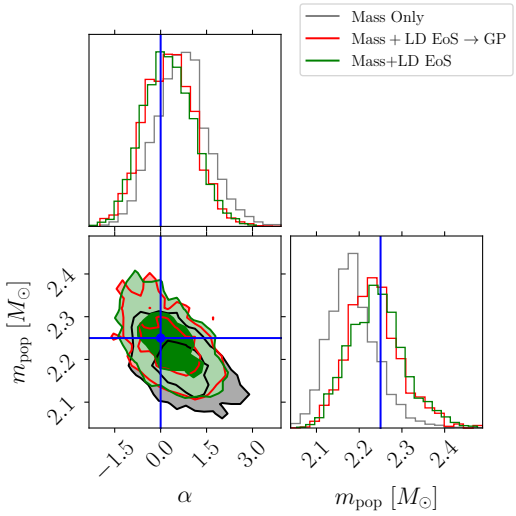

Figure 17.9: One- and two-dimensional posteriors for the mass distribution slope and maximum mass from 23 simulated BNSs. We plot mass-only population inference (grey) which defaults to the individual-event-inference prior on the tidal deformability, joint mass-EoS inference using the lower-dimensional EoS model (green) and the full mass-EoS joint inference with the GP EoS model (red). The reweighting scheme corrects the bias from inferring the mass distribution alone.

Appendix 17.C Method validation

We demonstrate the validity of the reweighting scheme described in Appendix 17.A with simulated GW data. We simulate BNS observations from a uniform mass distribution with $\alpha = 0$ between 1 $M_\odot$ and $M_{\rm pop,GW} = 2.25\,M_\odot$, assigning positions and orientations isotropically, and distances according to a merger rate uniform in the frame of the source across redshifts. Spins are distributed isotropically with uniform magnitudes up to 0.05. Tidal deformabilities are simulated according to a pre-selected EoS from the GP prior with $M_{\rm TOV} = 2.34\,M_\odot$ and $R_{1.4} = 12.5\,$km. After filtering for events that pass a detectability threshold of signal-to-noise ratio above 8, we obtain posterior samples using `bilby` [67]. We then follow the procedure of Appendix 17.A to compute the joint posterior for the mass distribution and the EoS.

In Figure 17.9 we show the inferred population hyperparameters under three analyses. The first (black) models only the mass distribution, which effectively means that the EoS model defaults to the tidal deformability prior used during sampling. This is selected to be uninformative to avoid restricting the posterior: flat between 0 and 1.5×10^3. Since this is not in reality how the tidal deformabilities of the analyzed objects are distributed, i.e., the follow a single EoS, mass inference is

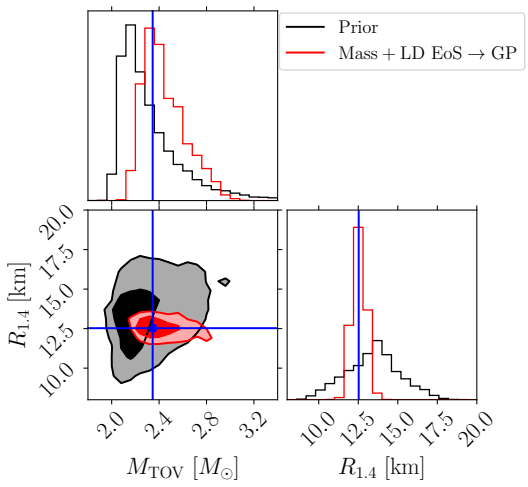

Figure 17.10: One- and two-dimensional posteriors for recovered EoS properties $M_{\rm TOV}$ and $R_{1.4}$ from 23 simulated BNSs. We plot the prior (black) and the result from reweighting to a full mass-EoS joint inference with the GP EoS model (red). The reweighting method is able to recover the true EoS (blue).

slightly biased [256]. The second analysis (green) corresponds to Eq. (17.11) that infers the mass distribution together with the lower-dimensional EoS model of Appendix 17.B. The inclusion of even this simple EoS model in the inference reduces the bias compared to the true parameters. This posterior is then used as a proposal to reweight to the final mass-EoS inference with the GP EoS model (red), which again agrees with the injected values. Figure 17.10 further shows that this procedure can infer the EoS parameters.

Appendix 17.D Effect of NICER observations

In this Appendix we quantify the impact of NICER observations on our inference. Specifically, we study the impact of J0030+0451 for which there is no concurrent radio-based mass measurement and the hotspot model has a large impact on inference [555] and J0437-4715 for which only one independent analysis is available [148]. We show results for $R_{1.4}$ in Figure 17.11. Removing any NICER pulsars leads to an increased uncertainty and a shift to lower radii (when removing J0030+0451) or larger radii (when removing J0437-4715). However, all results are consistent with each other at the 90% credible level. Using no NICER data leads to $R_{1.4} = 11.9^{+1.7}_{-1.6}$ km, no J0030+0451 data to $R_{1.4} = 11.6^{+1.3}_{-0.9}$ km, and no J0437-4715 data to $R_{1.4} = 12.5^{+1.0}_{-0.9}$ km.

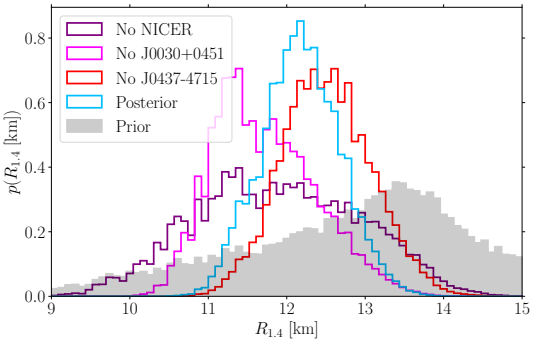

Figure 17.11: The effect of NICER constraints on EoS inference. We plot the prior (grey) and posterior for $R_{1.4}$, the radius of a $1.4\,M_\odot$ NS with different subsets of NICER data: all 3 pulsars (blue; main text analysis), excluding J0030+0451 (pink), excluding J0437-4715 (red), and excluding all NICER observations (purple).

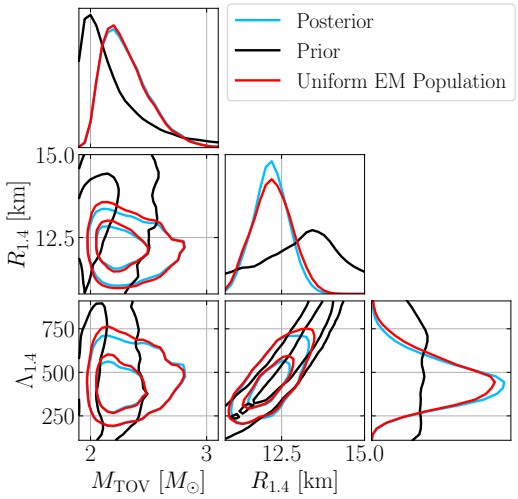

Figure 17.12: Impact of the EM population mass modeling on EoS inference. We plot the prior (black), the posterior from the full analysis (blue; same as Figure 17.3), and the posterior when the EM mass distribution is uniform and independnet of the EoS for J0030+0451 and J0437-4571 and uniform up to the TOV maximum mass of the EoS for J0740+6620 and J0348+0432. The posteriors are similar.

Appendix 17.E Uniform pulsar population

Since selection effects for pulsar radio surveys are not well quantified, it is not clear how the observed distribution of NS masses differs from the true distribution. To examine the impact of the observed EM population inference, we repeat the analysis using the approach of Ref. [332] for the EM population: it depends only on the EoS, and not on additional population hyperparameters. The GW population is still modeled with a truncated powerlaw per Sec. 17.2.3.2. We neglect all pulsars that

do not contribute directly to the EoS (due to low mass) as well as spider pulsars for consistency with Ref. [332]. The EM data now include only J0030+0451 and J0437-4715 [377, 148] with a uniform mass distribution in $[1.0{-}1.9]\,M_\odot$, and J0740+6620 and J0438+0432 [378, 61] with a uniform mass distribution in $[1.0{-}M_{\rm TOV}]\,M_\odot$, with $M_{\rm TOV}$ given by the EoS model. This choice corresponds to a uniform distribution up to the maximum mass allowed by the EoS. Because of this choice, EoSs that predict a larger TOV mass are penalized by an *Occam* penalty for the two high-mass pulsars.

Results are shown in Figure 17.12, where we find small changes to the inferred EoS quantities. In particular, $M_{\rm TOV}$ is relatively unchanged, $M_{\rm TOV} = 2.27^{+0.41}_{-0.20}\,M_\odot$ under the fixed population, which we attribute to the cancellation of two effects. One the one hand, the Occam penalty favors lower values of $M_{\rm TOV}$ under a fixed population. On the other hand, under the fixed-population scheme, the mass of the heaviest pulsars is not informed by lower-mass pulsars, and therefore ends up higher, which in turn results in a higher $M_{\rm TOV}$. The effect of the Occam penalty and the population-informed mass estimates in practice cancel out. The radius and tidal deformability change somewhat more, $R_{1.4} = 12.2^{+0.9}_{-1.0}\,$km, with a $\sim$10% larger uncertainty than the inferred-population case, and $\Lambda_{1.4} = 450^{+247}_{-175}$ being slightly larger than the inferred-population case.

Overall, inferring the EM mass distribution leads to marginally higher $M_{\rm TOV}$ and lower $R_{1.4}$. Put differently, the high-density EoS is marginally stiffer and the low-density EoS is marginally softer. As a consequence, the maximum sound-speed is higher in order to connect the soft(er) low-density EoS to a stiff(er) high-density EoS. This leads to increased support for violation of the conformal limit, $c_s^2 > 1/3$. The natural logarithm of the Bayes factor in favor of conformal violation is $\ln \mathcal{B}^{c_s^2>1/3}_{c_s^2<1/3} = 5.85 \pm 0.30$ for the fixed population model, and $\ln \mathcal{B}^{c_s^2>1/3}_{c_s^2<1/3} = 7.39 \pm 0.52$ when the mass distribution of EM pulsars is inferred.

Appendix 17.F Low spin assumption for GW190425

Assumptions about the spin of GW190425 have an effect on the inferred component masses [11]. In the main text, we assume that the NSs in GW190425 can have dimensionless spin magnitudes up to 0.4. However, other studies assume NSs have spins 0.05, motivated by the spin distribution of pulsars in binary systems expected to merge within a Hubble time [599]. In Figure 17.13, we present results with a low-spin assumption for GW190425, enforcing the same assumption in the sensitivity

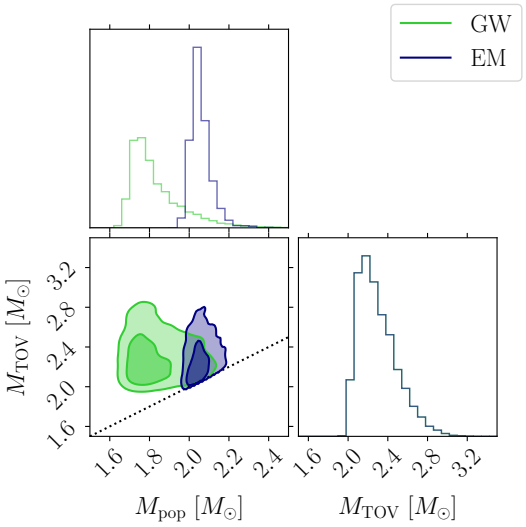

Figure 17.13: Similar to Figure 17.8 but with a low-spin assumption for GW190425 of < 0.05.

estimates as well. We find $M_{\rm TOV} = 2.26^{+0.39}_{-0.21}\,M_\odot$ and $M_{\rm pop,GW} = 1.79^{+0.32}_{-0.1}\,M_\odot$. As GW190425 is not the main observation informing $M_{\rm TOV}$, it values is consistent with the main analysis. However, as the low-spin restriction lowers the estimated masses of GW190425 due to the mass-spin correlation, we obtain a lower value for $M_{\rm pop,GW}$, though still consistent with $M_{\rm TOV}$.

Chapter 18

USING EQUATION OF STATE CONSTRAINTS TO CLASSIFY LOW-MASS COMPACT BINARY MERGERS

Jacob Golomb, Isaac Legred, Katerina Chatziioannou, et al. "Using equation of state constraints to classify low-mass compact binary mergers". In: *Phys. Rev. D* 110.6 (2024), p. 063014. DOI: 10.1103/PhysRevD.110.063014. arXiv: 2403.07697 [astro-ph.HE].

Abstract

Compact objects observed via gravitational waves are classified as black holes or neutron stars primarily based on their inferred mass with respect to stellar evolution expectations. However, astrophysical expectations for the lowest mass range, $\lesssim 1.2\,M_\odot$, are uncertain. If such low-mass compact objects exist, ground-based gravitational wave detectors may observe them in binary mergers. Lacking astrophysical expectations for classifying such observations, we go beyond the mass and explore the role of tidal effects. We evaluate how combined mass and tidal inference can inform whether each binary component is a black hole or a neutron star based on consistency with the supranuclear-density equation of state. Low-mass neutron stars experience a large tidal deformation; its observational identification (or lack thereof) can therefore aid in determining the nature of the binary components. Using simulated data, we find that the presence of a sub-solar mass neutron star (black hole) can be established with odds $\sim 100:1$ when two neutron stars (black holes) merge and emit gravitational waves at signal-to-noise ratio ~ 20. For the same systems, the absence of a black hole (neutron star) can be established with odds $\sim 10:1$. For mixed neutron star-black hole binaries, we can establish that the system contains a neutron star with odds $\gtrsim 5:1$. Establishing the presence of a black hole in mixed neutron star-black hole binaries is more challenging, except for the case of a $\lesssim 1\,M_\odot$ black hole with a $\gtrsim 1\,M_\odot$ neutron star companion. On the other hand, classifying each individual binary component suffers from an inherent labeling ambiguity.

18.1 Introduction

Astronomical observations have revealed a diversity in compact objects with masses $\lesssim 3\,M_\odot$. Classifying these observations as black holes (BHs), neutron stars (NSs), or white dwarfs (WDs), requires identifying observational signatures that are unique to each type. For example, pulsars are identified as NSs [254], while unique electromagnetic spectrum or emission signatures can distinguish between NSs and BHs even if the mass is unknown, as is the case for accreting X-ray binaries [452, 188, 537]. On the gravitational-wave (GW) side, classification is simplified by the fact that ground-based GW detectors are only sensitive to objects that do not disrupt or collide before reaching the detector sensitive band $\gtrsim 10$ Hz. For example, a pair of maximum compactness WDs each with mass $1.3\,M_\odot$ and radius 1700 km collide at a GW frequency of ≈ 1 Hz, see Appendix 18.A for calculation details. However, even after excluding WDs, distinguishing between NSs and BHs is challenging because, unlike electromagnetic emission, their GW emission is more similar, as it is primarily determined by the object's mass.

GW mass measurements in conjunction with astrophysical and nuclear physics can lead to preliminary classification indications. Causality limits NS masses $\lesssim 3\,M_\odot$ [460, 300]; more massive objects observed in GWs must be BHs. Astronomical and nuclear constraints suggest that NSs do not reach this theoretical maximum, however. Estimates of the maximum mass of stable nonrotating NSs [541, 407] range $2.0 - 2.5\,M_\odot$ [332, 459, 177, 413, 439, 235]; rigidly rotating NS can be $\sim 20\%$ more massive [153]. Based on these constraints, Refs. [11, 197, 23] argued that the GW190425 [11] primary was likely a NS, while the GW190814 [23] secondary was a BH. However, it is unclear if stellar evolution creates NSs up to the maximum mass allowed by nuclear physics; little evidence for or against this scenario is observationally available [28].

Switching to the full mass distribution, Galactic observations indicate that the observed NS population is strongly peaked at $\sim 1.4\,M_\odot$, with a lower (upper) truncation near $1.1\,(2.0)\,M_\odot$ [58, 214]. The Galactic BNS population is narrower and peaked at $1.4\,M_\odot$ [58, 494], though the impact of selection effects on these results is unclear. Neither result is consistent with the GW-observed NS mass distribution that displays no prominent peak at $1.4\,M_\odot$ [141, 327, 28]. Electromagnetic observations suggest a scarcity or even absence of sub-$5\,M_\odot$ BHs [409, 318, 217, 495], though candidates, subject to debate [277, 533, 73], exist [532, 294]. The $2.6\,M_\odot$ secondary in GW190814 [23] as well as galactic observations [83, 145] indicate that if a mass

gap between NSs and BHs does exist, it is not empty [28]. In the absence of unambiguous classification for $\sim 2-3\,M_\odot$ objects, Refs. [224, 28, 208] modeled the mass distribution of all sub-10 $M_\odot$ objects and identified a feature at $\sim 2.4\,M_\odot$. Under the assumption of nonoverlapping NS and BH distributions, such a feature could signal the transition from the NS to the BH population.

In contrast to these astrophysics- and nuclear physics-informed considerations about the high end of the NS mass range, the low end remains uncharted. No widely-accepted *astrophysical* process results in stellar remnants of either type with masses $\lesssim 1.2\,M_\odot$ [514, 329, 331], although *physically* cold NSs remain stable down to $O(10^{-1})\,M_\odot$ [329, 331].[1] Radio and X-ray observations have led to NS candidates with masses $\sim 1.17\,M_\odot$ [368] and $\sim 0.8\,M_\odot$ [182]. Additionally, masses and eccentricities of *Gaia* binaries suggest the existence of $\sim 1\,M_\odot$ NSs [494]. As for BHs, while sub-1 $M_\odot$ BHs do not form through stellar collapse, early-universe density fluctuations and sufficiently dissipative dark matter could collapse into primordial BHs with masses in this range [131, 400]. Searches for subsolar mass compact objects with GWs have as of yet yielded no detections [31, 32, 399]. If such BHs do exist, they may be detectable by current and future GW detectors, and properties such as their masses and spins may be measurable [573, 385].

Given these uncertainties, classification of potential sub-1.2 $M_\odot$ GW candidates requires an additional unique signature: matter effects.[2] GWs from mergers involving NSs carry the imprint of tidal interactions in the signal phase evolution [138, 282, 232]. To leading order[3], the effect is quantified by the dimensionless tidal deformability which depends on the nuclear equation of state (EoS) ($c = G = 1$):

$$\Lambda \equiv \frac{2}{3} k_2 C^{-5}\,, \tag{18.1}$$

where k_2 is the quadrupole tidal love number, and $C = m/R$ is the compactness, the ratio of the NS mass m to its radius R. Tidal interactions enter the GW phase to leading $5th$ Post-Newtonian (PN) order [232, 220] through $\tilde{\Lambda}$, a mass-weighted combination of the component tidal deformabilities. BHs in General Relativity have

[1] The minimum mass of a hot proto-NS is however likely larger than that of a cold NS [498, 331, 512].

[2] On the electromagnetic side, matter effects manifest as counterparts, such as with GW170817 [13], proving the presence of at least one NS and a 10:4 preference for two [280, 158, 159]. Absence of a counterpart does not necessarily rule out NSs, as detectability may be limited by beaming or prompt collapse [11].

[3] Higher-order effects, such as dynamical tides [281, 431, 245], also affect the waveform and can aid in distinguishing NSs and BHs.

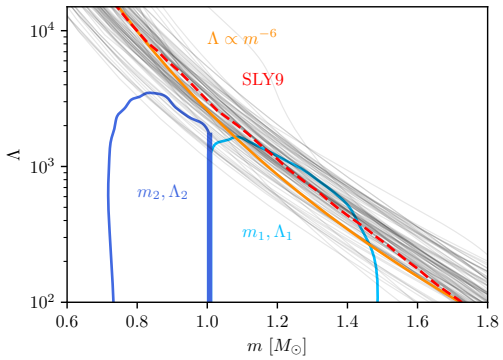

Figure 18.1: The $m - \Lambda$ relation for draws from the EoS posterior from [332] (gray lines). A red dashed line denotes the SLY9 EoS. An orange solid line indicates the $\Lambda \propto m^{-6}$ trend. The posteriors of the masses and tidal deformabilties of the primary and secondary component of a BBH simulated signal are shown in light blue and dark blue, respectively. Despite poorer tidal constraints, the secondary is less consistent with the EoSs, suggestive of a BH. While this demonstration does not capture the full 4-dimensional mass-Λ correlations, it sketches the main classification idea.

vanishing k_2, making Λ a unique signature of the compact object nature [96, 147] [4]. Tidal information has previously suggested the presence of at least one NS in GW170817 based on disfavoring zero tides [15], EoS-independent relations [19] and consistency of the tidal measurement with EoS inference [198]. Furthermore, Ref. [146] showed that lack of tidal signature can be used to identify $\sim 1 - 2\,M_\odot$ BHs if they exist, though distinguishing between NSBHs and BBHs is more challenging if the BH has a higher mass [115].

Tidal deformability becomes an increasingly better discriminator between BHs and NSs as the object's mass decreases. For $m \gtrsim 1\,M_\odot$, k_2 scales as $k_2 \sim m^{-1}$ [595], resulting in $\Lambda(m) \sim m^{-6}$, see Figure 18.1, assuming an approximately constant radius.[5] The lowest-mass NSs therefore exhibit the strongest tidal signatures and differ the most from BHs [162], with $\Lambda \sim O(10^4)$ for $m \sim 1\,M_\odot$, compared to $\Lambda \sim O(10)$ for $m \gtrapprox 2\,M_\odot$.

In this work, we leverage the expected large tidal deformabilities of low-mass NSs, combined with astrophysically-informed EoS constraints to classify compact objects as either NSs or BHs. Our classification is based on the fact that a compact object's tidal deformability must be consistent with the EoS prediction if it is a NS, see the

[4] Beyond static tides and Λ, Kerr BHs have nonvanishing dynamical tides [422].

[5] This is a good approximation excluding EoSs with phase transitions [262, 329].

$m - \Lambda$ relation in Figure 18.1, or zero if it is a BH. While the true EoS is unknown, astronomical observations have placed constraints, giving independent predictions for the tidal deformability of a NS of a given mass, e.g., [177, 326, 332, 413, 439]. This method expands upon efforts to identify NSs through a $\tilde{\Lambda} > 0$ condition [15], as we additionally require Λ to be consistent with predictions from the dense-matter EoS, similar to the GW170817 classification of [198]. In other words, our analysis combines the discriminatory power of two conditions: BHs are consistent with $\Lambda = 0$ and NSs are consistent with $\Lambda = \Lambda(m)$ as predicted by the EoS.

We test our classification approach with simulated data from low-mass sources with signal-to-noise ratios (SNRs) of 20 and 12 at advanced detector sensitivity. Lower (upper) limits on Λ allow us to rule out a BH-BH (NS-NS) origin when at least one of the binary components is a NS (BH). Figure 18.1 shows a demonstration of this idea in the BH-BH case. Though this plot is restricted to two dimensions and does not capture the strong correlations between Λ_1 and Λ_2, c.f., Figure 18.4, the full-dimensional posterior structure is leveraged in the classification scheme laid out in Sec. 18.2.2. In systems with sufficiently unequal masses, $m_2/m_1 \lesssim 0.8$, it might be possible to conclude that there is only a single NS. We also discuss an ambiguity in labeling individual objects that makes it difficult to identify the NS in a single-NS system.

The rest of the paper is organized as follows. In Sec. 18.2, we overview the parameter estimation methodology and source classification procedure. We present parameter estimation results on simulated signals in Sec. 18.3. Using these results, we quantify the evidence of BHs and NSs in Secs. 18.4 and 18.5, respectively. We conclude in Sec. 18.6.

18.2 Methods

In this section, we describe the classification procedure and the methods for demonstrating its effectiveness. In Sec. 18.2.1, we describe the simulated low-mass signals and parameter estimation. In Sec. 18.2.2, we fold in EoS information to quantify the probability of each source type.

18.2.1 Classification-agnostic Parameter Estimation

We simulate data for binaries with all unique configurations of source-frame masses $(m_1, m_2) \in \{0.8, 0.9, 1.0, 1.1, 1.2\}\, M_\odot$ with $m_1 \geq m_2$ and source type NS-NS,

BH-NS, NS-BH, and BH-BH, where the first (second) initial corresponds to the primary (secondary). The lower mass is selected both for computational reasons and because distinguishability is easier for even lower-mass systems. This results in 55 total configurations.[6] For brevity, we refer to BH-BH as BBH and NS-NS as BNS. We simulate sources with no spins and two network SNRs, one high-SNR set with $\rho_{\rm net} \approx 20$ and another lower-SNR set with $\rho_{\rm net} \approx 12$. The former corresponds to an optimistic detection scenario, although still quieter than GW170817 [10], while the latter is representative of the bulk of detections. Further details are provided in Appendix 18.B. BHs are simulated with vanishing Λ. For NSs, we assign $\Lambda(m)$ according to their mass m and the EoS SLY9 [263], chosen as a representative EoS that is consistent with current astronomical data [332], see Figure 18.1. We adopt standard priors for all parameters, detailed in Appendix 18.B. We remain agnostic on source type and adopt a uniform prior between 0 and 20×10^3 for the tidal deformabilities for all simulated signals.

We simulate data observed by the LIGO-Virgo detector network [16, 1, 38] with a zero noise realization, corresponding to a geometric mean of many noise realizations [393]. For the noise Power Spectral Densities (PSDs), we use the LIGO O4 low-sensitivity and O3 Virgo noise curves [16, 117, 39]. Signals are simulated and modeled with IMRPhenomXAS_NRTidalv3 [4], a phenomenological, frequency-domain waveform model for the dominant GW emission from the coalescence of BNS mergers with aligned spin components. The model is based on a BBH GW model [429], which is then augmented with a closed-form tidal expression [176, 4]. The model incorporates dynamical tidal effects [281] and is calibrated to a suite of numerical-relativity simulations. Two of these simulations are unequal-mass systems with a subsolar mass secondary ($0.98\,M_\odot$ and $0.90\,M_\odot$, with tidal deformabilities ~ 2600 and ~ 4600, respectively). The model has also been compared against an unequal-mass system with a subsolar mass component $\sim 0.94\,M_\odot$ and a tidal deformability of ~ 9300 [545]. Its reliability has been checked within $m_{1,2} \in [0.5, 3.0]\,M_\odot$ and $\Lambda_{1,2} \in [0, 20000]$, a range well-suited for our study.

For illustrative purposes, we show relevant frequencies around the binary merger as a function of mass in Figure 18.2, see Appendix 18.A for a detailed definition. We include the merger frequency, defined as the frequency of peak strain [258], the contact frequency, defined from a binary separation equal to the sum of the components' radii, and $f_{6M} \equiv (6^{3/2}(m_1 + m_2))^{-1}/(2\pi)$, an approximation for the

[6] The total number of possible systems is 100. Enforcing $m_1 > m_2$ and taking into account that equal-mass NS-BH and BH-NS systems are identical reduces this to 55.

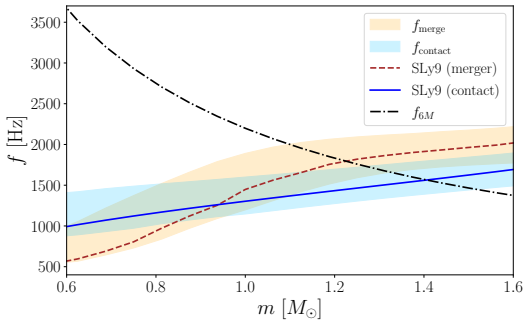

Figure 18.2: Relevant frequencies for late-inspiral signals: merger (peak strain, tan) and contact (orbital separation corresponding to objects touching, light blue) of NSs in equal-mass systems as a function of component mass. Shaded regions correspond to marginalization over the EoS posterior from [332]. Colored lines correspond to the SLy9 EoS [183, 263], which we use to simulate data. Lastly, we display an approximation for the plunge frequency of a comparable mass BBH f_{6M} with a black dash-dot line.

plunge frequency of BBHs. In the mass range of interest, all frequencies are between $\sim 1 - 3\,\mathrm{kHz}$.

18.2.2 Classifying Compact Binaries using EoS Information

The possible source classes for each detected binary are (T_1, T_2) one of $\{(\mathrm{BH}, \mathrm{BH}), (\mathrm{NS}, \mathrm{BH}), (\mathrm{BH}, \mathrm{NS}), (\mathrm{NS}, \mathrm{NS})\}$, where T_1 and T_2 refer to the source type (BH or NS) of the primary (more massive) or secondary (less massive) object, respectively. For each event, the likelihood given an EoS ϵ and source type T_1, T_2 is obtained by marginalizing over the binary masses and tidal deformabilities:

$$\begin{aligned}\mathcal{L}(d|\epsilon, T_1, T_2) = \int & dm_1\, dm_2\, d\Lambda_1\, d\Lambda_2\, \mathcal{L}(d|m_1, m_2, \Lambda_1, \Lambda_2) \\ & \times \pi(m_1, m_2)\pi(\Lambda_1, \Lambda_2|\epsilon, m_1, m_2, T_1, T_2)\,,\end{aligned} \tag{18.2}$$

where $\mathcal{L}(d|m_1, m_2, \Lambda_1, \Lambda_2)$ is the GW likelihood over the masses and tidal deformabilities, $\pi(m_1, m_2)$ is the prior on masses, and $\pi(\Lambda_1, \Lambda_2|\epsilon, m_1, m_2, T_1, T_2)$ is the prescription for computing the tidal deformabilities. For EoSs with a single stable branch[7]

$$\pi(\Lambda_i|\epsilon, m_i, T_i) = \begin{cases} \delta(\Lambda_i - \Lambda(m_i|\epsilon))\,, & \text{if } T_i = \mathrm{NS} \\ \delta(\Lambda_i)\,, & \text{if } T_i = \mathrm{BH} \end{cases}. \tag{18.3}$$

[7]If there are multiple stable branches we use a prior $\pi(\Lambda_i) = \sum_{j=0}^{N} \frac{1}{N}\delta(\Lambda_i - \Lambda(m_i|\epsilon, j))$, where j indexes stable branches and $\Lambda(m_i|\epsilon, j)$ is the tidal deformability on the $j-$th branch. A NS of a given mass is equally likely to be formed on any stable branch.

Equation (18.3) corresponds to the following prior on Λ_i: under the T_i = NS hypothesis, Λ_i is determined by the EoS ϵ and m_i, whereas under the T_i = BH hypothesis, the object has a vanishing tidal deformability. Equation (18.2) is independent of the prior on Λ_i and m_i used in the original single-event analysis of Sec. 18.2.1 as it only depends on the single-event likelihood. The Λ_i prior in Eq. (18.2) is instead the EoS-informed prior of Eq. (18.3).

The mass prior is encoded in $\pi(m_1, m_2)$, which is selected to be uniform in the joint source-frame component mass space, with $m_1, m_2 \in [0.5, 1.8]\, M_\odot$. This uniform prior is chosen for simplicity, as no constraints exist on the mass distribution of $\leq 1.2\, M_\odot$ NSs and BHs. It is nonetheless consistent with constraints on the $\sim 1 - 2\, M_\odot$ mass distribution [327, 28]. If a population of low-mass binaries were discovered, the mass prior would also be inferred via an extension of Eq. (18.2), e.g, [256, 584].

Whereas Eq. (18.2) is conditioned on a single EoS ϵ, the true EoS is unknown. We instead marginalize over the EoS and compute the likelihood for each classification:

$$P(d|T_1, T_2) = \int \mathcal{L}(d|\epsilon, T_1, T_2)\pi(\epsilon|d_{\rm aux})d\epsilon\,, \tag{18.4}$$

where $\pi(\epsilon|d_{\rm aux})$ is a distribution over EoSs informed by auxiliary data $d_{\rm aux}$. We adopt the posterior from Ref. [332] computed using a model-agnostic prior on the EoS based on a Gaussian process [325, 198, 333] and informed by radio-pulsar measurements [235, 61], X-ray pulse-profile [377, 461, 378, 462], and GW observations [10, 15, 11]. The EoS posterior is consistent with chiral effective field theory calculations at densities $\lesssim 1.5\, \rho_{\rm nuc}$ (where $\rho_{\rm nuc}$ is nuclear saturation density) [565, 272, 527, 184], comparable to the central densities of $\sim$ 1–1.5 $M_\odot$ NSs, though it does not explicitly incorporate this information [200]. It is also consistent with the existence of strong phase transitions [199].

The main physically relevant questions are

1. whether a source contains at least one BH,
2. whether a source contains at least one NS,
3. and, if so, whether it contains two NSs.

Due to the lack of constraints on the merger rates of different source types in the relevant mass range we assign equal prior probability on 3 hypotheses $\mathcal{H}$: (i) the

system has two NSs (BNS), (ii) the system has exactly one NS (OneNS), and (iii) the system has no NSs (BBH).

The marginal likelihood[8] of $\mathcal{H}$ is obtained by integrating over the relevant constituent source types:

$$\mathcal{Z}_{\mathcal{H}} \equiv \int p(d|T_1, T_2)\pi(T_1, T_2|\mathcal{H})\, dT_1\, dT_2\,, \tag{18.5}$$

where $p(d|T_1, T_2)$ is given in Eq. (18.4), and $\pi(T_1, T_2|H)$ is the normalized prior on the source types. The hypotheses $\mathcal{H}$ = BNS and $\mathcal{H}$ = BBH contain a single source type each, with the trivial priors $\pi(\mathrm{NS}, \mathrm{NS}\,|\mathrm{BNS}) = 1$, and $\pi(\mathrm{BH}, \mathrm{BH}\,|\mathrm{BBH}) = 1$ respectively. The hypothesis $\mathcal{H}$ = OneNS encompasses two source types, NSBH and BHNS, which we take to be equally likely *a priori*, $\pi(\mathrm{NS}, \mathrm{BH}\,|\mathrm{OneNS}) = \pi(\mathrm{BH}, \mathrm{NS}\,|\mathrm{OneNS}) = 1/2$.

The marginal likelihood for whether the system contains at least one NS ("HasNS") is then

$$\begin{aligned}\mathcal{Z}_{\mathrm{HasNS}} = {} & \mathcal{Z}_{\mathrm{OneNS}}\, \pi(\mathrm{OneNS}|\mathrm{HasNS}) \\ & + \mathcal{Z}_{\mathrm{BNS}}\, \pi(\mathrm{BNS}|\mathrm{HasNS})\,,\end{aligned} \tag{18.6}$$

where $\pi(\mathrm{OneNS}|\mathrm{HasNS}) = \pi(\mathrm{BNS}|\mathrm{HasNS}) = 1/2$, meaning under the assumption the system has at least one NS, we assign an equal prior probability that it has one or two NSs. The marginal likelihood for whether the system contains at least on BH ("HasBH") is Eq. (18.6), with BNS $\rightarrow$ BBH.

In what follows, we present odds ratios between two hypotheses $\mathcal{H}_1$ and $\mathcal{H}_2$:

$$\mathcal{O}^{\mathcal{H}_1}_{\mathcal{H}_2} = \frac{\mathcal{Z}_{\mathcal{H}_1}}{\mathcal{Z}_{\mathcal{H}_2}} \frac{\pi(\mathcal{H}_1)}{\pi(\mathcal{H}_2)}\,, \tag{18.7}$$

where $\pi(\mathcal{H})$ is the prior on the hypothesis $\mathcal{H}$, with $\pi(\mathrm{HasNS}) = \pi(\mathrm{HasBH}) = 2\pi(\mathrm{BNS}) = 2/3$.

18.3 Measuring the Masses and Tides of Low-mass Compact Binaries

In this section, we present posteriors from simulated signals. We do not assume we know whether each component is a NS or BH *a priori*. Throughout, we present results from simulations with $\rho = 20$.

[8]The marginal likelihood is also commonly referred to as the "evidence", though we use this term in its colloquial meaning.

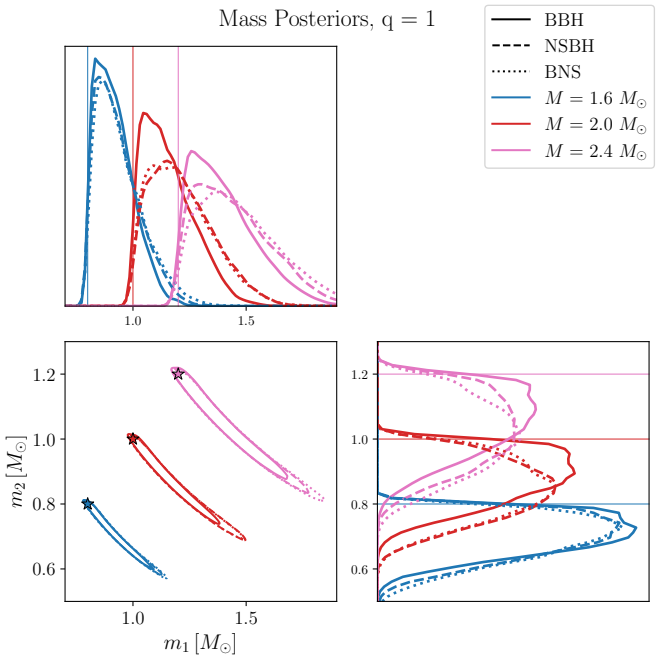

Figure 18.3: One- and two-dimensional marginalized source-frame mass posteriors for the $q \equiv m_2/m_1 = 1$ signals. Same-color lines denote systems with varying total mass M with true values marked. For a given mass, varying line styles denote BBH, NSBH, and BNS systems. Contours represent two-dimensional 2-σ regions. Given a simulated mass, similar posteriors across source types shows the subdominant effect of tides on the inferred masses.

The dominant intrinsic feature of a GW signal is the mass. In Figure 18.3, we present marginal posteriors for the source-frame masses for select equal-mass systems. Measurement uncertainties are consistent with those of Ref. [573], c.f., their Figs. 1 and 2, at the same SNR. Same-color lines denote systems with the same total mass, while varying line styles denote simulated source types. Same-mass signals result in similar mass posteriors, regardless of the source type, with a minor trend for longer tails as the tidal effects increase. This is due to the fact that the mass is primarily measured by the long inspiral phase (thousands of cycles), while tidal effect are relevant for the last $\sim$ 20 cycles. We obtain qualitatively similar posteriors for non-equal mass signals.

Having established that the presence of tides does not strongly impact mass inference, we now turn to tidal inference. Figures 18.4 and 18.5 show marginal posteriors for systems with fixed q and M, respectively, with colors denoting the source type. The top rows show the marginal $q - \tilde{\Lambda}$ posteriors. All posteriors are consistent with the true (simulated) values. Within each panel, i.e., for configurations of the same

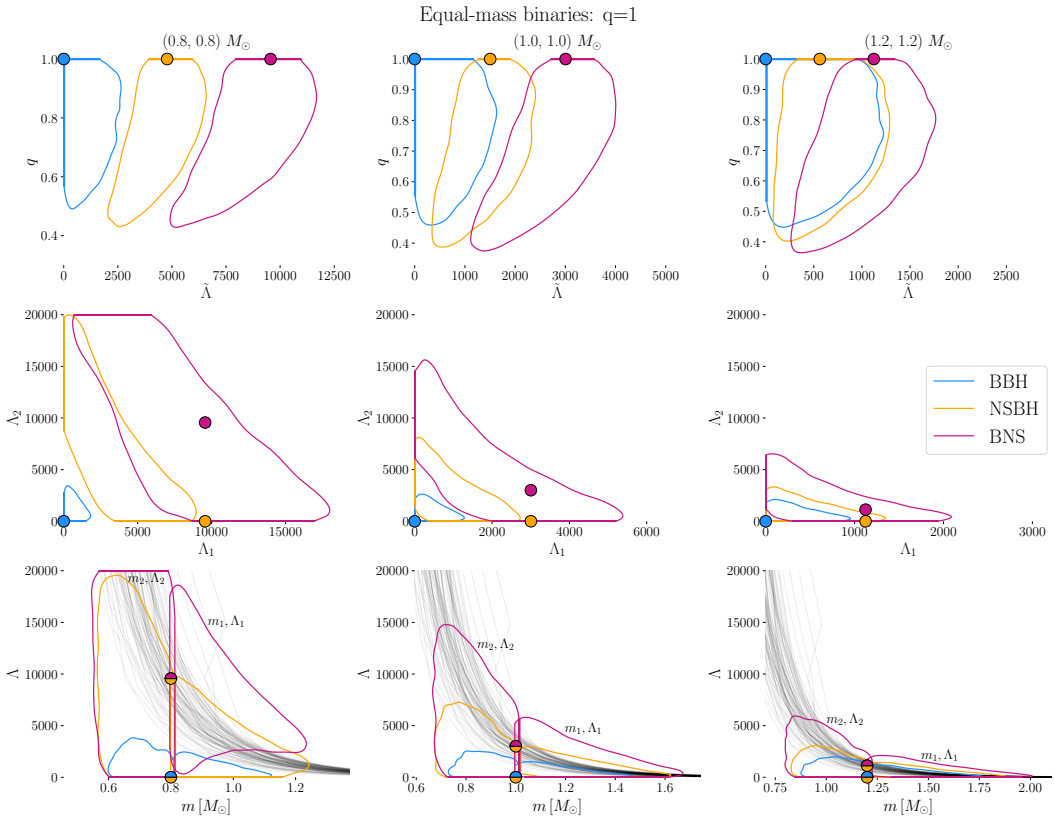

Figure 18.4: Two dimensional marginal posteriors for select parameters for systems with $q = 1$, with each column referring to a different simulated total mass. Blue, yellow, and magenta lines outline the 2-σ contours of the posterior for the BBH, NSBH, and BNS systems, respectively. We omit the BHNS configuration as it is identical to NSBH for equal-mass simulations. The left (right) halves of the third row plots are the posterior of the primary (secondary), and include draws from the EoS distribution [332] for reference. A decreasing total mass increases the tidal signature and correspondingly affects all posteriors.

mass, the posterior moves to higher values as the system contains more NSs and tidal effects become stronger. The posteriors further show a positive correlation between q and $\tilde{\Lambda}$ which becomes stronger as $\tilde{\Lambda}$ increases in value, consistent with [15]. An outcome of the increasing correlation strength is that the uncertainty also increases as the posterior is more extended both in the q, see also Figure 18.3, and $\tilde{\Lambda}$ directions.

The $q - \tilde{\Lambda}$ posterior offers the first evidence about the presence/absence of tides and thus source classification. For all mass configurations, the BBH signals are consistent with the true value $\tilde{\Lambda} = 0$, and the posteriors are similar for different masses, c.f., blue contours in Figs. 18.4 and 18.5, left to right. For NS-containing systems, the posteriors move away from $\tilde{\Lambda} = 0$, signaling the presence of tides. As

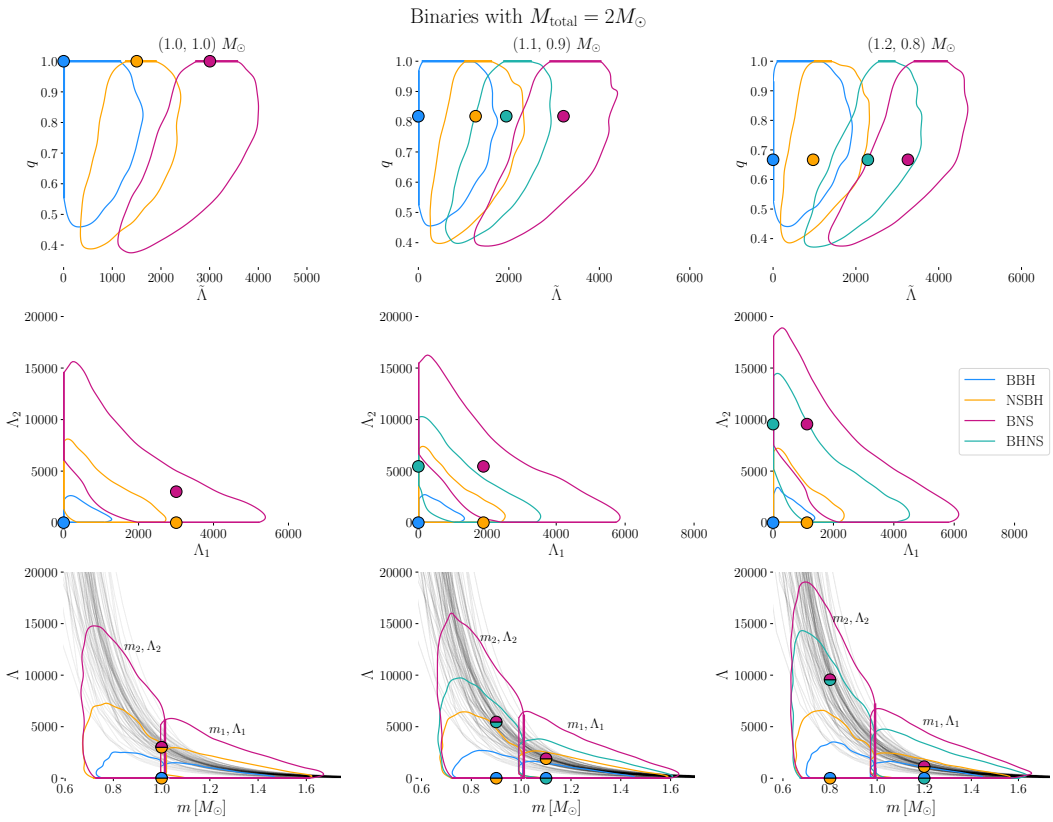

Figure 18.5: Similar to Figure 18.4 but for systems with the same simulated total mass $M = 2\,M_\odot$, with each column referring to a different simulated the mass ratio. When relevant, we also include BHNS configurations in green. The posteriors of all parameters are, weakly sensitive to the true mass ratio, with the exception of the BHNS cases.

expected, signals from lower-M systems can rule out $\tilde{\Lambda} = 0$ with higher credibility due to their higher true $\tilde{\Lambda}$ value, c.f., yellow and magenta contours in Figure 18.4, left to right. At a fixed M, the dependence of $\tilde{\Lambda}$ on the mass ratio is less pronounced, resulting in similar posteriors and thus ability to detect tides, c.f., yellow, green, and magenta contours in Figure 18.5, left to right.

Going beyond $\tilde{\Lambda}$, we turn to the tidal deformability of the individual binary components. The second row of Figs. 18.4 and 18.5 shows posteriors for $\Lambda_1 - \Lambda_2$. The posteriors span much of the prior and show a strong anticorrelation consistent with [142, 15, 9]. The direction of the anticorrelation is approximately a constant $\tilde{\Lambda}$ suggesting that almost all tidal information comes from measuring $\tilde{\Lambda}$, with limited higher-order information [563, 220]. This is further demonstrated in Appendix 18.C. Consequently, $\Lambda_1 - \Lambda_2$ (second row) does not offer much additional information about the source type beyond $q - \tilde{\Lambda}$ (first row): exclusion of $\tilde{\Lambda} = 0$

amounts to exclusion of $\Lambda_1 = \Lambda_2 = 0$. Crucially for source classification, all component tidal deformabilities are individually consistent with $\Lambda_i = 0$.[9] Effectively, a $\tilde{\Lambda}$ measurement is "spread" between Λ_1 and Λ_2 and the posterior for *both* parameters is consistent with high values when *either parameter* has a high true value.

In the final row of Figs. 18.4 and 18.5, we show the component $\Lambda_i - m_i$ posteriors, where gray lines are draws from the EoS posterior. As expected from the second row, even in cases where $\tilde{\Lambda} = 0$ is confidently ruled out, the posteriors are consistent with $\Lambda_i = 0$. More information can however be obtained by comparing the *upper limit* on Λ_i to EoS expectations at the relevant mass. As expected, all BNS posteriors (magenta) are consistent with the EoS draws in both (m_1, Λ_1) and (m_2, Λ_2). Switching to the NSBH signals (yellow), the primary is always consistent with being a NS: for all masses nearly all the EoS draws fall within the yellow posteriors. In contrast and again for all mass configurations, about half the EoS draws fall within the posterior for the secondary binary component, indicating decreasing support for a NS interpretation. Interestingly, this is despite the fact that the upper limit on Λ_1 is lower than that of Λ_2. The expected tidal deformability increases so rapidly for lower masses that Λ_1 is more consistent with the EoSs than Λ_2. The BHNS posteriors (green contours in Figure 18.5) fully overlap with the EoS draws for all masses. This is because BHNSs have a larger $\tilde{\Lambda}$ than NSBHs for the same mass, pushing all upper limits to high enough values that are consistent with EoS predictions.

Finally, for BBH signals the posteriors for both components show some tension with EoS draws, which decreases with the total mass, c.f., blue contours of Figure 18.4, left to right. For the lowest mass configuration, c.f., left-most panel of Figure 18.4, neither binary component overlaps with hardly any EoS draw. In these cases, the GW data can constrain the tides to values that are too low compared to viable EoSs. The binary mass ratio, on the other hand, does not strongly impact the overlap between the posterior and the EoSs, c.f., blue contours in Figure 18.5, left to right. This is because the Λ_i posterior does not strongly depend on the system mass, what changes is the EoS prediction which is a strong function of the total mass.

18.4 Determining if a System Contains a Black Hole

Astronomical observations and nuclear physics considerations cannot directly motivate the nature of potential $\lesssim 1.2\,M_\odot$ GW detections such as the ones studied in

[9] The only seeming exception is the lowest-mass BNS in Figure 18.4 but this is due to a posterior railing against the prior upper bound.

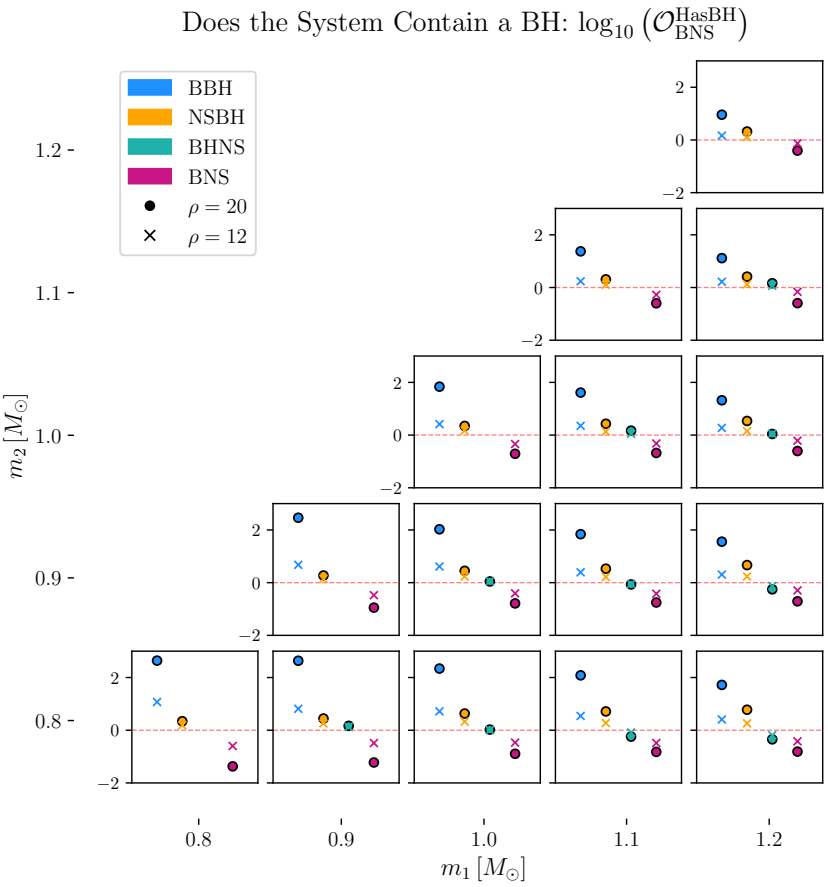

Figure 18.6: Base-10 logarithm of the odds ratio for each system containing at least one BH. Monte-Carlo errors for the odds ratios are too small to be visible in the scale of the figure. Panels correspond to the system source-frame masses and colors correspond to source type. The equal-mass panels do not contain BHNS systems as they are identical to the NSBH ones. Dots (crosses) denote signals with SNR 20(12). Points above $\log_{10}(\mathcal{O}_{\mathrm{BNS}}^{\mathrm{HasBH}}) = 0$ (red dashed line) denote support for the presence of at least one BH in the binary.

Sec. 18.3. We undertake signal classification with the fundamental question: does the signal provide evidence for the *presence* of a BH, thus establishing the existence of BHs below the expected astrophysical minimum mass?

We quantify this with the odds ratio $\mathcal{O}^{\rm HasBH}_{\rm BNS}$, where the "HasBH" hypothesis consists of the BBH, NSBH, and BHNS source types with equal prior probabilities. The alternative hypothesis is that the system is a BNS and thus the inferred masses and tides of both objects must be consistent with the EoS. In practice, the test comes down to whether the upper bound on the tidal effects is constraining enough to be in tension with the EoS prediction. We present the base-10 logarithm of the odds ratio, $\log_{10}\mathcal{O}^{\rm HasBH}_{\rm BNS}$ in Figure 18.6 for the $\rho = 20$ (solid dots) and the $\rho = 12$ signals (crosses). Below we focus on the $\rho = 20$ results; we obtain qualitatively similar though weaker constraints when $\rho = 12$.

The BBH signals (blue) show evidence for the presence of a BH, with odds $\gtrsim$10:1 for all masses. The evidence is stronger for lower-mass systems, with the odds ratio increasing from 10:1 to 100:1 between masses $1.2-1.2\,M_\odot$ and $0.8-0.8\,M_\odot$. This can be understood in the context of the EoS predictions; even though the Λ posteriors are similar for all masses, c.f., blue contours in Figure 18.4, bottom row, right to left, the EoS predicts that less massive NSs have much higher Λ values. As the mass decreases, the EoS predictions move away from the (m, Λ) posterior support; this brings the data from less massive systems into more tension with the BNS hypothesis.

NSBH signals (yellow) result in odds ranging between a few to $\sim 10:1$. For a given m_2, as m_1 increases (left to right), the odds ratio increases and we can more confidently infer the presence of a BH. This happens because both the true and inferred value of $\tilde{\Lambda}$ are smaller as m_1 increases. Both Λ_1 and Λ_2 are thus inferred to be smaller, but the estimate for m_2 is essentially unchanged, therefore, the secondary becomes more consistent with being a BH as the primary mass increases. This contrasts with the case of increasing the total mass at constant mass ratio (bottom left to top right) where the inferred value of Λ_2 decreases and the inferred value of m_2 increases, so consistency with EoS predictions remains unchanged.

Turning to the BHNS signals (green), we obtain near-equal odds for the presence of a BH for all masses. This is likely due to the larger tidal effects compared to the NSBH case (since now the secondary is a NS) and the corresponding higher upper limits on tidal parameters, *c.f.*, Figure 18.5, allowing both objects to agree with the EoS predictions. The odds for the presence of a BH decrease as the primary (BH)

mass increases (left to right), as BHs and NSs become less distinguishable.

Finally, BNSs (magenta) always yield evidence against the presence of a BH, which decreases with the mass.

18.5 Determining the Neutron Star Content of a System

The complementary question is whether a system contains at least one NS and if yes, whether it contains two. Here, the evidence comes from both consistency of each object with EoS predictions and the exclusion of $\tilde{\Lambda} = 0$.

18.5.1 Does the System Contain a Neutron Star?

The evidence for whether there is at least one NS in a system is quantified with the odds ratio $\mathcal{O}^{\mathrm{HasNS}}_{\mathrm{BBH}}$, Eq. (18.7). This is not equivalent to solely determining if the binary contains any matter; we further require the inferred tidal deformabilities to be consistent with the EoS.

In Figure 18.7, we show $\log_{10} \mathcal{O}^{\mathrm{HasNS}}_{\mathrm{BBH}}$. We again focus on the $\rho = 20$ results as $\rho = 12$ gives qualitatively similar, though less constraining, conclusions. The log odds ratios for BBHs are negative, indicating that the data favor the absence of any NSs. As the mass decreases, so does the odds ratio from $\mathcal{O}^{\mathrm{HasNS}}_{\mathrm{BBH}} \approx 1/50$ for 1.2–1.2 $M_\odot$ to $\approx 2/3$ for 0.8–0.8 $M_\odot$. It becomes less plausible for the lowest-mass BBH systems to contain a NS as the signals lack the strong tidal signature that the EoSs predict for these masses, c.f., blue contours in Figure 18.4 bottom left compared to bottom right panel. All NS-containing systems yield $\log_{10} \mathcal{O}^{\mathrm{HasNS}}_{\mathrm{BBH}} > 0$ though again the evidence decreases as the NS mass increases. For example, the odds ratio for $m_1 = 1.2\,M_\odot$, $m_2 = 0.8\,M_\odot$ is $\mathcal{O}^{\mathrm{HasNS}}_{\mathrm{BBH}} \approx 4$, much lower than the $m_1 = m_2 = 0.8\,M_\odot$ case which has $\mathcal{O}^{\mathrm{HasNS}}_{\mathrm{BBH}} > 100$. At all masses, there is more evidence for a NS in BHNSs than NSBHs. This is because the predicted tidal deformability of the primary is smaller than for the secondary, and thus a NS primary is more indistinguishable from a BH than a NS secondary. For systems containing exactly one $\lesssim 1\,M_\odot$ NS, we obtain $\mathcal{O}^{\mathrm{HasNS}}_{\mathrm{BBH}} \gtrapprox 10$. The strongest evidence is obtained for the presence of a NS in the BNS systems, all of which have $\log_{10} \mathcal{O}^{\mathrm{HasNS}}_{\mathrm{BBH}} \gtrapprox 2$. This is consistent with the BNS posteriors of Figure 18.4 and 18.5 that always rule out $\tilde{\Lambda} = 0$.

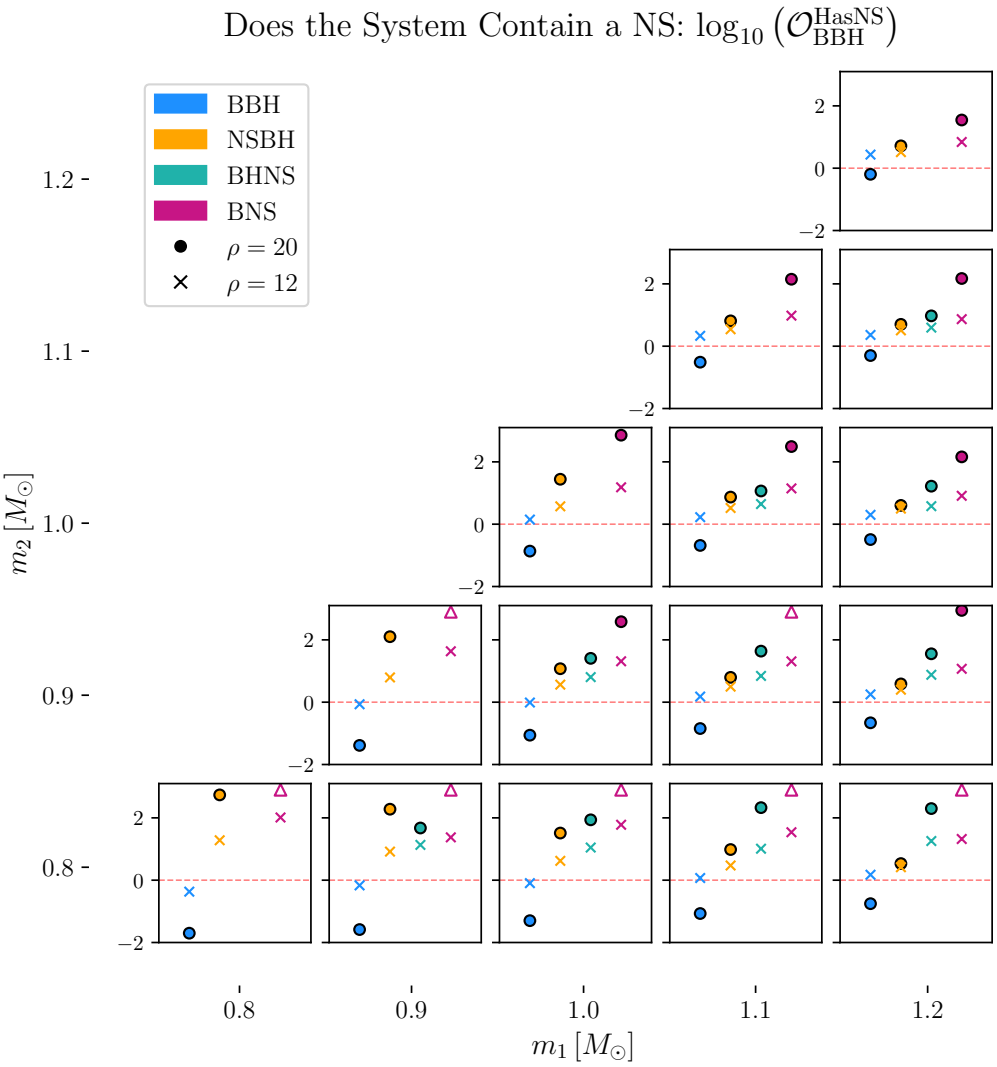

Figure 18.7: Similar to Figure 18.6 but for the odds ratio for each system containing at least one NS. Points above $\log_{10}(\mathcal{O}^{\mathrm{HasNS}}_{\mathrm{BBH}}) = 0$ (red dashed line) denote support for the presence of at least one NS in the binary. Triangular markers indicate that the odds ratio lies somewhere above the y-axis limit.

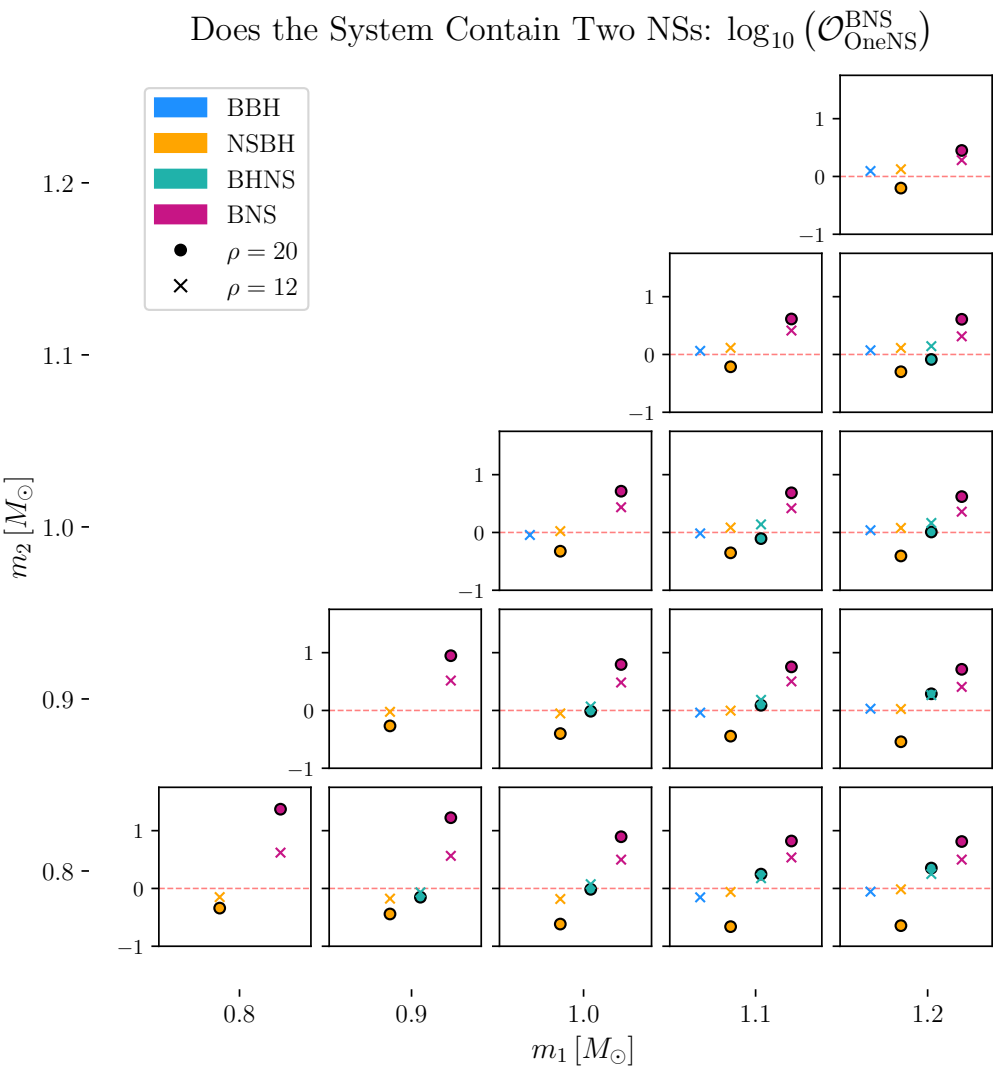

Figure 18.8: Similar to Figure 18.6 but for the odds ratio for each system containing exactly two NSs versus one NS. We only present results for systems with evidence of at least one NS in Figure 18.7 which includes all NS-containing systems. Points above $\log_{10}(O^{\mathrm{BNS}}_{\mathrm{OneNS}}) = 0$ (red dashed line) correspond to systems that are more likely to have two NSs than one.

18.5.2 Does the System Contain Two Neutron Stars?

Having established the presence of a NS, the next question is whether the source is a BNS or it contains only one NS. We compare these two hypotheses with the odds ratio $\mathcal{O}^{\rm BNS}_{\rm OneNS}$.

We show results in Figure 18.8, restricting to systems with evidence for at least one NS in Figure 18.7 which in practice is all the NS-containing systems and a few BBHs with marginal evidence. We again focus on the $\rho = 20$ results. BNS signals (pink) favor the presence of two NSs for all masses. As before, this evidence is stronger for less massive systems with odds $\gtrapprox 10:1$ when both components are $\lesssim 1\,M_\odot$. NSBHs (yellow) provide stronger evidence against the presence of two NSs than BHNSs. This is again because determining the nature of the secondary (least massive) is easier than primary (most massive) component.

However, neither BHNS nor NSBH signals result in odds greater than 10:1 against the BNS hypothesis; the strongest evidence is obtained for the 1.2–0.8 $M_\odot$ NSBH binary with $\mathcal{O}^{\rm BNS}_{\rm OneNS} \sim 1/8$. The reason refers back to the posteriors in Figs. 18.4 and 18.5. The BNS hypothesis requires that the EoS draws overlap with both the (m_1, Λ_1) and (m_2, Λ_2) posteriors. The bottom row of Figs. 18.4 and 18.5 show that the EoS draws completely overlap the primary posterior for all NSBH (yellow) and BHNS (green) signals. What is more, the posterior for the secondary is also fully (BHNS; green) or partially (NSBH; yellow) consistent with the EoS draws.

18.5.3 If the System Contains One Neutron Star, is it the primary or the secondary?

Though establishing the presence of exactly one NS is challenging at current sensitivity, we look forward to higher-SNR signals and consider how to identify which binary component it is. Most analyses label objects based on relative mass, e.g., primary and secondary, hence the most straightforward approach is to examine whether the primary is a NS or a BH:

$$\mathcal{O}^{\rm NSBH}_{\rm BHNS} = \frac{\mathcal{Z}_{\rm NSBH}}{\mathcal{Z}_{\rm BHNS}}\,. \tag{18.8}$$

However, this suffers from a labeling ambiguity. For example, an equal-mass NSBH system is equally-well described by assigning the tides on either component. This is due to the ambiguity in distinguishing binary components based on a property that is symmetric, i.e., the mass, and also plagues the component spins [99].

This ambiguity can be resolved by instead labeling the binary components with a unique property of each object that breaks this symmetry. For example, labeling binary components based on their tidal deformability would allow us to explore the properties of the stiffer and softer objects that reflect the NS and BH, respectively. Such an approach is of course only applicable for systems with *measurable* tidal asymmetry. For example, for BNSs, this approach would identify a "stiff" and a "soft" component, even if the tidal deformabilities are similar. More generally, there is no guarantee that objects are in fact distinguishable, e.g., an equal-mass and nonspinning BBH, there is thus no generic strategy for extracting individual component properties.

18.6 Conclusions

We have explored source classification for low-mass, $\leq 1.2\,M_\odot$ compact binary mergers based on the GW signal they emit and external information about the dense-matter EoS. The classification is based on the fact that the inferred component mass and tidal deformability must be consistent with EoS expectations if the object is a NS. A tidal measurement that is inconsistent with EoS predictions provides evidence that the object is not a NS, while $\Lambda = 0$ provides evidence for the object being a BH. The method's distinguishing power increases with decreasing mass, due to the fact that EoS predictions are a steep function of the mass, $\Lambda \sim m^{-6}$, and NSs become indistinguishable to BHs as the mass increases. Similarly, distinguishability is easier if the true EoS is stiffer as it would predict larger NS tidal deformabilities for all masses; here we have considered SLy9 that is consistent with the astrophysical data we employ.

We generally find it is easier to confirm the presence of a BH or NS than to refute it. For systems with subsolar-mass BHs, their presence can be identified at SNR $\rho = 20$. In contrast, BNSs strongly disfavor the presence of a BH, with the evidence growing with decreasing masses. Complimentarily, signals from $\lesssim 1\,M_\odot$ NS-containing binaries can reveal the NS presence based on compatibility of the mass-tidal measurement with EoS predictions. In contrast, if the binary *does not* contain a NS, its presence is disfavored with the evidence again growing as the mass decreases. Finally, identifying which object in a binary is a NS (or a BH) is subject to a labeling ambiguity that could be mitigated by labeling components based on relative tides rather than mass. Higher-SNR signals due to detector upgrades [16] or tighter EoS predictions thanks to future data will further strengthen distinguishability.

If subsolar-mass binaries exist and merge, combined mass and tidal information can aid in identifying the component nature and lead to constraints on primordial BH and NS physics. This prospect further motivates numerical simulations [367] and developing waveform models that can faithfully capture the large tidal effects of low-mass NSs. It further motivates studies of alternative possibilities to BHs and standard NSs such as dark matter admixed NSs with lower tidal deformability [284]. Tidal-based classification, as previously explored for higher-mass objects such as GW170817 [15, 198, 19], is especially promising for sub-solar mass objects whose nature is not otherwise astrophysically informed.

As this study was nearing completion, a preprint [160] that reached similar conclusions about the distinguishability of sub-solar mass BNS systems from BBHs appeared. Our methods differ in a few ways. The authors of [160] use Fisher matrix estimates (complemented with select full parameter estimation) and a modified TaylorF2 approximant to account for NS disruption, as compared to our use of full parameter estimation (with priors that keep Λ_1 and Λ_2 positive) with the NRTidalv3 waveform that includes appropriate termination conditions. Classification also differs: while Ref. [160] compares the upper limits on tidal inference to a fixed NS EoS, we form relevant hypotheses and marginalize over current uncertainty in the EoS to compute odds ratios. Additionally, we consider mixed NS-BH binaries, as opposed to only BNS and BBH systems. On the other hand, Ref. [160] also considers exotic compact objects. Regardless, both studies find that we can tell apart a sub-solar mass BBH from a BNS at SNR $\gtrsim 12$.

Acknowledgements

We thank Kareem El-Badry for helpful discussions about astronomical observations of compact objects. We also thank Jocelyn Read for useful comments on the manuscript. I.L. and K.C. acknowledge support from the Department of Energy under award number DE-SC0023101 and the Sloan Foundation. J.G. acknowledges funding from NSF Grant PHY-2207758. The project was supported by the European Union (ERC, SMArt, 101076369). Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. The authors are grateful for computational

resources provided by the LIGO Laboratory and supported by National Science Foundation Grants PHY-0757058 and PHY-0823459.

Software: `bilby` [67, 470], `dynesty` [507], `scipy` [556], `numpy` [270], `matplotlib` [289], `lwp` [196].

Appendix 18.A Limiting Frequencies

Compact binary inspirals terminate when the objects merge, disrupt each other, or their surfaces contact. In this Appendix, we quantify how compact binary components need to be in order to avoid disruption and contact and thus emit GWs in the sensitive band of ground-based detectors, see, e.g., [588] for a similar calculation.

The onset of merger is not precisely defined, but a separation of $r = 6M = 6(m_1+m_2)$ gives an order-of-magnitude estimate and a Keplerian frequency

$$f_{6M} = \frac{1}{\pi}\sqrt{\frac{M}{(6M)^3}}\,, \tag{18.9}$$

plotted in Figure 18.2; for $m_1 = m_2 = 1\,M_\odot$, $f_{6M} \sim 2\,\mathrm{kHz}$. Solar-mass compact objects therefore enter the LIGO-Virgo sensitive band before merger.

However, finite sizes might terminate the inspiral earlier if the objects contact each other before reaching $r = 6M$. For objects with radii R_1 and R_2, contact $r = R_1 + R_2$ occurs at a Keplerian frequency

$$f_{\rm cont} = \sqrt{\frac{G(m_1+m_2)}{4\pi^2(R_1+R_2)^3}}\,, \tag{18.10}$$

also plotted in Figure 18.2. For a BNS with $m_1 = m_2 = 1\,M_\odot$ and $R_1 = R_2 = 12\,\mathrm{km}$, $f_{\rm cont} \sim 1.5\,\mathrm{kHz}$. But for a NS-WD binary with an Earth-sized WD, $f_{\rm cont} \sim 0.2\,\mathrm{Hz}$, two orders of magnitude below the relevant frequency band.

Another possibility that prematurely ends an inspiral is disruption. The Newtonian tidal force felt by the secondary binary component due to the primary is

$$F_{21} = \frac{Gm_1m_2(r_{\rm out}-r_{\rm in})(r_{\rm out}+r_{\rm in})}{(r_{\rm out}r_{\rm in})^2}\,, \tag{18.11}$$

where $r_{\rm in} = r - R_2/2$ and $r_{\rm out} = r + R_2/2$ correspond to the distance between the primary and the outer and inner edge of the secondary, respectively. In the limit of wide orbital separation, $r \gg R_2$, Eq. (18.11) simplifies to

$$F_{21} \approx \frac{2Gm_1m_2R_2}{r^3}\,. \tag{18.12}$$

The secondary disrupts when F_{21} is comparable to its gravitational binding (self-)force

$$F_{21} \approx \frac{Gm_2^2}{R_2^2}\,, \tag{18.13}$$

which occurs at

$$r \approx \left(2\frac{m_1 R_2^3}{m_2}\right)^{1/3} , \tag{18.14}$$

corresponding to a Keplerian orbital frequency of

$$f_{\rm dis} \approx \sqrt{\frac{G m_2 (m_1 + m_2)}{8\pi^2 m_1 R_2^3}} . \tag{18.15}$$

Therefore,

$$\left(\frac{f_{\rm dis}}{f_{\rm cont}}\right)^2 = \frac{m_2 (R_1 + R_2)^3}{2 m_1 R_2^3} . \tag{18.16}$$

For compact objects with comparable radii and masses, $f_{\rm dis} \approx 2 f_{\rm cont}$ and thus the binary contacts before disruption. For a highly compact primary, for example a NS-WD binary with $R_1 \ll R_2$, $f_{\rm dis} < f_{\rm cont}$ and thus the binary disrupts before contact. In any case, for binaries involving WDs, both of these frequencies are well below the LIGO sensitive band.

Appendix 18.B Injection Properties

In this Appendix we provide more details for the parameter estimation analysis of Sec. 18.2.1. In Table Appendix 18.B we list the extrinsic parameters of the simulated signals. We select the luminosity distance unique to each system by scaling it to reach a target SNR, either 20 or 12.

For the single-event analyses, we sample the parameter posterior using Dynesty [507] as implemented in Bilby [67, 470], with a prior that is uniform in component detector-frame masses and aligned spin components. We adopt standard isotropic priors for position and inclination parameters, and a luminosity distance prior that is uniform in comoving volume [470]. The prior on the component tidal deformabilities is uniform and ranges from $\Lambda = 0$ to $\Lambda = 2 \times 10^4$, the maximum value the waveform was validated on [4]. In some cases, the Λ posterior distribution rails against this upper limit, but the simulated values for Λ are always within in the prior bounds.

We use a multibanding likelihood [384] and analyze 512 or 256 s of data (depending on the mass) at 8 kHz with lower and upper frequency cutoffs of 20 Hz and 3.5 kHz, respectively. The upper cutoff is above the inherent waveform termination [4, 258].

Parameter	Label	Value
Phase at 20 Hz	ϕ	0.24 rad
Right Ascension	α	0.18 rad
Declination	δ	0.62 rad
Inclination	ι	2.7 rad
Polarization Angle	ψ	0.58 rad
Merger time at geocenter	t_c	0 sec (GPS)

Table 18.1: Values for extrinsic parameters used for simulating the data.

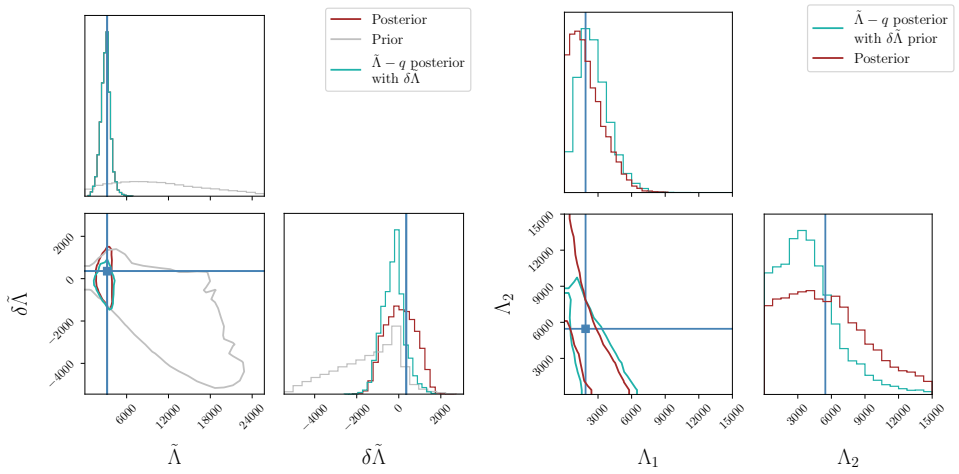

Figure 18.9: Marginal posterior (in brown) for tidal parameters from the BNS signal with $(m_1, m_2) = (1.1, 0.9)\, M_\odot$. (Left) Tidal parameters $\tilde{\Lambda}$ and $\delta\tilde{\Lambda}$, with the prior plotted in grey. (Right) Component tidal deformabilties Λ_1 and Λ_2. In both panels, the turquoise distribution corresponds to the posterior assuming that there is no information about $\delta\tilde{\Lambda}$. We find that information about $\delta\tilde{\Lambda}$ is nonnegligible, though insufficient to break the degeneracy between Λ_1 and Λ_2.

Appendix 18.C Impact of measurements of $\delta\tilde{\Lambda}$

In order to constrain the component tidal deformabilities, measurement of an additional parameter beyond $\tilde{\Lambda}$, such as $\delta\tilde{\Lambda}$, is required. The parameter $\delta\tilde{\Lambda}$ represents the tidal contributions to the frequency-domain phase which appear at 6PN and are not proportional to $\tilde{\Lambda}$; intuitively it is a measure of the asymmetry in the tidal contributions from the two components [563]. We examine the impact of the constraints on $\delta\tilde{\Lambda}$ in the tidal parameters from the $(m_1, m_2) = (1.1, 0.9)\, M_\odot$ BNS signal in Figure 18.9. In the left panel, we present the induced prior, see Sec. 18.2.1, and the recovered marginal posterior for $\tilde{\Lambda}$ and $\delta\tilde{\Lambda}$. We obtain a symmetric 90% credible interval for $\tilde{\Lambda} \in (1804, 4131)$ with respect to a prior that covers $0 < \tilde{\Lambda} \lesssim 26000$.

In order to break the degeneracy between Λ_1 and Λ_2, we must measure additional parameters. However, $\delta\tilde{\Lambda}$ is relatively poorly measured at current sensitivity. The left panel of Figure 18.9 shows that, even though the 1-d marginal posterior for $\delta\tilde{\Lambda}$ (red) appears to be well constrained relative to the prior (gray), this is primarily driven by $\tilde{\Lambda}$, c.f., the $2-d$ marginal posterior.

In order to investigate how information about $\delta\tilde{\Lambda}$ impacts the component tidal deformabilities, we approximate an inference where no information about $\delta\tilde{\Lambda}$ exists. We draw $(q, \tilde{\Lambda})$ samples from the full posterior and combine them with samples of $\delta\tilde{\Lambda}$ from its effective prior implied by the given $(q, \tilde{\Lambda})$, subject to the condition $\Lambda_i(q, \tilde{\Lambda}, \delta\tilde{\Lambda}) > 0$. We display the marginal distribution in the left panel panel of Figure 18.9 (teal). We compare this to the full marginal posterior on $\Lambda_1 - \Lambda_2$ (red). We find that while knowledge of $\delta\tilde{\Lambda}$ does change the distribution on Λ_1–Λ_2, this information does not substantially change the correlation structure. As expected for a well-measured parameter, this procedure leaves the $\tilde{\Lambda}$ posterior unaffected (left). The measurement of $\delta\tilde{\Lambda}$ itself favors higher values of $\delta\tilde{\Lambda}$ (left), which correspond to higher values of Λ_2 and lower values for Λ_1 (right).

Part V

Conclusions

Chapter 19

CONCLUSIONS

Being a part of a new field has been a particularly exciting journey for me, largely because of the opportunity to contribute to the discovery science. I have been fortunate to work at the forefront of gravitational wave physics, a field whose scientific output has grown tremendously during my graduate years. Gravitational waves are a fascinating messenger from deep space, due to their largely unfamiliar nature and unique effects. On top of that, I have been fortunate enough to be able to use these novel tools to study some of the most interesting objects in the universe, so far removed from human experience: black holes and neutron stars. I have only been able to do this thanks to extraordinary developments of instrumentation and data analysis techniques over the past several decades that have recently brought us into the era of gravitational wave astronomy. The field is still in its infancy, and there are many exciting opportunities for future work.

Work in this field demands expertise across many areas. To measure source properties and link them to underlying physics, I have had to learn and refine my knowledge of topics ranging from general relativity and astrophysics to applied mathematics and statistics.

In this thesis, I have explored some of the topics I have focused on in graduate school, weaving three intertwined threads: event-level inference, population constraints, and effects of dense-matter physics.

After reviewing some of the theory behind gravitational waves, I discussed measuring the properties of compact objects involved in the mergers we detect. In Chapter 3, I provided an overview of the Bayesian parameter estimation techniques and waveform models we employ to make the measurements. I then turned to science results I have been contributing to during the fourth observing run, and, in Chapter 4, presented brand new results from observations in the fourth Gravitational Wave Transient Catalog (GWTC4), containing detailed analysis of the properties of almost 100 compact binary mergers. Results from GWTC4 are set to be released three months from now, in August of 2025. In Chapter 5, I presented a case study demonstrating how neglecting subtleties in the data quality can lead to incorrect conclusions from the analysis of a gravitational wave signal.

With an understanding of the physical properties of each of our sources and the sensitivity of our instruments, we can learn a lot of astrophysics and cosmology by combining observations and studying them on a population level. I provided an overview in Chapter 6 of population inference, particularly including the use of hierarchical Bayesian analysis for measuring global properties from the distribution of compact objects. In Chapter 7 and Chapter 8, I presented the astrophysical distributions and population properties of binary black holes observed through GWTC3 and GWTC4, respectively. These results revealed several novel aspects of the binary black hole population, including surprising features in the mass distribution such as peaks in unexpected places, as well as the evolution of the merger rate across redshifts. These findings refine our picture of the astrophysical processes and channels that underpin the formation and evolution of binary black hole systems.

I included a brief description of a population inference software library I have been involved in developing in Chapter 9; these tools have been employed for obtaining headline results in LVK collaboration studies, as well as results presented in my own work and from many other authors. Then, in Chapter 10, I presented a study of the spin distribution of binary black holes with relaxed prior assumptions for the shape of the distribution. I showed the that some spin-distribution features may arise from model misspecification or systematic biases. I then related these back to the astrophysical conclusions we can draw. I followed this in Chapter 11 with work on directly relating population measurements to constraints on underlying physics, specifically applied to the pulsational pair instability mechanism relevant for high-mass stellar progenitors of binary black hole systems. This work demonstrated how features from underlying stellar processes manifest as signatures in the binary black hole mass distribution, and cast serious doubt on the idea that the overdensity of binary black holes at $\sim 30 M_\odot$ is a signature of pulsational pair instability in stellar progenitors.

I applied population-level constraints to other questions in astrophysics, including projections of the gravitational wave stochastic background and large-scale structure of the Universe in Chapters 12 and 13, respectively. I concluded this part in Chapter 14 with a detailed investigation into the biases associated with the Monte Carlo methods we use to infer population properties, and a brief presentation in Chapter 15 evaluating potential biases from incorrect estimation of detector sensitivities—crucial considerations for robust inference as the catalog grows.

Finally, I turned to both exploiting and learning properties of dense matter in

gravitational wave observations of neutron stars. In Chapter 16, I demonstrated the need for simultaneous inference of the nuclear equation of state and neutron star mass distribution and presented a new technique to do so. I applied this technique with real data, including from radio and X-ray observations of neutron stars, in Chapter 17, obtaining constraints on the behavior of nuclear matter and the distribution of neutron star masses, and investigating how these two properties may be related. I demonstrated in Chapter 18 how we can use our knowledge of properties of dense nuclear matter to determine whether the compact objects in a low-mass merger are neutron stars or black holes. This result showed that, at reasonable signal-to-noise ratios, we can confirm or rule out a neutron-star component in sub-solar-mass binaries.

The future of this new, dark side of astrophysics is bright. The data currently being collected in O4 and in the next observing run will contain numerous observations that will dramatically increase our understanding of properties of black holes and neutron stars. Furthermore, as the ever-expanding catalog grows, we will refine our knowledge of astrophysical context of compact binaries, fundamental physics, and cosmology. The methods and results in this thesis will be applied and refined by others in the field in future generations of gravitational wave discoveries.

BIBLIOGRAPHY

[1] J. Aasi et al. "Advanced LIGO". In: *Class. Quant. Grav.* 32 (2015), p. 074001. DOI: 10.1088/0264-9381/32/7/074001. arXiv: 1411.4547 [gr-qc].

[2] A. G. Abac et al. "Observation of Gravitational Waves from the Coalescence of a 2.5–4.5 $M_\odot$ Compact Object and a Neutron Star". In: *Astrophys. J. Lett.* 970.2 (2024), p. L34. DOI: 10.3847/2041-8213/ad5beb. arXiv: 2404.04248 [astro-ph.HE].

[3] A. G. Abac et al. "Population Properties of Merging Compact Objects from the Fourth LIGO-Virgo-KAGRA Gravitational-Wave Transient Catalog". In: *Astrophys. J.* (2025 in prep.).

[4] Adrian Abac, Tim Dietrich, Alessandra Buonanno, et al. "New and robust gravitational-waveform model for high-mass-ratio binary neutron star systems with dynamical tidal effects". In: *Phys. Rev. D* 109.2 (2024), p. 024062. DOI: 10.1103/PhysRevD.109.024062. arXiv: 2311.07456 [gr-qc].

[5] B. P. Abbott et al. "A Gravitational-wave Measurement of the Hubble Constant Following the Second Observing Run of Advanced LIGO and Virgo". In: *Astrophys. J.* 909.2 (2021), p. 218. DOI: 10.3847/1538-4357/abdcb7. arXiv: 1908.06060 [astro-ph.CO].

[6] B. P. Abbott et al. "Binary Black Hole Mergers in the first Advanced LIGO Observing Run". In: *Phys. Rev. X* 6.4 (2016). [Erratum: Phys.Rev.X 8, 039903 (2018)], p. 041015. DOI: 10.1103/PhysRevX.6.041015. arXiv: 1606.04856 [gr-qc].

[7] B. P. Abbott et al. "Binary Black Hole Population Properties Inferred from the First and Second Observing Runs of Advanced LIGO and Advanced Virgo". In: *Astrophys. J. Lett.* 882.2 (2019), p. L24. DOI: 10.3847/2041-8213/ab3800. arXiv: 1811.12940 [astro-ph.HE].

[8] B. P. Abbott et al. "Characterization of transient noise in Advanced LIGO relevant to gravitational wave signal GW150914". In: *Class. Quant. Grav.* 33.13 (2016), p. 134001. DOI: 10.1088/0264-9381/33/13/134001. arXiv: 1602.03844 [gr-qc].

[9] B. P. Abbott et al. "GW170817: Measurements of neutron star radii and equation of state". In: *Phys. Rev. Lett.* 121.16 (2018), p. 161101. DOI: 10.1103/PhysRevLett.121.161101. arXiv: 1805.11581 [gr-qc].

[10] B. P. Abbott et al. "GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral". In: *Phys. Rev. Lett.* 119.16 (2017), p. 161101. DOI: 10.1103/PhysRevLett.119.161101. arXiv: 1710.05832 [gr-qc].

[11] B. P. Abbott et al. "GW190425: Observation of a Compact Binary Coalescence with Total Mass $\sim 3.4 M_{\odot}$". In: *Astrophys. J. Lett.* 892.1 (2020), p. L3. DOI: 10.3847/2041-8213/ab75f5. arXiv: 2001.01761 [astro-ph.HE].

[12] 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". In: *Phys. Rev. X* 9.3 (2019), p. 031040. DOI: 10.1103/PhysRevX.9.031040. arXiv: 1811.12907 [astro-ph.HE].

[13] B. P. Abbott et al. "Multi-messenger Observations of a Binary Neutron Star Merger". In: *Astrophys. J. Lett.* 848.2 (2017), p. L12. DOI: 10.3847/2041-8213/aa91c9. arXiv: 1710.05833 [astro-ph.HE].

[14] B. P. Abbott et al. "Observation of Gravitational Waves from a Binary Black Hole Merger". In: *Phys. Rev. Lett.* 116.6 (2016), p. 061102. DOI: 10.1103/PhysRevLett.116.061102. arXiv: 1602.03837 [gr-qc].

[15] B. P. Abbott et al. "Properties of the binary neutron star merger GW170817". In: *Phys. Rev. X* 9.1 (2019), p. 011001. DOI: 10.1103/PhysRevX.9.011001. arXiv: 1805.11579 [gr-qc].

[16] B. P. Abbott et al. "Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA". In: *Living Rev. Rel.* 19 (2016), p. 1. DOI: 10.1007/s41114-020-00026-9. arXiv: 1304.0670 [gr-qc].

[17] B. P. Abbott et al. "Tests of General Relativity with the Binary Black Hole Signals from the LIGO-Virgo Catalog GWTC-1". In: *Phys. Rev. D* 100.10 (2019), p. 104036. DOI: 10.1103/PhysRevD.100.104036. arXiv: 1903.04467 [gr-qc].

[18] Benjamin P Abbott et al. "A guide to LIGO–Virgo detector noise and extraction of transient gravitational-wave signals". In: *Class. Quant. Grav.* 37.5 (2020), p. 055002. DOI: 10.1088/1361-6382/ab685e. arXiv: 1908.11170 [gr-qc].

[19] Benjamin P Abbott et al. "Model comparison from LIGO–Virgo data on GW170817's binary components and consequences for the merger remnant". In: *Class. Quant. Grav.* 37.4 (2020), p. 045006. DOI: 10.1088/1361-6382/ab5f7c. arXiv: 1908.01012 [gr-qc].

[20] R. Abbott et al. "All-sky, all-frequency directional search for persistent gravitational waves from Advanced LIGO's and Advanced Virgo's first three observing runs". In: *Phys. Rev. D* 105.12 (2022), p. 122001. DOI: 10.1103/PhysRevD.105.122001. arXiv: 2110.09834 [gr-qc].

[21] R. Abbott et al. "Constraints on the Cosmic Expansion History from GWTC–3". In: *Astrophys. J.* 949.2 (2023), p. 76. DOI: 10.3847/1538-4357/ac74bb. arXiv: 2111.03604 [astro-ph.CO].

[22] R. Abbott et al. “GW190521: A Binary Black Hole Merger with a Total Mass of $150M_{\odot}$”. In: *Phys. Rev. Lett.* 125.10 (2020), p. 101102. DOI: 10.1103/PhysRevLett.125.101102. arXiv: 2009.01075 [gr-qc].

[23] 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”. In: *Astrophys. J. Lett.* 896.2 (2020), p. L44. DOI: 10.3847/2041-8213/ab960f. arXiv: 2006.12611 [astro-ph.HE].

[24] R. Abbott et al. “GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run”. In: *Phys. Rev. X* 11 (2021), p. 021053. DOI: 10.1103/PhysRevX.11.021053. arXiv: 2010.14527 [gr-qc].

[25] 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”. In: *Phys. Rev. D* 109.2 (2024), p. 022001. DOI: 10.1103/PhysRevD.109.022001. arXiv: 2108.01045 [gr-qc].

[26] R. Abbott et al. “GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo during the Second Part of the Third Observing Run”. In: *Phys. Rev. X* 13.4 (2023), p. 041039. DOI: 10.1103/PhysRevX.13.041039. arXiv: 2111.03606 [gr-qc].

[27] R. Abbott et al. “Observation of Gravitational Waves from Two Neutron Star–Black Hole Coalescences”. In: *Astrophys. J. Lett.* 915.1 (2021), p. L5. DOI: 10.3847/2041-8213/ac082e. arXiv: 2106.15163 [astro-ph.HE].

[28] R. Abbott et al. “Population of Merging Compact Binaries Inferred Using Gravitational Waves through GWTC-3”. In: *Phys. Rev. X* 13.1 (2023), p. 011048. DOI: 10.1103/PhysRevX.13.011048. arXiv: 2111.03634 [astro-ph.HE].

[29] R. Abbott et al. “Population Properties of Compact Objects from the Second LIGO-Virgo Gravitational-Wave Transient Catalog”. In: *Astrophys. J. Lett.* 913.1 (2021), p. L7. DOI: 10.3847/2041-8213/abe949. arXiv: 2010.14533 [astro-ph.HE].

[30] R. Abbott et al. “Search for anisotropic gravitational-wave backgrounds using data from Advanced LIGO and Advanced Virgo’s first three observing runs”. In: *Phys. Rev. D* 104.2 (2021), p. 022005. DOI: 10.1103/PhysRevD.104.022005. arXiv: 2103.08520 [gr-qc].

[31] R. Abbott et al. “Search for Subsolar-Mass Binaries in the First Half of Advanced LIGO’s and Advanced Virgo’s Third Observing Run”. In: *Phys. Rev. Lett.* 129.6 (2022), p. 061104. DOI: 10.1103/PhysRevLett.129.061104. arXiv: 2109.12197 [astro-ph.CO].

[32] R. Abbott et al. “Search for subsolar-mass black hole binaries in the second part of Advanced LIGO’s and Advanced Virgo’s third observing run”. In: *Mon. Not. Roy. Astron. Soc.* 524.4 (2023). [Erratum: Mon.Not.Roy.Astron.Soc. 526, 6234 (2023)], pp. 5984–5992. DOI: 10.1093/mnras/stad588. arXiv: 2212.01477 [astro-ph.HE].

[33] R. Abbott et al. “Tests of general relativity with binary black holes from the second LIGO-Virgo gravitational-wave transient catalog”. In: *Phys. Rev. D* 103.12 (2021), p. 122002. DOI: 10.1103/PhysRevD.103.122002. arXiv: 2010.14529 [gr-qc].

[34] R. Abbott et al. “Tests of General Relativity with GWTC-3”. In: (Dec. 2021). arXiv: 2112.06861 [gr-qc].

[35] R. Abbott et al. “Upper limits on the isotropic gravitational-wave background from Advanced LIGO and Advanced Virgo’s third observing run”. In: *Phys. Rev. D* 104.2 (2021), p. 022004. DOI: 10.1103/PhysRevD.104.022004. arXiv: 2101.12130 [gr-qc].

[36] Rich Abbott et al. “Open data from the first and second observing runs of Advanced LIGO and Advanced Virgo”. In: *SoftwareX* 13 (2021), p. 100658. DOI: 10.1016/j.softx.2021.100658. arXiv: 1912.11716 [gr-qc].

[37] T. Accadia et al. “Noise from scattered light in Virgo’s second science run data”. In: *Class. Quant. Grav.* 27 (2010). Ed. by Fulvio Ricci, p. 194011. DOI: 10.1088/0264-9381/27/19/194011.

[38] F. Acernese et al. “Advanced Virgo: a second-generation interferometric gravitational wave detector”. In: *Class. Quant. Grav.* 32.2 (2015), p. 024001. DOI: 10.1088/0264-9381/32/2/024001. arXiv: 1408.3978 [gr-qc].

[39] F. Acernese et al. “Increasing the Astrophysical Reach of the Advanced Virgo Detector via the Application of Squeezed Vacuum States of Light”. In: *Phys. Rev. Lett.* 123.23 (2019), p. 231108. DOI: 10.1103/PhysRevLett.123.231108.

[40] F. Acernese et al. “Virgo Detector Characterization and Data Quality during the O3 run”. In: (May 2022). arXiv: 2205.01555 [gr-qc].

[41] T. Adams, D. Buskulic, V. Germain, et al. “Low-latency analysis pipeline for compact binary coalescences in the advanced gravitational wave detector era”. In: *Class. Quant. Grav.* 33.17 (2016), p. 175012. DOI: 10.1088/0264-9381/33/17/175012. arXiv: 1512.02864 [gr-qc].

[42] P. A. R. Ade et al. “Planck 2015 results. XIII. Cosmological parameters”. In: *Astron. Astrophys.* 594 (2016), A13. DOI: 10.1051/0004-6361/201525830. arXiv: 1502.01589 [astro-ph.CO].

[43] D. Adhikari et al. "Accurate Determination of the Neutron Skin Thickness of ^{208}Pb through Parity-Violation in Electron Scattering". In: *Phys. Rev. Lett.* 126.17 (2021), p. 172502. DOI: 10.1103/PhysRevLett.126.172502. arXiv: 2102.10767 [nucl-ex].

[44] D. Adhikari et al. "Precision Determination of the Neutral Weak Form Factor of Ca48". In: *Phys. Rev. Lett.* 129.4 (2022), p. 042501. DOI: 10.1103/PhysRevLett.129.042501. arXiv: 2205.11593 [nucl-ex].

[45] Michalis Agathos, Jeroen Meidam, Walter Del Pozzo, et al. "Constraining the neutron star equation of state with gravitational wave signals from coalescing binary neutron stars". In: *Phys. Rev. D* 92.2 (2015), p. 023012. DOI: 10.1103/PhysRevD.92.023012. arXiv: 1503.05405 [gr-qc].

[46] N. Aghanim et al. "Planck 2018 results. VI. Cosmological parameters". In: *Astron. Astrophys.* 641 (2020). [Erratum: Astron.Astrophys. 652, C4 (2021)], A6. DOI: 10.1051/0004-6361/201833910. arXiv: 1807.06209 [astro-ph.CO].

[47] P. Ajith et al. "A Template bank for gravitational waveforms from coalescing binary black holes. I. Non-spinning binaries". In: *Phys. Rev. D* 77 (2008). [Erratum: Phys.Rev.D 79, 129901 (2009)], p. 104017. DOI: 10.1103/PhysRevD.77.104017. arXiv: 0710.2335 [gr-qc].

[48] P. Ajith. "Addressing the spin question in gravitational-wave searches: Waveform templates for inspiralling compact binaries with nonprecessing spins". In: *Phys. Rev. D* 84 (2011), p. 084037. DOI: 10.1103/PhysRevD.84.084037. arXiv: 1107.1267 [gr-qc].

[49] P. Ajith et al. "Inspiral-merger-ringdown waveforms for black-hole binaries with non-precessing spins". In: *Phys. Rev. Lett.* 106 (2011), p. 241101. DOI: 10.1103/PhysRevLett.106.241101. arXiv: 0909.2867 [gr-qc].

[50] P. Ajith, N. Fotopoulos, S. Privitera, et al. "Effectual template bank for the detection of gravitational waves from inspiralling compact binaries with generic spins". In: *Phys. Rev. D* 89.8 (2014), p. 084041. DOI: 10.1103/PhysRevD.89.084041. arXiv: 1210.6666 [gr-qc].

[51] Parameswaran Ajith et al. "Phenomenological template family for black-hole coalescence waveforms". In: *Class. Quant. Grav.* 24 (2007). Ed. by B. Krishnan, M. A. Papa, and Bernard F. Schutz, S689–S700. DOI: 10.1088/0264-9381/24/19/S31. arXiv: 0704.3764 [gr-qc].

[52] A. Akmal, V. R. Pandharipande, and D. G. Ravenhall. "The Equation of state of nucleon matter and neutron star structure". In: *Phys. Rev. C* 58 (1998), pp. 1804–1828. DOI: 10.1103/PhysRevC.58.1804. arXiv: nucl-th/9804027.

[53] T. Akutsu et al. "KAGRA: 2.5 Generation Interferometric Gravitational Wave Detector". In: *Nature Astron.* 3.1 (2019), pp. 35–40. DOI: 10.1038/s41550-018-0658-y. arXiv: 1811.08079 [gr-qc].

[54] T. Akutsu et al. "Overview of KAGRA: Detector design and construction history". In: *PTEP* 2021.5 (2021), 05A101. DOI: 10.1093/ptep/ptaa125. arXiv: 2005.05574 [physics.ins-det].

[55] Bruce Allen. "χ^2 time-frequency discriminator for gravitational wave detection". In: *Phys. Rev. D* 71 (2005), p. 062001. DOI: 10.1103/PhysRevD.71.062001. arXiv: gr-qc/0405045.

[56] Bruce Allen, Warren G. Anderson, Patrick R. Brady, et al. "FINDCHIRP: An Algorithm for detection of gravitational waves from inspiraling compact binaries". In: *Phys. Rev. D* 85 (2012), p. 122006. DOI: 10.1103/PhysRevD.85.122006. arXiv: gr-qc/0509116.

[57] Bruce Allen, Wen-sheng Hua, and Adrian C. Ottewill. "Automatic cross talk removal from multichannel data". In: (Sept. 1999). arXiv: gr-qc/9909083.

[58] Justin Alsing, Hector O. Silva, and Emanuele Berti. "Evidence for a maximum mass cut-off in the neutron star mass distribution and constraints on the equation of state". In: *Mon. Not. Roy. Astron. Soc.* 478.1 (2018), pp. 1377–1391. DOI: 10.1093/mnras/sty1065. arXiv: 1709.07889 [astro-ph.HE].

[59] Warren G. Anderson, Patrick R. Brady, Jolien D. E. Creighton, et al. "An Excess power statistic for detection of burst sources of gravitational radiation". In: *Phys. Rev. D* 63 (2001), p. 042003. DOI: 10.1103/PhysRevD.63.042003. arXiv: gr-qc/0008066.

[60] Jeff J. Andrews, Julianne Cronin, Vicky Kalogera, et al. "Targeted Modeling of GW150914's Binary Black Hole Source with Dart_board". In: *Astrophys. J. Lett.* 914.2 (2021), p. L32. DOI: 10.3847/2041-8213/ac00a6. arXiv: 2011.13918 [astro-ph.HE].

[61] John Antoniadis et al. "A Massive Pulsar in a Compact Relativistic Binary". In: *Science* 340 (2013), p. 6131. DOI: 10.1126/science.1233232. arXiv: 1304.6875 [astro-ph.HE].

[62] John Antoniadis, Thomas M. Tauris, Feryal Ozel, et al. "The millisecond pulsar mass distribution: Evidence for bimodality and constraints on the maximum neutron star mass". In: (May 2016). arXiv: 1605.01665 [astro-ph.HE].

[63] Fabio Antonini and Mark Gieles. "Merger rate of black hole binaries from globular clusters: theoretical error bars and comparison to gravitational wave data from GWTC-2". In: *Phys. Rev. D* 102 (2020), p. 123016. DOI: 10.1103/PhysRevD.102.123016. arXiv: 2009.01861 [astro-ph.HE].

[64] Fabio Antonini, Mark Gieles, Fani Dosopoulou, et al. "Coalescing black hole binaries from globular clusters: mass distributions and comparison to gravitational wave data from GWTC-3". In: *Mon. Not. Roy. Astron. Soc.* 522.1 (2023), pp. 466–476. DOI: 10.1093/mnras/stad972. arXiv: 2208.01081 [astro-ph.HE].

[65] Theocharis A. Apostolatos, Curt Cutler, Gerald J. Sussman, et al. “Spin induced orbital precession and its modulation of the gravitational wave forms from merging binaries”. In: *Phys. Rev. D* 49 (1994), pp. 6274–6297. DOI: 10.1103/PhysRevD.49.6274.

[66] K. G. Arun, Bala R. Iyer, B. S. Sathyaprakash, et al. “Parameter estimation of inspiralling compact binaries using 3.5 post-Newtonian gravitational wave phasing: The Non-spinning case”. In: *Phys. Rev. D* 71 (2005). [Erratum: Phys.Rev.D 72, 069903 (2005)], p. 084008. DOI: 10.1103/PhysRevD.71.084008. arXiv: gr-qc/0411146.

[67] Gregory Ashton et al. “BILBY: A user-friendly Bayesian inference library for gravitational-wave astronomy”. In: *Astrophys. J. Suppl.* 241.2 (2019), p. 27. DOI: 10.3847/1538-4365/ab06fc. arXiv: 1811.02042 [astro-ph.IM].

[68] Gregory Ashton and Colm Talbot. “B ilby-MCMC: an MCMC sampler for gravitational-wave inference”. In: *Mon. Not. Roy. Astron. Soc.* 507.2 (2021), pp. 2037–2051. DOI: 10.1093/mnras/stab2236. arXiv: 2106.08730 [gr-qc].

[69] Gregory Ashton, Sarah Thiele, Yannick Lecoeuche, et al. “Parameterised population models of transient non-Gaussian noise in the LIGO gravitational-wave detectors”. In: *Class. Quant. Grav.* 39.17 (2022), p. 175004. DOI: 10.1088/1361-6382/ac8094. arXiv: 2110.02689 [gr-qc].

[70] Abbas Askar, Magdalena Szkudlarek, Dorota Gondek-Rosińska, et al. “MOCCA-SURVEY Database – I. Coalescing binary black holes originating from globular clusters”. In: *Mon. Not. Roy. Astron. Soc.* 464.1 (2017), pp. L36–L40. DOI: 10.1093/mnrasl/slw177. arXiv: 1608.02520 [astro-ph.HE].

[71] F. Aubin et al. “The MBTA pipeline for detecting compact binary coalescences in the third LIGO–Virgo observing run”. In: *Class. Quant. Grav.* 38.9 (2021), p. 095004. DOI: 10.1088/1361-6382/abe913. arXiv: 2012.11512 [gr-qc].

[72] Kareem El-Badry et al. “A population of neutron star candidates in wide orbits from Gaia astrometry”. In: (July 2024). DOI: 10.33232/001c.121261.

[73] Kareem El-Badry, Rhys Seeburger, Tharindu Jayasinghe, et al. “Unicorns and giraffes in the binary zoo: stripped giants with subgiant companions”. In: *Mon. Not. Roy. Astron. Soc.* 512.4 (2022), pp. 5620–5641. DOI: 10.1093/mnras/stac815. arXiv: 2203.06348 [astro-ph.SR].

[74] Kareem El-Badry, Joshua D. Simon, Henrique Reggiani, et al. “A 1.9 solar-mass neutron star candidate in a 2-year orbit”. In: *The Open Journal of Astrophysics* 7, 27 (Apr. 2024), p. 27. DOI: 10.33232/001c.116675. arXiv: 2402.06722 [astro-ph.SR].

[75] A. Makai Baker, Paul D. Lasky, Eric Thrane, et al. “Significant challenges for astrophysical inference with next-generation gravitational-wave observatories”. In: (Mar. 2025). arXiv: 2503.04073 [gr-qc].

[76] John Baker et al. “High angular resolution gravitational wave astronomy”. In: *Exper. Astron.* 51.3 (2021), pp. 1441–1470. DOI: 10.1007/s10686-021-09712-0. arXiv: 1908.11410 [astro-ph.HE].

[77] John G. Baker, Joan Centrella, Dae-Il Choi, et al. “Gravitational wave extraction from an inspiraling configuration of merging black holes”. In: *Phys. Rev. Lett.* 96 (2006), p. 111102. DOI: 10.1103/PhysRevLett.96.111102. arXiv: gr-qc/0511103.

[78] Sharan Banagiri, Vuk Mandic, Claudia Scarlata, et al. “Measuring angular N-point correlations of binary black hole merger gravitational-wave events with hierarchical Bayesian inference”. In: *Phys. Rev. D* 102.6 (2020), p. 063007. DOI: 10.1103/PhysRevD.102.063007. arXiv: 2006.00633 [astro-ph.CO].

[79] Sambaran Banerjee. “Binary black hole mergers from young massive clusters in the pair-instability supernova mass gap”. In: *Astron. Astrophys.* 665 (2022), A20. DOI: 10.1051/0004-6361/202142331. arXiv: 2109.14612 [astro-ph.HE].

[80] Sambaran Banerjee. “Merger rate density of stellar-mass binary black holes from young massive clusters, open clusters, and isolated binaries: Comparisons with LIGO-Virgo-KAGRA results”. In: *Phys. Rev. D* 105.2 (2022), p. 023004. DOI: 10.1103/PhysRevD.105.023004. arXiv: 2108.04250 [astro-ph.HE].

[81] Sambaran Banerjee. “Stellar-mass black holes in young massive and open stellar clusters – V. comparisons with LIGO-Virgo merger rate densities”. In: *Mon. Not. Roy. Astron. Soc.* 503.3 (2021), pp. 3371–3385. DOI: 10.1093/mnras/stab591. arXiv: 2011.07000 [astro-ph.HE].

[82] Kyle Barbary. *Nestle*. Version 0.2.0. 2016. URL: https://github.com/kbarbary/nestle.

[83] Ewan D. Barr et al. “A pulsar in a binary with a compact object in the mass gap between neutron stars and black holes”. In: *Science* 383.6680 (2024), pp. 275–279. DOI: 10.1126/science.adg3005. arXiv: 2401.09872 [astro-ph.HE].

[84] Jim W. Barrett, Sebastian M. Gaebel, Coenraad J. Neijssel, et al. “Accuracy of inference on the physics of binary evolution from gravitational-wave observations”. In: *Mon. Not. Roy. Astron. Soc.* 477.4 (2018), pp. 4685–4695. DOI: 10.1093/mnras/sty908. arXiv: 1711.06287 [astro-ph.HE].

[85] Simone S. Bavera et al. “The impact of mass-transfer physics on the observable properties of field binary black hole populations”. In: *Astron. Astrophys.* 647 (2021), A153. DOI: 10.1051/0004-6361/202039804. arXiv: 2010.16333 [astro-ph.HE].

[86] Simone S. Bavera, Tassos Fragos, Ying Qin, et al. “The origin of spin in binary black holes: Predicting the distributions of the main observables of Advanced LIGO”. In: *Astron. Astrophys.* 635 (2020), A97. DOI: 10.1051/0004-6361/201936204. arXiv: 1906.12257 [astro-ph.HE].

[87] Eric J. Baxter, Djuna Croon, Samuel D. McDermott, et al. “Find the Gap: Black Hole Population Analysis with an Astrophysically Motivated Mass Function”. In: *Astrophys. J. Lett.* 916.2 (2021), p. L16. DOI: 10.3847/2041-8213/ac11fc. arXiv: 2104.02685 [astro-ph.CO].

[88] Paulo Bedaque and Andrew W. Steiner. “Sound velocity bound and neutron stars”. In: *Phys. Rev. Lett.* 114.3 (2015), p. 031103. DOI: 10.1103/PhysRevLett.114.031103. arXiv: 1408.5116 [nucl-th].

[89] K. Belczynski et al. “Evolutionary roads leading to low effective spins, high black hole masses, and O1/O2 rates for LIGO/Virgo binary black holes”. In: *Astron. Astrophys.* 636 (2020), A104. DOI: 10.1051/0004-6361/201936528. arXiv: 1706.07053 [astro-ph.HE].

[90] Krzysztof Belczynski, Daniel E. Holz, Tomasz Bulik, et al. “The first gravitational-wave source from the isolated evolution of two 40-100 Msun stars”. In: *Nature* 534 (2016), p. 512. DOI: 10.1038/nature18322. arXiv: 1602.04531 [astro-ph.HE].

[91] Darsan S. Bellie, Sharan Banagiri, Zoheyr Doctor, et al. “Unresolved stochastic background from compact binary mergers detectable by next-generation ground-based gravitational-wave observatories”. In: *Phys. Rev. D* 110.2 (2024), p. 023006. DOI: 10.1103/PhysRevD.110.023006. arXiv: 2310.02517 [gr-qc].

[92] Nicola Bellomo, Daniele Bertacca, Alexander C. Jenkins, et al. “CLASS_GWB: robust modeling of the astrophysical gravitational wave background anisotropies”. In: *JCAP* 06.06 (2022), p. 030. DOI: 10.1088/1475-7516/2022/06/030. arXiv: 2110.15059 [gr-qc].

[93] Daniele Bertacca, Angelo Ricciardone, Nicola Bellomo, et al. “Projection effects on the observed angular spectrum of the astrophysical stochastic gravitational wave background”. In: *Phys. Rev. D* 101.10 (2020), p. 103513. DOI: 10.1103/PhysRevD.101.103513. arXiv: 1909.11627 [astro-ph.CO].

[94] L. Bildsten and C. Cutler. “Tidal interactions of inspiraling compact binaries”. In: *Astrophys. J.* 400 (1992), pp. 175–180. DOI: 10.1086/171983.

[95] Eli Bingham, Jonathan P. Chen, Martin Jankowiak, et al. “Pyro: Deep Universal Probabilistic Programming”. In: *J. Mach. Learn. Res.* 20 (2019), 28:1–28:6. URL: http://jmlr.org/papers/v20/18-403.html.

[96] Taylor Binnington and Eric Poisson. "Relativistic theory of tidal Love numbers". In: *Phys. Rev. D* 80 (2009), p. 084018. DOI: 10.1103/PhysRevD.80.084018. arXiv: 0906.1366 [gr-qc].

[97] Sylvia Biscoveanu, Thomas A. Callister, Carl-Johan Haster, et al. "The Binary Black Hole Spin Distribution Likely Broadens with Redshift". In: *Astrophys. J. Lett.* 932.2 (2022), p. L19. DOI: 10.3847/2041-8213/ac71a8. arXiv: 2204.01578 [astro-ph.HE].

[98] Sylvia Biscoveanu, Maximiliano Isi, Vijay Varma, et al. "Measuring the spins of heavy binary black holes". In: *Phys. Rev. D* 104.10 (2021), p. 103018. DOI: 10.1103/PhysRevD.104.103018. arXiv: 2106.06492 [gr-qc].

[99] Sylvia Biscoveanu, Maximiliano Isi, Salvatore Vitale, et al. "New Spin on LIGO-Virgo Binary Black Holes". In: *Phys. Rev. Lett.* 126.17 (2021), p. 171103. DOI: 10.1103/PhysRevLett.126.171103. arXiv: 2007.09156 [astro-ph.HE].

[100] Sylvia Biscoveanu, Philippe Landry, and Salvatore Vitale. "Population properties and multimessenger prospects of neutron star–black hole mergers following GWTC-3". In: *Mon. Not. Roy. Astron. Soc.* 518.4 (2022), pp. 5298–5312. DOI: 10.1093/mnras/stac3052. arXiv: 2207.01568 [astro-ph.HE].

[101] Sylvia Biscoveanu, Colm Talbot, and Salvatore Vitale. "The effect of spin mismodelling on gravitational-wave measurements of the binary neutron star mass distribution". In: *Mon. Not. Roy. Astron. Soc.* 511.3 (2022), pp. 4350–4359. DOI: 10.1093/mnras/stac347. arXiv: 2111.13619 [astro-ph.HE].

[102] Bhaskar Biswas and Stephan Rosswog. "Simultaneously Constraining the Neutron Star Equation of State and Mass Distribution through Multimessenger Observations and Nuclear Benchmarks". In: (Aug. 2024). arXiv: 2408.15192 [astro-ph.HE].

[103] Luc Blanchet, Thibault Damour, Bala R. Iyer, et al. "Gravitational radiation damping of compact binary systems to second postNewtonian order". In: *Phys. Rev. Lett.* 74 (1995), pp. 3515–3518. DOI: 10.1103/PhysRevLett.74.3515. arXiv: gr-qc/9501027.

[104] R. Blandford and S. A. Teukolsky. "Arrival-time analysis for a pulsar in a binary system." In: *ApJ* 205 (Apr. 1976), pp. 580–591. DOI: 10.1086/154315.

[105] C. I. Bliss. "THE METHOD OF PROBITS". In: *Science* 79.2037 (Jan. 1934), pp. 38–39. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.79.2037.38.

[106] Slavko Bogdanov et al. "Constraining the Neutron Star Mass–Radius Relation and Dense Matter Equation of State with *NICER*. I. The Millisecond Pulsar X-Ray Data Set". In: *Astrophys. J. Lett.* 887.1 (2019), p. L25. DOI: 10.3847/2041-8213/ab53eb. arXiv: 1912.05706 [astro-ph.HE].

[107] Slavko Bogdanov et al. “Constraining the Neutron Star Mass–Radius Relation and Dense Matter Equation of State with *NICER*. II. Emission from Hot Spots on a Rapidly Rotating Neutron Star”. In: *Astrophys. J. Lett.* 887.1 (2019), p. L26. DOI: 10.3847/2041-8213/ab5968. arXiv: 1912.05707 [astro-ph.HE].

[108] Alejandro Bohé et al. “Improved effective-one-body model of spinning, nonprecessing binary black holes for the era of gravitational-wave astrophysics with advanced detectors”. In: *Phys. Rev. D* 95.4 (2017), p. 044028. DOI: 10.1103/PhysRevD.95.044028. arXiv: 1611.03703 [gr-qc].

[109] Nicola Borghi, Michele Mancarella, Michele Moresco, et al. “Cosmology and Astrophysics with Standard Sirens and Galaxy Catalogs in View of Future Gravitational Wave Observations”. In: *Astrophys. J.* 964.2 (2024), p. 191. DOI: 10.3847/1538-4357/ad20eb. arXiv: 2312.05302 [astro-ph.CO].

[110] Michael Boyle et al. “The SXS Collaboration catalog of binary black hole simulations”. In: *Class. Quant. Grav.* 36.19 (2019), p. 195006. DOI: 10.1088/1361-6382/ab34e2. arXiv: 1904.04831 [gr-qc].

[111] James Bradbury, Roy Frostig, Peter Hawkins, et al. *JAX: composable transformations of Python+NumPy programs*. Version 0.3.13. 2018. URL: http://github.com/google/jax.

[112] Cosima Breu and Luciano Rezzolla. “Maximum mass, moment of inertia and compactness of relativistic stars”. In: *Mon. Not. Roy. Astron. Soc.* 459.1 (2016), pp. 646–656. DOI: 10.1093/mnras/stw575. arXiv: 1601.06083 [gr-qc].

[113] M. M. Briel, H. F. Stevance, and J. J. Eldridge. “Understanding the high-mass binary black hole population from stable mass transfer and super-Eddington accretion in bpass”. In: *Mon. Not. Roy. Astron. Soc.* 520.4 (2023), pp. 5724–5745. DOI: 10.1093/mnras/stad399. arXiv: 2206.13842 [astro-ph.HE].

[114] Duncan A. Brown, Prayush Kumar, and Alexander H. Nitz. “Template banks to search for low-mass binary black holes in advanced gravitational-wave detectors”. In: *Phys. Rev. D* 87 (2013), p. 082004. DOI: 10.1103/PhysRevD.87.082004. arXiv: 1211.6184 [gr-qc].

[115] Stephanie M. Brown, Collin D. Capano, and Badri Krishnan. “Using Gravitational Waves to Distinguish between Neutron Stars and Black Holes in Compact Binary Mergers”. In: *Astrophys. J.* 941.1 (2022), p. 98. DOI: 10.3847/1538-4357/ac98fe. arXiv: 2105.03485 [gr-qc].

[116] J. Buchner, A. Georgakakis, K. Nandra, et al. “X-ray spectral modelling of the AGN obscuring region in the CDFS: Bayesian model selection and catalogue”. In: *Astron. Astrophys.* 564 (2014), A125. DOI: 10.1051/0004-6361/201322971. arXiv: 1402.0004 [astro-ph.HE].

[117] Aaron Buikema et al. “Sensitivity and performance of the Advanced LIGO detectors in the third observing run”. In: *Phys. Rev. D* 102.6 (2020), p. 062003. DOI: 10.1103/PhysRevD.102.062003. arXiv: 2008.01301 [astro-ph.IM].

[118] A. Buonanno and T. Damour. “Effective one-body approach to general relativistic two-body dynamics”. In: *Phys. Rev. D* 59 (1999), p. 084006. DOI: 10.1103/PhysRevD.59.084006. arXiv: gr-qc/9811091.

[119] Alessandra Buonanno, Yan-bei Chen, Yi Pan, et al. “A Quasi-physical family of gravity-wave templates for precessing binaries of spinning compact objects. 2. Application to double-spin precessing binaries”. In: *Phys. Rev. D* 70 (2004). [Erratum: Phys.Rev.D 74, 029902 (2006)], p. 104003. DOI: 10.1103/PhysRevD.74.029902. arXiv: gr-qc/0405090.

[120] Alessandra Buonanno, Yan-bei Chen, and Michele Vallisneri. “Detecting gravitational waves from precessing binaries of spinning compact objects: Adiabatic limit”. In: *Phys. Rev. D* 67 (2003). [Erratum: Phys.Rev.D 74, 029904 (2006)], p. 104025. DOI: 10.1103/PhysRevD.67.104025. arXiv: gr-qc/0211087.

[121] Alessandra Buonanno, Bala Iyer, Evan Ochsner, et al. “Comparison of post-Newtonian templates for compact binary inspiral signals in gravitational-wave detectors”. In: *Phys. Rev. D* 80 (2009), p. 084043. DOI: 10.1103/PhysRevD.80.084043. arXiv: 0907.0700 [gr-qc].

[122] Thomas Callister. *Binary black hole merger rate constraints using GWTC-3 and full-O3 stochastic background constraints*. July 2024. DOI: 10.5281/zenodo.12608641. URL: https://zenodo.org/records/12608641.

[123] Thomas Callister, Maya Fishbach, Daniel Holz, et al. “Shouts and Murmurs: Combining Individual Gravitational-Wave Sources with the Stochastic Background to Measure the History of Binary Black Hole Mergers”. In: *Astrophys. J. Lett.* 896.2 (2020), p. L32. DOI: 10.3847/2041-8213/ab9743. arXiv: 2003.12152 [astro-ph.HE].

[124] Thomas Callister, Letizia Sammut, Shi Qiu, et al. “The limits of astrophysics with gravitational-wave backgrounds”. In: *Phys. Rev. X* 6.3 (2016), p. 031018. DOI: 10.1103/PhysRevX.6.031018. arXiv: 1604.02513 [gr-qc].

[125] Thomas A. Callister and Will M. Farr. “Parameter-Free Tour of the Binary Black Hole Population”. In: *Phys. Rev. X* 14.2 (2024), p. 021005. DOI: 10.1103/PhysRevX.14.021005. arXiv: 2302.07289 [astro-ph.HE].

[126] Thomas A. Callister, Carl-Johan Haster, Ken K. Y. Ng, et al. “Who Ordered That? Unequal-mass Binary Black Hole Mergers Have Larger Effective Spins”. In: *Astrophys. J. Lett.* 922.1 (2021), p. L5. DOI: 10.3847/2041-8213/ac2ccc. arXiv: 2106.00521 [astro-ph.HE].

[127] Thomas A. Callister, Simona J. Miller, Katerina Chatziioannou, et al. "No Evidence that the Majority of Black Holes in Binaries Have Zero Spin". In: *Astrophys. J. Lett.* 937.1 (2022), p. L13. DOI: 10.3847/2041-8213/ac847e. arXiv: 2205.08574 [astro-ph.HE].

[128] Manuela Campanelli, C. O. Lousto, P. Marronetti, et al. "Accurate evolutions of orbiting black-hole binaries without excision". In: *Phys. Rev. Lett.* 96 (2006), p. 111101. DOI: 10.1103/PhysRevLett.96.111101. arXiv: gr-qc/0511048.

[129] Manuela Campanelli, C. O. Lousto, and Y. Zlochower. "Spinning-black-hole binaries: The orbital hang up". In: *Phys. Rev. D* 74 (2006), p. 041501. DOI: 10.1103/PhysRevD.74.041501. arXiv: gr-qc/0604012.

[130] Kipp Cannon et al. "GstLAL: A software framework for gravitational wave discovery". In: (Oct. 2020). arXiv: 2010.05082 [astro-ph.IM].

[131] Bernard J. Carr and S. W. Hawking. "Black holes in the early Universe". In: *Mon. Not. Roy. Astron. Soc.* 168 (1974), pp. 399–415. DOI: 10.1093/mnras/168.2.399.

[132] Sean M. Carroll. *Spacetime and Geometry: An Introduction to General Relativity*. Cambridge University Press, July 2019. ISBN: 978-0-8053-8732-2, 978-1-108-48839-6, 978-1-108-77555-7. DOI: 10.1017/9781108770385.

[133] Sean M. Carroll. *Spacetime and geometry. An introduction to general relativity*. Cambridge University Press, 2004.

[134] Zack Carson, Andrew W. Steiner, and Kent Yagi. "Constraining nuclear matter parameters with GW170817". In: *Phys. Rev. D* 99.4 (2019), p. 043010. DOI: 10.1103/PhysRevD.99.043010. arXiv: 1812.08910 [gr-qc].

[135] Marco Cavaglia and Ashini Modi. "Two-dimensional correlation function of binary black hole coalescences". In: *Universe* 6.7 (2020), p. 93. DOI: 10.3390/universe6070093. arXiv: 2005.06004 [astro-ph.HE].

[136] S. Chatterji, L. Blackburn, G. Martin, et al. "Multiresolution techniques for the detection of gravitational-wave bursts". In: *Class. Quant. Grav.* 21 (2004), S1809–S1818. DOI: 10.1088/0264-9381/21/20/024. arXiv: gr-qc/0412119.

[137] K. Chatziioannou, T. Dent, M. Fishbach, et al. "Compact binary coalescences: gravitational-wave astronomy with ground-based detectors". In: (Sept. 2024). arXiv: 2409.02037 [gr-qc].

[138] Katerina Chatziioannou. "Neutron star tidal deformability and equation of state constraints". In: *Gen. Rel. Grav.* 52.11 (2020), p. 109. DOI: 10.1007/s10714-020-02754-3. arXiv: 2006.03168 [gr-qc].

[139] Katerina Chatziioannou, Neil Cornish, Marcella Wijngaarden, et al. "Modeling compact binary signals and instrumental glitches in gravitational wave data". In: *Phys. Rev. D* 103.4 (2021), p. 044013. DOI: 10.1103/PhysRevD.103.044013. arXiv: 2101.01200 [gr-qc].

[140] Katerina Chatziioannou, H. Thankful Cromartie, Stefano Gandolfi, et al. "Neutron stars and the dense matter equation of state: from microscopic theory to macroscopic observations". In: (July 2024). arXiv: 2407.11153 [nucl-th].

[141] Katerina Chatziioannou and Will M. Farr. "Inferring the maximum and minimum mass of merging neutron stars with gravitational waves". In: *Phys. Rev. D* 102.6 (2020), p. 064063. DOI: 10.1103/PhysRevD.102.064063. arXiv: 2005.00482 [astro-ph.HE].

[142] Katerina Chatziioannou, Carl-Johan Haster, and Aaron Zimmerman. "Measuring the neutron star tidal deformability with equation-of-state-independent relations and gravitational waves". In: *Phys. Rev. D* 97.10 (2018), p. 104036. DOI: 10.1103/PhysRevD.97.104036. arXiv: 1804.03221 [gr-qc].

[143] Katerina Chatziioannou, Antoine Klein, Neil Cornish, et al. "Analytic Gravitational Waveforms for Generic Precessing Binary Inspirals". In: *Phys. Rev. Lett.* 118.5 (2017), p. 051101. DOI: 10.1103/PhysRevLett.118.051101. arXiv: 1606.03117 [gr-qc].

[144] Katerina Chatziioannou, Antoine Klein, Nicolás Yunes, et al. "Constructing Gravitational Waves from Generic Spin-Precessing Compact Binary Inspirals". In: *Phys. Rev. D* 95.10 (2017), p. 104004. DOI: 10.1103/PhysRevD.95.104004. arXiv: 1703.03967 [gr-qc].

[145] Zu-Cheng Chen and Lang Liu. "Is PSR J0514–4002E in a PBH-NS binary?" In: *Sci. China Phys. Mech. Astron.* 68.6 (2025), p. 260411. DOI: 10.1007/s11433-025-2615-3. arXiv: 2401.12889 [astro-ph.HE].

[146] Hsin-Yu Chen and Katerina Chatziioannou. "Distinguishing Binary Neutron Star from Neutron Star–Black Hole Mergers with Gravitational Waves". In: *Astrophys. J. Lett.* 893.2 (2020), p. L41. DOI: 10.3847/2041-8213/ab86bc. arXiv: 1903.11197 [astro-ph.HE].

[147] Horng Sheng Chia. "Tidal deformation and dissipation of rotating black holes". In: *Phys. Rev. D* 104.2 (2021), p. 024013. DOI: 10.1103/PhysRevD.104.024013. arXiv: 2010.07300 [gr-qc].

[148] Devarshi Choudhury et al. "A NICER View of the Nearest and Brightest Millisecond Pulsar: PSR J0437–4715". In: *Astrophys. J. Lett.* 971.1 (2024), p. L20. DOI: 10.3847/2041-8213/ad5a6f. arXiv: 2407.06789 [astro-ph.HE].

[149] Martyna Chruślińska. "Chemical Evolution of the Universe and its Consequences for Gravitational-Wave Astrophysics". In: *Annalen Phys.* 536.2 (2024), p. 2200170. DOI: 10.1002/andp.202200170. arXiv: 2206.10622 [astro-ph.GA].

[150] Qi Chu et al. "SPIIR online coherent pipeline to search for gravitational waves from compact binary coalescences". In: *Phys. Rev. D* 105.2 (2022), p. 024023. DOI: 10.1103/PhysRevD.105.024023. arXiv: 2011.06787 [gr-qc].

[151] LIGO Scientific Collaboration Collaboration, Virgo Collaboration, and KAGRA. *GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run — Parameter estimation data release*. Nov. 2021. DOI: 10.5281/zenodo.5546663. URL: https://zenodo.org/records/5546663.

[152] Marta Colleoni, Felip A. Ramis Vidal, Cecilio García-Quirós, et al. "Fast frequency-domain gravitational waveforms for precessing binaries with a new twist". In: (Dec. 2024). arXiv: 2412.16721 [gr-qc].

[153] Gregory B. Cook, Stuart L. Shapiro, and Saul A. Teukolsky. "Rapidly rotating polytropes in general relativity". In: *Astrophys. J.* 422 (1994), pp. 227–242.

[154] Neil J. Cornish and Tyson B. Littenberg. "BayesWave: Bayesian Inference for Gravitational Wave Bursts and Instrument Glitches". In: *Class. Quant. Grav.* 32.13 (2015), p. 135012. DOI: 10.1088/0264-9381/32/13/135012. arXiv: 1410.3835 [gr-qc].

[155] Neil J. Cornish, Tyson B. Littenberg, Bence Bécsy, et al. "BayesWave analysis pipeline in the era of gravitational wave observations". In: *Phys. Rev. D* 103.4 (2021), p. 044006. DOI: 10.1103/PhysRevD.103.044006. arXiv: 2011.09494 [gr-qc].

[156] Sean M. Couch. "The mechanism(s) of core-collapse supernovae". In: *Philosophical Transactions of the Royal Society of London Series A* 375.2105, 20160271 (Sept. 2017), p. 20160271. DOI: 10.1098/rsta.2016.0271.

[157] Michael W. Coughlin et al. "Constraints on the neutron star equation of state from AT2017gfo using radiative transfer simulations". In: *Mon. Not. Roy. Astron. Soc.* 480.3 (2018), pp. 3871–3878. DOI: 10.1093/mnras/sty2174. arXiv: 1805.09371 [astro-ph.HE].

[158] Michael W. Coughlin and Tim Dietrich. "Can a black hole–neutron star merger explain GW170817, AT2017gfo, and GRB170817A?" In: *Phys. Rev. D* 100.4 (2019), p. 043011. DOI: 10.1103/PhysRevD.100.043011. arXiv: 1901.06052 [astro-ph.HE].

[159] Michael W. Coughlin, Tim Dietrich, Ben Margalit, et al. "Multimessenger Bayesian parameter inference of a binary neutron star merger". In: *Mon. Not. Roy. Astron. Soc.* 489.1 (2019), pp. L91–L96. DOI: 10.1093/mnrasl/slz133. arXiv: 1812.04803 [astro-ph.HE].

[160] Francesco Crescimbeni, Gabriele Franciolini, Paolo Pani, et al. "Can we identify primordial black holes? Tidal tests for subsolar-mass gravitational-wave observations". In: *Phys. Rev. D* 109.12 (2024), p. 124063. DOI: 10.1103/PhysRevD.109.124063. arXiv: 2402.18656 [astro-ph.HE].

[161] H. T. Cromartie et al. "Relativistic Shapiro delay measurements of an extremely massive millisecond pulsar". In: *Nature Astron.* 4.1 (2019), pp. 72–76. DOI: 10.1038/s41550-019-0880-2. arXiv: 1904.06759 [astro-ph.HE].

[162] Torrey Cullen, Ian Harry, Jocelyn Read, et al. "Matter Effects on LIGO/Virgo Searches for Gravitational Waves from Merging Neutron Stars". In: *Class. Quant. Grav.* 34.24 (2017), p. 245003. DOI: 10.1088/1361-6382/aa9424. arXiv: 1708.04359 [gr-qc].

[163] 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?" In: *Phys. Rev. D* 49 (1994), pp. 2658–2697. DOI: 10.1103/PhysRevD.49.2658. arXiv: gr-qc/9402014.

[164] Virginia D'Emilio, Rhys Green, and Vivien Raymond. "Density estimation with Gaussian processes for gravitational wave posteriors". In: *Mon. Not. Roy. Astron. Soc.* 508.2 (2021), pp. 2090–2097. DOI: 10.1093/mnras/stab2623. arXiv: 2104.05357 [gr-qc].

[165] Tito Dal Canton, Alexander H. Nitz, Bhooshan Gadre, et al. "Real-time Search for Compact Binary Mergers in Advanced LIGO and Virgo's Third Observing Run Using PyCBC Live". In: *Astrophys. J.* 923.2 (2021), p. 254. DOI: 10.3847/1538-4357/ac2f9a. arXiv: 2008.07494 [astro-ph.HE].

[166] Thibault Damour, Bala R. Iyer, and B. S. Sathyaprakash. "A Comparison of search templates for gravitational waves from binary inspiral - 3.5PN update". In: *Phys. Rev. D* 66 (2002), p. 027502. DOI: 10.1103/PhysRevD.66.027502. arXiv: gr-qc/0207021.

[167] Gareth S. Davies, Thomas Dent, Márton Tápai, et al. "Extending the PyCBC search for gravitational waves from compact binary mergers to a global network". In: *Phys. Rev. D* 102.2 (2020), p. 022004. DOI: 10.1103/PhysRevD.102.022004. arXiv: 2002.08291 [astro-ph.HE].

[168] D. Davis et al. *Data Quality Vetoes Applied to the Analysis of LIGO Data from the Third Observing Run*. Tech. rep. DCC-T2100045. LIGO, 2021. URL: https://dcc.ligo.org/LIGO-T2100045/public.

[169] D. Davis, T. B. Littenberg, I. M. Romero-Shaw, et al. "Subtracting glitches from gravitational-wave detector data during the third LIGO-Virgo observing run". In: *Class. Quant. Grav.* 39.24 (2022), p. 245013. DOI: 10.1088/1361-6382/aca238. arXiv: 2207.03429 [astro-ph.IM].

[170] D. Davis, T. J. Massinger, A. P. Lundgren, et al. "Improving the Sensitivity of Advanced LIGO Using Noise Subtraction". In: *Class. Quant. Grav.* 36.5 (2019), p. 055011. DOI: 10.1088/1361-6382/ab01c5. arXiv: 1809.05348 [astro-ph.IM].

[171] Derek Davis et al. "LIGO detector characterization in the second and third observing runs". In: *Class. Quant. Grav.* 38.13 (2021), p. 135014. DOI: 10.1088/1361-6382/abfd85. arXiv: 2101.11673 [astro-ph.IM].

[172] R. J. deBoer et al. "The $^{12}C(\alpha,\gamma)^{16}O$ reaction and its implications for stellar helium burning". In: *Rev. Mod. Phys.* 89.3 (2017), p. 035007. DOI: 10.1103/RevModPhys.89.035007. arXiv: 1709.03144 [nucl-ex].

[173] Walter Del Pozzo, Tjonnie G. F. Li, Michalis Agathos, et al. "Demonstrating the feasibility of probing the neutron star equation of state with second-generation gravitational wave detectors". In: *Phys. Rev. Lett.* 111.7 (2013), p. 071101. DOI: 10.1103/PhysRevLett.111.071101. arXiv: 1307.8338 [gr-qc].

[174] Ugo N. Di Carlo, Michela Mapelli, Yann Bouffanais, et al. "Binary black holes in the pair-instability mass gap". In: *Mon. Not. Roy. Astron. Soc.* 497.1 (2020), pp. 1043–1049. DOI: 10.1093/mnras/staa1997. arXiv: 1911.01434 [astro-ph.HE].

[175] Tim Dietrich et al. "Matter imprints in waveform models for neutron star binaries: Tidal and self-spin effects". In: *Phys. Rev. D* 99.2 (2019), p. 024029. DOI: 10.1103/PhysRevD.99.024029. arXiv: 1804.02235 [gr-qc].

[176] Tim Dietrich, Sebastiano Bernuzzi, and Wolfgang Tichy. "Closed-form tidal approximants for binary neutron star gravitational waveforms constructed from high-resolution numerical relativity simulations". In: *Phys. Rev. D* 96.12 (2017), p. 121501. DOI: 10.1103/PhysRevD.96.121501. arXiv: 1706.02969 [gr-qc].

[177] Tim Dietrich, Michael W. Coughlin, Peter T. H. Pang, et al. "Multimessenger constraints on the neutron-star equation of state and the Hubble constant". In: *Science* 370.6523 (2020), pp. 1450–1453. DOI: 10.1126/science.abb4317. arXiv: 2002.11355 [astro-ph.HE].

[178] Tim Dietrich, Anuradha Samajdar, Sebastian Khan, et al. "Improving the NRTidal model for binary neutron star systems". In: *Phys. Rev. D* 100.4 (2019), p. 044003. DOI: 10.1103/PhysRevD.100.044003. arXiv: 1905.06011 [gr-qc].

[179] Zoheyr Doctor, Daniel Wysocki, Richard O'Shaughnessy, et al. "Black Hole Coagulation: Modeling Hierarchical Mergers in Black Hole Populations". In: (Nov. 2019). DOI: 10.3847/1538-4357/ab7fac. arXiv: 1911.04424 [astro-ph.HE].

[180] Michal Dominik, Krzysztof Belczynski, Christopher Fryer, et al. "Double Compact Objects II: Cosmological Merger Rates". In: *Astrophys. J.* 779 (2013), p. 72. DOI: 10.1088/0004-637X/779/1/72. arXiv: 1308.1546 [astro-ph.HE].

[181] Michal Dominik, Emanuele Berti, Richard O'Shaughnessy, et al. "Double Compact Objects III: Gravitational Wave Detection Rates". In: *Astrophys. J.* 806.2 (2015), p. 263. DOI: 10.1088/0004-637X/806/2/263. arXiv: 1405.7016 [astro-ph.HE].

[182] Victor Doroshenko, Valery Suleimanov, Gerd Pühlhofer, et al. "A strangely light neutron star within a supernova remnant". In: *Nature Astron.* 6.12 (2022), pp. 1444–1451. DOI: 10.1038/s41550-022-01800-1.

[183] F. Douchin and P. Haensel. "A unified equation of state of dense matter and neutron star structure". In: *Astron. Astrophys.* 380 (2001), p. 151. DOI: 10.1051/0004-6361:20011402. arXiv: astro-ph/0111092.

[184] C. Drischler, J. W. Holt, and C. Wellenhofer. "Chiral Effective Field Theory and the High-Density Nuclear Equation of State". In: *Ann. Rev. Nucl. Part. Sci.* 71 (2021), pp. 403–432. DOI: 10.1146/annurev-nucl-102419-041903. arXiv: 2101.01709 [nucl-th].

[185] S. Duane, A. D. Kennedy, B. J. Pendleton, et al. "Hybrid Monte Carlo". In: *Phys. Lett. B* 195 (1987), pp. 216–222. DOI: 10.1016/0370-2693(87)91197-X.

[186] Bruce Edelman, Zoheyr Doctor, Jaxen Godfrey, et al. "Ain't No Mountain High Enough: Semiparametric Modeling of LIGO–Virgo's Binary Black Hole Mass Distribution". In: *Astrophys. J.* 924.2 (2022), p. 101. DOI: 10.3847/1538-4357/ac3667. arXiv: 2109.06137 [astro-ph.HE].

[187] Bruce Edelman, Ben Farr, and Zoheyr Doctor. "Cover Your Basis: Comprehensive Data-driven Characterization of the Binary Black Hole Population". In: *Astrophys. J.* 946.1 (2023), p. 16. DOI: 10.3847/1538-4357/acb5ed. arXiv: 2210.12834 [astro-ph.HE].

[188] J. van den Eijnden et al. "A new radio census of neutron star X-ray binaries". In: *Mon. Not. Roy. Astron. Soc.* 507.3 (2021), pp. 3899–3922. DOI: 10.1093/mnras/stab1995. arXiv: 2107.05286 [astro-ph.HE].

[189] Michael R. Elliott and Richard Valliant. "Inference for Nonprobability Samples". In: *Statistical Science* 32.2 (2017), pp. 249–264. DOI: 10.1214/16-STS598. URL: https://doi.org/10.1214/16-STS598.

[190] Reed Essick. "Calibration uncertainty's impact on gravitational-wave observations". In: *Phys. Rev. D* 105.8 (2022), p. 082002. DOI: 10.1103/PhysRevD.105.082002. arXiv: 2202.00823 [astro-ph.IM].

[191] Reed Essick. "Selection Effects in Periodic X-Ray Data from Maximizing Detection Statistics". In: *Astrophys. J.* 927.2 (2022), p. 195. DOI: 10.3847/1538-4357/ac517c. arXiv: 2111.04244 [astro-ph.HE].

[192] Reed Essick. "Semianalytic sensitivity estimates for catalogs of gravitational-wave transients". In: *Phys. Rev. D* 108.4 (2023), p. 043011. DOI: 10.1103/PhysRevD.108.043011. arXiv: 2307.02765 [gr-qc].

[193] Reed Essick and Will Farr. "Precision Requirements for Monte Carlo Sums within Hierarchical Bayesian Inference". In: (Apr. 2022). arXiv: 2204.00461 [astro-ph.IM].

[194] Reed Essick, Will M. Farr, Maya Fishbach, et al. "(An)isotropy measurement with gravitational wave observations". In: *Phys. Rev. D* 107.4 (2023), p. 043016. DOI: 10.1103/PhysRevD.107.043016. arXiv: 2207.05792 [astro-ph.HE].

[195] Reed Essick and Maya Fishbach. "Ensuring Consistency between Noise and Detection in Hierarchical Bayesian Inference". In: *Astrophys. J.* 962.2 (2024), p. 169. DOI: 10.3847/1538-4357/ad1604. arXiv: 2310.02017 [gr-qc].

[196] Reed Essick, Phil Landry, Katerina Chatziioannou, et al. *lwp*. URL: https://git.ligo.org/reed.essick/lwp.

[197] Reed Essick and Philippe Landry. "Discriminating between Neutron Stars and Black Holes with Imperfect Knowledge of the Maximum Neutron Star Mass". In: *Astrophys. J.* 904.1 (2020), p. 80. DOI: 10.3847/1538-4357/abbd3b. arXiv: 2007.01372 [astro-ph.HE].

[198] Reed Essick, Philippe Landry, and Daniel E. Holz. "Nonparametric Inference of Neutron Star Composition, Equation of State, and Maximum Mass with GW170817". In: *Phys. Rev. D* 101.6 (2020), p. 063007. DOI: 10.1103/PhysRevD.101.063007. arXiv: 1910.09740 [astro-ph.HE].

[199] Reed Essick, Isaac Legred, Katerina Chatziioannou, et al. "Phase transition phenomenology with nonparametric representations of the neutron star equation of state". In: *Phys. Rev. D* 108.4 (2023), p. 043013. DOI: 10.1103/PhysRevD.108.043013. arXiv: 2305.07411 [astro-ph.HE].

[200] Reed Essick, Ingo Tews, Philippe Landry, et al. "Direct Astrophysical Tests of Chiral Effective Field Theory at Supranuclear Densities". In: *Phys. Rev. C* 102.5 (2020), p. 055803. DOI: 10.1103/PhysRevC.102.055803. arXiv: 2004.07744 [astro-ph.HE].

[201] Héctor Estellés et al. “A Detailed Analysis of GW190521 with Phenomenological Waveform Models”. In: *Astrophys. J.* 924.2 (2022), p. 79. DOI: 10.3847/1538-4357/ac33a0. arXiv: 2105.06360 [gr-qc].

[202] Héctor Estellés, Marta Colleoni, Cecilio García-Quirós, et al. “New twists in compact binary waveform modeling: A fast time-domain model for precession”. In: *Phys. Rev. D* 105.8 (2022), p. 084040. DOI: 10.1103/PhysRevD.105.084040. arXiv: 2105.05872 [gr-qc].

[203] Jose María Ezquiaga and Daniel E. Holz. “Spectral Sirens: Cosmology from the Full Mass Distribution of Compact Binaries”. In: *Phys. Rev. Lett.* 129.6 (2022), p. 061102. DOI: 10.1103/PhysRevLett.129.061102. arXiv: 2202.08240 [astro-ph.CO].

[204] Stephen Fairhurst, Rhys Green, Charlie Hoy, et al. “Two-harmonic approximation for gravitational waveforms from precessing binaries”. In: *Phys. Rev. D* 102.2 (2020), p. 024055. DOI: 10.1103/PhysRevD.102.024055. arXiv: 1908.05707 [gr-qc].

[205] Yi-Zhong Fan, Ming-Zhe Han, Jin-Liang Jiang, et al. “Maximum gravitational mass MTOV=$2.25 - 0.07 + 0.08 M_{\odot}$ inferred at about 3% precision with multimessenger data of neutron stars”. In: *Phys. Rev. D* 109.4 (2024), p. 043052. DOI: 10.1103/PhysRevD.109.043052. arXiv: 2309.12644 [astro-ph.HE].

[206] Ebraheem Farag, Mathieu Renzo, Robert Farmer, et al. “Resolving the Peak of the Black Hole Mass Spectrum”. In: *Astrophys. J.* 937.2 (2022), p. 112. DOI: 10.3847/1538-4357/ac8b83. arXiv: 2208.09624 [astro-ph.HE].

[207] Amanda M. Farah, Bruce Edelman, Michael Zevin, et al. “Things That Might Go Bump in the Night: Assessing Structure in the Binary Black Hole Mass Spectrum”. In: *Astrophys. J.* 955.2 (2023), p. 107. DOI: 10.3847/1538-4357/aced02. arXiv: 2301.00834 [astro-ph.HE].

[208] Amanda M. Farah, Maya Fishbach, Reed Essick, et al. “Bridging the Gap: Categorizing Gravitational-wave Events at the Transition between Neutron Stars and Black Holes”. In: *Astrophys. J.* 931.2 (2022), p. 108. DOI: 10.3847/1538-4357/ac5f03. arXiv: 2111.03498 [astro-ph.HE].

[209] Amanda M. Farah, Maya Fishbach, and Daniel E. Holz. “Two of a Kind: Comparing Big and Small Black Holes in Binaries with Gravitational Waves”. In: *Astrophys. J.* 962.1 (2024), p. 69. DOI: 10.3847/1538-4357/ad0558. arXiv: 2308.05102 [astro-ph.HE].

[210] R. Farmer, M. Renzo, S. E. de Mink, et al. “Mind the gap: The location of the lower edge of the pair instability supernovae black hole mass gap”. In: (Oct. 2019). DOI: 10.3847/1538-4357/ab518b. arXiv: 1910.12874 [astro-ph.SR].

[211] Robert Farmer, Mathieu Renzo, Selma de Mink, et al. "Constraints from gravitational wave detections of binary black hole mergers on the $^{12}C(\alpha,\gamma)^{16}O$ rate". In: *Astrophys. J. Lett.* 902.2 (2020), p. L36. DOI: 10.3847/2041-8213/abbadd. arXiv: 2006.06678 [astro-ph.HE].

[212] B. Farr and W. M. Farr. *kombine: a kernel-density-based, embarrassingly parallel ensemble sampler*. 2015. URL: https://github.com/bfarr/kombine.

[213] Will M. Farr. "Accuracy Requirements for Empirically-Measured Selection Functions". In: *Research Notes of the AAS* 3.5 (2019), p. 66. DOI: 10.3847/2515-5172/ab1d5f. arXiv: 1904.10879 [astro-ph.IM].

[214] Will M. Farr and Katerina Chatziioannou. "A Population-Informed Mass Estimate for Pulsar J0740+6620". In: *Research Notes of the American Astronomical Society* 4.5, 65 (May 2020), p. 65. DOI: 10.3847/2515-5172/ab9088. arXiv: 2005.00032 [astro-ph.GA].

[215] Will M. Farr, Maya Fishbach, Jiani Ye, et al. "A Future Percent-Level Measurement of the Hubble Expansion at Redshift 0.8 With Advanced LIGO". In: *Astrophys. J. Lett.* 883.2 (2019), p. L42. DOI: 10.3847/2041-8213/ab4284. arXiv: 1908.09084 [astro-ph.CO].

[216] Will M. Farr, Jonathan R. Gair, Ilya Mandel, et al. "Counting And Confusion: Bayesian Rate Estimation With Multiple Populations". In: *Phys. Rev. D* 91.2 (2015), p. 023005. DOI: 10.1103/PhysRevD.91.023005. arXiv: 1302.5341 [astro-ph.IM].

[217] Will M. Farr, Niharika Sravan, Andrew Cantrell, et al. "The Mass Distribution of Stellar-Mass Black Holes". In: *Astrophys. J.* 741 (2011), p. 103. DOI: 10.1088/0004-637X/741/2/103. arXiv: 1011.1459 [astro-ph.GA].

[218] Will M. Farr, Simon Stevenson, M. Coleman Miller, et al. "Distinguishing Spin-Aligned and Isotropic Black Hole Populations With Gravitational Waves". In: *Nature* 548 (2017), p. 426. DOI: 10.1038/nature23453. arXiv: 1706.01385 [astro-ph.HE].

[219] Nicholas Farrow, Xing-Jiang Zhu, and Eric Thrane. "The mass distribution of Galactic double neutron stars". In: *Astrophys. J.* 876.1 (2019), p. 18. DOI: 10.3847/1538-4357/ab12e3. arXiv: 1902.03300 [astro-ph.HE].

[220] Marc Favata. "Systematic parameter errors in inspiraling neutron star binaries". In: *Phys. Rev. Lett.* 112 (2014), p. 101101. DOI: 10.1103/PhysRevLett.112.101101. arXiv: 1310.8288 [gr-qc].

[221] Lee S. Finn. "Detection, measurement and gravitational radiation". In: *Phys. Rev. D* 46 (1992), pp. 5236–5249. DOI: 10.1103/PhysRevD.46.5236. arXiv: gr-qc/9209010.

[222] Lee Samuel Finn and David F. Chernoff. "Observing binary inspiral in gravitational radiation: One interferometer". In: *Phys. Rev. D* 47 (1993), pp. 2198–2219. DOI: 10.1103/PhysRevD.47.2198. arXiv: gr-qc/9301003.

[223] Maya Fishbach, Zoheyr Doctor, Thomas Callister, et al. "When Are LIGO/Virgo's Big Black Hole Mergers?" In: *Astrophys. J.* 912.2 (2021), p. 98. DOI: 10.3847/1538-4357/abee11. arXiv: 2101.07699 [astro-ph.HE].

[224] Maya Fishbach, Reed Essick, and Daniel E. Holz. "Does Matter Matter? Using the mass distribution to distinguish neutron stars and black holes". In: *Astrophys. J. Lett.* 899 (2020), p. L8. DOI: 10.3847/2041-8213/aba7b6. arXiv: 2006.13178 [astro-ph.HE].

[225] Maya Fishbach and Daniel E. Holz. "Picky Partners: The Pairing of Component Masses in Binary Black Hole Mergers". In: *Astrophys. J. Lett.* 891.1 (2020), p. L27. DOI: 10.3847/2041-8213/ab7247. arXiv: 1905.12669 [astro-ph.HE].

[226] Maya Fishbach and Daniel E. Holz. "Where Are LIGO's Big Black Holes?" In: *Astrophys. J. Lett.* 851.2 (2017), p. L25. DOI: 10.3847/2041-8213/aa9bf6. arXiv: 1709.08584 [astro-ph.HE].

[227] Maya Fishbach, Daniel E. Holz, and Ben Farr. "Are LIGO's Black Holes Made From Smaller Black Holes?" In: *Astrophys. J. Lett.* 840.2 (2017), p. L24. DOI: 10.3847/2041-8213/aa7045. arXiv: 1703.06869 [astro-ph.HE].

[228] Maya Fishbach, Daniel E. Holz, and Will M. Farr. "Does the Black Hole Merger Rate Evolve with Redshift?" In: *Astrophys. J. Lett.* 863.2 (2018), p. L41. DOI: 10.3847/2041-8213/aad800. arXiv: 1805.10270 [astro-ph.HE].

[229] Maya Fishbach and Vicky Kalogera. "The Time Delay Distribution and Formation Metallicity of LIGO-Virgo's Binary Black Holes". In: *Astrophys. J. Lett.* 914.2 (2021), p. L30. DOI: 10.3847/2041-8213/ac05c4. arXiv: 2105.06491 [astro-ph.HE].

[230] Maya Fishbach, Chase Kimball, and Vicky Kalogera. "Limits on Hierarchical Black Hole Mergers from the Most Negative χ_{eff} Systems". In: *Astrophys. J. Lett.* 935.2 (2022), p. L26. DOI: 10.3847/2041-8213/ac86c4. arXiv: 2207.02924 [astro-ph.HE].

[231] Maya Fishbach and Lieke van Son. "LIGO–Virgo–KAGRA's Oldest Black Holes: Probing Star Formation at Cosmic Noon With GWTC-3". In: *Astrophys. J. Lett.* 957.2 (2023), p. L31. DOI: 10.3847/2041-8213/ad0560. arXiv: 2307.15824 [astro-ph.GA].

[232] Eanna E. Flanagan and Tanja Hinderer. "Constraining neutron star tidal Love numbers with gravitational wave detectors". In: *Phys. Rev. D* 77 (2008), p. 021502. DOI: 10.1103/PhysRevD.77.021502. arXiv: 0709.1915 [astro-ph].

[233] Eanna E. Flanagan and Scott A. Hughes. "The Basics of gravitational wave theory". In: *New J. Phys.* 7 (2005), p. 204. DOI: 10.1088/1367-2630/7/1/204. arXiv: gr-qc/0501041.

[234] Ryan J. Foley, David A. Coulter, Charles D. Kilpatrick, et al. "Updated Parameter Estimates for GW190425 Using Astrophysical Arguments and Implications for the Electromagnetic Counterpart". In: *Mon. Not. Roy. Astron. Soc.* 494.1 (2020), pp. 190–198. DOI: 10.1093/mnras/staa725. arXiv: 2002.00956 [astro-ph.HE].

[235] E. Fonseca et al. "Refined Mass and Geometric Measurements of the High-mass PSR J0740+6620". In: *Astrophys. J. Lett.* 915.1 (2021), p. L12. DOI: 10.3847/2041-8213/ac03b8. arXiv: 2104.00880 [astro-ph.HE].

[236] K. E. Saavik Ford and Barry McKernan. "Binary black hole merger rates in AGN discs versus nuclear star clusters: loud beats quiet". In: *Mon. Not. Roy. Astron. Soc.* 517.4 (2022), pp. 5827–5834. DOI: 10.1093/mnras/stac2861. arXiv: 2109.03212 [astro-ph.HE].

[237] Daniel Foreman-Mackey, David W. Hogg, Dustin Lang, et al. "emcee: The MCMC Hammer". In: *Publ. Astron. Soc. Pac.* 125 (2013), pp. 306–312. DOI: 10.1086/670067. arXiv: 1202.3665 [astro-ph.IM].

[238] Daniel Foreman-Mackey, David W. Hogg, and Timothy D. Morton. "EXOPLANET POPULATION INFERENCE AND THE ABUNDANCE OF EARTH ANALOGS FROM NOISY, INCOMPLETE CATALOGS". In: *The Astrophysical Journal* 795.1 (Oct. 2014), p. 64. ISSN: 1538-4357. DOI: 10.1088/0004-637x/795/1/64. URL: http://dx.doi.org/10.1088/0004-637X/795/1/64.

[239] William A. Fowler and F. Hoyle. "Neutrino Processes and Pair Formation in Massive Stars and Supernovae". In: *Astrophys. J. Suppl.* 9 (1964), pp. 201–319. DOI: 10.1086/190103.

[240] Paulo Cesar Carvalho Freire, Scott M. Ransom, Steve Begin, et al. "Eight New Millisecond Pulsars in NGC 6440 and NGC 6441". In: *Astrophys. J.* 675 (2008), p. 670. DOI: 10.1086/526338. arXiv: 0711.0925 [astro-ph].

[241] Jim Fuller and Linhao Ma. "Most Black Holes are Born Very Slowly Rotating". In: *Astrophys. J. Lett.* 881.1 (2019), p. L1. DOI: 10.3847/2041-8213/ab339b. arXiv: 1907.03714 [astro-ph.SR].

[242] Jim Fuller, Anthony L. Piro, and Adam S. Jermyn. "Slowing the spins of stellar cores". In: *Mon. Not. Roy. Astron. Soc.* 485.3 (2019), pp. 3661–3680. DOI: 10.1093/mnras/stz514. arXiv: 1902.08227.

[243] Shanika Galaudage et al. "Building Better Spin Models for Merging Binary Black Holes: Evidence for Nonspinning and Rapidly Spinning Nearly Aligned Subpopulations". In: *Astrophys. J. Lett.* 921.1 (2021). [Erratum: Astrophys.J.Lett. 936, L18 (2022), Erratum: Astrophys.J. 936, L18 (2022)], p. L15. DOI: 10.3847/2041-8213/ac2f3c. arXiv: 2109.02424 [gr-qc].

[244] Shanika Galaudage, Christian Adamcewicz, Xing-Jiang Zhu, et al. "Heavy Double Neutron Stars: Birth, Midlife, and Death". In: *Astrophys. J. Lett.* 909.2 (2021), p. L19. DOI: 10.3847/2041-8213/abe7f6. arXiv: 2011.01495 [astro-ph.HE].

[245] Rossella Gamba and Sebastiano Bernuzzi. "Resonant tides in binary neutron star mergers: Analytical-numerical relativity study". In: *Phys. Rev. D* 107.4 (2023), p. 044014. DOI: 10.1103/PhysRevD.107.044014. arXiv: 2207.13106 [gr-qc].

[246] Cecilio García-Quirós, Marta Colleoni, Sascha Husa, et al. "Multimode frequency-domain model for the gravitational wave signal from nonprecessing black-hole binaries". In: *Phys. Rev. D* 102.6 (2020), p. 064002. DOI: 10.1103/PhysRevD.102.064002. arXiv: 2001.10914 [gr-qc].

[247] Davide Gerosa and Emanuele Berti. "Are merging black holes born from stellar collapse or previous mergers?" In: *Phys. Rev. D* 95.12 (2017), p. 124046. DOI: 10.1103/PhysRevD.95.124046. arXiv: 1703.06223 [gr-qc].

[248] Davide Gerosa, Emanuele Berti, Richard O'Shaughnessy, et al. "Spin orientations of merging black holes formed from the evolution of stellar binaries". In: *Phys. Rev. D* 98.8 (2018), p. 084036. DOI: 10.1103/PhysRevD.98.084036. arXiv: 1808.02491 [astro-ph.HE].

[249] Davide Gerosa and Maya Fishbach. "Hierarchical mergers of stellar-mass black holes and their gravitational-wave signatures". In: *Nature Astron.* 5.8 (2021), pp. 749–760. DOI: 10.1038/s41550-021-01398-w. arXiv: 2105.03439 [astro-ph.HE].

[250] Davide Gerosa, Michael Kesden, Ulrich Sperhake, et al. "Multi-timescale analysis of phase transitions in precessing black-hole binaries". In: *Phys. Rev. D* 92 (2015), p. 064016. DOI: 10.1103/PhysRevD.92.064016. arXiv: 1506.03492 [gr-qc].

[251] G. Ghirlanda et al. "Short gamma-ray bursts at the dawn of the gravitational wave era". In: *Astron. Astrophys.* 594 (2016), A84. DOI: 10.1051/0004-6361/201628993. arXiv: 1607.07875 [astro-ph.HE].

[252] Shaon Ghosh, Xiaoshu Liu, Jolien Creighton, et al. "Rapid model comparison of equations of state from gravitational wave observation of binary neutron star coalescences". In: *Phys. Rev. D* 104.8 (2021), p. 083003. DOI: 10.1103/PhysRevD.104.083003. arXiv: 2104.08681 [gr-qc].

[253] Nicola Giacobbo and Michela Mapelli. "The progenitors of compact-object binaries: impact of metallicity, common envelope and natal kicks". In: *Mon. Not. Roy. Astron. Soc.* 480.2 (2018), pp. 2011–2030. DOI: 10.1093/mnras/sty1999. arXiv: 1806.00001 [astro-ph.HE].

[254] Thomas Gold. "Rotating neutron stars and the nature of pulsars". In: *Nature* 221 (1969), pp. 25–27. DOI: 10.1038/221025a0.

[255] Jacob Golomb, Maximiliano Isi, and Will M. Farr. "Physical Models for the Astrophysical Population of Black Holes: Application to the Bump in the Mass Distribution of Gravitational-wave Sources". In: *Astrophys. J.* 976.1 (2024), p. 121. DOI: 10.3847/1538-4357/ad8572. arXiv: 2312.03973 [astro-ph.HE].

[256] Jacob Golomb and Colm Talbot. "Hierarchical Inference of Binary Neutron Star Mass Distribution and Equation of State with Gravitational Waves". In: *Astrophys. J.* 926.1 (2022), p. 79. DOI: 10.3847/1538-4357/ac43bc. arXiv: 2106.15745 [astro-ph.HE].

[257] Jacob Golomb and Colm Talbot. "Searching for structure in the binary black hole spin distribution". In: *Phys. Rev. D* 108.10 (2023), p. 103009. DOI: 10.1103/PhysRevD.108.103009. arXiv: 2210.12287 [astro-ph.HE].

[258] Alejandra Gonzalez et al. "Second release of the CoRe database of binary neutron star merger waveforms". In: *Class. Quant. Grav.* 40.8 (2023), p. 085011. DOI: 10.1088/1361-6382/acc231. arXiv: 2210.16366 [gr-qc].

[259] K. M. Górski, E. Hivon, A. J. Banday, et al. "HEALPix - A Framework for high resolution discretization, and fast analysis of data distributed on the sphere". In: *Astrophys. J.* 622 (2005), pp. 759–771. DOI: 10.1086/427976. arXiv: astro-ph/0409513.

[260] Rachel Gray et al. "Joint cosmological and gravitational-wave population inference using dark sirens and galaxy catalogues". In: *JCAP* 12 (2023), p. 023. DOI: 10.1088/1475-7516/2023/12/023. arXiv: 2308.02281 [astro-ph.CO].

[261] S. K. Greif, K. Hebeler, J. M. Lattimer, et al. "Equation of state constraints from nuclear physics, neutron star masses, and future moment of inertia measurements". In: *Astrophys. J.* 901.2 (2020), p. 155. DOI: 10.3847/1538-4357/abaf55. arXiv: 2005.14164 [astro-ph.HE].

[262] Sebastien Guillot, Mathieu Servillat, Natalie A. Webb, et al. "Measurement of the Radius of Neutron Stars with High S/N Quiescent Low-mass X-ray Binaries in Globular Clusters". In: *Astrophys. J.* 772 (2013), p. 7. DOI: 10.1088/0004-637X/772/1/7. arXiv: 1302.0023 [astro-ph.HE].

[263] F. Gulminelli and Ad. R. Raduta. "Unified treatment of subsaturation stellar matter at zero and finite temperature". In: *Phys. Rev. C* 92.5 (2015), p. 055803. DOI: 10.1103/PhysRevC.92.055803. arXiv: 1504.04493 [nucl-th].

[264] Ming-Zhe Han, Shao-Peng Tang, Yi-Ming Hu, et al. "Is GW190425 consistent with being a neutron star–black hole merger?" In: *Astrophys. J. Lett.* 891.1 (2020), p. L5. DOI: 10.3847/2041-8213/ab745a. arXiv: 2001.07882 [astro-ph.HE].

[265] Chad Hanna et al. “Fast evaluation of multidetector consistency for real-time gravitational wave searches”. In: *Phys. Rev. D* 101.2 (2020), p. 022003. DOI: 10.1103/PhysRevD.101.022003. arXiv: 1901.02227 [gr-qc].

[266] Mark Hannam et al. “General-relativistic precession in a black-hole binary”. In: *Nature* 610.7933 (2022), pp. 652–655. DOI: 10.1038/s41586-022-05212-z. arXiv: 2112.11300 [gr-qc].

[267] Mark Hannam, Patricia Schmidt, Alejandro Bohé, et al. “Simple Model of Complete Precessing Black-Hole-Binary Gravitational Waveforms”. In: *Phys. Rev. Lett.* 113.15 (2014), p. 151101. DOI: 10.1103/PhysRevLett.113.151101. arXiv: 1308.3271 [gr-qc].

[268] O. A. Hannuksela, K. Haris, K. K. Y. Ng, et al. “Search for gravitational lensing signatures in LIGO-Virgo binary black hole events”. In: *Astrophys. J. Lett.* 874.1 (2019), p. L2. DOI: 10.3847/2041-8213/ab0c0f. arXiv: 1901.02674 [gr-qc].

[269] K. Haris, Ajit Kumar Mehta, Sumit Kumar, et al. “Identifying strongly lensed gravitational wave signals from binary black hole mergers”. In: (July 2018). arXiv: 1807.07062 [gr-qc].

[270] Charles R. Harris et al. “Array programming with NumPy”. In: *Nature* 585.7825 (2020), pp. 357–362. DOI: 10.1038/s41586-020-2649-2. arXiv: 2006.10256 [cs.MS].

[271] Ian Harry, Stephen Privitera, Alejandro Bohé, et al. “Searching for Gravitational Waves from Compact Binaries with Precessing Spins”. In: *Phys. Rev. D* 94.2 (2016), p. 024012. DOI: 10.1103/PhysRevD.94.024012. arXiv: 1603.02444 [gr-qc].

[272] K. Hebeler, J. M. Lattimer, C. J. Pethick, et al. “Constraints on neutron star radii based on chiral effective field theory interactions”. In: *Phys. Rev. Lett.* 105 (2010), p. 161102. DOI: 10.1103/PhysRevLett.105.161102. arXiv: 1007.1746 [nucl-th].

[273] Alexander Heger, S. E. Woosley, and H. C. Spruit. “Presupernova evolution of differentially rotating massive stars including magnetic fields”. In: *Astrophys. J.* 626 (2005), pp. 350–363. DOI: 10.1086/429868. arXiv: astro-ph/0409422.

[274] D. D. Hendriks, L. A. C. van Son, M. Renzo, et al. “Pulsational pair-instability supernovae in gravitational-wave and electromagnetic transients”. In: *Mon. Not. Roy. Astron. Soc.* 526.3 (2023), pp. 4130–4147. DOI: 10.1093/mnras/stad2857. arXiv: 2309.09339 [astro-ph.HE].

[275] Francisco Hernandez Vivanco, Rory Smith, Eric Thrane, et al. “Measuring the neutron star equation of state with gravitational waves: The first forty binary neutron star merger observations”. In: *Phys. Rev. D* 100.10 (2019), p. 103009. DOI: 10.1103/PhysRevD.100.103009. arXiv: 1909.02698 [gr-qc].

[276] Jason W. T. Hessels, Scott M. Ransom, Ingrid H. Stairs, et al. “A radio pulsar spinning at 716-hz”. In: *Science* 311 (2006), pp. 1901–1904. DOI: 10.1126/science.1123430. arXiv: astro-ph/0601337.

[277] P. J. van den Heuvel and Thomas M. Tauris. “Comment on ”A non-interacting low-mass black hole – giant star binary system””. In: (May 2020). DOI: 10.1126/science.aba3282. arXiv: 2005.04896 [astro-ph.SR].

[278] Kotaro Hijikawa, Ataru Tanikawa, Tomoya Kinugawa, et al. “On the population III binary black hole mergers beyond the pair-instability mass gap”. In: *Mon. Not. Roy. Astron. Soc.* 505.1 (2021), pp. L69–L73. DOI: 10.1093/mnrasl/slab052. arXiv: 2104.13384 [astro-ph.HE].

[279] Ian Hinder, Birjoo Vaishnav, Frank Herrmann, et al. “Universality and final spin in eccentric binary black hole inspirals”. In: *Phys. Rev. D* 77 (2008), p. 081502. DOI: 10.1103/PhysRevD.77.081502. arXiv: 0710.5167 [gr-qc].

[280] Tanja Hinderer et al. “Distinguishing the nature of comparable-mass neutron star binary systems with multimessenger observations: GW170817 case study”. In: *Phys. Rev. D* 100.6 (2019), p. 06321. DOI: 10.1103/PhysRevD.100.063021. arXiv: 1808.03836 [astro-ph.HE].

[281] Tanja Hinderer et al. “Effects of neutron-star dynamic tides on gravitational waveforms within the effective-one-body approach”. In: *Phys. Rev. Lett.* 116.18 (2016), p. 181101. DOI: 10.1103/PhysRevLett.116.181101. arXiv: 1602.00599 [gr-qc].

[282] Tanja Hinderer. “Tidal Love numbers of neutron stars”. In: *Astrophys. J.* 677 (2008). [Erratum: Astrophys.J. 697, 964 (2009)], pp. 1216–1220. DOI: 10.1086/533487. arXiv: 0711.2420 [astro-ph].

[283] Tanja Hinderer, Benjamin D. Lackey, Ryan N. Lang, et al. “Tidal deformability of neutron stars with realistic equations of state and their gravitational wave signatures in binary inspiral”. In: *Phys. Rev. D* 81 (2010), p. 123016. DOI: 10.1103/PhysRevD.81.123016. arXiv: 0911.3535 [astro-ph.HE].

[284] Maurício Hippert, Emily Dillingham, Hung Tan, et al. “Dark matter or regular matter in neutron stars? How to tell the difference from the coalescence of compact objects”. In: *Phys. Rev. D* 107.11 (2023), p. 115028. DOI: 10.1103/PhysRevD.107.115028. arXiv: 2211.08590 [astro-ph.HE].

[285] Fabian Hofmann, Enrico Barausse, and Luciano Rezzolla. “The final spin from binary black holes in quasi-circular orbits”. In: *Astrophys. J. Lett.* 825.2 (2016), p. L19. DOI: 10.3847/2041-8205/825/2/L19. arXiv: 1605.01938 [gr-qc].

[286] Sophie Hourihane, Katerina Chatziioannou, Marcella Wijngaarden, et al. "Accurate modeling and mitigation of overlapping signals and glitches in gravitational-wave data". In: *Phys. Rev. D* 106.4 (2022), p. 042006. DOI: 10.1103/PhysRevD.106.042006. arXiv: 2205.13580 [gr-qc].

[287] Sophie Hourihane, Patrick Meyers, Aaron Johnson, et al. "Accurate characterization of the stochastic gravitational-wave background with pulsar timing arrays by likelihood reweighting". In: *Phys. Rev. D* 107.8 (2023), p. 084045. DOI: 10.1103/PhysRevD.107.084045. arXiv: 2212.06276 [gr-qc].

[288] Qian Hu and John Veitch. "Assessing the model waveform accuracy of gravitational waves". In: *Phys. Rev. D* 106.4 (2022), p. 044042. DOI: 10.1103/PhysRevD.106.044042. arXiv: 2205.08448 [gr-qc].

[289] John D. Hunter. "Matplotlib: A 2D Graphics Environment". In: *Comput. Sci. Eng.* 9.3 (2007), pp. 90–95. DOI: 10.1109/MCSE.2007.55.

[290] Sascha Husa, Sebastian Khan, Mark Hannam, et al. "Frequency-domain gravitational waves from nonprecessing black-hole binaries. I. New numerical waveforms and anatomy of the signal". In: *Phys. Rev. D* 93.4 (2016), p. 044006. DOI: 10.1103/PhysRevD.93.044006. arXiv: 1508.07250 [gr-qc].

[291] Richard A. Isaacson. "Gravitational Radiation in the Limit of High Frequency. II. Nonlinear Terms and the Ef fective Stress Tensor". In: *Phys. Rev.* 166 (1968), pp. 1272–1279. DOI: 10.1103/PhysRev.166.1272.

[292] Maximiliano Isi. "Parametrizing gravitational-wave polarizations". In: *Class. Quant. Grav.* 40.20 (2023), p. 203001. DOI: 10.1088/1361-6382/acf28c. arXiv: 2208.03372 [gr-qc].

[293] Maximiliano Isi and Will M. Farr. "Analyzing black-hole ringdowns". In: (July 2021). arXiv: 2107.05609 [gr-qc].

[294] T. Jayasinghe et al. "A unicorn in monoceros: the 3 $M_\odot$ dark companion to the bright, nearby red giant V723 Mon is a non-interacting, mass-gap black hole candidate". In: *Mon. Not. Roy. Astron. Soc.* 504.2 (2021), pp. 2577–2602. DOI: 10.1093/mnras/stab907. arXiv: 2101.02212 [astro-ph.SR].

[295] Alexander C. Jenkins, Richard O'Shaughnessy, Mairi Sakellariadou, et al. "Anisotropies in the astrophysical gravitational-wave background: The impact of black hole distributions". In: *Phys. Rev. Lett.* 122.11 (2019), p. 111101. DOI: 10.1103/PhysRevLett.122.111101. arXiv: 1810.13435 [astro-ph.CO].

[296] Alexander C. Jenkins and Mairi Sakellariadou. "Shot noise in the astrophysical gravitational-wave background". In: *Phys. Rev. D* 100.6 (2019), p. 063508. DOI: 10.1103/PhysRevD.100.063508. arXiv: 1902.07719 [astro-ph.CO].

[297] Alexander C. Jenkins, Mairi Sakellariadou, Tania Regimbau, et al. "Anisotropies in the astrophysical gravitational-wave background: Predictions for the detection of compact binaries by LIGO and Virgo". In: *Phys. Rev. D* 98.6 (2018), p. 063501. DOI: 10.1103/PhysRevD.98.063501. arXiv: 1806.01718 [astro-ph.CO].

[298] Vassiliki Kalogera. "Orbital characteristics of binary systems after asymmetric supernova explosions". In: *Astrophys. J.* 471 (1996), p. 352. DOI: 10.1086/177974. arXiv: astro-ph/9605186.

[299] Vassiliki Kalogera. "Spin orbit misalignment in close binaries with two compact objects". In: *Astrophys. J.* 541 (2000), pp. 319–328. DOI: 10.1086/309400. arXiv: astro-ph/9911417.

[300] Vassiliki Kalogera and Gordon Baym. "The maximum mass of a neutron star". In: *Astrophys. J. Lett.* 470 (1996), pp. L61–L64. DOI: 10.1086/310296. arXiv: astro-ph/9608059.

[301] Christos Karathanasis, Suvodip Mukherjee, and Simone Mastrogiovanni. "Binary black holes population and cosmology in new lights: signature of PISN mass and formation channel in GWTC-3". In: *Mon. Not. Roy. Astron. Soc.* 523.3 (2023), pp. 4539–4555. DOI: 10.1093/mnras/stad1373. arXiv: 2204.13495 [astro-ph.CO].

[302] Robert E. Kass and Adrian E. Raftery. "Bayes Factors". In: *J. Am. Statist. Assoc.* 90.430 (1995), pp. 773–795. DOI: 10.1080/01621459.1995.10476572.

[303] Michael Kesden, Davide Gerosa, Richard O'Shaughnessy, et al. "Effective potentials and morphological transitions for binary black-hole spin precession". In: *Phys. Rev. Lett.* 114.8 (2015), p. 081103. DOI: 10.1103/PhysRevLett.114.081103. arXiv: 1411.0674 [gr-qc].

[304] Sebastian Khan, Katerina Chatziioannou, Mark Hannam, et al. "Phenomenological model for the gravitational-wave signal from precessing binary black holes with two-spin effects". In: *Phys. Rev. D* 100.2 (2019), p. 024059. DOI: 10.1103/PhysRevD.100.024059. arXiv: 1809.10113 [gr-qc].

[305] Sebastian Khan, Sascha Husa, Mark Hannam, et al. "Frequency-domain gravitational waves from nonprecessing black-hole binaries. II. A phenomenological model for the advanced detector era". In: *Phys. Rev. D* 93.4 (2016), p. 044007. DOI: 10.1103/PhysRevD.93.044007. arXiv: 1508.07253 [gr-qc].

[306] Lawrence E. Kidder. "Coalescing binary systems of compact objects to postNewtonian 5/2 order. 5. Spin effects". In: *Phys. Rev. D* 52 (1995), pp. 821–847. DOI: 10.1103/PhysRevD.52.821. arXiv: gr-qc/9506022.

[307] R. Weizmann Kiendrebeogo et al. "Updated Observing Scenarios and Multimessenger Implications for the International Gravitational-wave Networks O4 and O5". In: *Astrophys. J.* 958.2 (2023), p. 158. DOI: 10.3847/1538-4357/acfcb1. arXiv: 2306.09234 [astro-ph.HE].

[308] Chase Kimball et al. "Evidence for Hierarchical Black Hole Mergers in the Second LIGO–Virgo Gravitational Wave Catalog". In: *Astrophys. J. Lett.* 915.2 (2021), p. L35. DOI: 10.3847/2041-8213/ac0aef. arXiv: 2011.05332 [astro-ph.HE].

[309] Chase Kimball, Colm Talbot, Christopher P. L. Berry, et al. "Black Hole Genealogy: Identifying Hierarchical Mergers with Gravitational Waves". In: *Astrophys. J.* 900.2 (2020), p. 177. DOI: 10.3847/1538-4357/aba518. arXiv: 2005.00023 [astro-ph.HE].

[310] Tomoya Kinugawa, Takashi Nakamura, and Hiroyuki Nakano. "Chirp Mass and Spin of Binary Black Holes from First Star Remnants". In: *Mon. Not. Roy. Astron. Soc.* 498.3 (2020), pp. 3946–3963. DOI: 10.1093/mnras/staa2511. arXiv: 2005.09795 [astro-ph.HE].

[311] Rudolf Kippenhahn, Alfred Weigert, and Achim Weiss. *Stellar structure and evolution*. Astronomy and Astrophysics Library. Springer, Aug. 2012. ISBN: 978-0-387-58013-5, 978-3-642-30255-8, 978-3-642-30304-3. DOI: 10.1007/978-3-642-30304-3.

[312] Leslie Kish. *Survey Sampling*. Third. Oxford, England: Wiley-Interscience, 1995. URL: https://search.worldcat.org/title/Survey-sampling/oclc/256017.

[313] S. Klimenko et al. "Method for detection and reconstruction of gravitational wave transients with networks of advanced detectors". In: *Phys. Rev. D* 93.4 (2016), p. 042004. DOI: 10.1103/PhysRevD.93.042004. arXiv: 1511.05999 [gr-qc].

[314] S. Klimenko and Guenakh Mitselmakher. "A wavelet method for detection of gravitational wave bursts". In: *Class. Quant. Grav.* 21 (2004), S1819–S1830. DOI: 10.1088/0264-9381/21/20/025.

[315] Oleg Komoltsev and Aleksi Kurkela. "How Perturbative QCD Constrains the Equation of State at Neutron-Star Densities". In: *Phys. Rev. Lett.* 128.20 (2022), p. 202701. DOI: 10.1103/PhysRevLett.128.202701. arXiv: 2111.05350 [nucl-th].

[316] Oleg Komoltsev, Rahul Somasundaram, Tyler Gorda, et al. "Equation of state at neutron-star densities and beyond from perturbative QCD". In: *Phys. Rev. D* 109.9 (2024), p. 094030. DOI: 10.1103/PhysRevD.109.094030. arXiv: 2312.14127 [nucl-th].

[317] Ely D. Kovetz, Ilias Cholis, Patrick C. Breysse, et al. "Black hole mass function from gravitational wave measurements". In: *Phys. Rev. D* 95.10 (2017), p. 103010. DOI: 10.1103/PhysRevD.95.103010. arXiv: 1611.01157 [astro-ph.CO].

[318] Laura Kreidberg, Charles D. Bailyn, Will M. Farr, et al. "Mass Measurements of Black Holes in X-Ray Transients: Is There a Mass Gap?" In: *Astrophys. J.* 757 (2012), p. 36. DOI: 10.1088/0004-637X/757/1/36. arXiv: 1205.1805 [astro-ph.HE].

[319] Pavel Kroupa. "On the variation of the initial mass function". In: *Mon. Not. Roy. Astron. Soc.* 322 (2001), p. 231. DOI: 10.1046/j.1365-8711.2001.04022.x. arXiv: astro-ph/0009005.

[320] Pavel Kroupa and Tereza Jerabkova. "The Salpeter IMF and its descendants". en. In: *Nature Astronomy* 3.6 (June 2019), pp. 482–484. ISSN: 2397-3366. DOI: 10.1038/s41550-019-0793-0. URL: https://www.nature.com/articles/s41550-019-0793-0.

[321] R. Kunz, M. Fey, M. Jaeger, et al. "Astrophysical Reaction Rate of ^{12}C(α, γ)^{16}O". In: *ApJ* 567.1 (Mar. 2002), pp. 643–650. DOI: 10.1086/338384.

[322] Doron Kushnir, Matias Zaldarriaga, Juna A. Kollmeier, et al. "GW150914: Spin based constraints on the merger time of the progenitor system". In: *Mon. Not. Roy. Astron. Soc.* 462.1 (2016), pp. 844–849. DOI: 10.1093/mnras/stw1684. arXiv: 1605.03839 [astro-ph.HE].

[323] Benjamin D. Lackey and Leslie Wade. "Reconstructing the neutron-star equation of state with gravitational-wave detectors from a realistic population of inspiralling binary neutron stars". In: *Phys. Rev. D* 91.4 (2015), p. 043002. DOI: 10.1103/PhysRevD.91.043002. arXiv: 1410.8866 [gr-qc].

[324] William G. Lamb and Stephen R. Taylor. "Spectral Variance in a Stochastic Gravitational-wave Background from a Binary Population". In: *Astrophys. J. Lett.* 971.1 (2024), p. L10. DOI: 10.3847/2041-8213/ad654a. arXiv: 2407.06270 [gr-qc].

[325] Philippe Landry and Reed Essick. "Nonparametric inference of the neutron star equation of state from gravitational wave observations". In: *Phys. Rev. D* 99.8 (2019), p. 084049. DOI: 10.1103/PhysRevD.99.084049. arXiv: 1811.12529 [gr-qc].

[326] Philippe Landry, Reed Essick, and Katerina Chatziioannou. "Nonparametric constraints on neutron star matter with existing and upcoming gravitational wave and pulsar observations". In: *Phys. Rev. D* 101.12 (2020), p. 123007. DOI: 10.1103/PhysRevD.101.123007. arXiv: 2003.04880 [astro-ph.HE].

[327] Philippe Landry and Jocelyn S. Read. "The Mass Distribution of Neutron Stars in Gravitational-wave Binaries". In: *Astrophys. J. Lett.* 921.2 (2021), p. L25. DOI: 10.3847/2041-8213/ac2f3e. arXiv: 2107.04559 [astro-ph.HE].

[328] Jacob Lange, Richard O'Shaughnessy, and Monica Rizzo. "Rapid and accurate parameter inference for coalescing, precessing compact binaries". In: (May 2018). arXiv: 1805.10457 [gr-qc].

[329] J. M. Lattimer and M. Prakash. "Neutron star structure and the equation of state". In: *Astrophys. J.* 550 (2001), p. 426. DOI: 10.1086/319702. arXiv: astro-ph/0002232.

[330] James M. Lattimer and Madappa Prakash. "The Equation of State of Hot, Dense Matter and Neutron Stars". In: *Phys. Rept.* 621 (2016), pp. 127–164. DOI: 10.1016/j.physrep.2015.12.005. arXiv: 1512.07820 [astro-ph.SR].

[331] James M. Lattimer and Maddapa Prakash. "Neutron Star Observations: Prognosis for Equation of State Constraints". In: *Phys. Rept.* 442 (2007), pp. 109–165. DOI: 10.1016/j.physrep.2007.02.003. arXiv: astro-ph/0612440.

[332] Isaac Legred, Katerina Chatziioannou, Reed Essick, et al. "Impact of the PSR J0740+6620 radius constraint on the properties of high-density matter". In: *Phys. Rev. D* 104.6 (2021), p. 063003. DOI: 10.1103/PhysRevD.104.063003. arXiv: 2106.05313 [astro-ph.HE].

[333] Isaac Legred, Katerina Chatziioannou, Reed Essick, et al. "Implicit correlations within phenomenological parametric models of the neutron star equation of state". In: *Phys. Rev. D* 105.4 (2022), p. 043016. DOI: 10.1103/PhysRevD.105.043016. arXiv: 2201.06791 [astro-ph.HE].

[334] 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". In: *Phys. Rev. D* 104.12 (2021), p. 123030. DOI: 10.1103/PhysRevD.104.123030. arXiv: 2109.09872 [astro-ph.IM].

[335] LIGO Scientific Collaboration. *LIGO Algorithm Library - LALSuite*. free software (GPL). 2018. DOI: 10.7935/GT1W-FZ16.

[336] LIGO Scientific Collaboration. "Parameter estimation sample release for GW190425". In: *https://dcc.ligo.org/LIGO-P2000026/public* (2020). URL: https://dcc.ligo.org/LIGO-P2000026/public.

[337] LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration. *GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run — O3 search sensitivity estimates*. 2023. DOI: 10.5281/ZENODO.7890437. URL: https://zenodo.org/record/7890437.

[338] LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration. *GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run — Parameter estimation data release*. 2021. DOI: 10.5281/ZENODO.5546663. URL: https://zenodo.org/record/5546663.

[339] LIGO Scientific Collaboration and Virgo Collaboration. In: *GCN* 26926 (2020). URL: https://gcn.gsfc.nasa.gov/other/S200129m.gcn3.

[340] LIGO Scientific Collaboration and Virgo Collaboration. *BayesWave, https://git.ligo.org/lscsoft/baye* 2018. URL: https://git.ligo.org/lscsoft/bayeswave.

[341] Lee Lindblom. "Determining the Nuclear Equation of State from Neutron-Star Masses and Radii". In: *ApJ* 398 (Oct. 1992), p. 569. DOI: 10.1086/171882.

[342] Lee Lindblom. "Spectral Representations of Neutron-Star Equations of State". In: *Phys. Rev. D* 82 (2010), p. 103011. DOI: 10.1103/PhysRevD.82.103011. arXiv: 1009.0738 [astro-ph.HE].

[343] Tyson B. Littenberg and Neil J. Cornish. "Bayesian inference for spectral estimation of gravitational wave detector noise". In: *Phys. Rev. D* 91.8 (2015), p. 084034. DOI: 10.1103/PhysRevD.91.084034. arXiv: 1410.3852 [gr-qc].

[344] An-Chiao Liu, Sander Scholtus, and Ton De Waal. "Correcting Selection Bias in Big Data by Pseudo-Weighting". In: *Journal of Survey Statistics and Methodology* (Dec. 2022), smac029. ISSN: 2325-0984. DOI: 10.1093/jssam/smac029. eprint: https://academic.oup.com/jssam/advance-article-pdf/doi/10.1093/jssam/smac029/48411623/smac029.pdf. URL: https://doi.org/10.1093/jssam/smac029.

[345] Lionel London, Sebastian Khan, Edward Fauchon-Jones, et al. "First higher-multipole model of gravitational waves from spinning and coalescing black-hole binaries". In: *Phys. Rev. Lett.* 120.16 (2018), p. 161102. DOI: 10.1103/PhysRevLett.120.161102. arXiv: 1708.00404 [gr-qc].

[346] Alessandro Longo, Stefano Bianchi, Wolfango Plastino, et al. "Scattered light noise characterisation at the Virgo interferometer with tvf-EMD adaptive algorithm". In: *Class. Quant. Grav.* 37.14 (2020), p. 145011. DOI: 10.1088/1361-6382/ab9719. arXiv: 2002.10529 [astro-ph.IM].

[347] Alessandro Longo, Stefano Bianchi, Guillermo Valdes, et al. "Daily monitoring of scattered light noise due to microseismic variability at the Virgo interferometer". In: *Class. Quant. Grav.* 39.3 (2022), p. 035001. DOI: 10.1088/1361-6382/ac4117. arXiv: 2112.06046 [astro-ph.IM].

[348] Thomas J. Loredo. "Accounting for source uncertainties in analyses of astronomical survey data". In: *AIP Conf. Proc.* 735.1 (2004). Ed. by Rainer Fischer, Roland Preuss, and Udo von Toussaint, pp. 195–206. DOI: 10.1063/1.1835214. arXiv: astro-ph/0409387.

[349] Jake Mac Uilliam, Sarp Akcay, and Jonathan E. Thompson. "Survey of four precessing waveform models for binary black hole systems". In: *Phys. Rev. D* 109.8 (2024), p. 084077. DOI: 10.1103/PhysRevD.109.084077. arXiv: 2402.06781 [gr-qc].

[350] Duncan M. Macleod, Joseph S. Areeda, Scott B. Coughlin, et al. "GWpy: A Python package for gravitational-wave astrophysics". In: *SoftwareX* 13, 100657 (Jan. 2021), p. 100657. DOI: 10.1016/j.softx.2021.100657.

[351] Piero Madau and Mark Dickinson. "Cosmic Star Formation History". In: *Ann. Rev. Astron. Astrophys.* 52 (2014), pp. 415–486. DOI: 10.1146/annurev-astro-081811-125615. arXiv: 1403.0007 [astro-ph.CO].

[352] Michele Maggiore. *Gravitational Waves. Vol. 1: Theory and Experiments*. Oxford University Press, 2007. ISBN: 978-0-19-171766-6, 978-0-19-852074-0. DOI: 10.1093/acprof:oso/9780198570745.001.0001.

[353] R. Maiolino et al. "AMAZE. I. The evolution of the mass-metallicity relation at z>3". In: *Astron. Astrophys.* 488 (2008), pp. 463–479. DOI: 10.1051/0004-6361:200809678. arXiv: 0806.2410 [astro-ph].

[354] Michele Mancarella and Edwin Genoud-Prachex. *CosmoStatGW/MGCosmoPop: v1.0.0*. Version v1.0.0. Mar. 2022. DOI: 10.5281/zenodo.6323173. URL: https://doi.org/10.5281/zenodo.6323173.

[355] Michele Mancarella, Francesco Iacovelli, and Davide Gerosa. "Inferring, not just detecting: Metrics for high-redshift sources observed with third-generation gravitational-wave detectors". In: *Phys. Rev. D* 107.10 (2023), p. L101302. DOI: 10.1103/PhysRevD.107.L101302. arXiv: 2303.16323 [gr-qc].

[356] R. N. Manchester. "Millisecond Pulsars, their Evolution and Applications". In: *J. Astrophys. Astron.* 38 (2017), p. 42. DOI: 10.1007/s12036-017-9469-2. arXiv: 1709.09434 [astro-ph.HE].

[357] Ilya Mandel and Alison Farmer. "Merging stellar-mass binary black holes". In: *Phys. Rept.* 955 (2022), pp. 1–24. DOI: 10.1016/j.physrep.2022.01.003. arXiv: 1806.05820 [astro-ph.HE].

[358] Ilya Mandel, Will M. Farr, Andrea Colonna, et al. "Model-independent inference on compact-binary observations". In: *Mon. Not. Roy. Astron. Soc.* 465.3 (2017), pp. 3254–3260. DOI: 10.1093/mnras/stw2883. arXiv: 1608.08223 [astro-ph.HE].

[359] Ilya Mandel, Will M. Farr, and Jonathan R. Gair. "Extracting distribution parameters from multiple uncertain observations with selection biases". In: *Mon. Not. Roy. Astron. Soc.* 486.1 (2019), pp. 1086–1093. DOI: 10.1093/mnras/stz896. arXiv: 1809.02063 [physics.data-an].

[360] Ilya Mandel and Tassos Fragos. "An alternative interpretation of GW190412 as a binary black hole merger with a rapidly spinning secondary". In: *Astrophys. J. Lett.* 895.2 (2020), p. L28. DOI: 10.3847/2041-8213/ab8e41. arXiv: 2004.09288 [astro-ph.HE].

[361] Ilya Mandel and Selma E. de Mink. "Merging binary black holes formed through chemically homogeneous evolution in short-period stellar binaries". In: *Mon. Not. Roy. Astron. Soc.* 458.3 (2016), pp. 2634–2647. DOI: 10.1093/mnras/stw379. arXiv: 1601.00007 [astro-ph.HE].

[362] Ilya Mandel and Richard O'Shaughnessy. "Compact Binary Coalescences in the Band of Ground-based Gravitational-Wave Detectors". In: *Class. Quant. Grav.* 27 (2010). Ed. by Sascha Husa and Badri Krishnan, p. 114007. DOI: 10.1088/0264-9381/27/11/114007. arXiv: 0912.1074 [astro-ph.HE].

[363] Michela Mapelli, Filippo Santoliquido, Yann Bouffanais, et al. "Mass and Rate of Hierarchical Black Hole Mergers in Young, Globular and Nuclear Star Clusters". In: *Symmetry* 13.9 (2021), p. 1678. DOI: 10.3390/sym13091678. arXiv: 2007.15022 [astro-ph.HE].

[364] Pablo Marchant, Norbert Langer, Philipp Podsiadlowski, et al. "A new route towards merging massive black holes". In: *Astron. Astrophys.* 588 (2016), A50. DOI: 10.1051/0004-6361/201628133. arXiv: 1601.03718 [astro-ph.SR].

[365] Pablo Marchant, Mathieu Renzo, Robert Farmer, et al. "Pulsational pair-instability supernovae in very close binaries". In: (Oct. 2018). DOI: 10.3847/1538-4357/ab3426. arXiv: 1810.13412 [astro-ph.HE].

[366] Ben Margalit and Brian D. Metzger. "The Multi-Messenger Matrix: the Future of Neutron Star Merger Constraints on the Nuclear Equation of State". In: *Astrophys. J. Lett.* 880.1 (2019), p. L15. DOI: 10.3847/2041-8213/ab2ae2. arXiv: 1904.11995 [astro-ph.HE].

[367] Ivan Markin, Anna Neuweiler, Adrian Abac, et al. "General-relativistic hydrodynamics simulation of a neutron star–sub-solar-mass black hole merger". In: *Phys. Rev. D* 108.6 (2023), p. 064025. DOI: 10.1103/PhysRevD.108.064025. arXiv: 2304.11642 [gr-qc].

[368] J. G. Martinez, K. Stovall, P. C. C. Freire, et al. "Pulsar J0453+1559: A Double Neutron Star System with a Large Mass Asymmetry". In: *Astrophys. J.* 812.2 (2015), p. 143. DOI: 10.1088/0004-637X/812/2/143. arXiv: 1509.08805 [astro-ph.HE].

[369] L. Martino, V. Elvira, and F. Louzada. "Effective Sample Size for Importance Sampling based on discrepancy measures". In: *arXiv e-prints*, arXiv:1602.03572 (Feb. 2016), arXiv:1602.03572. arXiv: 1602.03572 [stat.CO].

[370] Simone Mastrogiovanni, Grégoire Pierra, Stéphane Perriès, et al. "ICAROGW: A python package for inference of astrophysical population properties of noisy, heterogeneous, and incomplete observations". In: *Astron. Astrophys.* 682 (2024), A167. DOI: 10.1051/0004-6361/202347007. arXiv: 2305.17973 [astro-ph.CO].

[371] B. McKernan, K. E. S. Ford, R. O'Shaughnessy, et al. "Monte Carlo simulations of black hole mergers in AGN discs: Low $\chi_{\rm eff}$ mergers and predictions for LIGO". In: *Mon. Not. Roy. Astron. Soc.* 494.1 (2020), pp. 1203–1216. DOI: 10.1093/mnras/staa740. arXiv: 1907.04356 [astro-ph.HE].

[372] Duncan Meacher, Michael Coughlin, Sean Morris, et al. "Mock data and science challenge for detecting an astrophysical stochastic gravitational-wave background with Advanced LIGO and Advanced Virgo". In: *Phys. Rev. D* 92.6 (2015), p. 063002. DOI: 10.1103/PhysRevD.92.063002. arXiv: 1506.06744 [astro-ph.HE].

[373] Duncan Meacher, Eric Thrane, and Tania Regimbau. "Statistical properties of astrophysical gravitational-wave backgrounds". In: *Phys. Rev. D* 89.8 (2014), p. 084063. DOI: 10.1103/PhysRevD.89.084063. arXiv: 1402.6231 [astro-ph.CO].

[374] Ajit Kumar Mehta, Alessandra Buonanno, Jonathan Gair, et al. "Observing Intermediate-mass Black Holes and the Upper Stellar-mass gap with LIGO and Virgo". In: *Astrophys. J.* 924.1 (2022), p. 39. DOI: 10.3847/1538-4357/ac3130. arXiv: 2105.06366 [gr-qc].

[375] Cody Messick et al. "Analysis Framework for the Prompt Discovery of Compact Binary Mergers in Gravitational-wave Data". In: *Phys. Rev. D* 95.4 (2017), p. 042001. DOI: 10.1103/PhysRevD.95.042001. arXiv: 1604.04324 [astro-ph.IM].

[376] Brian D. Metzger. "Kilonovae". In: *Living Rev. Rel.* 23.1 (2020), p. 1. DOI: 10.1007/s41114-019-0024-0. arXiv: 1910.01617 [astro-ph.HE].

[377] M. C. Miller et al. "PSR J0030+0451 Mass and Radius from *NICER* Data and Implications for the Properties of Neutron Star Matter". In: *Astrophys. J. Lett.* 887.1 (2019), p. L24. DOI: 10.3847/2041-8213/ab50c5. arXiv: 1912.05705 [astro-ph.HE].

[378] M. C. Miller et al. "The Radius of PSR J0740+6620 from NICER and XMM-Newton Data". In: *Astrophys. J. Lett.* 918.2 (2021), p. L28. DOI: 10.3847/2041-8213/ac089b. arXiv: 2105.06979 [astro-ph.HE].

[379] M. C. Miller, F. K. Lamb, A. J. Dittmann, et al. "The Radius of PSR J0740+6620 from NICER and XMM-Newton Data". In: *arXiv:2105.06979 [astro-ph, physics:gr-qc, physics:nucl-ex, physics:nucl-th]* (May 2021). arXiv: 2105.06979. URL: http://arxiv.org/abs/2105.06979.

[380] M. Coleman Miller, Cecilia Chirenti, and Frederick K. Lamb. "Constraining the equation of state of high-density cold matter using nuclear and astronomical measurements". In: (Apr. 2019). DOI: 10.3847/1538-4357/ab4ef9. arXiv: 1904.08907 [astro-ph.HE].

[381] M. Coleman Miller and Douglas P. Hamilton. "Four-body effects in globular cluster black hole coalescence". In: *Astrophys. J.* 576 (2002), p. 894. DOI: 10.1086/341788. arXiv: astro-ph/0202298.

[382] Simona Miller, Thomas A. Callister, and Will Farr. "The Low Effective Spin of Binary Black Holes and Implications for Individual Gravitational-Wave Events". In: *Astrophys. J.* 895.2 (2020), p. 128. DOI: 10.3847/1538-4357/ab80c0. arXiv: 2001.06051 [astro-ph.HE].

[383] Charles W. Misner, K. S. Thorne, and J. A. Wheeler. *Gravitation*. San Francisco: W. H. Freeman, 1973. ISBN: 978-0-7167-0344-0, 978-0-691-17779-3. URL: https://inspirehep.net/literature/95654.

[384] Soichiro Morisaki. "Accelerating parameter estimation of gravitational waves from compact binary coalescence using adaptive frequency resolutions". In: *Phys. Rev. D* 104.4 (2021), p. 044062. DOI: 10.1103/PhysRevD.104.044062. arXiv: 2104.07813 [gr-qc].

[385] Gonzalo Morras et al. "Analysis of a subsolar-mass compact binary candidate from the second observing run of Advanced LIGO". In: *Phys. Dark Univ.* 42 (2023), p. 101285. DOI: 10.1016/j.dark.2023.101285. arXiv: 2301.11619 [gr-qc].

[386] Elias R. Most, L. Jens Papenfort, Lukas R. Weih, et al. "A lower bound on the maximum mass if the secondary in GW190814 was once a rapidly spinning neutron star". In: *Mon. Not. Roy. Astron. Soc.* 499.1 (2020), pp. L82–L86. DOI: 10.1093/mnrasl/slaa168. arXiv: 2006.14601 [astro-ph.HE].

[387] Matthew Mould, Davide Gerosa, Floor S. Broekgaarden, et al. "Which black hole formed first? Mass-ratio reversal in massive binary stars from gravitational-wave data". In: *Mon. Not. Roy. Astron. Soc.* 517.2 (2022), pp. 2738–2745. DOI: 10.1093/mnras/stac2859. arXiv: 2205.12329 [astro-ph.HE].

[388] Suvodip Mukherjee and Joseph Silk. "Can we distinguish astrophysical from primordial black holes via the stochastic gravitational wave background?" In: *Mon. Not. Roy. Astron. Soc.* 506.3 (2021), pp. 3977–3985. DOI: 10.1093/mnras/stab1932. arXiv: 2105.11139 [gr-qc].

[389] Toshiya Namikawa. "Analyzing clustering of astrophysical gravitational-wave sources: Luminosity-distance space distortions". In: *JCAP* 01 (2021), p. 036. DOI: 10.1088/1475-7516/2021/01/036. arXiv: 2007.04359 [astro-ph.CO].

[390] Coenraad J. Neijssel, Alejandro Vigna-Gómez, Simon Stevenson, et al. "The effect of the metallicity-specific star formation history on double compact object mergers". In: *Mon. Not. Roy. Astron. Soc.* 490.3 (2019), pp. 3740–3759. DOI: 10.1093/mnras/stz2840. arXiv: 1906.08136 [astro-ph.SR].

[391] Ken K. Y. Ng, Salvatore Vitale, Will M. Farr, et al. "Probing multiple populations of compact binaries with third-generation gravitational-wave detectors". In: *Astrophys. J. Lett.* 913.1 (2021), p. L5. DOI: 10.3847/2041-8213/abf8be. arXiv: 2012.09876 [astro-ph.CO].

[392] Atsushi Nishizawa, Atsushi Taruya, Kazuhiro Hayama, et al. "Probing non-tensorial polarizations of stochastic gravitational-wave backgrounds with ground-based laser interferometers". In: *Phys. Rev. D* 79 (2009), p. 082002. DOI: 10.1103/PhysRevD.79.082002. arXiv: 0903.0528 [astro-ph.CO].

[393] Samaya Nissanke, Daniel E. Holz, Scott A. Hughes, et al. "Exploring short gamma-ray bursts as gravitational-wave standard sirens". In: *Astrophys. J.* 725 (2010), pp. 496–514. DOI: 10.1088/0004-637X/725/1/496. arXiv: 0904.1017 [astro-ph.CO].

[394] Alex Nitz et al. *gwastro/pycbc: PyCBC Release v1.16.13*. doi.org/10.5281/zenodo.4309869. Version v1.16.13. Dec. 2020. DOI: 10.5281/zenodo.4309869.

[395] Alexander H. Nitz, Tito Dal Canton, Derek Davis, et al. "Rapid detection of gravitational waves from compact binary mergers with PyCBC Live". In: *Phys. Rev. D* 98.2 (2018), p. 024050. DOI: 10.1103/PhysRevD.98.024050. arXiv: 1805.11174 [gr-qc].

[396] Alexander H. Nitz, Thomas Dent, Tito Dal Canton, et al. "Detecting binary compact-object mergers with gravitational waves: Understanding and Improving the sensitivity of the PyCBC search". In: *Astrophys. J.* 849.2 (2017), p. 118. DOI: 10.3847/1538-4357/aa8f50. arXiv: 1705.01513 [gr-qc].

[397] Alexander H. Nitz, Thomas Dent, Gareth S. Davies, et al. "2-OGC: Open Gravitational-wave Catalog of binary mergers from analysis of public Advanced LIGO and Virgo data". In: *Astrophys. J.* 891 (Mar. 2020), p. 123. DOI: 10.3847/1538-4357/ab733f. arXiv: 1910.05331 [astro-ph.HE].

[398] Alexander H. Nitz, Sumit Kumar, Yi-Fan Wang, et al. "4-OGC: Catalog of Gravitational Waves from Compact Binary Mergers". In: *Astrophys. J.* 946.2 (2023), p. 59. DOI: 10.3847/1538-4357/aca591. arXiv: 2112.06878 [astro-ph.HE].

[399] Alexander H. Nitz and Yi-Fan Wang. "Broad search for gravitational waves from subsolar-mass binaries through LIGO and Virgo's third observing run". In: *Phys. Rev. D* 106.2 (2022), p. 023024. DOI: 10.1103/PhysRevD.106.023024. arXiv: 2202.11024 [astro-ph.HE].

[400] I. D. Novikov, A. G. Polnarev, A. A. Starobinskii, et al. “Primordial black holes”. In: *A&A* 80.1 (Nov. 1979), pp. 104–109.

[401] R. O’Shaughnessy, L. London, J. Healy, et al. “Precession during merger: Strong polarization changes are observationally accessible features of strong-field gravity during binary black hole merger”. In: *Phys. Rev. D* 87.4 (2013), p. 044038. DOI: 10.1103/PhysRevD.87.044038. arXiv: 1209.3712 [gr-qc].

[402] Richard O’Shaughnessy, Davide Gerosa, and Daniel Wysocki. “Inferences about supernova physics from gravitational-wave measurements: GW151226 spin misalignment as an indicator of strong black-hole natal kicks”. In: *Phys. Rev. Lett.* 119.1 (2017), p. 011101. DOI: 10.1103/PhysRevLett.119.011101. arXiv: 1704.03879 [astro-ph.HE].

[403] Ryosuke Okuta, Yuya Unno, Daisuke Nishino, et al. “CuPy: A NumPy-Compatible Library for NVIDIA GPU Calculations”. In: *Proceedings of Workshop on Machine Learning Systems (LearningSys) in The Thirty-first Annual Conference on Neural Information Processing Systems (NIPS)*. 2017. URL: http://learningsys.org/nips17/assets/papers/paper_16.pdf.

[404] Aleksandra Olejak and Krzysztof Belczynski. “The Implications of High BH Spins on the Origin of BH–BH Mergers”. In: *Astrophys. J. Lett.* 921.1 (2021), p. L2. DOI: 10.3847/2041-8213/ac2f48. arXiv: 2109.06872 [astro-ph.HE].

[405] Aleksandra Olejak, Chris L. Fryer, Krzysztof Belczynski, et al. “The role of supernova convection for the lower mass gap in the isolated binary formation of gravitational wave sources”. In: *Mon. Not. Roy. Astron. Soc.* 516.2 (2022), pp. 2252–2271. DOI: 10.1093/mnras/stac2359. arXiv: 2204.09061 [astro-ph.HE].

[406] Seth Olsen, Tejaswi Venumadhav, Jonathan Mushkin, et al. “New binary black hole mergers in the LIGO-Virgo O3a data”. In: *Phys. Rev. D* 106.4 (2022), p. 043009. DOI: 10.1103/PhysRevD.106.043009. arXiv: 2201.02252 [astro-ph.HE].

[407] J. R. Oppenheimer and G. M. Volkoff. “On massive neutron cores”. In: *Phys. Rev.* 55 (1939), pp. 374–381. DOI: 10.1103/PhysRev.55.374.

[408] Serguei Ossokine et al. “Multipolar Effective-One-Body Waveforms for Precessing Binary Black Holes: Construction and Validation”. In: *Phys. Rev. D* 102.4 (2020), p. 044055. DOI: 10.1103/PhysRevD.102.044055. arXiv: 2004.09442 [gr-qc].

[409] Feryal Ozel, Dimitrios Psaltis, Ramesh Narayan, et al. “The Black Hole Mass Distribution in the Galaxy”. In: *Astrophys. J.* 725 (2010), pp. 1918–1927. DOI: 10.1088/0004-637X/725/2/1918. arXiv: 1006.2834 [astro-ph.GA].

[410] Feryal Özel and Paulo Freire. "Masses, Radii, and the Equation of State of Neutron Stars". In: *Ann. Rev. Astron. Astrophys.* 54 (2016), pp. 401–440. DOI: 10.1146/annurev-astro-081915-023322. arXiv: 1603.02698 [astro-ph.HE].

[411] Yi Pan, Alessandra Buonanno, Andrea Taracchini, et al. "Inspiral-merger-ringdown waveforms of spinning, precessing black-hole binaries in the effective-one-body formalism". In: *Phys. Rev. D* 89.8 (2014), p. 084006. DOI: 10.1103/PhysRevD.89.084006. arXiv: 1307.6232 [gr-qc].

[412] Peter T. H. Pang et al. "An updated nuclear-physics and multi-messenger astrophysics framework for binary neutron star mergers". In: *Nature Commun.* 14.1 (2023), p. 8352. DOI: 10.1038/s41467-023-43932-6. arXiv: 2205.08513 [astro-ph.HE].

[413] Peter T. H. Pang, Ingo Tews, Michael W. Coughlin, et al. "Nuclear Physics Multimessenger Astrophysics Constraints on the Neutron Star Equation of State: Adding NICER's PSR J0740+6620 Measurement". In: *Astrophys. J.* 922.1 (2021), p. 14. DOI: 10.3847/1538-4357/ac19ab. arXiv: 2105.08688 [astro-ph.HE].

[414] Bill Paxton et al. "Modules for Experiments in Stellar Astrophysics (MESA): Pulsating Variable Stars, Rotation, Convective Boundaries, and Energy Conservation". In: *Astrophys. J. Suppl.* 243.1 (2019), p. 10. DOI: 10.3847/1538-4365/ab2241. arXiv: 1903.01426 [astro-ph.SR].

[415] Ethan Payne, Sharan Banagiri, Paul Lasky, et al. "Searching for anisotropy in the distribution of binary black hole mergers". In: *Phys. Rev. D* 102.10 (2020), p. 102004. DOI: 10.1103/PhysRevD.102.102004. arXiv: 2006.11957 [astro-ph.CO].

[416] Ethan Payne, Sophie Hourihane, Jacob Golomb, et al. *Data Release for the curious case of GW200129: interplay between spin-precession inference and data-quality issues*. 2022. DOI: 10.5281/zenodo.7259655. URL: https://zenodo.org/record/7259655.

[417] Ethan Payne, Maximiliano Isi, Katerina Chatziioannou, et al. "Fortifying gravitational-wave tests of general relativity against astrophysical assumptions". In: *Phys. Rev. D* 108.12 (2023), p. 124060. DOI: 10.1103/PhysRevD.108.124060. arXiv: 2309.04528 [gr-qc].

[418] Ethan Payne, Colm Talbot, Paul D. Lasky, et al. "Gravitational-wave astronomy with a physical calibration model". In: *Phys. Rev. D* 102 (2020), p. 122004. DOI: 10.1103/PhysRevD.102.122004. arXiv: 2009.10193 [astro-ph.IM].

[419] Ethan Payne, Colm Talbot, and Eric Thrane. "Higher order gravitational-wave modes with likelihood reweighting". In: *Phys. Rev. D* 100.12 (2019), p. 123017. DOI: 10.1103/PhysRevD.100.123017. arXiv: 1905.05477 [astro-ph.IM].

[420] Ethan Payne and Eric Thrane. "Model exploration in gravitational-wave astronomy with the maximum population likelihood". In: *Phys. Rev. Res.* 5.2 (2023), p. 023013. DOI: 10.1103/PhysRevResearch.5.023013. arXiv: 2210.11641 [astro-ph.IM].

[421] Fabian Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *J. Machine Learning Res.* 12 (2011), pp. 2825–2830. arXiv: 1201.0490 [cs.LG].

[422] Malcolm Perry and Maria J. Rodriguez. "Dynamical Love Numbers for Kerr Black Holes". In: (Oct. 2023). arXiv: 2310.03660 [gr-qc].

[423] Polina Petrov, Leo P. Singer, Michael W. Coughlin, et al. "Data-driven Expectations for Electromagnetic Counterpart Searches Based on LIGO/Virgo Public Alerts". In: *Astrophys. J.* 924.2 (2022), p. 54. DOI: 10.3847/1538-4357/ac366d. arXiv: 2108.07277 [astro-ph.HE].

[424] Du Phan, Neeraj Pradhan, and Martin Jankowiak. "Composable Effects for Flexible and Accelerated Probabilistic Programming in NumPyro". In: Dec. 2019. arXiv: 1912.11554 [stat.ML].

[425] E. S. Phinney. "A Practical theorem on gravitational wave backgrounds". In: (July 2001). arXiv: astro-ph/0108028.

[426] Eric Poisson and Clifford M. Will. "Gravitational waves from inspiraling compact binaries: Parameter estimation using second postNewtonian wave forms". In: *Phys. Rev. D* 52 (1995), pp. 848–855. DOI: 10.1103/PhysRevD.52.848. arXiv: gr-qc/9502040.

[427] Lorenzo Pompili et al. "Laying the foundation of the effective-one-body waveform models SEOBNRv5: Improved accuracy and efficiency for spinning nonprecessing binary black holes". In: *Phys. Rev. D* 108.12 (2023), p. 124035. DOI: 10.1103/PhysRevD.108.124035. arXiv: 2303.18039 [gr-qc].

[428] Geraint Pratten et al. "Computationally efficient models for the dominant and subdominant harmonic modes of precessing binary black holes". In: *Phys. Rev. D* 103.10 (2021), p. 104056. DOI: 10.1103/PhysRevD.103.104056. arXiv: 2004.06503 [gr-qc].

[429] Geraint Pratten, Sascha Husa, Cecilio Garcia-Quiros, et al. "Setting the cornerstone for a family of models for gravitational waves from compact binaries: The dominant harmonic for nonprecessing quasicircular black holes". In: *Phys. Rev. D* 102.6 (2020), p. 064001. DOI: 10.1103/PhysRevD.102.064001. arXiv: 2001.11412 [gr-qc].

[430] Geraint Pratten, Patricia Schmidt, Riccardo Buscicchio, et al. "Measuring precession in asymmetric compact binaries". In: *Phys. Rev. Res.* 2.4 (2020), p. 043096. DOI: 10.1103/PhysRevResearch.2.043096. arXiv: 2006.16153 [gr-qc].

[431] Geraint Pratten, Patricia Schmidt, and Natalie Williams. "Impact of Dynamical Tides on the Reconstruction of the Neutron Star Equation of State". In: *Phys. Rev. Lett.* 129.8 (2022), p. 081102. DOI: 10.1103/PhysRevLett.129.081102. arXiv: 2109.07566 [astro-ph.HE].

[432] Frans Pretorius. "Evolution of binary black hole spacetimes". In: *Phys. Rev. Lett.* 95 (2005), p. 121101. DOI: 10.1103/PhysRevLett.95.121101. arXiv: gr-qc/0507014.

[433] 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*". In: *Astrophys. J.* 935.2 (2022), p. 167. DOI: 10.3847/1538-4357/ac7c74. arXiv: 2206.14220 [astro-ph.IM].

[434] Michael Pürrer. "Frequency domain reduced order model of aligned-spin effective-one-body waveforms with generic mass-ratios and spins". In: *Phys. Rev. D* 93.6 (2016), p. 064041. DOI: 10.1103/PhysRevD.93.064041. arXiv: 1512.02248 [gr-qc].

[435] Michael Pürrer and Carl-Johan Haster. "Gravitational waveform accuracy requirements for future ground-based detectors". In: *Phys. Rev. Res.* 2.2 (2020), p. 023151. DOI: 10.1103/PhysRevResearch.2.023151. arXiv: 1912.10055 [gr-qc].

[436] Y. Qin, T. Fragos, G. Meynet, et al. "The spin of the second-born black hole in coalescing binary black holes". In: *Astron. Astrophys.* 616 (2018), A28. DOI: 10.1051/0004-6361/201832839. arXiv: 1802.05738 [astro-ph.SR].

[437] Ying Qin, Pablo Marchant, Tassos Fragos, et al. "On the Origin of Black-Hole Spin in High-Mass X-ray Binaries". In: *Astrophys. J. Lett.* 870.2 (2019), p. L18. DOI: 10.3847/2041-8213/aaf97b. arXiv: 1810.13016 [astro-ph.SR].

[438] G. Raaijmakers et al. "Constraining the dense matter equation of state with joint analysis of NICER and LIGO/Virgo measurements". In: *Astrophys. J. Lett.* 893.1 (2020), p. L21. DOI: 10.3847/2041-8213/ab822f. arXiv: 1912.11031 [astro-ph.HE].

[439] G. Raaijmakers, S. K. Greif, K. Hebeler, et al. "Constraints on the Dense Matter Equation of State and Neutron Star Properties from NICER's Mass–Radius Estimate of PSR J0740+6620 and Multimessenger Observations". In: *Astrophys. J. Lett.* 918.2 (2021), p. L29. DOI: 10.3847/2041-8213/ac089a. arXiv: 2105.06981 [astro-ph.HE].

[440] Etienne Racine. "Analysis of spin precession in binary black hole systems including quadrupole-monopole interaction". In: *Phys. Rev. D* 78 (2008), p. 044021. DOI: 10.1103/PhysRevD.78.044021. arXiv: 0803.1820 [gr-qc].

[441] David Radice, Albino Perego, Francesco Zappa, et al. "GW170817: Joint Constraint on the Neutron Star Equation of State from Multimessenger Observations". In: *Astrophys. J. Lett.* 852.2 (2018), p. L29. DOI: 10.3847/2041-8213/aaa402. arXiv: 1711.03647 [astro-ph.HE].

[442] Ninoy Rahman, Hans-Thomas Janka, Georg Stockinger, et al. "Pulsational pair-instability supernovae: gravitational collapse, black hole formation, and beyond". In: *Mon. Not. Roy. Astron. Soc.* 512.3 (2022), pp. 4503–4540. DOI: 10.1093/mnras/stac758. arXiv: 2112.09707 [astro-ph.HE].

[443] G. Rakavy and G. Shaviv. "Instabilities in Highly Evolved Stellar Models". In: *ApJ* 148 (June 1967), p. 803. DOI: 10.1086/149204.

[444] Antoni Ramos-Buades, Alessandra Buonanno, Héctor Estellés, et al. "Next generation of accurate and efficient multipolar precessing-spin effective-one-body waveforms for binary black holes". In: *Phys. Rev. D* 108.12 (2023), p. 124037. DOI: 10.1103/PhysRevD.108.124037. arXiv: 2303.18046 [gr-qc].

[445] Antoni Ramos-Buades, Patricia Schmidt, Geraint Pratten, et al. "Validity of common modeling approximations for precessing binary black holes with higher-order modes". In: *Phys. Rev. D* 101.10 (2020), p. 103014. DOI: 10.1103/PhysRevD.101.103014. arXiv: 2001.10936 [gr-qc].

[446] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian processes for machine learning*. Adaptive computation and machine learning. MIT Press, 2006. ISBN: 9780262182539.

[447] Anarya Ray, Ignacio Magaña Hernandez, Siddharth Mohite, et al. "Nonparametric Inference of the Population of Compact Binaries from Gravitational-wave Observations Using Binned Gaussian Processes". In: *Astrophys. J.* 957.1 (2023), p. 37. DOI: 10.3847/1538-4357/acf452. arXiv: 2304.08046 [gr-qc].

[448] Jocelyn S. Read, Benjamin D. Lackey, Benjamin J. Owen, et al. "Constraints on a phenomenologically parameterized neutron-star equation of state". In: *Phys. Rev. D* 79 (2009), p. 124032. DOI: 10.1103/PhysRevD.79.124032. arXiv: 0812.2163 [astro-ph].

[449] Tania Regimbau. "Stochastic background from inspiralling double neutron stars". In: *Phys. Rev. D* 75 (2007), p. 043002. DOI: 10.1103/PhysRevD.75.043002. arXiv: astro-ph/0701004.

[450] Tania Regimbau. "The astrophysical gravitational wave stochastic background". In: *Res. Astron. Astrophys.* 11 (2011), pp. 369–390. DOI: 10.1088/1674-4527/11/4/001. arXiv: 1101.2762 [astro-ph.CO].

[451] Tania Regimbau. "The Quest for the Astrophysical Gravitational-Wave Background with Terrestrial Detectors". In: *Symmetry* 14.2 (2022), p. 270. DOI: 10.3390/sym14020270.

[452] Ronald A. Remillard and Jeffrey E. McClintock. "X-ray Properties of Black-Hole Binaries". In: *Ann. Rev. Astron. Astrophys.* 44 (2006), pp. 49–92. DOI: 10.1146/annurev.astro.44.051905.092532. arXiv: astro-ph/0606352.

[453] Arianna Renzini and Carlo Contaldi. "Improved limits on a stochastic gravitational-wave background and its anisotropies from Advanced LIGO O1 and O2 runs". In: *Phys. Rev. D* 100.6 (2019), p. 063527. DOI: 10.1103/PhysRevD.100.063527. arXiv: 1907.10329 [gr-qc].

[454] Arianna I. Renzini, Tom Callister, Katerina Chatziioannou, et al. "Background information: A study on the sensitivity of astrophysical gravitational-wave background searches". In: *Phys. Rev. D* 110.2 (2024), p. 023014. DOI: 10.1103/PhysRevD.110.023014. arXiv: 2403.14793 [astro-ph.HE].

[455] Arianna I. Renzini, Boris Goncharov, Alexander C. Jenkins, et al. "Stochastic Gravitational-Wave Backgrounds: Current Detection Efforts and Future Prospects". In: *Galaxies* 10.1 (2022), p. 34. DOI: 10.3390/galaxies10010034. arXiv: 2202.00178 [gr-qc].

[456] Arianna I. Renzini, Joseph D. Romano, Carlo R. Contaldi, et al. "Comparison of maximum-likelihood mapping methods for gravitational-wave backgrounds". In: *Phys. Rev. D* 105.2 (2022), p. 023519. DOI: 10.1103/PhysRevD.105.023519. arXiv: 2107.02292 [gr-qc].

[457] M. Renzo, R. Farmer, S. Justham, et al. "Predictions for the hydrogen-free ejecta of pulsational pair-instability supernovae". In: *Astron. Astrophys.* 640 (2020), A56. DOI: 10.1051/0004-6361/202037710. arXiv: 2002.05077 [astro-ph.SR].

[458] M. Renzo, R. J. Farmer, S. Justham, et al. "Sensitivity of the lower-edge of the pair instability black hole mass gap to the treatment of time dependent convection". In: *Mon. Not. Roy. Astron. Soc.* 493.3 (2020), pp. 4333–4341. DOI: 10.1093/mnras/staa549. arXiv: 2002.08200 [astro-ph.SR].

[459] Luciano Rezzolla, Elias R. Most, and Lukas R. Weih. "Using gravitational-wave observations and quasi-universal relations to constrain the maximum mass of neutron stars". In: *Astrophys. J. Lett.* 852.2 (2018), p. L25. DOI: 10.3847/2041-8213/aaa401. arXiv: 1711.00314 [astro-ph.HE].

[460] Clifford E. Rhoades Jr. and Remo Ruffini. "Maximum mass of a neutron star". In: *Phys. Rev. Lett.* 32 (1974), pp. 324–327. DOI: 10.1103/PhysRevLett.32.324.

[461] Thomas E. Riley et al. "A *NICER* View of PSR J0030+0451: Millisecond Pulsar Parameter Estimation". In: *Astrophys. J. Lett.* 887.1 (2019), p. L21. DOI: 10.3847/2041-8213/ab481c. arXiv: 1912.05702 [astro-ph.HE].

[462] Thomas E. Riley et al. "A NICER View of the Massive Pulsar PSR J0740+6620 Informed by Radio Timing and XMM-Newton Spectroscopy". In: *Astrophys. J. Lett.* 918.2 (2021), p. L27. DOI: 10.3847/2041-8213/ac0a81. arXiv: 2105.06980 [astro-ph.HE].

[463] Stefano Rinaldi and Walter Del Pozzo. "(H)DPGMM: a hierarchy of Dirichlet process Gaussian mixture models for the inference of the black hole mass function". In: *Mon. Not. Roy. Astron. Soc.* 509.4 (2021), pp. 5454–5466. DOI: 10.1093/mnras/stab3224. arXiv: 2109.05960 [astro-ph.IM].

[464] Mallory S. E. Roberts. "Surrounded by Spiders! New Black Widows and Redbacks in the Galactic Field". In: *IAU Symp.* 291 (2013). Ed. by Joeri van Leeuwen, pp. 127–132. DOI: 10.1017/S174392131202337X. arXiv: 1210.6903 [astro-ph.HE].

[465] Florent Robinet, Nicolas Arnaud, Nicolas Leroy, et al. "Omicron: a tool to characterize transient noise in gravitational-wave detectors". In: *SoftwareX* 12 (2020), p. 100620. DOI: 10.1016/j.softx.2020.100620. arXiv: 2007.11374 [astro-ph.IM].

[466] Carl L. Rodriguez, Pau Amaro-Seoane, Sourav Chatterjee, et al. "Post-Newtonian Dynamics in Dense Star Clusters: Highly-Eccentric, Highly-Spinning, and Repeated Binary Black Hole Mergers". In: *Phys. Rev. Lett.* 120.15 (2018), p. 151101. DOI: 10.1103/PhysRevLett.120.151101. arXiv: 1712.04937 [astro-ph.HE].

[467] Carl L. Rodriguez, Michael Zevin, Pau Amaro-Seoane, et al. "Black holes: The next generation—repeated mergers in dense star clusters and their gravitational-wave properties". In: *Phys. Rev. D* 100.4 (2019), p. 043027. DOI: 10.1103/PhysRevD.100.043027. arXiv: 1906.10260 [astro-ph.HE].

[468] Carl L. Rodriguez, Michael Zevin, Chris Pankow, et al. "Illuminating Black Hole Binary Formation Channels with Spins in Advanced LIGO". In: *Astrophys. J. Lett.* 832.1 (2016), p. L2. DOI: 10.3847/2041-8205/832/1/L2. arXiv: 1609.05916 [astro-ph.HE].

[469] Joseph D. Romano and Neil J. Cornish. "Detection methods for stochastic gravitational-wave backgrounds: a unified treatment". In: *Living Rev. Rel.* 20.1 (2017), p. 2. DOI: 10.1007/s41114-017-0004-1. arXiv: 1608.06889 [gr-qc].

[470] 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". In: *Mon. Not. Roy. Astron. Soc.* 499.3 (2020), pp. 3295–3319. DOI: 10.1093/mnras/staa2850. arXiv: 2006.00714 [astro-ph.IM].

[471] Isobel M. Romero-Shaw, Nicholas Farrow, Simon Stevenson, et al. "On the origin of GW190425". In: *Mon. Not. Roy. Astron. Soc.* 496.1 (2020),

pp. L64–L69. DOI: 10.1093/mnrasl/slaa084. arXiv: 2001.06492 [astro-ph.HE].

[472] Isobel M. Romero-Shaw, Eric Thrane, and Paul D. Lasky. "When models fail: An introduction to posterior predictive checks and model misspecification in gravitational-wave astronomy". In: *Publ. Astron. Soc. Austral.* 39 (2022), e025. DOI: 10.1017/pasa.2022.24. arXiv: 2202.05479 [astro-ph.IM].

[473] Murray Rosenblatt. "Remarks on Some Nonparametric Estimates of a Density Function". In: *The Annals of Mathematical Statistics* 27.3 (Sept. 1956), pp. 832–837. ISSN: 0003-4851, 2168-8990. DOI: 10.1214/aoms/1177728190.

[474] Javier Roulet, Horng Sheng Chia, Seth Olsen, et al. "Distribution of effective spins and masses of binary black holes from the LIGO and Virgo O1–O3a observing runs". In: *Phys. Rev. D* 104.8 (2021), p. 083010. DOI: 10.1103/PhysRevD.104.083010. arXiv: 2105.10580 [astro-ph.HE].

[475] Christian Rover, Renate Meyer, and Nelson Christensen. "Modelling coloured residual noise in gravitational-wave signal processing". In: *Class. Quant. Grav.* 28 (2011), p. 015010. DOI: 10.1088/0264-9381/28/1/015010. arXiv: 0804.3853 [stat.ME].

[476] Nathan Rutherford et al. "Constraining the Dense Matter Equation of State with New NICER Mass–Radius Measurements and New Chiral Effective Field Theory Inputs". In: *Astrophys. J. Lett.* 971.1 (2024), p. L19. DOI: 10.3847/2041-8213/ad5f02. arXiv: 2407.06790 [astro-ph.HE].

[477] 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". In: (Jan. 2019). arXiv: 1901.08580 [gr-qc].

[478] Surabhi Sachdev, Tania Regimbau, and B. S. Sathyaprakash. "Subtracting compact binary foreground sources to reveal primordial gravitational-wave backgrounds". In: *Phys. Rev. D* 102.2 (2020), p. 024051. DOI: 10.1103/PhysRevD.102.024051. arXiv: 2002.05365 [gr-qc].

[479] Jayana A. Saes, Raissa F. P. Mendes, and Nicolás Yunes. "Approximately universal I-Love-<cs2> relations for the average neutron star stiffness". In: *Phys. Rev. D* 110.2 (2024), p. 024011. DOI: 10.1103/PhysRevD.110.024011. arXiv: 2402.05997 [gr-qc].

[480] Mohammadtaher Safarzadeh and Will M. Farr. "The Impact of metallicity evolution of the universe on the maximum mass of LIGO binary black holes". In: *Astrophys. J. Lett.* 883.1 (2019), p. L24. DOI: 10.3847/2041-8213/ab40bd. arXiv: 1909.01356 [astro-ph.HE].

[481] Mohammadtaher Safarzadeh, Enrico Ramirez-Ruiz, and Edo Berger. "Does GW190425 require an alternative formation pathway than a fast-merging channel?" In: *Astrophys. J.* 900.1 (2020), p. 13. DOI: 10.3847/1538-4357/aba596. arXiv: 2001.04502 [astro-ph.HE].

[482] Alexander Saffer, Nicolas Yunes, and Kent Yagi. "The gravitational wave stress–energy (pseudo)-tensor in modified gravity". In: *Class. Quant. Grav.* 35.5 (2018), p. 055011. DOI: 10.1088/1361-6382/aaa7de. arXiv: 1710.08863 [gr-qc].

[483] Mohit Raj Sah and Suvodip Mukherjee. "Non-stationary astrophysical stochastic gravitational-wave background: a new probe to the high-redshift population of binary black holes". In: *Monthly Notices of the Royal Astronomical Society* 527.2 (Nov. 2023), pp. 4100–4111. ISSN: 0035-8711. DOI: 10.1093/mnras/stad3365. eprint: https://academic.oup.com/mnras/article-pdf/527/2/4100/53789042/stad3365.pdf. URL: https://doi.org/10.1093/mnras/stad3365.

[484] A. L. Sallaska, C. Iliadis, A. E. Champagne, et al. "STARLIB: A Next-Generation Reaction-Rate Library for Nuclear Astrophysics". In: *Astrophys. J. Suppl.* 207 (2013), p. 18. DOI: 10.1088/0067-0049/207/1/18. arXiv: 1304.7811 [astro-ph.SR].

[485] Edwin E. Salpeter. "The Luminosity function and stellar evolution". In: *Astrophys. J.* 121 (1955), pp. 161–167. DOI: 10.1086/145971.

[486] L. Santamaria et al. "Matching post-Newtonian and numerical relativity waveforms: systematic errors and a new phenomenological model for non-precessing black hole binaries". In: *Phys. Rev. D* 82 (2010), p. 064016. DOI: 10.1103/PhysRevD.82.064016. arXiv: 1005.3306 [gr-qc].

[487] Filippo Santoliquido, Michela Mapelli, Nicola Giacobbo, et al. "The cosmic merger rate density of compact objects: impact of star formation, metallicity, initial mass function and binary evolution". In: *Mon. Not. Roy. Astron. Soc.* 502.4 (2021), pp. 4877–4889. DOI: 10.1093/mnras/stab280. arXiv: 2009.03911 [astro-ph.HE].

[488] 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". In: *Phys. Rev. D* 86 (2012), p. 104063. DOI: 10.1103/PhysRevD.86.104063. arXiv: 1207.3088 [gr-qc].

[489] Patricia Schmidt, Mark Hannam, Sascha Husa, et al. "Tracking the precession of compact binaries from their gravitational-wave signal". In: *Phys. Rev. D* 84 (2011), p. 024046. DOI: 10.1103/PhysRevD.84.024046. arXiv: 1012.2879 [gr-qc].

[490] 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". In: *Phys. Rev. D* 91.2 (2015), p. 024043. DOI: 10.1103/PhysRevD.91.024043. arXiv: 1408.1810 [gr-qc].

[491] F. R. N. Schneider, H. Sana, C. J. Evans, et al. "An excess of massive stars in the local 30 Doradus starburst". In: *Science* 359.6371 (Jan. 2018), pp. 69–71. DOI: 10.1126/science.aan0106. arXiv: 1801.03107 [astro-ph.SR].

[492] Bernard F. Schutz. *A FIRST COURSE IN GENERAL RELATIVITY*. Cambridge, UK: Cambridge Univ. Pr., 1985. ISBN: 978-0-511-98418-1. DOI: 10.1017/CBO9780511984181.

[493] Alberto Sesana, Alberto Vecchio, and Carlo Nicola Colacino. "The stochastic gravitational-wave background from massive black hole binary systems: implications for observations with Pulsar Timing Arrays". In: *Mon. Not. Roy. Astron. Soc.* 390 (2008), p. 192. DOI: 10.1111/j.1365-2966.2008.13682.x. arXiv: 0804.4476 [astro-ph].

[494] Sahar Shahaf, Dolev Bashi, Tsevi Mazeh, et al. "Triage of the Gaia DR3 astrometric orbits – I. A sample of binaries with probable compact companions". In: *Mon. Not. Roy. Astron. Soc.* 518.2 (2022), pp. 2991–3003. DOI: 10.1093/mnras/stac3290. arXiv: 2209.00828 [astro-ph.SR].

[495] Yong Shao. "On the Neutron Star/Black Hole Mass Gap and Black Hole Searches". In: *Res. Astron. Astrophys.* 22.12 (2022), p. 122002. DOI: 10.1088/1674-4527/ac995e. arXiv: 2210.00425 [astro-ph.HE].

[496] Yong Shao and Xiang-Dong Li. "Population Synthesis of Black Hole Binaries with Compact Star Companions". In: *Astrophys. J.* 920.2 (2021), p. 81. DOI: 10.3847/1538-4357/ac173e. arXiv: 2107.03565 [astro-ph.HE].

[497] Hector O. Silva, A. Miguel Holgado, Alejandro Cárdenas-Avendaño, et al. "Astrophysical and theoretical physics implications from multimessenger neutron star observations". In: *Phys. Rev. Lett.* 126.18 (2021), p. 181101. DOI: 10.1103/PhysRevLett.126.181101. arXiv: 2004.01253 [gr-qc].

[498] Hector O. Silva, Hajime Sotani, and Emanuele Berti. "Low-mass neutron stars: universal relations, the nuclear symmetry energy and gravitational radiation". In: *Mon. Not. Roy. Astron. Soc.* 459.4 (2016), pp. 4378–4388. DOI: 10.1093/mnras/stw969. arXiv: 1601.03407 [astro-ph.HE].

[499] Leo P. Singer and Larry R. Price. "Rapid Bayesian position reconstruction for gravitational-wave transients". In: *Phys. Rev. D* 93.2 (2016), p. 024013. DOI: 10.1103/PhysRevD.93.024013. arXiv: 1508.03634 [gr-qc].

[500] John Skilling. "Nested sampling for general Bayesian computation". In: *Bayesian Analysis* 1.4 (2006), pp. 833–859. DOI: 10.1214/06-BA127.

[501] Rory Smith, Scott E. Field, Kent Blackburn, et al. "Fast and accurate inference on gravitational waves from precessing compact binaries". In: *Phys. Rev. D* 94.4 (2016), p. 044031. DOI: 10.1103/PhysRevD.94.044031. arXiv: 1604.08253 [gr-qc].

[502] Rory J. E. Smith, Gregory Ashton, Avi Vajpeyi, et al. "Massively parallel Bayesian inference for transient gravitational-wave astronomy". In: *Mon. Not. Roy. Astron. Soc.* 498.3 (2020), pp. 4492–4502. DOI: 10.1093/mnras/staa2483. arXiv: 1909.11873 [gr-qc].

[503] Rahul Somasundaram, Ingo Tews, and Jérôme Margueron. "Perturbative QCD and the neutron star equation of state". In: *Phys. Rev. C* 107.5 (2023), p. L052801. DOI: 10.1103/PhysRevC.107.L052801. arXiv: 2204.14039 [nucl-th].

[504] L. A. C. van Son, S. E. de Mink, T. Callister, et al. "The Redshift Evolution of the Binary Black Hole Merger Rate: A Weighty Matter". In: *Astrophys. J.* 931.1 (2022), p. 17. DOI: 10.3847/1538-4357/ac64a3. arXiv: 2110.01634 [astro-ph.HE].

[505] Hanlin Song, Dicong Liang, Ziming Wang, et al. "Impact of spin in compact binary foreground subtraction for estimating the residual stochastic gravitational-wave background in ground-based detectors". In: *Phys. Rev. D* 109.12 (2024), p. 123014. DOI: 10.1103/PhysRevD.109.123014. arXiv: 2401.00984 [gr-qc].

[506] S. Soni et al. "Reducing scattered light in LIGO's third observing run". In: *Class. Quant. Grav.* 38.2 (2020), p. 025016. DOI: 10.1088/1361-6382/abc906. arXiv: 2007.14876 [astro-ph.IM].

[507] Joshua S. Speagle. "dynesty: a dynamic nested sampling package for estimating Bayesian posteriors and evidences". In: *Mon. Not. Roy. Astron. Soc.* 493.3 (2020), pp. 3132–3158. DOI: 10.1093/mnras/staa278. arXiv: 1904.02180 [astro-ph.IM].

[508] Mario Spera, Michela Mapelli, Nicola Giacobbo, et al. "Merging black hole binaries with the SEVN code". In: *Mon. Not. Roy. Astron. Soc.* 485.1 (2019), pp. 889–907. DOI: 10.1093/mnras/stz359. arXiv: 1809.04605 [astro-ph.HE].

[509] Leo C. Stein. "qnm: A Python package for calculating Kerr quasinormal modes, separation constants, and spherical-spheroidal mixing coefficients". In: *J. Open Source Softw.* 4.42 (2019), p. 1683. DOI: 10.21105/joss.01683. arXiv: 1908.10377 [gr-qc].

[510] Simon Stevenson, Christopher P. L. Berry, and Ilya Mandel. "Hierarchical analysis of gravitational-wave measurements of binary black hole spin–orbit misalignments". In: *Mon. Not. Roy. Astron. Soc.* 471.3 (2017), pp. 2801–2811. DOI: 10.1093/mnras/stx1764. arXiv: 1703.06873 [astro-ph.HE].

[511] Simon Stevenson, Matthew Sampson, Jade Powell, et al. "The impact of pair-instability mass loss on the binary black hole mass distribution". In: (Apr. 2019). DOI: 10.3847/1538-4357/ab3981. arXiv: 1904.02821 [astro-ph.HE].

[512] Klaus Strobel and Manfred K. Weigel. "On the minimum and maximum mass of neutron stars and the delayed collapse". In: *Astron. Astrophys.* 367 (2001), p. 582. DOI: 10.1051/0004-6361:20000428. arXiv: astro-ph/0012321.

[513] Ling Sun et al. "Characterization of systematic error in Advanced LIGO calibration". In: *Class. Quant. Grav.* 37.22 (2020), p. 225008. DOI: 10.1088/1361-6382/abb14e. arXiv: 2005.02531 [astro-ph.IM].

[514] Yudai Suwa, Takashi Yoshida, Masaru Shibata, et al. "On the minimum mass of neutron stars". In: *Mon. Not. Roy. Astron. Soc.* 481.3 (2018), pp. 3305–3312. DOI: 10.1093/mnras/sty2460. arXiv: 1808.02328 [astro-ph.HE].

[515] Hiromichi Tagawa, Zoltan Haiman, Imre Bartos, et al. "Spin Evolution of Stellar-mass Black Hole Binaries in Active Galactic Nuclei". In: *Astrophys. J.* 899.1 (2020), p. 26. DOI: 10.3847/1538-4357/aba2cc. arXiv: 2004.11914 [astro-ph.HE].

[516] Hiromichi Tagawa, Bence Kocsis, Zoltan Haiman, et al. "Mass-gap Mergers in Active Galactic Nuclei". In: *Astrophys. J.* 908.2 (2021), p. 194. DOI: 10.3847/1538-4357/abd555. arXiv: 2012.00011 [astro-ph.HE].

[517] Colm Talbot. "GWPopulation pipe". In: (Nov. 2021). DOI: 10.5281/zenodo.5654673. URL: https://git.ligo.org/RatesAndPopulations/gwpopulation_pipe.

[518] Colm Talbot, Amanda Farah, Shanika Galaudage, et al. "GWPopulation: Hardware agnostic population inference for compact binaries and beyond". In: *J. Open Source Softw.* (Sept. 2024). DOI: 10.21105/joss.07753. arXiv: 2409.14143 [astro-ph.IM].

[519] Colm Talbot and Jacob Golomb. "Growing pains: understanding the impact of likelihood uncertainty on hierarchical Bayesian inference for gravitational-wave astronomy". In: *Mon. Not. Roy. Astron. Soc.* 526.3 (2023), pp. 3495–3503. DOI: 10.1093/mnras/stad2968. arXiv: 2304.06138 [astro-ph.IM].

[520] Colm Talbot, Rory Smith, Eric Thrane, et al. "Parallelized Inference for Gravitational-Wave Astronomy". In: *Phys. Rev. D* 100.4 (2019), p. 043030. DOI: 10.1103/PhysRevD.100.043030. arXiv: 1904.02863 [astro-ph.IM].

[521] Colm Talbot and Eric Thrane. "Determining the population properties of spinning black holes". In: *Phys. Rev. D* 96.2 (2017), p. 023012. DOI: 10.1103/PhysRevD.96.023012. arXiv: 1704.08370 [astro-ph.HE].

[522] Colm Talbot and Eric Thrane. "Flexible and Accurate Evaluation of Gravitational-wave Malmquist Bias with Machine Learning". In: *Astrophys. J.* 927.1 (2022), p. 76. DOI: 10.3847/1538-4357/ac4bc0. arXiv: 2012.01317 [gr-qc].

[523] Colm Talbot and Eric Thrane. "Gravitational-wave astronomy with an uncertain noise power spectral density". In: *Phys. Rev. Res.* 2.4 (2020), p. 043298. DOI: 10.1103/PhysRevResearch.2.043298. arXiv: 2006.05292 [astro-ph.IM].

[524] Colm Talbot and Eric Thrane. "Measuring the binary black hole mass spectrum with an astrophysically motivated parameterization". In: *Astrophys. J.* 856.2 (2018), p. 173. DOI: 10.3847/1538-4357/aab34c. arXiv: 1801.02699 [astro-ph.HE].

[525] Andrea Taracchini et al. "Effective-one-body model for black-hole binaries with generic mass ratios and spins". In: *Phys. Rev. D* 89.6 (2014), p. 061502. DOI: 10.1103/PhysRevD.89.061502. arXiv: 1311.2544 [gr-qc].

[526] The pandas development team. *pandas-dev/pandas: Pandas*. Version latest. Feb. 2020. DOI: 10.5281/zenodo.3509134. URL: https://doi.org/10.5281/zenodo.3509134.

[527] Ingo Tews, Joseph Carlson, Stefano Gandolfi, et al. "Constraining the speed of sound inside neutron stars with chiral effective field theory interactions and observations". In: *Astrophys. J.* 860.2 (2018), p. 149. DOI: 10.3847/1538-4357/aac267. arXiv: 1801.01923 [nucl-th].

[528] The LIGO Scientific Collaboration, The Virgo Collaboration, and The KAGRA Collaboration. *GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run — Data behind the figures*. Mar. 2022. DOI: 10.5281/zenodo.6368595. URL: https://doi.org/10.5281/zenodo.6368595.

[529] The LIGO Scientific Collaboration, the Virgo Collaboration, and the KAGRA Collaboration. *GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run — O1+O2+O3 Search Sensitivity Estimates*. Nov. 2021. DOI: 10.5281/zenodo.5636816.

[530] The LIGO Scientific Collaboration, the Virgo Collaboration, and the KAGRA Collaboration. *GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run — O3 search sensitivity estimates*. Nov. 2021. DOI: 10.5281/zenodo.5546676.

[531] The LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration. *GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run — O3 search sensitivity estimates*. LIGO Laboratory and Advanced LIGO are funded by the United States National Science Foundation (NSF) as well as the Science and Technology Facilities Council (STFC) of the United Kingdom, the Max-Planck-Society (MPS), and the State of Niedersachsen/Germany for support of the construction of Advanced LIGO and construction and operation of the GEO600 detector. Additional support for Advanced LIGO

was provided by the Australian Research Council. Virgo is funded, through the European Gravitational Observatory (EGO), by the French Centre National de Recherche Scientifique (CNRS), the Italian Istituto Nazionale di Fisica Nucleare (INFN) and the Dutch Nikhef, with contributions by institutions from Belgium, Germany, Greece, Hungary, Ireland, Japan, Monaco, Poland, Portugal, Spain. The construction and operation of KAGRA are funded by Ministry of Education, Culture, Sports, Science and Technology (MEXT), and Japan Society for the Promotion of Science (JSPS), National Research Foundation (NRF) and Ministry of Science and ICT (MSIT) in Korea, Academia Sinica (AS) and the Ministry of Science and Technology (MoST) in Taiwan. Nov. 2021. DOI: 10.5281/zenodo.5546676. URL: https://doi.org/10.5281/zenodo.5546676.

[532] Todd A. Thompson et al. "Discovery of a Candidate Black Hole - Giant Star Binary System in the Galactic Field". In: (June 2018). DOI: 10.1126/science.aau4005. arXiv: 1806.02751 [astro-ph.HE].

[533] Todd A. Thompson, Christopher S. Kochanek, Krzysztof Z. Stanek, et al. "Response to Comment on "A Non-Interacting Low-Mass Black Hole – Giant Star Binary System"". In: (May 2020). DOI: 10.1126/science.aba4356. arXiv: 2005.07653 [astro-ph.HE].

[534] Kip S. Thorne. "Tidal stabilization of rigidly rotating, fully relativistic neutron stars". In: *Phys. Rev. D* 58 (1998), p. 124031. DOI: 10.1103/PhysRevD.58.124031. arXiv: gr-qc/9706057.

[535] Eric Thrane and Joseph D. Romano. "Sensitivity curves for searches for gravitational-wave backgrounds". In: *Phys. Rev. D* 88.12 (2013), p. 124032. DOI: 10.1103/PhysRevD.88.124032. arXiv: 1310.5300 [astro-ph.IM].

[536] Eric Thrane and Colm Talbot. "An introduction to Bayesian inference in gravitational-wave astronomy: parameter estimation, model selection, and hierarchical models". In: *Publ. Astron. Soc. Austral.* 36 (2019). [Erratum: Publ.Astron.Soc.Austral. 37, e036 (2020)], e010. DOI: 10.1017/pasa.2019.2. arXiv: 1809.02293 [astro-ph.IM].

[537] Lev Titarchuk and Elena Seifina. "How to distinguish white dwarf and neutron star X-ray binaries during their X-ray outbursts?" In: (Nov. 2023). arXiv: 2311.12982 [astro-ph.HE].

[538] Vaibhav Tiwari. "Estimation of the Sensitive Volume for Gravitational-wave Source Populations Using Weighted Monte Carlo Integration". In: *Class. Quant. Grav.* 35.14 (2018), p. 145009. DOI: 10.1088/1361-6382/aac89d. arXiv: 1712.00482 [astro-ph.HE].

[539] Vaibhav Tiwari. "VAMANA: modeling binary black hole population with minimal assumptions". In: *Class. Quant. Grav.* 38.15 (2021), p. 155007. DOI: 10.1088/1361-6382/ac0b54. arXiv: 2006.15047 [astro-ph.HE].

[540] Vaibhav Tiwari and Stephen Fairhurst. "The Emergence of Structure in the Binary Black Hole Mass Distribution". In: *Astrophys. J. Lett.* 913.2 (2021), p. L19. DOI: 10.3847/2041-8213/abfbe7. arXiv: 2011.04502 [astro-ph.HE].

[541] Richard C. Tolman. "Static solutions of Einstein's field equations for spheres of fluid". In: *Phys. Rev.* 55 (1939), pp. 364–373. DOI: 10.1103/PhysRev.55.364.

[542] Hui Tong, Shanika Galaudage, and Eric Thrane. "Population properties of spinning black holes using the gravitational-wave transient catalog 3". In: *Phys. Rev. D* 106.10 (2022), p. 103019. DOI: 10.1103/PhysRevD.106.103019. arXiv: 2209.02206 [astro-ph.HE].

[543] L. K. Tsui and Pui-Tang Leung. "Probing the interior of neutron stars with gravitational waves". In: *Phys. Rev. Lett.* 95 (2005), p. 151101. DOI: 10.1103/PhysRevLett.95.151101. arXiv: astro-ph/0506681.

[544] Kevin Turbang, Max Lalleman, Thomas A. Callister, et al. "The Metallicity Dependence and Evolutionary Times of Merging Binary Black Holes: Combined Constraints from Individual Gravitational-wave Detections and the Stochastic Background". In: *Astrophys. J.* 967.2 (2024), p. 142. DOI: 10.3847/1538-4357/ad3d5c. arXiv: 2310.17625 [astro-ph.HE].

[545] Maximiliano Ujevic, Alireza Rashti, Henrique Gieg, et al. "High-accuracy high-mass-ratio simulations for binary neutron stars and their comparison to existing waveform models". In: *Phys. Rev. D* 106.2 (2022), p. 023029. DOI: 10.1103/PhysRevD.106.023029. arXiv: 2202.09343 [gr-qc].

[546] Samantha A. Usman et al. "The PyCBC search for gravitational waves from compact binary coalescence". In: *Class. Quant. Grav.* 33.21 (2016), p. 215004. DOI: 10.1088/0264-9381/33/21/215004. arXiv: 1508.02357 [gr-qc].

[547] E. Vangioni, K. A. Olive, T. Prestegard, et al. "The Impact of Star Formation and Gamma-Ray Burst Rates at High Redshift on Cosmic Chemical Evolution and Reionization". In: *Mon. Not. Roy. Astron. Soc.* 447 (2015), p. 2575. DOI: 10.1093/mnras/stu2600. arXiv: 1409.2462 [astro-ph.GA].

[548] Vijay Varma, Sylvia Biscoveanu, Tousif Islam, et al. "Evidence of Large Recoil Velocity from a Black Hole Merger Signal". In: *Phys. Rev. Lett.* 128.19 (2022), p. 191102. DOI: 10.1103/PhysRevLett.128.191102. arXiv: 2201.01302 [astro-ph.HE].

[549] Vijay Varma, Scott E. Field, Mark A. Scheel, et al. "Surrogate model of hybridized numerical relativity binary black hole waveforms". In: *Phys. Rev. D* 99.6 (2019), p. 064045. DOI: 10.1103/PhysRevD.99.064045. arXiv: 1812.07865 [gr-qc].

[550] Vijay Varma, Scott E. Field, Mark A. Scheel, et al. "Surrogate models for precessing binary black hole simulations with unequal masses". In: *Phys. Rev. Research.* 1 (2019), p. 033015. DOI: 10.1103/PhysRevResearch.1.033015. arXiv: 1905.09300 [gr-qc].

[551] Vijay Varma, Davide Gerosa, Leo C. Stein, et al. "High-accuracy mass, spin, and recoil predictions of generic black-hole merger remnants". In: *Phys. Rev. Lett.* 122.1 (2019), p. 011101. DOI: 10.1103/PhysRevLett.122.011101. arXiv: 1809.09125 [gr-qc].

[552] J. Veitch et al. "Parameter estimation for compact binaries with ground-based gravitational-wave observations using the LALInference software library". In: *Phys. Rev. D* 91.4 (2015), p. 042003. DOI: 10.1103/PhysRevD.91.042003. arXiv: 1409.7215 [gr-qc].

[553] Tejaswi Venumadhav, Barak Zackay, Javier Roulet, et al. "New binary black hole mergers in the second observing run of Advanced LIGO and Advanced Virgo". In: *Phys. Rev. D* 101.8 (2020), p. 083030. DOI: 10.1103/PhysRevD.101.083030. arXiv: 1904.07214 [astro-ph.HE].

[554] Doğa Veske, Imre Bartos, Zsuzsa Márka, et al. "Characterizing the Observation Bias in Gravitational-wave Detections and Finding Structured Population Properties". In: *Astrophys. J.* 922.2 (2021), p. 258. DOI: 10.3847/1538-4357/ac27ac. arXiv: 2105.13983 [gr-qc].

[555] Serena Vinciguerra et al. "An Updated Mass–Radius Analysis of the 2017–2018 NICER Data Set of PSR J0030+0451". In: *Astrophys. J.* 961.1 (2024), p. 62. DOI: 10.3847/1538-4357/acfb83. arXiv: 2308.09469 [astro-ph.HE].

[556] Pauli Virtanen et al. "SciPy 1.0–Fundamental Algorithms for Scientific Computing in Python". In: *Nature Meth.* 17 (2020), p. 261. DOI: 10.1038/s41592-019-0686-2. arXiv: 1907.10121 [cs.MS].

[557] Salvatore Vitale, Sylvia Biscoveanu, and Colm Talbot. "Spin it as you like: The (lack of a) measurement of the spin tilt distribution with LIGO-Virgo-KAGRA binary black holes". In: *Astron. Astrophys.* 668 (2022), p. L2. DOI: 10.1051/0004-6361/202245084. arXiv: 2209.06978 [astro-ph.HE].

[558] Salvatore Vitale, Walter Del Pozzo, Tjonnie G. F. Li, et al. "Effect of calibration errors on Bayesian parameter estimation for gravitational wave signals from inspiral binary systems in the Advanced Detectors era". In: *Phys. Rev. D* 85 (2012), p. 064034. DOI: 10.1103/PhysRevD.85.064034. arXiv: 1111.3044 [gr-qc].

[559] Salvatore Vitale, Will M. Farr, Ken Ng, et al. "Measuring the star formation rate with gravitational waves from binary black holes". In: *Astrophys. J. Lett.* 886.1 (2019), p. L1. DOI: 10.3847/2041-8213/ab50c0. arXiv: 1808.00901 [astro-ph.HE].

[560] Salvatore Vitale, Davide Gerosa, Will M. Farr, et al. "Inferring the properties of a population of compact binaries in presence of selection effects". In: (July 2020). DOI: 10.1007/978-981-15-4702-7_45-1. arXiv: 2007.05579 [astro-ph.IM].

[561] Salvatore Vitale, Davide Gerosa, Carl-Johan Haster, et al. "Impact of Bayesian Priors on the Characterization of Binary Black Hole Coalescences". In: *Phys. Rev. Lett.* 119.25 (2017), p. 251103. DOI: 10.1103/PhysRevLett.119.251103. arXiv: 1707.04637 [gr-qc].

[562] Salvatore Vitale, Carl-Johan Haster, Ling Sun, et al. "Physical approach to the marginalization of LIGO calibration uncertainties". In: *Phys. Rev. D* 103.6 (2021), p. 063016. DOI: 10.1103/PhysRevD.103.063016. arXiv: 2009.10192 [gr-qc].

[563] Leslie Wade, Jolien D. E. Creighton, Evan Ochsner, et al. "Systematic and statistical errors in a bayesian approach to the estimation of the neutron-star equation of state using advanced gravitational wave detectors". In: *Phys. Rev. D* 89.10 (2014), p. 103012. DOI: 10.1103/PhysRevD.89.103012. arXiv: 1402.5156 [gr-qc].

[564] Yuan-Zhu Wang, Yin-Jie Li, Jorick S. Vink, et al. "Potential Subpopulations and Assembling Tendency of the Merging Black Holes". In: *Astrophys. J. Lett.* 941.2 (2022), p. L39. DOI: 10.3847/2041-8213/aca89f. arXiv: 2208.11871 [astro-ph.HE].

[565] Steven Weinberg. "Phenomenological Lagrangians". In: *Physica A* 96.1-2 (1979). Ed. by S. Deser, pp. 327–340. DOI: 10.1016/0378-4371(79)90223-1.

[566] Linqing Wen and Yanbei Chen. "Geometrical Expression for the Angular Resolution of a Network of Gravitational-Wave Detectors". In: *Phys. Rev. D* 81 (2010), p. 082001. DOI: 10.1103/PhysRevD.81.082001. arXiv: 1003.2504 [astro-ph.CO].

[567] John Whelan. *The Geometry of Gravitational Wave Detection*. July 2014. URL: https://dcc.ligo.org/LIGO-T1300666/public.

[568] Martin J. White, Douglas Scott, and Joseph Silk. "Anisotropies in the cosmic microwave background". In: *Ann. Rev. Astron. Astrophys.* 32 (1994), pp. 319–370. DOI: 10.1146/annurev.aa.32.090194.001535.

[569] Nathan Whitehorn, Jakob van Santen, and Sven Lafebre. "Penalized Splines for Smooth Representation of High-dimensional Monte Carlo Datasets". In: *Comput. Phys. Commun.* 184 (2013), pp. 2214–2220. DOI: 10.1016/j.cpc.2013.04.008. arXiv: 1301.2184 [physics.data-an].

[570] Marcella Wijngaarden, Katerina Chatziioannou, Andreas Bauswein, et al. "Probing neutron stars with the full premerger and postmerger gravitational wave signal from binary coalescences". In: *Phys. Rev. D* 105.10 (2022),

p. 104019. DOI: 10.1103/PhysRevD.105.104019. arXiv: 2202.09382 [gr-qc].

[571] Grzegorz Wiktorowicz, Łukasz Wyrzykowski, Martyna Chruslinska, et al. "Populations of stellar mass Black holes from binary systems". In: (July 2019). DOI: 10.3847/1538-4357/ab45e6. arXiv: 1907.11431 [astro-ph.HE].

[572] Robert B. Wiringa, V. Fiks, and A. Fabrocini. "Equation of state for dense nucleon matter". In: *Phys. Rev. C* 38 (1988), pp. 1010–1037. DOI: 10.1103/PhysRevC.38.1010.

[573] Noah E. Wolfe, Salvatore Vitale, and Colm Talbot. "Too small to fail: characterizing sub-solar mass black hole mergers with gravitational waves". In: *JCAP* 11 (2023), p. 039. DOI: 10.1088/1475-7516/2023/11/039. arXiv: 2305.19907 [astro-ph.HE].

[574] Kaze W. K. Wong, Katelyn Breivik, Will M. Farr, et al. "Backward Population Synthesis: Mapping the Evolutionary History of Gravitational-wave Progenitors". In: *Astrophys. J.* 950.2 (2023), p. 181. DOI: 10.3847/1538-4357/acc863. arXiv: 2206.04062 [astro-ph.HE].

[575] Kaze W. K. Wong, Katelyn Breivik, Kyle Kremer, et al. "Joint constraints on the field-cluster mixing fraction, common envelope efficiency, and globular cluster radii from a population of binary hole mergers via deep learning". In: *Phys. Rev. D* 103.8 (2021), p. 083021. DOI: 10.1103/PhysRevD.103.083021. arXiv: 2011.03564 [astro-ph.HE].

[576] S. E. Woosley. "Pulsational Pair-Instability Supernovae". In: *Astrophys. J.* 836.2 (2017), p. 244. DOI: 10.3847/1538-4357/836/2/244. arXiv: 1608.08939 [astro-ph.HE].

[577] S. E. Woosley. "The Evolution of Massive Helium Stars, Including Mass Loss". In: *Astrophys. J.* 878.1 (2019), p. 49. DOI: 10.3847/1538-4357/ab1b41.

[578] S. E. Woosley and Alexander Heger. "The Pair-Instability Mass Gap for Black Holes". In: *Astrophys. J. Lett.* 912.2 (2021), p. L31. DOI: 10.3847/2041-8213/abf2c4. arXiv: 2103.07933 [astro-ph.SR].

[579] C. Wu, V. Mandic, and T. Regimbau. "Accessibility of the Gravitational-Wave Background due to Binary Coalescences to Second and Third Generation Gravitational-Wave Detectors". In: *Phys. Rev. D* 85 (2012), p. 104024. DOI: 10.1103/PhysRevD.85.104024. arXiv: 1112.1898 [gr-qc].

[580] D. Wysocki, R. O'Shaughnessy, Jacob Lange, et al. "Accelerating parameter inference with graphics processing units". In: *Phys. Rev. D* 99.8 (2019), p. 084026. DOI: 10.1103/PhysRevD.99.084026. arXiv: 1902.04934 [astro-ph.IM].

[581] Daniel Wysocki, Davide Gerosa, Richard O'Shaughnessy, et al. "Explaining LIGO's observations via isolated binary evolution with natal kicks". In: *Phys. Rev. D* 97.4 (2018), p. 043014. DOI: 10.1103/PhysRevD.97.043014. arXiv: 1709.01943 [astro-ph.HE].

[582] Daniel Wysocki, Jacob Lange, and Richard O'Shaughnessy. "Reconstructing phenomenological distributions of compact binaries via gravitational wave observations". In: *Phys. Rev. D* 100.4 (2019), p. 043012. DOI: 10.1103/PhysRevD.100.043012. arXiv: 1805.06442 [gr-qc].

[583] Daniel Wysocki and Richard O'Shaughnessy. *Bayesian Parametric Population Models*. [Online; accessed 2024-07-24]. 2017–. URL: bayesian-parametric-population-models.readthedocs.io.

[584] Daniel Wysocki, Richard O'Shaughnessy, Leslie Wade, et al. "Inferring the neutron star equation of state simultaneously with the population of merging neutron stars". In: (Jan. 2020). arXiv: 2001.01747 [gr-qc].

[585] Kent Yagi and Nicolás Yunes. "Approximate Universal Relations for Neutron Stars and Quark Stars". In: *Phys. Rept.* 681 (2017), pp. 1–72. DOI: 10.1016/j.physrep.2017.03.002. arXiv: 1608.02582 [gr-qc].

[586] Kent Yagi and Nicolas Yunes. "I-Love-Q". In: *Science* 341 (2013), pp. 365–368. DOI: 10.1126/science.1236462. arXiv: 1302.4499 [gr-qc].

[587] Kent Yagi and Nicolas Yunes. "I-Love-Q Relations in Neutron Stars and their Applications to Astrophysics, Gravitational Waves and Fundamental Physics". In: *Phys. Rev. D* 88.2 (2013), p. 023009. DOI: 10.1103/PhysRevD.88.023009. arXiv: 1303.1528 [gr-qc].

[588] Takahiro S. Yamamoto, Ryoto Inui, Yuichiro Tada, et al. "Prospects of detection of subsolar mass primordial black hole and white dwarf binary mergers". In: *Phys. Rev. D* 109.10 (2024), p. 103514. DOI: 10.1103/PhysRevD.109.103514. arXiv: 2401.00044 [gr-qc].

[589] Y. Yang, I. Bartos, Z. Haiman, et al. "Cosmic Evolution of Stellar-mass Black Hole Merger Rate in Active Galactic Nuclei". In: *Astrophys. J.* 896.2 (2020), p. 138. DOI: 10.3847/1538-4357/ab91b4. arXiv: 2003.08564 [astro-ph.HE].

[590] Barak Zackay, Liang Dai, Tejaswi Venumadhav, et al. "Detecting gravitational waves with disparate detector responses: Two new binary black hole mergers". In: *Phys. Rev. D* 104.6 (2021), p. 063030. DOI: 10.1103/PhysRevD.104.063030. arXiv: 1910.09528 [astro-ph.HE].

[591] Barak Zackay, Tejaswi Venumadhav, Liang Dai, et al. "Highly spinning and aligned binary black hole merger in the Advanced LIGO first observing run". In: *Phys. Rev. D* 100.2 (2019), p. 023007. DOI: 10.1103/PhysRevD.100.023007. arXiv: 1902.10331 [astro-ph.HE].

[592] Michael Zevin and Simone S. Bavera. "Suspicious Siblings: The Distribution of Mass and Spin across Component Black Holes in Isolated Binary Evolution". In: *Astrophys. J.* 933.1 (2022), p. 86. DOI: 10.3847/1538-4357/ac6f5d. arXiv: 2203.02515 [astro-ph.HE].

[593] Michael Zevin, Simone S. Bavera, Christopher P. L. Berry, et al. "One Channel to Rule Them All? Constraining the Origins of Binary Black Holes Using Multiple Formation Pathways". In: *Astrophys. J.* 910.2 (2021), p. 152. DOI: 10.3847/1538-4357/abe40e. arXiv: 2011.10057 [astro-ph.HE].

[594] Michael Zevin, Chris Pankow, Carl L. Rodriguez, et al. "Constraining Formation Models of Binary Black Holes with Gravitational-Wave Observations". In: *Astrophys. J.* 846.1 (2017), p. 82. DOI: 10.3847/1538-4357/aa8408. arXiv: 1704.07379 [astro-ph.HE].

[595] Tianqi Zhao and James M. Lattimer. "Tidal Deformabilities and Neutron Star Mergers". In: *Phys. Rev. D* 98.6 (2018), p. 063020. DOI: 10.1103/PhysRevD.98.063020. arXiv: 1808.02858 [astro-ph.HE].

[596] Bei Zhou, Luca Reali, Emanuele Berti, et al. "Subtracting compact binary foregrounds to search for subdominant gravitational-wave backgrounds in next-generation ground-based observatories". In: *Phys. Rev. D* 108.6 (2023), p. 064040. DOI: 10.1103/PhysRevD.108.064040. arXiv: 2209.01310 [gr-qc].

[597] Xing-Jiang Zhu and Gregory Ashton. "Characterizing Astrophysical Binary Neutron Stars with Gravitational Waves". In: *Astrophys. J. Lett.* 902.1 (July 2020), p. L12. DOI: 10.3847/2041-8213/abb6ea. arXiv: 2007.08198 [astro-ph.HE].

[598] Xing-Jiang Zhu, E. Howell, T. Regimbau, et al. "Stochastic Gravitational Wave Background from Coalescing Binary Black Holes". In: *Astrophys. J.* 739 (2011), p. 86. DOI: 10.1088/0004-637X/739/2/86. arXiv: 1104.3565 [gr-qc].

[599] Xingjiang Zhu, Eric Thrane, Stefan Oslowski, et al. "Inferring the population properties of binary neutron stars with gravitational-wave measurements of spin". In: *Phys. Rev. D* 98 (2018), p. 043002. DOI: 10.1103/PhysRevD.98.043002. arXiv: 1711.09226 [astro-ph.HE].

[600] Andrea Zonca, Leo Singer, Daniel Lenz, et al. "healpy: equal area pixelization and spherical harmonics transforms for data on the sphere in Python". In: *Journal of Open Source Software* 4.35 (2019), p. 1298. DOI: 10.21105/joss.01298.