University of Glasgow

Colloms, Storm (2026) *Looking back from the stellar graveyard: inferring the origins of binary black holes with gravitational waves.* PhD thesis.

https://theses.gla.ac.uk/86285/

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

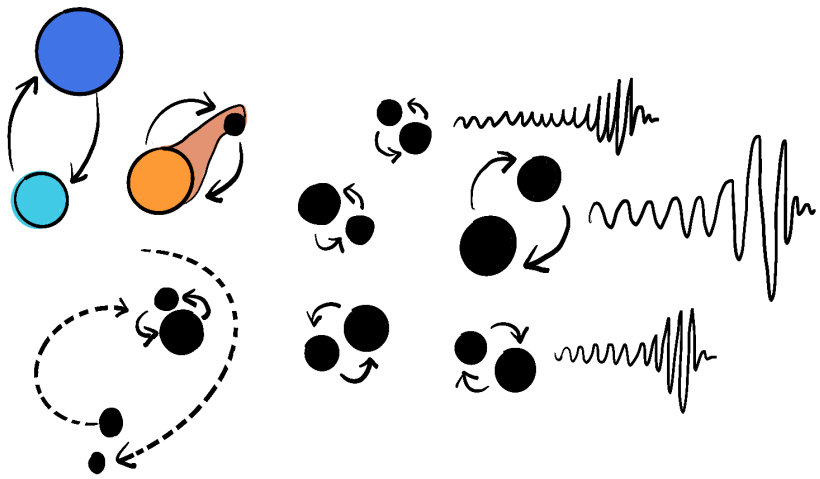
When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses
https://theses.gla.ac.uk/
research-enlighten@glasgow.ac.uk

Looking Back from the Stellar Graveyard: Inferring the Origins of Binary Black Holes with Gravitational Waves

Storm Colloms

2026



Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy.

School of Physics & Astronomy
College of Science and Engineering
University of Glasgow

Abstract

Gravitational waves (GWs) have transformed our understanding of binary black holes (BBHs), with hundreds of observations now published since the first direct detection in 2015. Measuring the population properties of many BBHs allows us to statistically infer the of the lives and deaths of their progenitors, which are subject to many overlapping uncertainties. The first source of these uncertainties arises from observational limitations: BBHs likely originate from massive, potentially low-metallicity stars, which are short-lived and hard to observe. Second, predicting the evolution of massive stars is limited by our theoretical understanding of multi-scale stellar evolutionary processes, which for BBH progenitors is further complicated by binary and dynamical interactions. Third, understanding the astrophysics over a population of BBH progenitors is limited by the computational expense of simulating the detailed physics of their progenitors using population synthesis models. The growing number of GW observations offer unique insight into these uncertainties, allowing us to understand the evolutionary pathways that lead to merging BBHs.

In this thesis, I investigate the origins of BBHs with new complementary modelling frameworks, exploring correlations between observed masses and mass ratios, and developing emulators of population synthesis simulations.

Motivated by evidence that BBHs prefer to merge with similar masses, I develop parametric population models to infer whether the minimum mass of the smaller black hole varies with the mass of the larger black hole. Using GW data from hundreds of observations, I show that big black holes can merge with the smallest black holes, contrary to expectations from previous studies of the BBH mass ratio distribution. I find that the interpretation of population outliers, such as unequal mass ratio observation GW190814, depends on the construction of our population models. This demonstrates how analysing GW observations using new population models furthers our understanding BBH observations and astrophysics.

Comparing the observed BBH population to population synthesis simulations allows for constraints on astrophysical parameters governing their evolution, and the relative contribution from different formation channels. These simulations are computationally expensive and rely on a wide range of assumptions,

therefore achieving robust predictions from population synthesis remains a key challenge in GW astronomy. I demonstrate the effectiveness of normalising flows, a type machine learning algorithm, to interpolate predictions from population synthesis and overcome the necessity of a dense grid of simulations. This framework enables previously intractable constraints over a continuous range of parameters governing the physics of BBH evolution, and probe degeneracies between key assumptions in population synthesis simulations.

Extending this framework to ensure consistency between the emulated populations and the observational selection effects, I find that differences in how selection effects are calculated lead to large discrepancies in inference results with $O(100)$ observations. These discrepancies arise from small differences between the emulated populations and the underlying population synthesis models, indicating that more accurate emulation of population synthesis is necessary avoid systematic bias in resulting astrophysical constraints. This is particularly important as we analyse an increasing number of observations, which will drive increasing differences between constraints under these two methods.

Showing how these analyses are subject to shrinking statistical and persistent systematic uncertainties with growing GW catalogues, I highlight the future work necessary for robust BBH population inference. Overcoming these systematics will enable the application of new population models, including more flexible parametric models of correlated distributions, and population synthesis models exploring the correlations between more astrophysical parameters and formation channels. The development of robust inference frameworks will allow us to confidently constrain multiple uncertain parameters in BBH evolution, and develop trustworthy simulations of BBH progenitor populations. As we advance further into the era of GW astronomy, these analysis provide crucial steps towards piecing together where and how BBHs form.

Contents

Abstract iii

Acknowledgements xiii

Declaration xv

1 Introduction 1
- 1.1 Gravitational-wave observations . . . 2
  - 1.1.1 Gravitational-wave signals . . . 2
  - 1.1.2 Gravitational-wave detection . . . 5
  - 1.1.3 Estimation of single event parameters . . . 10
- 1.2 Binary black hole astrophysics . . . 13
  - 1.2.1 Single massive stars . . . 14
  - 1.2.2 Binary black holes from isolated binary stellar evolution 18
  - 1.2.3 Binary black holes from dynamical interactions . . . 22
  - 1.2.4 Binary black holes from exotic sources: AGN, triples, primordial black holes . . . 24
  - 1.2.5 Simulating populations of binary black holes . . . 25
- 1.3 Gravitational-wave population inference . . . 29
  - 1.3.1 Statistical analysis of binary black hole observations . . 30
  - 1.3.2 Constructing a population model . . . 35
  - 1.3.3 Current understanding of the binary black hole population 37
- 1.4 Outline of new work . . . 41

2 Can Big Black Holes Merge with the Smallest Black Holes? 43
- 2.1 Abstract . . . 43
- 2.2 Introduction . . . 43
- 2.3 Methods . . . 46
  - 2.3.1 Variable minimum-mass models . . . 46
  - 2.3.2 Statistical framework . . . 48
  - 2.3.3 Gravitational-wave observations . . . 50
- 2.4 Results . . . 50
  - 2.4.1 Evolution of minimum secondary mass . . . 50

2.4.2 Mass-ratio and primary-mass distributions 55
2.5 Conclusions 57

3 Exploring the evolution of gravitational-wave emitters with efficient emulation: Constraining the origins of binary black holes using normalising flows 61
3.1 Abstract 61
3.2 Introduction 62
3.3 Methods 65
3.3.1 Population models 65
3.3.2 Statistical framework 67
3.3.3 Normalising flows 69
3.3.4 Observational data 72
3.4 Results 73
3.4.1 Emulation and interpolation performance 73
3.4.2 Testing normalising flows with discrete inference 77
3.4.3 Continuous inference with gravitational-wave observations 81
3.4.4 Comparing the inferred population to parametric models 85
3.5 Conclusion 87

4 Inferring the origins of binary black holes with growing gravitational wave catalogues 91
4.1 Abstract 91
4.2 Introduction 91
4.3 Measuring a variable minimum mass ratio in successive catalogues 94
4.3.1 Methods 94
4.3.2 Results with GWTC-4.0 95
4.3.3 Comparison across catalogues 98
4.4 Inferring binary black hole astrophysics from population synthesis models in successive catalogues 101
4.4.1 Method 101
4.4.2 Results 105
4.5 Updating detection efficiency interpolation 108
4.5.1 Outline of methods 108
4.5.2 Training normalising flows for SNR emulation 109
4.5.3 Detection efficiency emulation 113
4.5.4 Inference with machine learning detection efficiency . . 117
4.6 Conclusions 119

5 Conclusions 123
5.1 Summary of results 123
5.1.1 Probing binary black hole correlations with parametric models 123

5.1.2 Inferring the origins of binary black holes with normalising flows 124
5.1.3 The effect of a growing catalogue on binary black hole population inference 125
5.2 Looking forward 127

Bibliography **133**

List of Figures

i Not a black hole xiv

1.1.1 The response of a ring of test particles to gravitational-wave plus and cross polarisations with time. 3
1.1.2 Example binary black hole waveform in the time-domain 4
1.1.3 Representative amplitude spectral density for LIGO Hanford and Livingston, Virgo, and KAGRA for each observing run. 7
1.1.4 Cumulative gravitational wave detections with time 9
1.1.5 Illustration of compact binary coalescence spin and inclination angles 12
1.2.1 Schematic diagram of Roche lobe overflow 19
1.2.2 Map of binary population synthesis codes 27
1.3.1 Primary mass distribution of binary black holes in GWTC-4.0 . . . 38
1.3.2 Mass ratio distribution of binary black holes in GWTC-4.0 39
1.3.3 Effective inspiral distribution of binary black holes in GWTC-4.0 . 40

2.3.1 Representative variable minimum secondary mass models 48
2.4.1 Inferred evolution of minimum secondary masses with primary masses under each variable minimum secondary mass model 51
2.4.2 Hyperposteriors on key hyperparameters for the Power Law minimum secondary mass model 52
2.4.3 Hyperposteriors on key hyperparameters for the Increasing Parabola minimum secondary mass model 53
2.4.4 Hyperposteriors on key hyperparameters for the Relaxed Parabola minimum secondary mass model 54
2.4.5 Inferred primary mass and mass ratio distributions for each variable minimum mass ratio model 56
2.4.6 Inferred mass ratio distributions at slices of primary mass for variable minimum mass ratio models 57

3.4.1 Distribution of gravitational-wave observables for the CE channel with $\chi_b = 0$ and $\alpha_{CE} = 2$, comparing the training samples, KDE, and normalising flow 75

3.4.2 The normalising-flow emulated distributions of chirp mass, mass ratio, effective inspiral spin and redshift for the common-envelope channel at five values of natal spin χ_b and common-envelope efficiency α_{CE}, showing the interpolation abilities of the flow. 76
3.4.3 Distribution of gravitational-wave observables learned by the normalising flow for a set of unseen test data unseen, compared to the original samples 78
3.4.4 The log likelihood ratio between the normalising flow and the KDE representation of the CE population model with $\chi_b = 0$ and $\alpha_{CE} = 5$, at a slice through chirp mass and mass ratio, with and without regularisation 79
3.4.5 Hyperposteriors on underlying branching fractions between discrete population synthesis models comparing normalising flows and KDEs 81
3.4.6 Hyperposteriors on astrophysical hyperparameters χ_b, α_{CE}, underlying and detectable branching fractions using normalising flows for continuous inference with GWTC-3.0 data 84
3.4.7 Distribution of inferred gravitational-wave observables marginalised over the hyperposterior from continuous inference with GWTC-3.0 85

4.3.1 The evolution of $m_2^{\min}$ with m_1 with and without GW190814 for each variable minimum secondary mass model and GWTC-4.0 observations 96
4.3.2 Inferred primary mass and mass ratio distributions with and without GW190814 for our three variable minimum secondary mass models and GWTC-4.0 observations 98
4.3.3 The evolution of $m_2^{\min}$ with m_1 for our three variable minimum secondary mass models with GWTC-2.1, GWTC-3.0, and GWTC-4.0 observations, with and without GW190814 99
4.3.4 The primary mass and mass ratio distributions found with our three variable minimum secondary mass models with and without GW190814, with GWTC-3.0 observations. 100
4.3.5 The primary mass and mass ratio distributions found with our three variable minimum secondary mass models including GW190814, with GWTC-2.1, GWTC-3.0, and GWTC-4.0 observations. 102
4.3.6 The number of observations versus the width of the 90% credible interval of $p(m_1)$ at $m_1 \sim 10\,M_\odot$ and $m_1 \sim 35\,M_\odot$, averaged over our three model variations with GW190814 and without GW190814. 103
4.4.1 The inferred natal spin, common envelope efficiency, underlying and detectable branching fractions from three successive gravitational wave catalogues, using continuous inference with normalising flows and `pchip` interpolation of the detection efficiency. 105
4.5.1 The performance of the normalising flows emulating SNR distributions, compared to three test populations 111

4.5.2 The percent errors on quantiles between the SNR normalising flows and the test data, including the errors as a function of mass ratio . . 112
4.5.3 Schematic diagram of the machine learning detection efficiency interpolation method 113
4.5.4 The detection efficiency interpolation across natal spin for five formation channels, comparing machine learning to a `pchip` interpolation. 114
4.5.5 The detection efficiency interpolation across common envelope efficiency for the common envelope channel, comparing machine learning to a `pchip` interpolation. 116
4.5.6 The inferred natal spin, common envelope efficiency, underlying and detectable branching fractions from three successive gravitational wave catalogues using continuous inference with normalising flows, and machine learning interpolation of the detection efficiency . . . 118
4.5.7 The inferred natal spin, common envelope efficiency, underlying and detectable branching fractions from GWTC-4.0 using continuous inference with normalising flows, comparing three different detection efficiency interpolation methods 120

List of Tables

2.3.1 Hyperparameter priors for each minimum secondary mass model . . 49
2.4.1 Log Bayes factors between variable minimum secondary mass models and fiducial default model 55

3.3.1 Network parameters of for each normalising flow used to emulate outputs from population synthesis 72
3.4.1 Inferred median and 90% symmetric credible interval for the underlying branching fractions from the discrete inference using normalising flows and using KDEs. 80
3.4.2 Inferred median and 90% symmetric credible intervals for the underlying and detectable branching fractions and the natal spin, and the 90% credible lower limit for CE efficiency from the continuous inference results using normalising flows. 82

4.3.1 Log Bayes factors between variable minimum secondary mass models and fiducial default model in GWTC-3.0 97
4.4.1 Credible intervals on the natal spin, common envelope efficiency, and branching fractions for five formation channels from three successive gravitational wave catalogues 107

Acknowledgements

First, a deep thank you to my supervisors Christopher Berry and John Veitch for their guidance, support, and for passing down their scientific knowledge and wisdom. I have learned a lot from them both.

I thank the members of the Institute of Gravitational Research for making my time there truly special, especially all who have come and left office 253 during the years. Thank you to Leigh, Jess, Ross, Federico, Victoria, Naren, Christian, and Thejas (though I do not forgive the endless coffee grinding); to Harry, Varun, Josh, Natalia, August, and Elizabeth. In the wider IGR, I thank the mentorship and expertise of Rachel, Chris, Siong, Daniel, and friendship of Thomas, Rhona, Dominika, Alex, Ardianna, Jen, Xiaofei, Shashwat, Jack, Sean, Scott and Zhuotao.

Thank you those who supported me during my time at Northwestern University, for both widening my scientific horizons and for extending their hospitality during my visit. In particular, I thank Zoheyr Doctor for his excellent mentorship, Sharan Banagiri, Sylvia Biscoveanu, Monica Gallegos Garcia, Chase Kimball, Camille Liotine, Kyle Rocha, Elizabeth Teng, and Michael Zevin. Thank you also to Huei Sears and Jessie Thwaites for their kindness during my travels, and to the extended members of Astrobites for the community and knowledge that I have received from them. In addition, I thank the LIGO magazine editorial team for their support, especially to Hannah Middleton and Anna Green.

I extend many thanks to my fellow organisers, including Amanda Farah, Simona Miller, Aditya Vijaykumar, Ward Amar, Michele Mancarella, Gregoire Pierra, and Ana Lorenzo Medina. Working alongside them has been one of the most rewarding parts of the last few years.

Last, I thank my friends and family for their boundless support. Thank you to Lucy Arditi, my parents, and my partner Joseph Murtagh for all the meals shared together, crises solved, and laughs together. Special thanks to my cat, Pico, who has provided both cuddles and chaos which were essential to my PhD experience.



Figure i: Not a black hole

Declaration

This thesis is the result of my own work under the supervision of Dr Christopher Berry and Dr John Veitch at the Institute for Gravitational Research at the University of Glasgow. Chapter 1 is entirely review material including citations where appropriate, and was written by me with feedback from Dr Christopher Berry. The following chapters include work published as journal articles:

- Chapter 2 presents work that was published in the *Astrophysical Journal*, as 'Can Big Black Holes Merge with the Smallest Black Holes?', in collaboration with Dr Zoheyr Doctor and Dr Christopher Berry (Colloms et al., 2025b). The chapter is presented almost verbatim to the published article, with small changes to wording, citations and additional reference to the Chapter 4.
- Chapter 3 presents work that was published in the *Astrophysical Journal*, as 'Exploring the evolution of gravitational-wave emitters with efficient emulation: Constraining the origins of binary black holes using normalising flows', in collaboration with Dr Christopher Berry, Dr John Veitch, and Dr Michael Zevin (Colloms et al., 2025a). The results from this chapter are additionally presented in 'Exploring the astrophysical origins of binary black holes using normalising flows', which was published in the *Journal of Physics: Conference Series*, in collaboration with Dr Christopher Berry, Dr John Veitch, and Dr Michael Zevin (Colloms et al., 2026). The chapter is similarly presented almost verbatim to the published article, though Appendix A as is in the paper is now included in the main body of the Chapter as Section 3.3.3, and the chapter includes the plot and an excerpt of text from Colloms et al. (2026).

Results presented in Chapter 4 are not published in any journal at the time of thesis submission; this work was conducted under the supervision of Dr Christopher Berry and Dr John Veitch and was written by me with guidance from Dr Christopher Berry. Chapter 5 summarises the results presented in this thesis, and was written by me with feedback from Dr Christopher Berry.

While not directly discussed as part of this thesis, I have contributed to other works whose results are drawn upon and have been cited throughout this thesis.

These include contributions to analysis and review for GWTC-4.0 data products and code, used in 'GWTC-4.0: Population Properties of Merging Compact Binaries' (Abac et al., 2025a, including S. Colloms), and 'Compact Binary Coalescence Sensitivity Estimates with Injection Campaigns during the LIGO-Virgo-KAGRA Collaborations' Fourth Observing Run' (Essick et al., 2025, including S. Colloms). Additionally, I was part of the paper writing team for 'GWTC-5.0: Population Properties of Merging Compact Binaries', leading the section on broad population correlations (Abac et al., 2026a, including S. Colloms).

I Introduction

Since the first direct detection of gravitational waves (GWs) in 2015 (Abbott et al., 2016a), hundreds of binary black holes (BBHs) have been detected with GWs (Abac et al., 2026b,c). These observations have revolutionised our understanding of the physics of their progenitors — massive, potentially low-metallicity stars, which are hard to observe otherwise due to their intrinsic rareness, and lifetimes of orders-of-magnitude shorter than lower mass stars (Kroupa, 2002). The feedback, chemical enrichment, and dynamical interactions of massive stars are important for galactic evolution, star formation, and stellar clusters (Mackey et al., 2008; Hopkins et al., 2014; Nomoto et al., 2013; Eldridge & Stanway, 2022; Ekström, 2026). Furthermore, the majority of massive stars will interact with a companion during their lifetimes (Sana et al., 2012, 2025; Offner et al., 2023). Despite significant progress in understanding observations of massive stellar remnants, there remains many open questions about BBH evolution and the lives and deaths of their progenitors (e.g., Mapelli, 2021; Mandel & Broekgaarden, 2022).

Combining observations of massive binary stars at multiple stages in their evolution, e.g. x-ray binaries, stripped-stars, and supernovae, is key for constraining the evolutionary processes that create BBHs (Marchant & Bodensteiner, 2024). The observational selection effects of GWs are particularly well understood, making these observations especially suitable for statistical studies of the properties of stellar remnants (Abac et al., 2026d; Essick et al., 2025). Measuring the population properties of GW observations gives us unique insight about the formation processes of stellar binaries, the life and death of massive stars, the dynamics of stellar clusters, and the role of these systems in our Universe.

In this chapter, I introduce the journey so far from theory to detection to GW astrophysics (Section 1.1). Next, I review BBH astrophysics, including single and binary stellar evolution, potential formation pathways, and the current state of population synthesis (Section 1.2). I outline the population inference methods used to constrain BBH astrophysics from GW observables, and inference results using current sets of observations (Section 1.3). Finally, I introduce the new work presented in the following chapters of this thesis (Section 1.4).

1.1 | Gravitational-wave observations

GWs were first directly detected on the 14th of September 2015 (Abbott et al., 2016a), almost one hundred years after they were predicted in Einstein's theory of general relativity (Einstein, 1918). The first signal, named GW150914 after the date of its observation, came from the coalescence of two black holes each with mass $\sim 30\,M_\odot$ and over one billion light-years away (Abbott et al., 2016a,b, 2024). Since 2015, hundreds of signals from compact binary coalescences (CBCs) have been detected, revealing the presence of such sources and their properties (Abbott et al., 2021a, 2023a; Abac et al., 2026b,c), transforming our knowledge of compact object formation (Abac et al., 2025a, 2026a). GW detections have proved the accuracy of the prediction of their radiation from general relativity (Abac et al., 2026e), and offered independent insights into the Universe's expansion (Abac et al., 2025b, 2026f). In particular, GWs allow us to study the physics of BBHs which are not directly observable through other messengers. Mapping GW measurements of BBHs back to the evolution of their progenitors gives unique insight into these systems, allowing us to infer the astrophysics of massive stars.

In this subsection, I introduce the theoretical prediction of GWs (Subsection 1.1.1), their detection (Subsection 1.1.2), and analysis of the properties of the signal (Subsection 1.1.3), outlining the journey from GW signals to GW astrophysics.

1.1.1 | Gravitational-wave signals

Our ability to detect GWs from BBHs is strengthened by the fact that the forms of the signals can be accurately predicted following general relativity. General relativity predicts that GWs arise due to perturbations to spacetime, caused by the asymmetric acceleration of massive objects. The mathematical derivation of GWs can be found by applying perturbations to Einstein's equations, which describe the curvature of spacetime by matter (Flanagan & Hughes, 2005). While any asymmetric acceleration leads to GW emission, the amplitude of a GW roughly depends on the the mass and typical velocity of the system (Sathyaprakash & Schutz, 2009). The GW emission $\bar{h}^{ij}$ can be expressed in terms of the quadrupole moment, the leading order term for weak field radiation (where the system's velocity is much smaller than the speed of light):

$$\bar{h}^{ij} = \frac{2G}{c^4}\frac{1}{D_\mathrm{L}}\ddot{Q}^{ij}, \tag{1.1}$$

where G is the gravitational constant, c is the speed of light, D_L is the luminosity distance to the source (in meters), and Q^{ij} is the quadrupole moment, which depends on the density and velocity fields of the system. The dependence on the second derivative of Q^{ij} shows the requirement for asymmetrically accelerating matter.

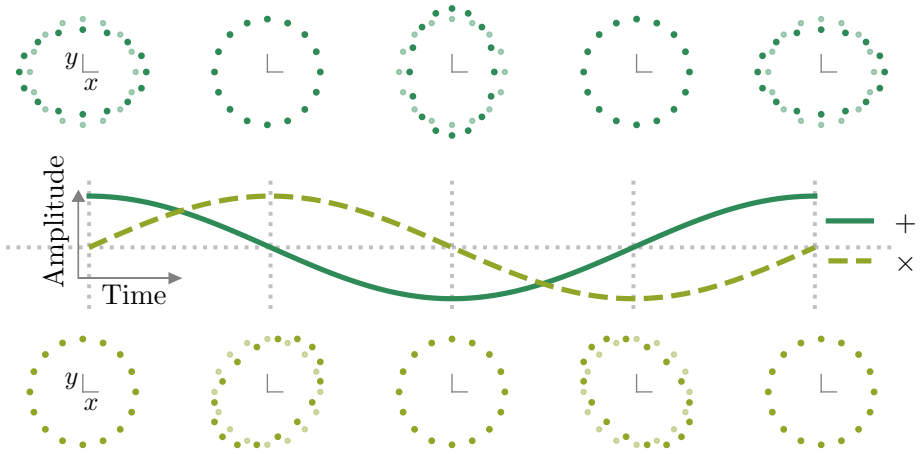


Figure 1.1.1: The response of a ring of test particles to GW plus (top) and cross (bottom) polarisations with time. The amplitude of each polarisation is shown in the middle panel, demonstrating the relative phase difference by $\pi/2$. Figure based off of Figure 2 in Isi (2023), using the accompanying code release Isi (2022).

GWs periodically alter the relative separation of free falling masses perpendicular to the waves' direction of propagation. For a ring of test particles, this can be seen in periodic stretching along one axis and squishing along the other, with two contributing polarisations, h_+ and $h_\times$, which are offset by 45° (Figure 1.1.1). However, the strength of distortion from GWs is very small. Assuming $\ddot{Q}^{ij}$ is equivalent to the kinetic energy of an object moving non-axisymmetrically, for example a dumbbell with masses each 500 kg on a 1 m rod spinning at 10 rads^{-1}, this would result in GWs of amplitude $\sim 10^{-40}$ at 10 m distance away. Therefore, very large masses travelling at relativistic speeds are required to produce strong GWs. This leads us to consider highly energetic astrophysical phenomena to be able to produce detectable signals.

The merging of compact binaries — merging binary neutron stars (BNSs), neutron star–black holes (NSBHs), and BBHs — is an example of such energetic phenomena. Orbiting binary systems emit GWs due to their acceleration (Peters & Mathews, 1963; Peters, 1964). If the binary's separation is sufficiently close, GWs will take away enough orbital energy to cause the binary to decrease in separation, inspiral, and merge. The GW signal emitted during this process can be divided into three stages: (1) inspiral, (2) merger, (3) ringdown (Figure. 1.1.2). During the inspiral, the two objects orbit each other and slowly lose energy to GWs; two bodies may be orbiting each other for billions of years before merging. As the objects get closer, the GW frequency and amplitude increases, until the merger when the GW amplitude peaks. At merger, a new black hole is created, which emits further GWs as it stabilises in the ringdown stage.

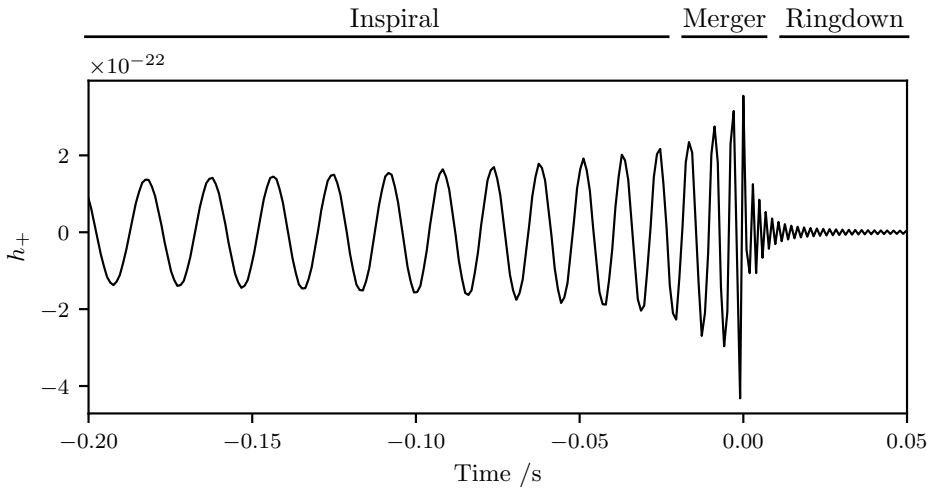


Figure 1.1.2: Example BBH waveform generated with the IMRPhenomXPHM waveform model (Pratten et al., 2021), showing the evolution of the plus polarisation GW amplitude with time, from black holes of component masses 20 $M_\odot$ and 15 $M_\odot$, spins of 0.2 and 0, and at a distance of 1 Gpc. At first, the black holes inspiral, with the GW amplitude and frequency increasing until merger, when the amplitude peaks. Post-merger, the newly created black hole rings down, dissipating high frequency GWs.

The exact frequency and amplitude evolution of GWs from a CBC is determined by the properties of the objects in the binary. For a BBH merging in a circular orbit the intrinsic waveform will be solely determined by the masses and spins of the black holes, e.g., higher mass BBHs merge at lower frequencies. Finding the GW waveform for a BBH requires solving Einstein's equations. However, analytical solutions are intractable for a BBH with arbitrary masses and spins. Instead, various methods are employed to model the GW waveform through simulation or approximation to the solution from general relativity. Numerical relativity can be used to find GWs signals through simulations (see e.g., Duez & Zlochower, 2019), though these simulations are the most computationally expensive. Phenomenological models or approximants are often used to find the GW waveform, especially for the earlier stages in the inspiral where these approximations are more accurate (e.g., Taracchini et al., 2014; Husa et al., 2016). Phenomenological models may be based on a post-Newtonian approximation, following a power law expansion in v/c as corrections to the Newtonian solution (Blanchet, 2002; Arun et al., 2005). Accurate waveform models allow us to evaluate GW waveforms efficiently for merging BBH with arbitrary parameters, and hence search for and understand the properties of these signals.

1.1.2 | Gravitational-wave detection

While GWs from CBCs cause strong changes to spacetime near their origin, the dependence of their amplitude on the inverse of the source's distance means that GW signals are very small by the time they reach Earth. For BBHs, the relative change in length is usually below 1 part in 10^{-20}. We require precise instruments to detect these GWs. Currently, a network of four ground-based laser interferometers are are used for detection. Each interferometer consists of a laser which is split at a beam splitter and travels down two perpendicular arms of equal length. The laser light is reflected off of a mirror at the end of each arm to return down the arms and recombine at the beam-splitter, with the output detected at a photodetector. In the case of equal arm lengths, the lasers destructively interfere upon recombination. As a GW passes through a detector, the relative path length between the two laser arms is altered, and the signal is detected as a fraction of the laser power reaching the photodetector. We then measure the relative change in length in the interferometer, called the GW strain.

The network of ground-based laser interferometers are made up of two Advanced laser interferometer gravitational-wave observatory (LIGO) detectors in Hanford, Washington, and Livingston, Louisiana (Aasi et al., 2015); the Advanced Virgo detector in Cascina, Italy (Acernese et al., 2015); and the KAGRA detector based at the Kamioka observatory in Japan (Aso et al., 2013). These detectors make up the instruments contributing to the LIGO–Virgo–KAGRA (LVK) Collaboration. The first observing run with an advanced detector configuration ran from September 2015 to January 2016, with four observing runs having completed in total. The first detection was made by both Advanced LIGO detectors in the first observing run (Abbott et al., 2016c,d), with Virgo joining during the second observing run (Abbott et al., 2019), and KAGRA joining at the end of the third observing run (Abe et al., 2023). The concurrent observation by multiple detectors allow us to verify the detection of a GW in multiple observatories and triangulate the origin of the signals (Fairhurst, 2009; Schutz, 2011). Each break between observing runs is used for commissioning and detector upgrades to increase the detectors' sensitivity to GWs.

Improvements to the interferometers are planned in order to progress towards the design sensitivity of the detectors (Abbott et al., 2020a; Acernese et al., 2026). A future LIGO detector in India is currently under construction, expected to be operational from 2030 (Saleem et al., 2022). Beyond the current generation of detectors, next-generation detectors Cosmic Explorer (Evans et al., 2021) and Einstein Telescope (Maggiore et al., 2020) are planned to start operating from the 2040s. These detectors will have a drastic increase in sensitivity, resulting in BBH detections up to redshifts of hundreds, and high-precision measurements on their properties (Kalogera et al., 2021). Over the next decades, the next-generation detectors are expected to probe BBHs throughout the history of the Universe, detecting the very first systems.

Current detectors are subject to multiple different noise sources, including quantum noise, seismic noise, and thermal noise (Maggiore, 2007a). The strengths of each noise source varies with frequency, with radiation pressure, seismic noise, and suspension thermal noise limiting at low frequencies, shot noise limiting at high frequencies, and coating thermal noise limiting at central frequencies (Aasi et al., 2015; Wang & Zhang, 2023). The cumulative effects of these noise sources mean that the most sensitive frequencies are in the range of $\sim$ 10–2000 Hz for each detector. The noise in each detector is characterised by the power spectral density (PSD), which is a measure of the power of the noise as a function of frequency, or the amplitude spectral density (ASD), defined to be the square-root of the PSD. Figure 1.1.3 shows the ASD in each detector averaged over each observing run. Detector improvements lower the average PSD with each detector upgrade, allowing us to detect more signals. While the PSD varies over each observing run, the average PSD describes the sensitivity for each detector during a particular time period.

The PSD characterises the noise sources that are (approximately) stationary, meaning that the power in the noise does not change significantly with time. Non-stationary noise also affects the detectors, with timescales of $O(\text{seconds})$ (Nuttall, 2018; Davis et al., 2021; Capote et al., 2025). These bursts of noise are called glitches. Distinguishing astrophysical signals from non-stationary noise that mimic or obscure GW signals is one of the key problems in GW detection.

Advanced search techniques are necessary to identify a GW signal amongst noise, and distinguish a signal from glitches. For CBCs, we can employ our knowledge of the expected waveform by searching for a match of a signal to a template waveform. The range of different signal morphologies from different BBH parameter configurations requires a range of templates, which make up a template bank. These templates are then compared to data in a process called matched filtering to find the cross-correlation between the data and the template (Hooper et al., 2012; Allen et al., 2012; Messick et al., 2017; Dal Canton et al., 2021; Alléné et al., 2025). This results in candidate signals identified with a corresponding detection statistic that can quantify the match to a template. If the candidate signal's detection statistic passes some threshold, we can claim detection. For GW data d found with template h_θ with parameters θ, the optimal detection statistic is

$$\langle d|h_\theta\rangle = 2\int_0^\infty \frac{\tilde{d}(f)\tilde{h}^*_\theta(f) + d^*(f)\tilde{h}_\theta(f)}{S(f)}\mathrm{d}f, \tag{1.2}$$

where $\tilde{d}(f)$ denotes the complex-valued Fourier transform of the time-domain data, $d(t)$; $\tilde{h}_\theta(f)$ that of the waveform; and $S(f)$ is the PSD (Creighton & Anderson, 2011; Abac et al., 2025c). The degree of match can be determined by the matched-filter signal-to-noise ratio (SNR):

$$\rho^2 = \frac{\langle d|h_\theta\rangle}{\sqrt{\langle h_\theta|h_\theta\rangle}}, \tag{1.3}$$

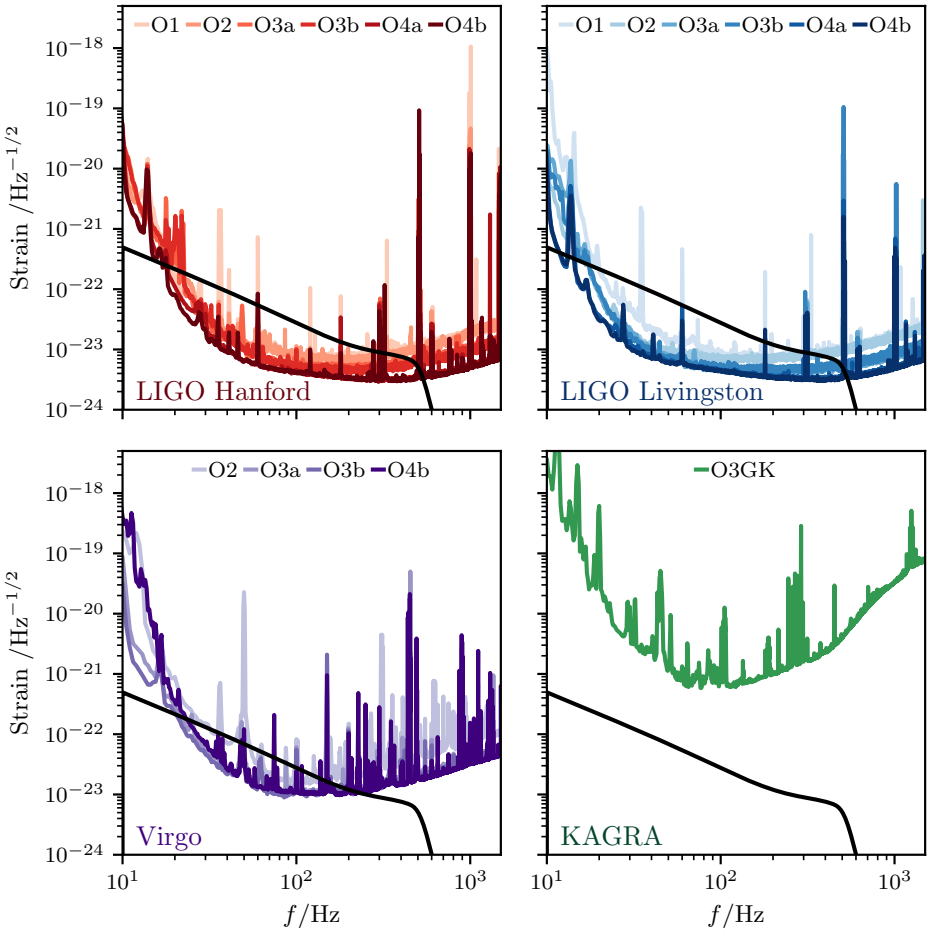


Figure 1.1.3: The average ASDs for LIGO Hanford and Livingston, and Virgo for each observing run from O1 to O4b. The average ASDs for KAGRA during O3GK, a joint observing run between KAGRA and the GEO 600 observatory continued after the end of O3b, is shown in the bottom right (Abe et al., 2023). The black line shows the frequency domain waveform for a BBH with component masses 20 $M_\odot$ and 15 $M_\odot$ at 1 Gpc.

where $\sqrt{\langle h_\theta | h_\theta \rangle}$ is the optimal SNR. This is the best match of the signal to the template; in matched-filter searches the signal may be recovered with a lower SNR due to an imperfect match to the template, or the presence of noise. For multiple detectors, the combined network SNR adds in quadrature,

$$\rho^2_{\rm net} = \sum_{det} \rho^2. \tag{1.4}$$

Glitches can lead to a high SNR even when they do not match a template well, especially in the case where they are high power. Alternative statistics are therefore necessary to rank and determine truly astrophysical candidate GWs. The χ^2 signal consistency statistic encapsulates the match of a signal to a template, and can weigh down spurious high SNR triggers caused by glitches (Allen, 2005). Alternatively, the false alarm rate (FAR) can be used to take into account false positive candidate signals from noise. The FAR represents the rate of background triggers over a whole template bank with a ranking statistic equal to or higher than the one observed for the event. This is calculated empirically, with search pipelines performing timeslides on data to randomise coincidence between detectors, finding all triggers, and determining their ranking statistics. The $p_{\rm astro}$ statistic takes into account the FAR as well as the true alarm rate, assuming we know the true population of signals (Guglielmetti et al., 2009; Farr et al., 2015). Practically, a combination of SNR, FAR and $p_{\rm astro}$ are used as a threshold to identify GW detections and to select events for follow up analyses.

Other search methods aside from matched filtering have been developed in order to overcome either the computational expense of searching using a large template bank, or to search for signals which do not originate from signals that closely match CBC waveforms (e.g., due to lensing, non-GR effects, or sources which are not CBCs). These include machine learning based approaches (Nagarajan & Messenger, 2025; Silver et al., 2025), and coherent energy searches (Klimenko et al., 2016; Martini et al., 2026). Currently the majority of CBC sources are detected via matched-filtering.

Confidently detected signals which meet a certain detection threshold make up GW catalogues and are studied in more detail. Catalogues are cumulative collections of observations: fourth Gravitational-Wave Transient Catalogue (GWTC-4.0) contains 218 events from the start of O1 to the end of the first part of the fourth observing run (O4a; Abac et al., 2026b), while fifth Gravitational-Wave Transient Catalogue (GWTC-5.0) contains 390 events up until the end of the second part of the fourth observing run (O4b; Abac et al., 2026c). The majority of these observations are from BBHs, while two are confidently from BNSs, and five NSBHs are included in GWTC-5.0 (Abac et al., 2026b,c), with three of these at high enough confidence for population analyses (Abac et al., 2025a). It is unclear whether GW230529 is an NSBH, it may be a heavy BNS system (Abac et al., 2024). An additional observation, GW190814, may be either a BBH or NSBH (Abbott et al., 2020b). The uncertainty on the source classification of observations is twofold.

First, the classification is based on the measured component masses of the binary, which have some intrinsic uncertainty. Second, the location of the boundary between black hole and neutron star masses is uncertain due to the dependence on unknown astrophysics (see Section 1.2.1; e.g., Kalogera & Baym, 1996; Fryer et al., 2012; Koehn et al., 2025). Only one event, GW170817, was detected with a confirmed accompanying electromagnetic counterpart, allowing a more confident classification as a BNS (Abbott et al., 2017; Nicholl & Andreoni, 2025).

The cumulative number of detections and candidate detections from each observing run is shown in Figure 1.1.4. The fourth observing run has taken the number of observations from ~ 100 to almost 500. Beyond the forth observing run, hundreds of BBH observations are expected per year of observing at O5 sensitivity (Muccillo et al., 2026). GW astronomy is now on the brink of having thousands of detections for downstream analyses.

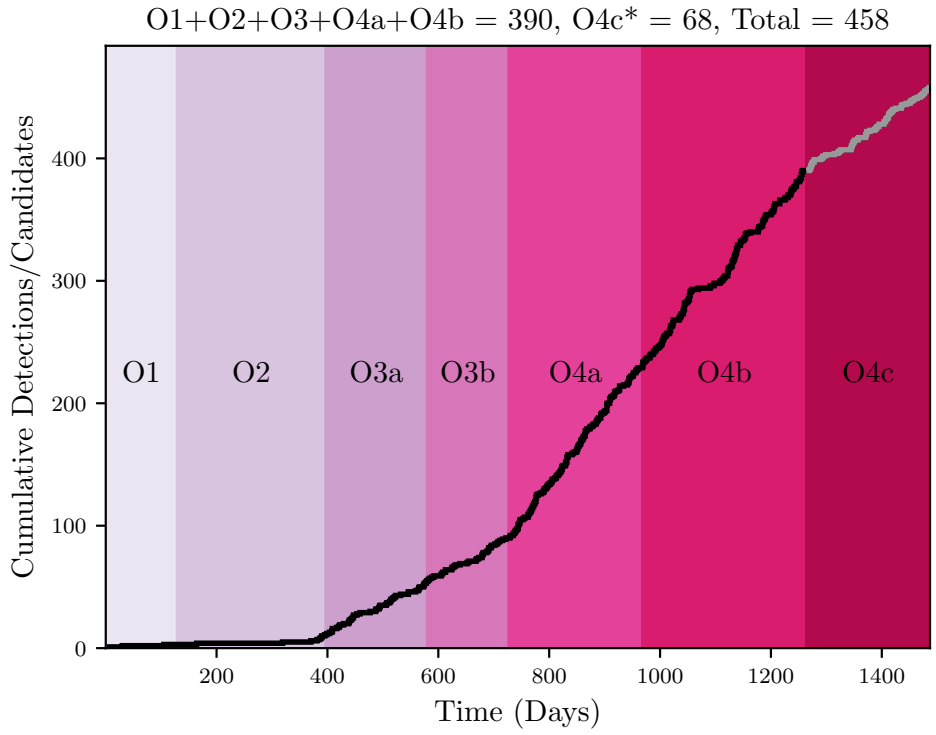


Figure 1.1.4: Cumulative GW detections with time, highlighting the observing runs from O1 to O4c. O1 through O4b highlights published events with FAR< 1yr^{-1} or $p_{\text{astro}} > 0.5$. O4c highlights the detection candidates with FAR< 1month^{-1} that have not yet been published in a catalogue. The rate of detection has increased with each observing run, with 218 published detections up until the end of O4a, 390 up until the end of O4b, and 458 combined detections and candidates from O1 to the end of O4.

1.1.3 | Estimation of single event parameters

Once a signal is confidentially identified in the data, parameter estimation is performed to measure the source's properties. A signal is fully parametrised by a set of intrinsic and extrinsic parameters, $\vec{\theta}$. The intrinsic properties of a BBH are solely comprised of the masses and spins of each component. A source's extrinsic properties are given by its distance, inclination, sky location, reference phase, merger time, and polarisation (Thrane & Talbot, 2019). These parameters are not trivially recovered from the data due to detector noise and degeneracies between parameters which obscure the true parameters of the signal. Parameter estimation uses Bayesian inference to measure the properties of signal, described by the probability distribution over $\vec{\theta}$, given the detected data and our prior belief in the probability of the signal's parameters (Cutler & Flanagan, 1994; Veitch & Vecchio, 2010).

We want to measure the posterior probability $p(\vec{\theta}|\vec{d})$, the probability of the signal's parameter's $\vec{\theta}$ given the GW data $\vec{d}$. Using Bayes' Theorem (Bayes & Price, 1763),

$$p(\vec{\theta}|\vec{d}) = \frac{p(\vec{d}|\vec{\theta})p(\vec{\theta})}{\mathcal{Z}(\vec{d})}, \tag{1.5}$$

where $p(\vec{d}|\vec{\theta})$ is the likelihood of observing the data $\vec{d}$ given parameters $\vec{\theta}$, $p(\vec{\theta})$ is the prior probability of $\vec{\theta}$, and $\mathcal{Z}(\vec{d})$ is the evidence. The prior is the current belief in observing a signal with true parameters $\vec{\theta}$. The evidence is the probability of observing data $\vec{d}$, marginalised over all possible combinations of true parameters θ, and normalises the posterior probability distribution such that $\int p(\vec{\theta}|\vec{d})\mathrm{d}\theta = 1$. The likelihood is usually evaluated by comparing the data to a theoretical model of the GW signal with parameters $\vec{\theta}$ (i.e. a template waveform), normalised by the noise in the detector (i.e. the PSD; Veitch & Vecchio, 2010).

The intrinsic parameters of a BBH can be described by the individual component masses and spin vectors, or combinations of intrinsic parameters. Certain combinations of parameters determine the leading order contribution to the GW waveform, and therefore give the best measurements of the system's properties.

For mass, the chirp mass, a reduced mass of the system, is the leading order term in determining the frequency and amplitude of the GW signal (Blanchet et al., 1995). The chirp mass is defined as:

$$\mathcal{M} = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}}, \tag{1.6}$$

where m_1 and m_2 are the primary and secondary component masses, respectively, with $m_2 < m_1$. The chirp mass together with the mass ratio, $q = m_2/m_1$ ($0 < q < 1$), can fully describe the mass of the system. Chirp mass is usually the best

measured parameter overall for CBCs, with lower mass sources that have longer signals having better measured masses (Vitale et al., 2017).

The spins of each component can be described by the dimensionless spin vector, $\vec{\chi}_{1,2} = \vec{S}_{1,2}/m_{1,2}^2$, where $\vec{S}_{1,2}$ is each object's spin vector. The spins can also be described by the dimensionless spin magnitudes, $|\vec{\chi}_{1,2}|$, spin tilt angles $\theta_{1,2}$, and spin azimuthal angles $\phi_{1,2}$, or their difference ϕ_{12}, shown in Figure 1.1.5. For a precessing system, the spin tilt angles and azimuthal angles change as the objects inspiral due to the conservation of angular momentum of both bodies and the total system, and are defined at a reference frequency during parameter estimation (Hannam, 2014). These angles are evolved back to infinite separation during final parameter estimation, in order to evaluate the pre-merger system configuration.

Spin effects are subdominant to mass effects on the waveform, meaning they are usually less well measured. Individual spin magnitudes and tilts are subject to degeneracies and large uncertainties (Vitale et al., 2014; Pürrer et al., 2017). Instead, effective spin parameters can be chosen which describe more dominant effects on the GW waveform, and are better measured (Ajith et al., 2011; Pürrer et al., 2013; Schmidt et al., 2015; Vitale et al., 2017).

The effective inspiral spin describes the degree of the component spins that are parallel with the orbital angular momentum, and can be in the range $[-1, 1]$. The effective inspiral spin is defined as

$$\chi_{\rm eff} = \frac{\left(m_1\vec{\chi}_1 + m_2\vec{\chi}_2\right)\cdot\vec{L}}{m_1 + m_2}, \tag{1.7}$$

where $\vec{L}$ is the instantaneous orbital angular momentum direction (Ajith et al., 2011). Effective inspiral spin close to zero suggests that the system has little spin, or that all spin is in-plane, i.e. perpendicular to the orbital angular momentum vector. One spin aligned component and one spin anti-aligned component may lead to effective inspiral spin close to zero, though this requires each component to have similar mass-weighted spin magnitudes. Large positive (negative) values of effective inspiral spin suggest that the black holes are highly spinning with spins aligned (anti-aligned) with the orbital angular momentum.

Any spin with a degree of misalignment to the orbital angular momentum causes precession of orbital plane and individual spins around total angular momentum (Apostolatos et al., 1994; Kidder, 1995). This spin modulates the waveform, and the degree of precession is determined by the amount of in-plane spin (Hannam, 2014; Schmidt et al., 2015). The spin parameter describing the degree of in-plane spin is called the effective precessing spin, $\chi_{\rm p}$. This is given by

$$\chi_{\rm p} = \max\left(\chi_1\sin\theta_1, \left(\frac{4q+3}{4+3q}\right)q\chi_2\sin\theta_2\right). \tag{1.8}$$

Both $\chi_{\rm eff}$ and $\chi_{\rm p}$ are best measured in the inspiral phase, and when there is unequal mass ratio. Usually $\chi_{\rm p}$ is less well measured than $\chi_{\rm eff}$, and precession is difficult to measure for signals with few observed cycles (Abac et al., 2026b).

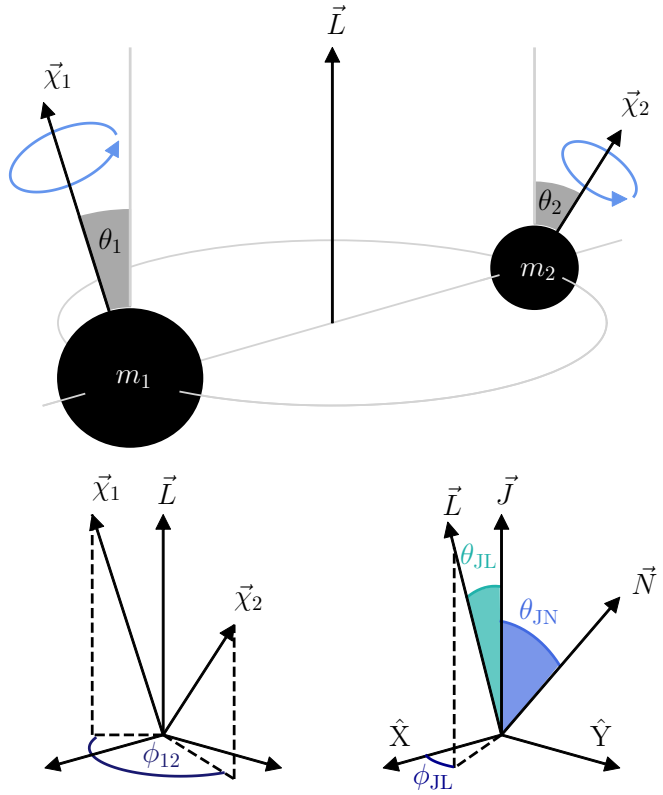


Figure 1.1.5: Top: Illustration of a BBH with primary mass m_1 and secondary mass m_2, highlighting the corresponding component spins $\vec{\chi}_1$ and $\vec{\chi}_2$ and spin tilt angles θ_1 and θ_2, in orbit with orbital angular momentum $\vec{L}$. Bottom left: The BBH's component spin vectors, showing the azimuthal spin angle, ϕ_{12}, between the projections of the component spins onto the orbital plane. Bottom right: The total angular momentum vector $\vec{J}$ and the line-of-sight vector $\vec{N}$, showing the orbital inclination, θ_{JN}, and ϕ_{JL}, the azimuthal angle between the projection of the orbital angular momentum onto the $\hat{\rm X}$–$\hat{\rm Y}$ plane and the $\hat{\rm X}$ axis (Biscoveanu et al., 2021).

Parameter estimation results in a measurement of the intrinsic and extrinsic properties of individual GW sources. The masses, spins and distances of events in particular are relevant for BBH population inference. By mapping the measurements of these parameters to our knowledge of BBH formation physics, we can decipher the origins of these sources.

1.2 | Binary black hole astrophysics

The observation of GWs proves that there are mechanisms through which BBHs can form and merge within the age of the Universe. The existence of merging BBHs evokes a question central to BBH astrophysics: *how* do two black holes get close enough together to merge? A binary merger requires sufficient hardening, meaning the reduction in the binary's orbital separation. GWs can only carry away a significant amount of energy (and therefore are only efficient at reducing orbital separation) for a binary already at small separation (Maggiore, 2007b; Mandel & Broekgaarden, 2022). For two objects with equal mass components M to merge within time T only through the emission of GWs, they must have a separation of

$$a \leq 3.7 R_\odot \left(\frac{M}{M_\odot}\right)^{3/4} \left(\frac{T}{14\,\mathrm{Gyr}}\right)^{1/4}, \tag{1.9}$$

where $R_\odot$ is the radius of the Sun (Maggiore, 2007b). For a BBH with both component masses equal to $10\,M_\odot$ to merge within the age of Universe, the required separation is $a \sim 20\,R_\odot$. A natural candidate for the progenitor of BBHs are massive binary stars which form black holes at the end of their lives. However, massive binary stars will expand to radii much greater than this limit during this lifetime (Podsiadlowski et al., 1992). It therefore seems physically impossible to achieve a merging BBH from binary stars. Observations of BBHs with GW prove that there are astrophysical processes that can transform massive stars to tightly bound black holes, via binary stellar evolution or otherwise. What these astrophysical processes are is currently one of the biggest unknowns in BBH astrophysics.

GW observations have not only allowed us to prove the existence of merging BBHs, but also allow us to measure the properties of their components. These measurements give us a unique probe into the final stages of massive stars that are impossible to observe through messengers other than GWs. By measuring the properties of merging black holes we gain hints towards understanding their formation. Collating multiple observations and analysing the BBH population allows us to study the pathways that can create merging black holes.

GW observations show preference for multiple formation pathways contributing to the population of BBHs (e.g., Zevin et al., 2021; Ray et al., 2023; Antonini et al., 2025; Banagiri et al., 2025b; Vijaykumar et al., 2026; Farah et al., 2026). For massive binary stars to contribute to the observed population, there must exist

evolutionary pathways from massive binary stars to tight BBH without disruption of the binary. Additionally, dynamical interactions in dense environments such as stellar clusters are predicted to lead to BBH formation (e.g., Rodriguez et al., 2022; Kremer, 2026). Other formation pathways such as stellar triples (e.g., Toonen et al., 2022), formation in the accretion disks of active galactic nuclei (AGN) (e.g., McKernan et al., 2012; Bartos et al., 2017; Sedda et al., 2023b), or primordial black holes (e.g., Bagui et al., 2025), may also lead to merging BBHs. In this section I introduce the current understanding and uncertainties of the physical processes in BBH formation for these different channels: isolated binaries, starting from the physics of single massive stars (Subsections 1.2.1 and 1.2.2); dynamical interactions (Subsection 1.2.3); and more exotic processes (Subsection 1.2.4). I then review the current ways of modelling each channel with population synthesis simulations (Subsection 1.2.5).

1.2.1 | Single massive stars

Massive stars are defined as stars that undergo core-collapse supernovae at the end of their lives, possibly forming a remnant neutron star or black hole. This requires the star to be born with a mass of greater than $\sim 8\,M_\odot$, or ~ 18–$20\,M_\odot$ for black hole formation, though these boundaries are dependent on stellar properties including internal structure, rotation, and metallicity (Poelarends et al., 2008; Smartt, 2009; Hirschi et al., 2025; Luo et al., 2025). Massive stars are intrinsically rare, and have lifetimes orders-of-magnitude shorter than lower mass stars (Kroupa, 2002; Ekström, 2026). Because of this, observations of massive stars are much rarer compared to lower mass stars. However, massive stars are extremely luminous, have strong stellar winds, and can undergo nuclear synthesis of elements up to iron which in turn enriches the interstellar medium. The impact of massive stars on their surroundings is therefore of great significance for the evolution of numerous other systems in our Universe. Despite their importance across various fields of astrophysics, many open questions remain about massive stellar evolution.

The pre-supernovae lives of massive single stars, including their mass loss, angular momentum evolution, and nuclear burning, are chiefly determined by their mass, rotation, and composition at birth (Langer, 2012). The composition can be described by the star's metallicity:

$$Z = \sum_{e>\mathrm{He}} \frac{m_e}{M_*} = 1 - X - Y, \tag{1.10}$$

where X and Y are the fraction of hydrogen and helium respectively, m_e is the mass of any heavier elements and M_* is the total mass of the star. Metallicity is often quoted in terms of our Sun's metallicity, $Z_\odot \sim 0.014$ (Asplund et al., 2009).

Massive stars undergo nuclear burning from hydrogen to iron, with fusion occurring in the stellar core, and the stellar envelope transporting energy to the

star's surface. Interior mixing controls the internal energy transport, and can supply additional fuel to the stellar core, increasing the effective stellar mass, and extending the life of the star (Maeder, 1975; Schroder et al., 1997; Temaj et al., 2024). Modelling mixing in stellar interiors is challenging due to the inherently 3D and random process of convection, the requirement to consider boundaries between convective and radiative transport zones, and the requirement to consider multiple timescales for these processes (Heger et al., 2000). Instead, usually stellar interiors are approximated by some 1D mixing length (Prandtl, 1925; Joyce & Tayar, 2023). A sizeable fraction of massive stars are rapid rotators (Ramírez-Agudelo et al., 2013), complicating the assumptions of spherical symmetry in stellar interiors necessary for mixing length approximations to be valid.

Internal mixing impacts the angular momentum transport inside the star, which is important for determining how the spin of the star's core changes throughout it's lifetime. The spin of the stellar core pre-supernova is expected to leave an imprint on the birth spin of the stellar remnant. Efficient angular momentum transport couples the spin of the core to the envelope, allowing angular momentum to be dissipated through stellar winds. This results in a slower spinning core, and therefore lower remnant birth spins (Fuller & Ma, 2019). Meanwhile, inefficient angular momentum transport is expected to result in higher birth spins (Heger et al., 2000; Périgois et al., 2023). Challenges in modelling the internal transport in stars result in uncertain predictions for black hole birth spins.

For single stars, mass loss is principally due to stellar winds: the continuous ejection of material from the star's surface. The winds of the massive stars are stronger than lower mass stars, and are due to radiation pressure injecting energy into the stellar atmosphere. This can drive winds with terminal flow speeds of ~ 1000–3000 km/s and mass loss rates of $\dot{M} \sim 10^{-10}$–$10^{-5}\, M_{\odot}\mathrm{yr}^{-1}$ (Lamers & M. Levesque, 2017). The effect of stellar winds on the star can be significant, with massive stars losing up to a large fraction (tens of percent) of their initial mass due to winds (Smith, 2014; Fichtner et al., 2022; Merritt et al., 2025). The rate of radiative mass loss due to winds depends on metallicity, with higher metallicity stars having higher rates of mass loss (Vink et al., 2001). Lower metallicity stars therefore retain more mass, and can create more massive remnants (Belczynski et al., 2010). The mass loss effects from stellar winds remains uncertain, and are dependent on rotation and internal mixing, requiring the joint study of multiple effects of stellar evolution (Hirschi et al., 2025; Romagnolo et al., 2026).

A massive star may undergo core-collapse supernova at the end of nuclear burning, when fusion is no longer provides sufficient thermal pressure, and gravity causes the core to collapse (Baade & Zwicky, 1934). In-falling material bounces off the collapsing core due to neutrino-powered shocks, and the envelope is ejected into the surrounding interstellar medium, causing a bright transient. A remnant black hole may form through partial fallback of the surrounding material; a non-negligible fraction of black holes can form via fallback after core-collapse

supernova (Kochanek, 2015). Not all massive stars are expected to undergo a successful supernova: only a sufficiently strong shock can eject the star's envelope, while a weaker shock causes the envelope to fall back and the direct collapse of the star into a black hole (Fryer, 1999; Kochanek et al., 2008; Smartt, 2009; Adams et al., 2017). A visible supernova is not expected for direct collapse, but for rapidly spinning stars, direct collapse may be observed as a long gamma-ray burst (LGRB): high-energy jets powered by an accretion disk (Woosley, 1993; MacFadyen & Woosley, 1999). Previously it was thought that higher mass progenitors led to more fallback, but recent work has found there may be 'islands of explodability' that depend on the properties of the stellar core right before merger (Sukhbold et al., 2016; Patton & Sukhbold, 2020; Schneider et al., 2021).

The final fate of the star is highly sensitive to the properties of its progenitor (e.g., its mass and spin), and the physics of the star and the explosion (e.g., the equation of state, neutrino transport, or general relativity; Fryer, 1999; Sukhbold et al., 2018; Burrows & Vartanyan, 2021). Due to the complexity of these dependencies, and the computational limitations of running 3D hydrodynamical supernovae simulations, many questions are still unanswered about when and how black holes form from core-collapse supernovae. Understanding which stars undergo supernovae versus direct collapse, which stars create black holes versus neutron stars, the dependence of remnant's mass on the birth mass of the star, and how the supernova impacts the remnant's spin and velocity are highly active areas of research (e.g., Janka, 2012; Burrows & Vartanyan, 2021).

Assumptions in supernova physics impact the low-mass regime of black hole formation, with some models resulting in a gap between the lowest mass black holes and the highest mass neutron stars (e.g., Fryer et al., 2012; Liu et al., 2021; Patton et al., 2022; Fryer et al., 2022). Observations of black holes in X-ray binaries suggested this mass gap was present in the compact object population (Bailyn et al., 1998; Ozel et al., 2010). However, observations of objects within the lower mass gap suggest it is somewhat populated (Thompson et al., 2019; Abbott et al., 2020b; Jayasinghe et al., 2021; Abbott et al., 2023a; Abac et al., 2024; Fishbach et al., 2026). While the difference between observations may be in part explained by observational selection effects, it is likely that our understanding of the physics of compact objects in this mass regime does not well describe observations (Siegel et al., 2023). As of yet, it is unclear how to reconcile low mass black hole observations and supernovae prescriptions.

Stars with even higher initial masses, $\gtrsim 100$–$260\,M_\odot$, are expected to undergo a (pulsational) pair instability supernova (PISN, Fowler & Hoyle, 1964; Rakavy & Shaviv, 1967; Woosley et al., 2002). Due to the mass of these stars, the stellar core is hot enough to spontaneously generate electron-positron pairs, reducing the radiation pressure supporting the star and causing a runaway explosion. For stars $\gtrsim 130\,M_\odot$, this explosion completely eviscerates the star, leaving no remnant. Below this mass, matter is shed in pulsations until the star reaches a stable mass. It is predicted that PISNe leave a dearth of BBHs between ~ 50–$130\,M_\odot$. The

lower edge of the PISN gap before total disintegration are highly sensitive to the carbon–oxygen reaction rate, metallicity, and rotation (Farmer et al., 2019; Vink et al., 2025).

To capture the complexity of supernova physics while avoiding computationally inhibitive 3D models, research has focussed on finding a simplified model to capture the explodability criteria. The mapping from pre-supernovae object to the remnant is known as a supernova prescription. A commonly implemented prescription by Fryer et al. (2012) finds that the minimum black hole mass depends on the time delay between the core bounce and the revival explosion. Recent developments in supernovae prescriptions have related the core compactness, the average density of the core, to the supernova outcome (O'Connor & Ott, 2011; Patton & Sukhbold, 2020; Schneider et al., 2021; Boccioli et al., 2025). The compactness depends on the final mass of the carbon–oxygen core, and the fraction of carbon. Highly compact cores are harder to explode and therefore more likely to form black holes. Core compactness is not a monotonic function of initial stellar mass (and depends on other properties of the star), rather than one threshold mass above which all stars become black holes. Other work has found that compactness is not a good proxy for explodability (Burrows et al., 2020). Work towards finding accurate supernovae prescriptions is ongoing.

Supernovae additionally impact two properties of the remnant: its spin and velocity, via a supernova kick. Black holes that form with (partially) ejected baryonic matter may receive kicks of tens, up to several hundreds of $\mathrm{km s^{-1}}$, provided that a high fraction of the envelope is ejected (Chan et al., 2018, 2020). The kick magnitude from direct collapse is predicted to be lower, at $\sim 10\ \mathrm{km\ s^{-1}}$, but could be up to $100\ \mathrm{km\ s^{-1}}$, depending on the asymmetry of neutrino emission (Vigna-Gómez, 2025). Modelling the kick magnitude and post-supernova black hole spin is challenging due to the dependence on explosion physics, such as the neutrino transport and equation of state (Popov et al., 2025). In addition, kicks may only be fully imparted up to several minutes after the supernova, significantly after the final time for most simulations. Gamma-ray burst observations suggest that black holes from core-collapse supernova are born slowly spinning (Gottlieb et al., 2023). Supernova kicks may significantly change the spin of the black hole from the initial core rotation (Chan et al., 2020; Tauris, 2022).

The lives of massive stars are non-trivial, depending on many uncertain aspects of stellar physics, including interior mixing, stellar winds, and supernova physics. The uncertain physics has a large impact on the properties of massive stellar remnants, with many co-dependencies between the different aspects of uncertain physics. Observations of massive stars and their remnants are key for guiding theory towards an understanding of massive stellar evolution. In particular, BBHs observations provide a novel view into the physics of massive stars through their remnants. However, the impact of the uncertainties in single stellar evolution on remnant black holes is further complicated by binary interactions, obscuring the understanding of massive stars from BBH observations.

1.2.2 | Binary black holes from isolated binary stellar evolution

The majority of massive stars will undergo interactions with a companion during their lifetimes (Sana et al., 2012, 2025; Offner et al., 2023). These interactions change the mass, composition, and rotation of stars during their evolution, causing the evolution of binary stars differ greatly to that of single stars (e.g., De Marco & Izzard, 2017; Eldridge, 2020). Binary interactions create systems which can rarely, if ever, be outcomes of single stellar evolution: stripped stars, stellar mergers, X-ray binaries, and merging compact objects. These observational phenomena allow us to study massive stars, stellar binaries, and their remnants. While electromagnetic observations can probe earlier stages in the evolution of massive binary stars, GWs offer a unique probe to the final stages of these systems' evolution when they no longer emit electromagnetic light.

Merging two black holes from binary stars require a finely tuned system: the remnants must be brought to sufficient separation to be able to merge within the age of the Universe, without being brought too close in earlier stages of binary evolution, resulting in a stellar merger prior to compact object formation. Furthermore, the binary needs to survive any disruption from two supernova explosions, which can unpair the binary due to supernova kicks. BBH mergers are therefore relatively rare products from massive binary stars — crude estimates for the fraction of binaries that lead to a double compact object give $O(10^{-6})$ mergers per stellar binary (Mandel & Broekgaarden, 2022).

One key difference in the lives of binary stars is the possibility of mass transfer (Langer, 2012; Marchant & Bodensteiner, 2024). Mass transfer may occur via wind mass transfer, or, more efficiently, through Roche lobe overflow (Paczyński, 1971). In Roche lobe overflow, one star in a binary overfills its Roche lobe, the volume in which matter remains gravitationally bound to the star (Figure 1.2.1). This is often unavoidable during a binary's lifetime due to each star's expansion during its evolution (Sana et al., 2012). When matter extends beyond the donor's Roche lobe, it enters the gravitational influence of the companion star, and can be accreted or lost from the system. Mass transfer can have a large impact on the binary system, changing the masses, mass ratio, compositions, separation and spins of objects in the binary (e.g. Dominik et al., 2012; de Mink et al., 2013; Neijssel et al., 2019; Bavera et al., 2021; van Son et al., 2022b; Briel et al., 2022b).

The impact of mass transfer depends on the response of the accretor and the impact of accretion on the binary separation. Mass transfer can be split into two categories: conservative mass transfer (i.e. all mass is accreted by the accretor, with conservation of both mass and angular momentum), or non-conservative mass transfer (i.e. some loss of matter and angular momentum through ejection of donated matter). Both cases may result in shrinking binary separation, though the impact on the orbital evolution and mass transfer stability between each mode will differ, resulting in different observable signatures on the remnants (Negu

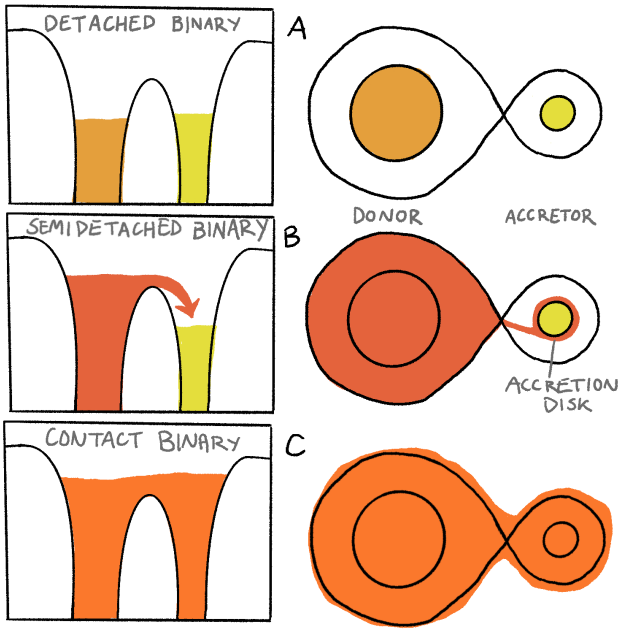


Figure 1.2.1: Left: the gravitational potential of binary system and the matter density for detached binary, semi-detached binary with stable mass transfer through Roche lobe overflow, and contact binary with a common envelope. Right: The Roche lobes for each case, showing no Roche lobe overflow for the detached binary, one full Roche lobe for stable mass transfer, and two full Roche lobes for common envelope evolution.

& Tessema, 2015; Willcox et al., 2023). Binaries may undergo multiple stages of mass transfer during their lifetimes, each with different physical responses and consequences for the system.

If the donor star expands less rapidly or shrinks faster than its Roche lobe, mass transfer will generally be stable and may continue for a long time. Stable mass transfer may gradually reduce the separation of the binary, leading to BBHs which can merge within a Hubble time (e.g., van den Heuvel et al., 2017; Gallegos-Garcia et al., 2024; Picco et al., 2024). Which binaries can become merging BBHs through only stable mass transfer depends on how much angular momentum loss there is during mass transfer, hardening requires sufficient orbital angular momentum loss through non-conservative mass transfer. This additionally depends on whether or not the accretion is Eddington-limited, which can change both the mass of the accreting black hole, and the rates of BBHs produced via stable mass transfer (van Son et al., 2020; Bavera et al., 2021).

If the accretor cannot accept transferred matter fast enough, or if the separa-

tion shrinks sufficiently fast, runaway unstable mass transfer is triggered. Unstable mass transfer leads to the filling of both objects' Roche lobes in a common envelope (Paczynski, 1976; Ivanova et al., 2013). A common envelope can be ejected if there is sufficient energy deposited into the envelope through transfer of orbital energy from the binary, or other energy sources such as radiative energy. Through the transfer of orbital energy, the separation of the objects can be drastically decreased. Common envelope has therefore been a therefore been a popular pathway to explain tight binaries, including BBHs (Paczynski, 1976; Dominik et al., 2012). However, this remains one of the most uncertain stages of binary evolution, owing to complex physics, and lack of observations of systems during the short common envelope stage.

Prescriptions can be used to characterise the energy that is supplied from the binary to eject the envelope. One prescription is to model the common envelope efficiency, $\alpha_{\rm CE}$, defined as

$$E_{\rm env,bind} = \alpha_{\rm CE} \Delta E_{\rm orb}, \tag{1.11}$$

where $E_{\rm env,bind}$ is the total binding energy of the envelope, and $\Delta E_{\rm orb}$ is the fraction of orbital energy released during the common envelope (van den Heuvel, 1976; Paczynski, 1976; Belczynski et al., 2001; Dominik et al., 2012). In terms of the properties of the binary objects, this becomes

$$\frac{GM_1 M_{1,\rm env}}{\lambda R_1} = \alpha_{\rm CE} \left(-\frac{G\left(M_{1,\rm core} + M_{1,\rm env}\right) M_2}{2a_i} + \frac{GM_{1,\rm core} M_2}{2a_f} \right), \tag{1.12}$$

where M_1 is the mass of the donor star with envelope mass $M_{1,\rm env}$ and core mass $M_{1,\rm core}$, M_2 is the mass of the accretor star, a_i and a_f are the initial and final separations of the binary, and λ is a parameter which describes the envelope structure. The $\alpha_{\rm CE}$ parameter quantifies the energy needed to eject the envelope as a fraction of the orbital energy. For a fixed envelope binding energy, lower $\alpha_{\rm CE}$ results in tighter binaries (with the risk of binaries merging within the envelope), while higher $\alpha_{\rm CE}$ results in less tight binaries (with the risk of not merging within a Hubble time). An $\alpha_{\rm CE}$ greater than one requires that energy other than orbital energy, such as radiative energy from the stellar objects, to be used in the envelope's ejection (Ivanova et al., 2013).

It is generally understood that the $\alpha_{\rm CE}$ prescription is is highly simplified and does not capture the complex physics of common envelope evolution. For example, $\alpha_{\rm CE}$ is often treated as global value, though as a better descriptor it is more likely to vary with the physics of the system (Davis et al., 2011). While 3D hydrodynamical simulations could provide more detailed common envelope prescriptions, they remain too computationally expensive for simulations at a population level (Law-Smith et al., 2020; Moreno et al., 2022; Wei et al., 2024).

The merger rates of BBHs through common envelope evolution are highly uncertain, owing to the complex physics and short timescales of such processes,

as well as the uncertainties of the stability criteria. The adopted criteria for mass transfer stability can have a large impact on which binaries undergo stable versus unstable mass transfer, and which areas of initial binary parameter space lead to merging BBHs (e.g., Gallegos-Garcia et al., 2021; Briel et al., 2022b; Dorozsmai & Toonen, 2024). In addition, the merger rates of isolated binaries are sensitive to metallicity and star formation rate (Chruslinska et al., 2019; Briel et al., 2022a; van Son et al., 2025). Combined with the uncertainties in single stellar evolution, it is clear that predictions of BBHs from isolated binary evolution are challenging.

Beyond mass transfer, tidal effects can have a large impact on binary evolution. Tides can create stars that are rapidly spinning until their death, leading to LGRBs (Izzard et al., 2004a); can circularise binaries pre-mass transfer (Hurley et al., 2002); and may suppress or enhance chemical mixing depending on how angular momentum transfer is modelled (Sciarini et al., 2026). Stars which are tidally synchronized during core hydrogen burning can rapidly spin, leading to their interiors becoming chemically homogeneous, such that they do not expand and fill their Roche lobes during the red-giant phase (Maeder, 1989; Langer, 1992; de Mink et al., 2008; Mandel & de Mink, 2016; du Buisson et al., 2020). This is known as chemically homogeneous evolution. If chemically homogeneous stars are born with close enough separations, they do not require mass transfer processes to harden the binary, and could lead to a BBH merger. However, achieving such highly spinning stars requires minimising the loss of angular momentum to stellar winds, and thus chemically homogeneous stars stars need to be low metallicity (e.g., Yoon et al., 2006). The requirements in progenitor metallicity, separation, and mass for BBHs created through chemically homogeneous evolution means that they should only span a narrow area of binary parameter space, but in turn this means they could produce BBHs with distinguishing features (Marchant et al., 2016). Due to the required parameter space for strong tidal interactions, BBHs created through chemically homogeneous evolution are predicted to be more massive than other isolated binaries, with more equal mass ratios (du Buisson et al., 2020; Riley et al., 2021).

While many predictions of the properties of BBHs from isolated binary evolution are still highly uncertain, a few signatures of isolated evolution are more confident. An upper mass gap, leading to a dearth of black holes between ~ 50–$130\,M_\odot$ is predicted due to the impact of PISNe (Woosley et al., 2002). The exact location of this gap is uncertain, and the lower edge could vary from ~ 35–$90\,M_\odot$ (Farmer et al., 2019; Costa et al., 2020; Farag et al., 2022). This gap is suspected to be a clear indicator of the PISN mechanism, though other binary interactions could mimic the presence of a build up of the lower edge of the gap (Briel et al., 2022b), or the gap itself (Olejak, 2026). Supernova physics may have an impact on structure in the mass distribution below the upper mass gap, though the way this structure manifests is model dependent (Schneider et al., 2023; Willcox et al., 2025). BBHs from isolated binary evolution are predicted to be preferentially spin-aligned, due to tidal effects realigning the binary's spin with the orbital angular

momentum post-supernova (Kalogera, 2000; Fragos et al., 2010; Gerosa et al., 2018). This requires small black hole natal kicks to avoid significant misalignment. In the case of large enough kicks, the misalignment of a black hole's spin vector could have no preference for alignment (Tauris, 2022; Baibhav & Kalogera, 2024). Identifying imprints from isolated binary evolution in GW data requires measuring these signatures in the context of multiple formation channels for unbiased astrophysical constraints.

1.2.3 | Binary black holes from dynamical interactions

An alternative channel to isolated binaries is the formation of BBHs in much denser environments than the galactic field, in particular, in stellar clusters. The stellar densities of clusters are $10–10^5$ pc^{-3}, compared to the stellar densities of the Milky Way's disk of 0.1 pc^{-3} (Kremer, 2026). Due to these high densities, there are expected to be frequent dynamical interactions between objects in the cluster, causing binary formation, disruption, and hardening. This lead to mass segregation, as the kinetic energies of objects tend to equalise during a dynamical interaction (Spitzer, 1969). Via conservation of momentum, less massive objects gain larger velocities during interactions (and are more likely to be ejected from the cluster), while more massive objects move towards the system's centre. The central regions of clusters are expected to be important locations for the formation of BBHs which are more massive than other objects in the cluster. Moreover, BBHs are vital to explain the abundance of stellar clusters: they inject energy into the cluster, delaying its collapse (Mackey et al., 2008; Breen & Heggie, 2013). BBHs which form through dynamical capture may avoid the requirement of hardening via stellar evolutionary processes: once captured they can already be at the required separation for merger via GW emission within a Hubble time, or can be brought to sufficiently close separations via dynamical hardening.

Binaries in stellar clusters may be made up of binaries that formed at birth (or 'primordial' binaries), or dynamically formed binaries. Primordial binaries in clusters may be indistinguishable from binaries formed in the field, assuming the same underlying distribution of initial stellar properties. Dynamically formed binaries can be produced by the interactions between > 2 objects, leading to gravitational capture, with $O(100)$ forming during a cluster lifetime (Aarseth & Heggie, 1976; Atallah et al., 2024; Kremer, 2026). After formation, binaries with separation closer than some strongly interacting limit will naturally harden further as a consequence of subsequent dynamical encounters (Kremer, 2026). This dynamical hardening also imparts a recoil kick on the binary, meaning sufficiently tight binaries are kicked out of the cluster, with higher mass clusters producing tighter ejected binaries. Mergers therefore happen through three processes in a cluster: post-ejection merger outside the cluster, merger within clusters without ejection due to higher eccentricities (Rodriguez et al., 2018; Kremer et al., 2020), and through GW capture resulting from sufficiently close dynamical encounters,

which happen on shorter timescales, requiring no hardening from a third body (Quinlan & Shapiro, 1987; Samsing et al., 2014, 2020; Rodriguez & Loeb, 2018).

Numerical relativity simulations have shown that merger products of BBHs will receive a post-merger kick of $\sim 100\,\mathrm{km\,s^{-1}}$ (Baker et al., 2006), though can be as high as $O(1000)\,\mathrm{km\,s^{-1}}$ for certain spin configurations (Campanelli et al., 2007; Bruegmann et al., 2008). Modelling kick velocities and their dependence on BBH properties is challenging due to the computational expense of detailed numerical relativity simulations (e.g. Islam & Wadekar, 2026). More certain is that remnant black holes are predicted to have spin magnitudes of ~ 0.67 (Pretorius, 2005; Gonzalez et al., 2007; Buonanno et al., 2008), though the retained remnant population may result in a wider range of spin magnitudes (Borchers et al., 2025).

If the velocity of the kick is higher than the escape velocity of the cluster, the resulting merger product is ejected, while if the kick is small or the escape velocity is high, the merger product can be retained in the cluster. Retained objects lead to the possibility of merger products to go on to merge again in a so-called hierarchical merger. Hierarchical mergers are expected to be a key tracer of dynamical black hole formation, creating black holes with masses impossible to create through isolated evolution due to the PISN mass gap, as well as signature spin magnitudes and mass ratios.

Beyond hierarchical mergers, the formation of BBH in stellar clusters is predicted to lead to several signatures properties of the signal. Due to the random pairing of BBHs in clusters, spin tilt directions are expected to be isotropic, rather than correlated with the orbital angular momentum as for isolated binaries (Rodriguez et al., 2016b). However, different assumptions of cluster physics could lead to preferential spin alignment (Kıroğlu et al., 2025). It is expected that a notable fraction of BBHs in clusters retain eccentricity until merger (O'Leary et al., 2009), unlike the majority of field binaries which circularise due to GW emission (Peters, 1964), making eccentricity an expected key tracer of dynamical formation. The history of cluster formation and evolution is uncertain, leading to uncertain predictions on rates of dynamical BBHs. The mass function of clusters, metallicity, and virial radius impact BBH formation (Rodriguez & Loeb, 2018; Kremer et al., 2020; Bruel et al., 2024).

There are multiple kinds of stellar clusters, all of which may host merging BBHs. Globular clusters are found in the halos of many galaxies, with > 150 globular clusters observed in the Milky Way (Harris, 1996). They host old stellar populations, and have masses 10^5–$10^6\,M_\odot$. There have been both direct and indirect observations of compact objects in globular clusters, while rough estimates for the rate of BBHs in globular clusters suggest that they could produce a significant fraction of observed mergers (Kremer, 2026). Nuclear star clusters have masses typically an order of magnitude higher, with majority $\geq 10^6\,M_\odot$ up to $\sim 10^9\,M_\odot$, and can have densities $> 10^6\,M_\odot/\mathrm{pc}^3$ (Neumayer et al., 2020). They may also produce a significant fraction of BBHs, and are expected produce a different distribution of BBH observables to globular clusters (Antonini & Rasio,

2016). Due to their larger masses, nuclear star clusters have a higher escape velocity than globular clusters, and higher rate of retention of binaries and BBH merger products (Miller & Lauburg, 2009). This could result in a higher fraction of higher generation BBH mergers in nuclear star clusters than in globular clusters. Nuclear star clusters are also expected to be provided with refreshed stellar populations from galaxy mergers or migration of star clusters, repopulating the BBH population (Antonini, 2014). Lastly, open clusters are clusters within galactic disks, typically are less massive 10^4–$10^5\,M_\odot$ (Banerjee, 2022). The most massive and young open clusters are called young massive clusters and are associated with increased rates of star formation. High-metallicity young massive clusters could produce a wider range of black hole masses, including a majority at $\sim 10\,M_\odot$ (Ye et al., 2026). Theoretical predictions of observational signatures from different types of stellar clusters may help us distinguish the origins of dynamically formed BBHs, though this requires disentangling these signatures from other contributing formation pathways.

1.2.4 | Binary black holes from exotic sources: AGN, triples, primordial black holes

While isolated binary evolution and dynamical evolution are the most established formation channels, other relatively less explored channels may create BBHs. These include BBHs which merge in the accretion disks of AGN, massive stellar triples, or primordial black holes.

BBHs may exist in disks of AGN, having migrated there due to dynamical friction (Morris, 1993; Rasio et al., 2003; O'Leary et al., 2009), been brought to the galactic centre via the infall of stellar clusters (Tremaine et al., 1975), galaxy mergers, or formed in-situ from massive stars in the disk (Goodman & Tan, 2004; Levin, 2007; McKernan et al., 2012). Here they may merge, facilitated by accretion from the disk or from dynamical interactions (McKernan et al., 2014; Bartos et al., 2017; Stone et al., 2017; Sedda et al., 2023b; McKernan et al., 2025). One signature of the AGN formation channel may be an increased abundance of massive black holes that can be be retained due to high escape velocities, avoiding ejection after recoil kicks (Tagawa et al., 2020). Furthermore, interactions between the disk and the binary may spin-up or spin-down black holes, driving aligned spins (Tagawa et al., 2020; Dittmann et al., 2024; McKernan & Ford, 2024). BBH mergers in disks may coincide with electromagnetic signatures due to interactions with the matter in the disk, allowing for multi-messenger analysis (Bartos et al., 2017; Graham et al., 2020).

Multiplicity of massive stars extends beyond the abundance of binaries; many, up to several tens of percent of, such massive stars are members of triple or even higher order systems (Sana et al., 2014; Moe & Di Stefano, 2017; Offner et al., 2023). The lives of triple systems are governed both by stellar interactions and

dynamical ones, depending on the stability of the system. Hierarchical triples, with a tighter inner binary and a third object surrounding it, are the most stable; while the most dynamically unstable systems lead to to a larger impact from the tertiary object. A higher fraction of triples undergo binary interactions (i.e., mass transfer) than binaries, due to the influence of the third object (Toonen et al., 2020), which may lead to BBHs which are tight enough to merge. Triple interactions may induce eccentricity visible in BBH mergers, but otherwise may not lead to a distinguishable BBH distribution from isolated evolution (Kummer et al., 2023).

Primordial black holes, formed from density fluctuations in the early Universe, may contribute to observations of GWs (Bagui et al., 2025; Sasaki, 2025; Andrés-Carcasona et al., 2026). They are of interest as a possible contribution to dark matter in the Universe (Carr & Kuhnel, 2020; Green & Kavanagh, 2021). The rates of stellar-mass primordial black holes and their properties if they exist are still highly uncertain, creating more uncertainty in the origins of the observed BBH population.

1.2.5 | Simulating populations of binary black holes

A clear description of the rates and properties of BBHs is challenging to produce, owing to the large array of free and uncertain parameters and multi-scale processes determining their evolution. While there has been significant progress in understanding the evolution of BBH progenitors, there are still many uncertainties of the physics of each channel, leading to a wide range of predictions for the rates and distributions of BBH properties even within the same channel (Mandel & Broekgaarden, 2022; Broekgaarden et al., 2026). Given the rarity of these systems in the Universe (hence lack of observations of their progenitors), simulations of these systems are necessary to better understand BBH astrophysics. For isolated binaries, a wide range of initial parameters of the systems (the intrinsic masses, metallicities of each star, as well as binary parameters such as eccentricity, separation) must be explored to account for the diverse physics of stellar binaries. For stars in clusters, interactions of millions of objects must be accounted for to study BBH formation.

Population synthesis codes allow modelling of populations of stellar systems using empirical prescriptions for their evolution. With population synthesis simulations, individual objects within stellar binaries can be evolved through evolutionary stages from zero age main sequence to merger, for millions of objects within a population. Alternatively, objects and their interactions (or statistical approximations thereof) within a stellar cluster can be evolved for a cluster lifetime. By varying the input physics to the simulations — such as prescriptions for supernovae, stellar winds, mass transfer, cluster mass, metallicity, etc. — multiple universes of stellar populations can be simulated with different assumptions. The

different simulations can be used to predict the resulting observations from each universe, and hence explore the uncertainties in BBH evolution.

For isolated binary (and triple) evolution, population synthesis codes simulate varying levels of detail, and therefore computational expense. Rapid population synthesis codes follow computationally efficient prescriptions describing single stars and binary interactions. Detailed codes model the stellar structure of each binary component along with their orbital evolution, requiring more computational resources to simulate the same number of systems. There is not a strict binary distinction between rapid and detailed codes: some detailed codes employ rapid prescriptions for (select) elements of the evolution, while some may interpolate detailed prescriptions for computational efficiency. Figure 1.2.2 shows the relationship between a selection of these codes.

Many rapid codes employ fitting formula (e.g. `SSE`, Hurley et al., 2000), or look-up tables (Kruckow et al., 2018) based on single-star models, and then model the binary interactions using separate prescriptions or parametrisations. The fitting formula from `SSE` were first expanded to include binary interactions in the `BSE` algorithm by Hurley et al. (2002). Following the release of `BSE`, many other rapid codes have been developed making use of fits from `SSE`, `BSE`, or otherwise, with variations to the evolutionary prescriptions. Examples of these include `binary_c` (Izzard et al., 2004b), `COMPAS` (Stevenson et al., 2017b), `COSMIC` (Breivik et al., 2020), `MOBSE` (Giacobbo et al., 2018), the `Scenario Machine` (Lipunov et al., 2009), `SeBa` (Portegies Zwart & Verbunt, 1996; Toonen et al., 2012), `StarTrack` (Belczynski et al., 2008) for binary evolution, and `MSE` (Hamers et al., 2021), `TSE` (Stegmann et al., 2022), and `TRES` (Toonen et al., 2016) for triples. While computationally efficient, rapid codes make a variety of assumptions about binary evolution. Due to the fixed assumptions from the fitting formulae, exploring varied stellar astrophysics, such as different internal energy transport or nuclear reaction rates is not possible. Furthermore, they are unable to resolve the effects of mass transfer on the interiors of either object in the binary (Breivik, 2026).

Some codes add further complexity to their models by interpolating between evolutionary tracks of single stars. Examples of these include `ComBinE` (Kruckow et al., 2018), `METISSE` (Agrawal et al., 2020), and `SEVN` (Iorio et al., 2022). These codes typically have a more detailed handling of single stellar evolution, though they incorporate binary evolution with the same assumptions as rapid codes.

Detailed codes directly models some aspects of binary evolution, and are therefore able to encompass some of the effects of mass transfer on the system. Examples include `BPASS` (Eldridge et al., 2017; Stanway & Eldridge, 2018), `POSYDON` (Fragos et al., 2022; Andrews et al., 2025), and The Brussels code (De Donder & Vanbeveren, 2004). `BPASS` incorporates the response of the accretor to mass transfer, but does not simultaneously evolve both binary objects. `POSYDON` uses grids of binary star models from `MESA` (Jermyn et al., 2023) for select stages of the binary's evolution, interpolating between simulations (Fragos et al., 2022;

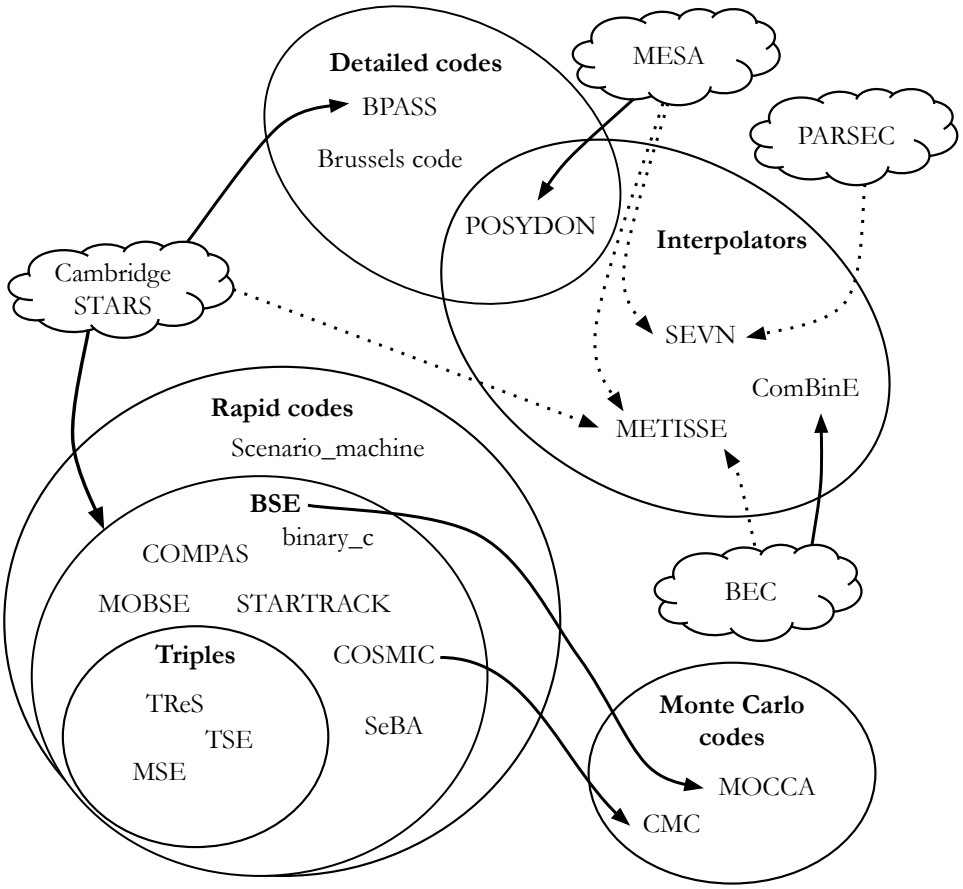


Figure 1.2.2: Map of the relationship between a selection of isolated binary population synthesis and Monte Carlo cluster codes used for BBH evolution. Circles represent groups of codes which broadly fall under the same category: detailed codes, interpolators, Monte Carlo codes, and rapid codes, which within it contains a group of BSE-based codes, and within that a group of codes which model triple systems. Clouds show codes which evaluate the single or binary stellar astrophysics, while arrows show the dependencies between these codes and the isolated binary population synthesis codes, or the isolated binary population synthesis codes and the cluster codes. Dotted arrows indicate the work has been done employing the given stellar physics package, but that the population synthesis code is not restricted to that code. Inspired by Iorio (2025).

Andrews et al., 2025). Varying assumptions about mass transfer in these codes remains challenging as it requires regenerating the detailed binary models, while varying supernova physics may be incorporated in post-processing. The most state-of-the-art detailed population synthesis models are highly computationally costly, taking $O(300)$ CPU hours for one population of $O(10^6)$ binaries (Andrews et al., 2025).

For simulations of dynamical systems in clusters, population synthesis codes can fall into three categories, in depending on the number of bodies required in the simulation (or equivalently the computational expense).

N-body codes capture the most detailed physics, directly modelling all objects in a system and their interactions. They find the sum of the gravitational force between each pair of objects, using this to compute the instantaneous acceleration of each body. Usually they model stellar evolution physics, using pre-computed stellar evolution tracks (e.g., `SSE` or `BSE`; Hurley et al., 2002). As the most computationally expensive dynamical code, N-body codes are limited to a low number of interactions and low binary fractions, with simulations of 10^6 objects requiring runtimes of ~ 1 yr even employing GPU parallelisation (Wang et al., 2015).

Monte Carlo codes approximates many two-body interactions with one effective interaction, statistically approximating each body's change in energy and momentum according to its location in the cluster (Hénon, 1971a,b). Examples include `CMC` (Rodriguez et al., 2022), which uses `COSMIC` for stellar evolution, and `MOCCA` (Hypki & Giersz, 2013), which uses `BSE`. Monte Carlo codes are still too computationally expensive for simulating clusters with the largest numbers of bodies, with more approximate methods necessary for clusters with $\gg 10^6$ objects.

For large dense clusters, e.g., nuclear star clusters, models can employ mass segregation and semi-analytic calculations. These semi-analytic models describe the interaction timescales analytically, then evolving the cluster through timesteps (e.g., Antonini & Rasio, 2016; Rose et al., 2022; Mapelli et al., 2021; Kritos et al., 2024). Recent works have used semi-analytical models and Monte Carlo methods for population synthesis of BBHs in gaseous environments, such as AGN (Bartos et al., 2017; Tagawa et al., 2020; Vaccaro et al., 2024; McKernan et al., 2025). While Monte Carlo and semi-analytical methods make varying assumptions about the dynamical interactions, they allow for more variation in the number of objects, and hence the types of host environments for BBHs in clusters and accretion disks.

Development of robust population synthesis predictions of BBHs remains a highly active area of research to better understand the abundance of evolutionary uncertainties. The differences in the handling of stellar evolution between codes for isolated binary evolution can have a significant impact on the observables of the resulting BBHs (Toonen et al., 2014; Gallegos-Garcia et al., 2021). Certain uncertainties, such as the fraction of mergers from the common envelope channel, are especially uncertain for BBHs, compared to BNSs (Broekgaarden et al., 2026).

Comparing simulations to observations is therefore essential to constrain inputs to population synthesis and the evolution of BBHs (e.g., Stevenson et al., 2017b; Neijssel et al., 2019; Zevin et al., 2021; Mastrogiovanni et al., 2022; Fishbach & Fragione, 2023). Investigating many physical assumptions is especially essential to probe correlations between uncertain population parameters. However, the computational expense of population synthesis makes direct comparisons to simulations covering a large population parameter space infeasible.

1.3 | Gravitational-wave population inference

The properties of BBHs revealed by GWs observations provide a crucial insights into the uncertainties in the evolution of massive binary stars, triples, stellar clusters, AGN, and primordial black holes. By understanding the population of GW sources, we can statistically probe the physics of BBHs. The growing abundance of BBH observations from GWs is of great advantage for studying the population, and determining the unclear astrophysics of their progenitors.

Understanding the uncertainties of the evolution of BBH progenitors from GWs requires an analysis framework to compare the observed BBHs population to a model of their formation or properties. This analysis is referred to as population inference, and is typically done in a hierarchical Bayesian framework for GW observations (Mandel et al., 2019; Thrane & Talbot, 2019; Vitale et al., 2020; Abac et al., 2025a). The main objective of population inference is to measure the free parameters describing the chosen population model, where these parameters are known as the model's hyperparameters. Population inference allows us to measure the astrophysical properties of BBHs, for example, the rate of mergers, the shape of the mass or spin distributions, or the fractional contributions from different formation processes.

Population inference requires three ingredients: a population model, measurements of individual BBH properties from parameter estimation, and an understanding of the observational selection effects (Mandel et al., 2019; Abac et al., 2026a). These selection effects describe the how the detected population relates to the astrophysical population: they allows us to transform the biased set of observations we detect to the underlying physics of BBH formation in the Universe. GW observations are especially suitable for statistical population inference due to their observational selection effects being particularly well understood.

In this section I outline the statistical framework behind population inference (Subsection 1.3.1), review the current methods of constructing a population model (Subsection 1.3.2), and review the current understanding of the BBH population from the latest sets of GW observations (Section 1.3.3).

1.3.1 | Statistical analysis of binary black hole observations

Hierarchical Bayesian inference is used to measure the most probable values of population hyperparameters given the observed BBH population. This is done by comparing the individual measurements (for example the masses, spins, mass ratios, and distances) of GW observations to a population model. Inference can be done to solely constrain the shape of the population distribution. Alternatively, incorporating the number of observations allows measurement of the overall rate of BBH mergers per given volume and timespan of the Universe. Here, I derive the probability distribution of the population hyperparameters given GW data, first to measure the shape of the population, then incorporating selection effects, and finally, incorporating the rate of observations.

After obtaining measurements on individual event parameters for N events, $\{\vec{\theta}\}$, from GW strain data $\{\vec{d}\}$, our objective is to measure the hyperparameters of the model, $\vec{\Lambda}$, given the GW data, e.g., $p(\vec{\Lambda}|\{\vec{d}\})$. To measure the hyperposterior $p(\vec{\Lambda}|\{\vec{d}\})$, we employ Bayes theorem (Bayes & Price, 1763):

$$p(\vec{\Lambda}|\{\vec{d}\}) = \frac{p(\{\vec{d}\}|\vec{\Lambda})\pi(\vec{\Lambda})}{p(\{\vec{d}\})}, \tag{1.13}$$

where $p(\{\vec{d}\}|\vec{\Lambda})$ is the hyperlikelihood of observing data $\{\vec{d}\}$ for a population described by hyperparameters $\vec{\Lambda}$; $\pi(\vec{\Lambda})$ is the hyperprior, our intrinsic belief in the existence of a population described by $\vec{\Lambda}$; and $p(\{\vec{d}\})$ is the hyperevidence, the probability of observing data $\{\vec{d}\}$ under any assumptions about the population, i.e. any value of $\vec{\Lambda}$ (Mandel et al., 2019; Thrane & Talbot, 2019; Vitale et al., 2020).

For multiple independent data $\vec{d}_i$, where i corresponds to each event, and $i \in [0, N]$, the hyperlikelihood $p(\{\vec{d}\}|\vec{\Lambda})$ is a product over the individual event probabilities (Mandel et al., 2019; Thrane & Talbot, 2019),

$$p(\{\vec{d}\}|\vec{\Lambda}) = \prod_i^N p(\vec{d}_i|\vec{\Lambda}). \tag{1.14}$$

As we do not know the true event parameters, the probability of observing the data of one event given $\vec{\Lambda}$, $p(\vec{d}_i|\vec{\Lambda})$, can be marginalised over all potential true parameters of the event, $\vec{\theta}_i$,

$$p(\vec{d}_i|\vec{\Lambda}) = \int \mathrm{d}\vec{\theta}_i p\left(\vec{d}_i|\vec{\theta}_i\right) p\left(\vec{\theta}_i|\vec{\Lambda}\right). \tag{1.15}$$

Here $p(\vec{d}_i|\vec{\theta}_i)$ is the individual event likelihood, and $p(\vec{\theta}_i|\vec{\Lambda})$ is the probability of the true event parameters given our population model, $p(\vec{\theta}|\vec{\Lambda})$, which determines the true underlying distribution of BBH properties given $\vec{\Lambda}$. Combining

equations 1.14 and 1.15, we have (Thrane & Talbot, 2019)

$$p(\{\vec{d}\}|\vec{\Lambda}) = \prod_i^N \int \mathrm{d}\vec{\theta}_i p\left(\vec{d}_i|\vec{\theta}_i\right) p\left(\vec{\theta}_i|\vec{\Lambda}\right). \tag{1.16}$$

Meanwhile, the hyperevidence $p(\{\vec{d}\})$ is the integral over all possible values of the multi-event likelihood multiplied by the hyperprior,

$$p(\{\vec{d}\}) = \int \mathrm{d}\vec{\Lambda} p(\{\vec{d}\}|\vec{\Lambda})\pi(\vec{\Lambda}). \tag{1.17}$$

We are therefore faced with a double integral across $aN + M$ dimensions where a is the dimensionality of $\vec{\theta}$ and M is that of the hyperparameters (Thrane & Talbot, 2019). We can avoid this using the samples from individual event posteriors, $p\left(\vec{\theta}_i|\vec{d}_i\right)$, which conveniently is the end product from individual event parameter estimation. (Mandel et al., 2019) This employs the equivalence of the expectation value of a function $f(x)$ with an average of samples from a posterior distribution,

$$\langle f(x)\rangle_{p(x)} = \int \mathrm{d}x p(x) f(x) \approx \frac{1}{n_s}\sum_k^{n_s} f\left(x_k\right). \tag{1.18}$$

In our case, $p(x)$ is the posterior distribution function from which we have samples, and the sum over k runs over n_s posterior samples. Furthermore, by choosing a fiducial population model described by $\varnothing$, giving a fiducial prior on each event's parameters, $\pi(\vec{\theta}_i|\varnothing)$, the individual event likelihood can be written using the rearranged Bayes' theorem as

$$p\left(\vec{\theta}_i|\vec{d}_i\right) = \mathcal{Z}_\varnothing(\vec{d}_i)\frac{p(\vec{\theta}_i|\vec{d}_i,\varnothing)}{\pi(\vec{\theta}_i|\varnothing)}. \tag{1.19}$$

Plugging this into eq. 1.16,

$$p(\{\vec{d}\}|\vec{\Lambda}) = \prod_i^N \int \mathrm{d}\vec{\theta}_i \mathcal{Z}_\varnothing(\vec{d}_i) p(\vec{\theta}_i|\vec{d}_i,\varnothing)\frac{p\left(\vec{\theta}_i|\vec{\Lambda}\right)}{\pi(\vec{\theta}_i|\varnothing)}, \tag{1.20}$$

and using eq. 1.18, the hyperlikelihood becomes (Mandel et al., 2019; Thrane & Talbot, 2019; Vitale et al., 2020)

$$p(\{\vec{d}\}|\vec{\Lambda}) = \prod_i^N \frac{\mathcal{Z}_\varnothing(\vec{d}_i)}{n_i}\sum_k^{n_i}\frac{p\left(\vec{\theta}_i^k|\vec{\Lambda}\right)}{\pi(\vec{\theta}_i^k|\varnothing)}. \tag{1.21}$$

Essentially, this allows for reweighting of posterior samples from GW events by new hyperparameters $\vec{\Lambda}$ for population inference. The fiducial prior can be

chosen to be any population model given support across the range of $\vec{\theta}$ of interest (Thrane & Talbot, 2019). The fiducial evidence is ultimately cancelled out in the final calculation and the hyperposterior becomes:

$$p(\vec{\Lambda}|\{\vec{d}\}) = \prod_i^N \frac{\sum_k^{n_i} \left[p\left(\vec{\theta}_i^k|\vec{\Lambda}\right) / \pi(\vec{\theta}_i^k|\varnothing) \right] \pi(\vec{\Lambda})}{\int \mathrm{d}\vec{\Lambda} \sum_k^{n_i} \left[p\left(\vec{\theta}_i^k|\vec{\Lambda}\right) / \pi(\vec{\theta}_i^k|\varnothing) \right] \pi(\vec{\Lambda})}, \tag{1.22}$$

where the denominator could be replaced by $\int \mathrm{d}\vec{\Lambda} p(\{\vec{d}\}|\vec{\Lambda})\pi(\vec{\Lambda})$, requiring an extra factor of $\mathcal{Z}_{\varnothing}(\vec{d}_i)/n_i$ in the numerator.

The above analysis does not consider the fact that some GW signals will be harder or impossible to detect, and thus the detected events will be biased by selection effects. Each event can be said to have some probability of being detected depending on the data, $p_{\mathrm{det}}(\vec{d}_i)$, which has some dependence on the true event parameters $\vec{\theta}_i$. The dependence on $\vec{\theta}_i$ can be written as an integral over all possible realisations of the data for a signal with true parameters $\vec{\theta}_i$:

$$p_{\mathrm{det}}(\vec{\theta}) = \int_{\vec{d}>\mathrm{threshold}} p(\vec{d}|\vec{\theta})\mathrm{d}\vec{d} = \int \mathcal{I}_{\mathrm{det}}(\vec{d})p(\vec{d}|\vec{\theta})\mathrm{d}\vec{d}, \tag{1.23}$$

where $\mathcal{I}_{\mathrm{det}}(\vec{d})$ is an indicator function with $\mathcal{I}_{\mathrm{det}}(\vec{d}) = 1$ for a detected signal, and 0 otherwise (Mandel et al., 2019). The probability of detection is hence the integral of $p(\vec{d}|\vec{\theta})$ over all data realisations for which the signal is detected (Mandel et al., 2019; Vitale et al., 2020). The detection probability depends on both intrinsic and extrinsic properties of the signal, as well as the properties of the noise in the detector.

In equation 1.16, we implicitly normalised by the population distribution $\int \mathrm{d}\vec{\theta}_i p(\vec{\theta}_i|\Lambda)$, which is equal to one. Now, in order to take into account the selection effects on the population, we include dependence on p_{det} in the integral over true event parameters. In the numerator, $p_{\mathrm{det}}(\vec{\theta}) = 1$ as we assume an event is detected when it enters our observations. Given observations which come from a population where each event has detection probability $p_{\mathrm{det}}(\vec{\theta})$, the hyperlikelihood (eq. 1.16) becomes normalised by an additional selection term (Mandel et al., 2019; Vitale et al., 2020),

$$p(\{\vec{d}\}|\vec{\Lambda}) = \prod_i^N \frac{\int \mathrm{d}\vec{\theta}_i p\left(d_i|\vec{\theta}_i\right) p\left(\vec{\theta}_i|\Lambda\right)}{\int \mathrm{d}\vec{\theta}_i p_{\mathrm{det}}(\vec{\theta}_i) p\left(\vec{\theta}_i|\Lambda\right)}. \tag{1.24}$$

The denominator is often indicated as the detection efficiency, $\xi(\vec{\Lambda})$. This detection efficiency can be described as the fraction of detected sources from a

population described by hyperparameters $\vec{\Lambda}$, i.e.

$$\tilde{\xi}(\vec{\Lambda}) \equiv \frac{N_{\rm det}}{N_{\rm pop}}, \tag{1.25}$$

where $N_{\rm pop}$ is the total number of sources that are produced by a population of BBHs in the Universe, while $N_{\rm det}$ is the number of detectable events in the same population, defined by

$$N_{\rm det} = \int \mathrm{d}\vec{\theta} \frac{\mathrm{d}N_{\rm pop}}{\mathrm{d}\vec{\theta}} p_{\rm det}(\{\vec{\theta}\}) = N_{\rm pop} \int \mathrm{d}\vec{\theta} p_{\rm det}(\vec{\theta}) p\left(\vec{\theta}|\Lambda\right), \tag{1.26}$$

therefore giving us (Vitale et al., 2020):

$$\tilde{\xi}(\vec{\Lambda}) = \int \mathrm{d}\vec{\theta} p_{\rm det}(\vec{\theta}) p\left(\vec{\theta}|\Lambda\right). \tag{1.27}$$

In the presence of selection effects, the hyperlikelihood is proportional to $\tilde{\xi}(\vec{\Lambda})^{-N}$:

$$\begin{aligned} p(\{\vec{d}\}|\vec{\Lambda}) &= \prod_i^N \frac{\int \mathrm{d}\vec{\theta}_i p\left(d_i|\vec{\theta}_i\right) p\left(\vec{\theta}_i|\Lambda\right)}{\tilde{\xi}(\vec{\Lambda})} \\ &= \frac{1}{\tilde{\xi}(\vec{\Lambda})^N} \prod_i^N \int \mathrm{d}\vec{\theta}_i p\left(d_i|\vec{\theta}_i\right) p\left(\vec{\theta}_i|\Lambda\right). \end{aligned} \tag{1.28}$$

Accurately describing the selection effects is important, especially with large observational catalogues, as for each $\vec{\Lambda}$, the same contribution to the detection efficiency is scaled to the power of N.

The detection probability can be determined in practice through different methods. Placing a cut on SNR can approximate the detection probability of a signal, where $p_{\rm det}(\theta) = 1$ for the SNR above some threshold value, and zero otherwise. However, this does not account for loud non-stationary glitches which can have reasonably high SNR. Instead, a cut can be made on the FAR of observations, which considers the rate of background glitches, mitigating pollution from non-astrophysical signals. In GW catalogues the main detection statistic used is $p_{\rm astro}$, the probability of an events being truly astrophysical (Abbott et al., 2023a). However, the $p_{\rm astro}$ statistic includes a true alarm rate assuming a population model, and therefore may induce bias in population inference, should this model not be an accurate description of the observed population. Additionally, the lack of a unified method for calculating $p_{\rm astro}$ across search pipelines complicates its use for population studies (Banagiri et al., 2023). Therefore $p_{\rm astro}$ is typically not used in population inference. Typically, a FAR cut of $1\,\mathrm{yr}^{-1}$ is used for population inference, resulting in 42, 69, 153 and 267 BBHs used for inference in deep extended second Gravitational-Wave Transient Catalogue (GWTC-2.1), GWTC-3.0, GWTC-4.0, and GWTC-5.0 respectively.

Injection campaigns can be used to empirically estimate the detection probability for a population of events for a given search pipeline by finding which events are recovered from the pipeline (Tiwari, 2018; Essick et al., 2025). The output from injection campaigns is a measure of the detection statistics for each injected signal (e.g., SNR, FAR). Injections recovered with a statistic above the threshold denote which injections are detected. The detection probability $p_{\rm det}(\theta)$ can then be found with a Monte-Carlo sum comparing the distribution detected injections to the original injected distribution (Essick et al., 2025). This is the most statistically robust method, as it allows us to account for biases from search pipelines, and real detector noise, which fluctuates over the observing run. However, injection campaigns requires a large computational expense, with $O(10^8)$ injections needed to cover the required parameter space.

The above derivation for $p(\{\vec{d}\}|\vec{\Lambda})$ lets us infer the hyperparameters governing the shape of the distribution of BBH parameters. Incorporating the number of detections and the length of observing time allows us to infer the overall rate of observations. This assumes that BBH mergers are a Poisson process, with the probability of actually observing N events, given that we expect to detect $N_{\rm det}$ events, given by

$$p(N|N_{\rm det}) = e^{-N_{\rm det}}\,(N_{\rm det})^N\,. \tag{1.29}$$

Accounting for the Poissonian rate of events and our prior on the total expected number of detectable and non-detectable sources, $\pi(N_{\rm pop})$, the hyperposterior on the hyperparameters including the rate becomes (Mandel et al., 2019; Vitale et al., 2020),

$$\begin{aligned} p(\vec{\Lambda}, N_{\rm pop}|\{\vec{d}\}) =& \frac{\pi(N_{\rm pop})e^{-N_{\rm det}}(N_{\rm det})^N}{\bar{\xi}(\vec{\Lambda})^N} \\ &\times \prod_i^N \frac{\sum_k^{n_i}\left[p\left(\vec{\theta}_i^k|\vec{\Lambda}\right)/\pi(\vec{\theta}_i^k|\varnothing)\right]\pi(\vec{\Lambda})}{\int \mathrm{d}\vec{\Lambda}\sum_k^{n_i}\left[p\left(\vec{\theta}_i^k|\vec{\Lambda}\right)/\pi(\vec{\theta}_i^k|\varnothing)\right]\pi(\vec{\Lambda})}. \end{aligned} \tag{1.30}$$

Evaluating the rate-included hyperposterior therefore requires us to consider the actual number of detected events N and the expected number of detected events $N_{\rm det}$, which may not be the same due to Poisson fluctuations. While the detection efficiency $\bar{\xi}(\vec{\Lambda})$ explicitly encodes selection effects, they are implicitly encoded in $N_{\rm det}$ (Vitale et al., 2020). If the prior $\pi(N_{\rm pop}) \propto 1/N_{\rm pop}$, then we can marginalise over $N_{\rm pop}$ to regain equation 1.22, aside from a normalisation constant that only depends on the number observed events N (Fishbach et al., 2018).

Due to the high dimensionality of $p(\vec{\Lambda}|\{\vec{d}\})$, it is almost always infeasible to evaluate the hyperposterior analytically. In practice, the hyperposterior is found by sampling the distribution over the prior range of $\vec{\Lambda}$. This generates samples from the hyperposterior rather than an evaluation of the probability distribution itself.

The most common sampling methods are Markov Chain Monte Carlo (MCMC) and nested sampling (Thrane & Talbot, 2019).

MCMC draws samples with a Markov Chain, with 'walkers' moving through hyperparameter space towards areas of higher probability (Metropolis et al., 1953; Hastings, 1970). The walker's next step is defined by a conditional probability governed by the underlying distribution. The walker's trace will converge to the target distribution after some burn-in period when the walker's location is defined by the random initialisation. By tracking the position of the walkers as they explore the hyperposterior space, we save draws from the target distribution, which follow the hyperposterior.

Nested sampling generates samples from the target distribution as a by-product of calculating the evidence (Skilling, 2004). To do this, a set of live points are distributed in the target space, and recursively the lowest likelihood point is removed, and replaced with a point from the prior with a higher likelihood. At each stage the evidence is calculated by finding the sum of the likelihood multiplied by each sample point's prior volume. Sampling is stopped once the evidence gets sufficiently close to a known upper limit, which is found by assuming the highest likelihood for the remaining prior volume. The positions of the live points trace the target hyperposterior after this condition is met.

Both MCMC and nested sampling require many hyperlikelihood evaluations. This requires a well-described and computationally efficient population model for population inference. The following subsection introduces current methods of constructing such population models.

1.3.2 | Constructing a population model

The population model, $p(\vec{\theta}|\vec{\Lambda})$, encodes the expected astrophysics and characteristics of the BBH population. The method of constructing population models has followed several paths. Each class of models can be characterised by their degree of assumptions about the form of the population and the underlying population astrophysics.

One class of method is parametric models, which fit the observed distributions of observables with a simple functional form (e.g., a power-law or Gaussians, Fishbach & Holz, 2017; Talbot & Thrane, 2018; Kimball et al., 2021; Callister et al., 2021). The functional forms may be motivated by astrophysical theory or convenience. Parametric models can be used to test expected assumptions about the shape of the underlying population, agnostic to assumptions in theory-dependent population synthesis. For example, they can be used to test a build-up of BBHs from PISNe (Golomb et al., 2024), the rate of highly spinning black holes as a signature of observation mergers (Antonini et al., 2025), or the agreement of the redshift evolution of BBHs with the global star formation rate (Abac et al., 2025a). Parametric models are often intuitive, and their analytical form make

them computationally inexpensive. However, requiring model features that do not truly exist can lead to model misspecification or overfitting, leading to a mischaracterisation of BBH astrophysics.

The most astrophysics-agnostic class is data-driven models, which do not assume a predefined shape of the population, rather seek to represent the population with maximum flexibility. They are also referred to as weakly-parametrised models as opposed to strongly-parametrised parametric models. There are various techniques for constructing a data-driven models, for example using Gaussian mixture models, splines, binning, or auto-regressive processes (Mandel et al., 2017; Rinaldi & Del Pozzo, 2021; Tiwari, 2021; Golomb & Talbot, 2022; Edelman et al., 2023; Heinzel et al., 2024; Callister & Farr, 2024; Ray et al., 2023). These analyses are less constrained by a pre-defined shape of the population model, and can therefore identify features in the data that were previously unaccounted for in parametric models. However, these models are typically governed by many more parameters than parametric models, leading to a higher computational expense and higher uncertainties.

Last, population synthesis models may be compared to GW observations (Stevenson et al., 2017b; Zevin et al., 2017; Neijssel et al., 2019; Bouffanais et al., 2021; Zevin et al., 2021; Bavera et al., 2021; Mastrogiovanni et al., 2022; Stevenson & Clarke, 2022; Iwaya et al., 2023; Chen et al., 2025; Gayathri et al., 2025). Directly comparing population synthesis to observations can give the strongest astrophysical constraints on the inputs to the simulations (e.g., direct constraints on metallicity evolution, supernova kick magnitude, or cluster mass function). However, these constraints strongly depend on the assumptions that have generated the population synthesis models. These assumptions may be very different between different population synthesis codes. Additionally, generating population synthesis simulations is much more computationally expensive than generating a parametric or data-driven model. Constructing a model representation from the population synthesis output samples is necessary to evaluate the population likelihood. The diversity of population synthesis predictions, lack of consistency between different codes predicting mergers from different channels, and the computational expense leads to challenges in varying the input physics from multiple channels to population synthesis models for comparison to data.

Altogether, a combination of multiple methods is necessary to understand the population. Data-driven models allow identification of new features, while parametric models allow us to intuitively probe those features, and comparison to population synthesis lets us directly constrain population astrophysics.

1.3.3 | Current understanding of the binary black hole population

This subsection outlines the current understanding of the BBH population from third Gravitational-Wave Transient Catalogue (GWTC-3.0) and GWTC-4.0. This includes results from marginal (i.e. one-dimensional) distributions and correlations between BBH parameters. Marginal distributions tell us about a particular property of interest, such as masses, or spins. Correlations between population parameters can encode additional information on BBH formation beyond the marginal distributions, but require more data to probe the additional parameter space. Understanding the physics of BBH formation requires studying features in both marginal and correlated distributions.

The primary masses of BBHs span $\sim 5\,M_\odot$ to $\sim 100\,M_\odot$, with more low-mass BBHs than high-mass BBHs (Figure 1.3.1 Abac et al., 2025a). The rate of low-mass black holes (below $\sim 8\,M_\odot$) is dependant on the thresholds on mass used to categorise BBHs from NSBHs for BBH only analyses, and therefore is not well constrained. BBH-only analyses set an arbitrary threshold requiring that both component masses are greater than $3\,M_\odot$ to $> 99\%$ certainty. Observations with uncertain classifications are typically omitted from BBHs only analyses, such as GW190814 which has $m_2 = 2.6^{+0.1}_{-0.1} M_\odot$, and is often considered an outlier to the population (Abbott et al., 2020b; Essick et al., 2022). Population inference of the full population of both black hole and neutron star-containing observations allow the study of low mass black holes, and objects with uncertain classifications (Farah et al., 2022).

The majority of BBHs have primary masses of $\sim 10\,M_\odot$. There is likely another feature in the primary mass distribution at $m_1 \sim 35\,M_\odot$, though it is not clear whether this is an over-density or a steeper drop off in masses after $m_1 \sim 40\,M_\odot$ (Abbott et al., 2023b; Abac et al., 2025a). Additional structure at $\sim 20\,M_\odot$ and $\sim 60\,M_\odot$ has been found, though these are more model-dependant and less statistically robust (Toubiana et al., 2023; Pierra & Papadopoulos, 2026). Below the predicted upper-mass gap, peaks in the mass distribution have been attributed to various formation processes (e.g., van Son et al., 2022a; Li et al., 2025; Ye et al., 2026; Legred et al., 2026). Observations of black holes between 50–$100\,M_\odot$ suggest that there are ways to create massive black holes in the pair instability mass gap, such as through stellar or black hole mergers.

Characterising the full BBH mass distribution can be done through modelling the mass ratio distribution jointly with the primary mass distribution. It is well established that BBHs prefer to merge with similar masses (Figure 1.3.2; Fishbach & Holz, 2020; Abbott et al., 2021b, 2023b; Abac et al., 2025a). Parametric models typically model mass ratio as a power-law distribution, requiring the mass ratio distribution to be monotonic, therefore requiring a maximum at $q = 1$ for positive power-law indices. Several data-driven methods have found that the mass ratio distribution peaks away from one, though still favouring pairings with similar

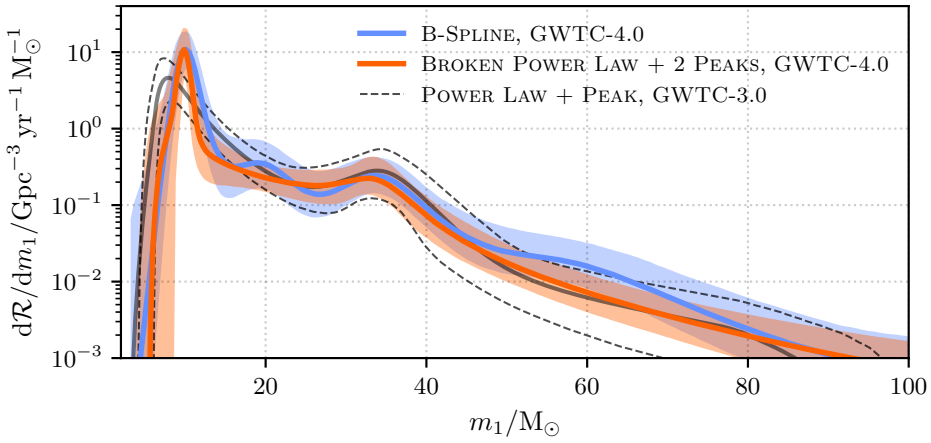


Figure 1.3.1: Primary mass distribution of BBHs in GWTC-4.0 using the weakly-parametrised B-Spline model (Edelman et al., 2023) and the strongly-parametrised Broken Power Law + 2 Peaks model Abac et al. (2025a), compared to the distribution of BBHs in GWTC-3.0 with the strongly-parametrised Power Law + Peak model (Talbot & Thrane, 2018). With both catalogues, the global maximum of the distribution is at $\sim 10\,M_\odot$, there is an additional feature at $\sim 35\,M_\odot$, and the high mass tail of the distribution extends up to $\sim 100\,M_\odot$. Figure modified from Abac et al. (2025a).

masses (Sadiq et al., 2024; Godfrey et al., 2023; Rinaldi et al., 2024; Callister & Farr, 2024; Rinaldi et al., 2025). Furthermore, the mass ratio distribution has been shown to vary with primary mass, with different features along the primary mass spectrum exhibiting different mass ratio distributions (Abbott et al., 2023b; Godfrey et al., 2023; Roy et al., 2025; Banagiri et al., 2025b). Analyses which jointly model the m_1–m_2 space agree with the findings that BBHs prefer to merge with equal masses, though show there may be different features in the secondary mass spectrum than primary masses (Farah et al., 2024). Combined, these results show that BBHs do not follow a random pairing, but that there is structure in their joint mass distribution arising from the physics of their formation.

Studies have found that most BBHs have small spins, with very little support at large spin magnitudes (Roulet & Zaldarriaga, 2019; Callister et al., 2022; Abbott et al., 2023b; Abac et al., 2025a). It is unclear whether all black holes have some degree of non-zero spin, or whether some fraction have negligible spins (Kimball et al., 2021; Callister et al., 2022; Tong et al., 2022; Adamcewicz et al., 2025). A subpopulation of BBHs with negligible spins would suggest that these black holes form from stars with efficient angular momentum transport and resist spin-up post formation (Fuller & Ma, 2019). There may be correlations between spin parameters that encode BBH formation physics (Hussain et al., 2026; Plunkett et al., 2026). The distribution of spin tilts indicate that there may be some prefer-

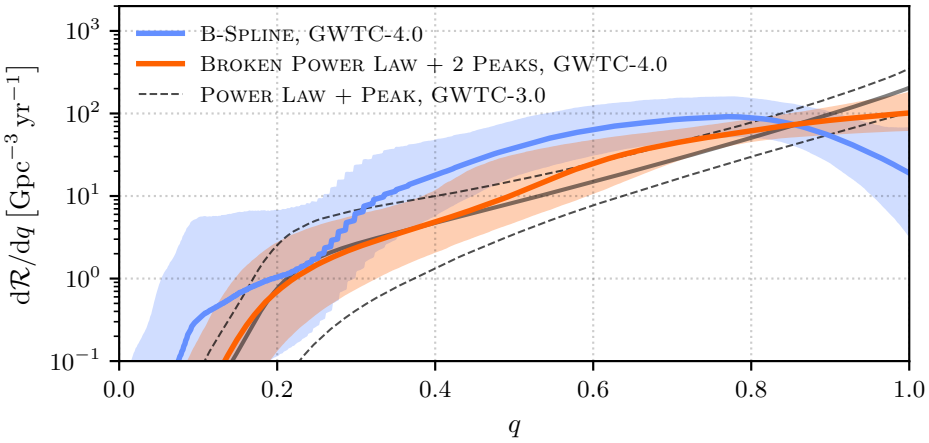


Figure 1.3.2: Mass ratio distribution of BBHs in GWTC-4.0 with parametric Power Law model (orange) and data-driven B-Spline model (blue), compared to the parametric result from GWTC-3.0 (dashed line). There is a preference for equal mass mergers, with support down to $q \sim 0.2$ with both models. The more flexible B-Spline model finds that the distribution peaks away from $q \sim 1$, closer to $q \sim 0.8$. Figure modified from Abac et al. (2025a).

ence for aligned spins, though there is support for a large range of tilts within the observed population (Baibhav et al., 2023; Vitale et al., 2022; Abbott et al., 2023b; Abac et al., 2025a).

Measuring the distribution of component spin magnitudes and tilts is challenging due to individual event uncertainties, and definitive population inference of spin magnitudes and tilts may not be possible even with $>$ 1000 observations (Miller et al., 2024; Vitale & Mould, 2025). Effective spin distributions can give more robust measurements of the BBH spins. The effective inspiral spin distribution exhibits some asymmetry around zero, suggesting that there is some preference for alignment (Figure 1.3.3). Furthermore, the distribution has non-zero support at negative effective inspiral spins, suggesting that there is a contribution from anti-aligned BBHs (Banagiri et al., 2025a; Abac et al., 2025a). This could indicate presence of dynamically formed BBHs, or binaries which have been misaligned due to supernova kicks.

The merger rate of BBHs has been shown to increase with increasing redshift with both parametric and non-parametric models (Fishbach et al., 2018; Abbott et al., 2023b; Callister & Farr, 2024; Ray et al., 2023; Abac et al., 2025a). Parametric models find that the redshift evolution is consistent with cosmic star formation rate density from Madau & Dickinson (2014), consistent with the intuition that BBHs trace the Universe's star formation. Future observations may unveil a turnover in the redshift evolution, expected due to the turnover in star formation at $z \sim 2$, though no such feature has been confidently observed within the

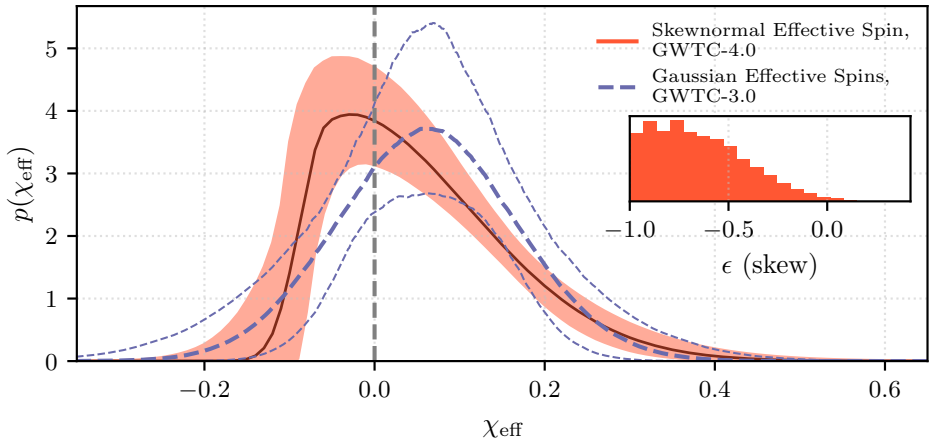


Figure 1.3.3: Effective inspiral spin distribution of BBHs in GWTC-4.0, measured with a parametric SKEWNORMAL model, compared to the result from GWTC-3.0 using a GAUSSIAN model. Both models exhibit a asymmetric distribution with a presence of BBHs with both positive and negative χ_{eff}, while the SKEWNORMAL model finds strong evidence for a skewed distribution. Figure modified from Abac et al. (2025a).

observing horizon of current detectors.

Beyond the study of marginal masses, spins, and redshifts, multiple correlations have been identified in the BBH population. These include structure between BBH masses and spins, mass ratio and spins, redshifts and spins.

Correlations between effective inspiral spin and mass ratio have been reported starting with GWTC-2 data, indicating that χ_{eff} increases and/or gets broader towards more unequal mass ratios (Callister et al., 2021; Adamcewicz & Thrane, 2022; Abbott et al., 2023b; Abac et al., 2025a). The underlying astrophysics causing this correlation is not yet clear, with explanations varying from mass ratio reversal in stable mass transfer, hierarchical mergers, to AGN (Olejak et al., 2024; McKernan et al., 2025; Wang et al., 2026; Vijaykumar et al., 2026).

High mass BBHs have higher spins, with a larger spread of χ_{eff} values, higher spin magnitudes and an isotropic spin tilt distribution (Biscoveanu et al., 2022; Franciolini & Pani, 2022; Heinzel et al., 2024; Antonini et al., 2025; Pierra et al., 2024; Tong et al., 2025; Tiwari, 2025; Ray & Kalogera, 2026). The indication that black holes with masses above the predicted PISN mass gap have higher spins may be a signature of hierarchical mergers which are predicted to have higher masses and larger, isotropically distributed spins. Strong evidence for this signature of hierarchical-like-mergers may be model dependant (Ray & Kalogera, 2026). Other studies have found that the effective inspiral spin distribution may broaden with redshift (Biscoveanu et al., 2022; Heinzel et al., 2024; Farah et al., 2026). Because we detect more massive BBHs at higher redshifts, correlations between masses

and spins versus redshifts and spins may not be distinguishable without studying the combined 3D correlation.

These correlations provide growing evidence for features in the BBH population that cannot be explained by a single formation channel. Before and during GWTC-3.0, observations of multiple correlations hinted at differing formation physics (Kimball et al., 2021; Baibhav et al., 2023; Ray et al., 2025; Galaudage & Lamberts, 2025), while comparisons to population synthesis showed evidence for a range of formation channels (Zevin et al., 2021; Cheng et al., 2023; Li et al., 2024b; Gayathri et al., 2025). With increased observations, evidence for multiple formation channels is growing in GWTC-4.0 (e.g., Banagiri et al., 2025b; Ray et al., 2026b). This evidence persists in GWTC-5.0, with the inclusion of $O(100)$ more observations allowing a more detailed view into the correlations between BBH parameters (Abac et al., 2026a). Understanding the relative contributions from different formation channels and pinning down the underlying astrophysics of each channel is a highly active area of research, following a range of methods to study the population of BBHs observed with GWs.

1.4 | Outline of new work

The evolution of BBH progenitors is highly uncertain, as it requires understanding multi-scale processes across multiple formation channels, for which there are few observations and computational challenges in accurate simulations. Population inference of GW observations provides a unique way to study these sources. The now hundreds of GWs offer increasing insight into the unknown astrophysics of BBH formation. In this thesis, I address these unknowns with new analyses of GW observations. I detail the development of new methods to analyse the observed BBH population using complementary techniques (Chapters 2 and 3), and apply these analyses to the successive catalogues of GWs (Chapter 4).

First, I use parametric population models to investigate new conjectures about correlations between black hole masses and mass ratios in Chapter 2. Following the observed preference for equal masses in the BBH population (Fishbach & Holz, 2020; Abbott et al., 2021b, 2023b), I test whether big black holes can merge with the smallest black holes with GWTC-3.0 observations (Abbott et al., 2023a). Second, I develop machine-learning emulators for population synthesis models, allowing direct measurements of the inputs to astrophysical simulations of BBH populations over a continuous range. Training the emulators on population synthesis simulations from five formation channels, I compare to a non-interpolating method, and show inference results with GWTC-3.0 observations. Applying these analyses to GWTC-2.1 (Abbott et al., 2024), GWTC-3.0, and GWTC-4.0 (Abac et al., 2026b), I show the constraining power of the growing number of observations, and the necessity of accurately modelling both the BBH population and selection effects. With upcoming GW data set to revolutionise our under-

standing of compact object astrophysics, this work provides the tools necessary to maximise insights into BBH formation with robust inference methods.

2 Can Big Black Holes Merge with the Smallest Black Holes?

2.1 | Abstract

Measurements of correlations between BBHs population properties provide insights into the evolution of merging binaries. We explore potential correlation between mass and mass ratio with phenomenological population models where the minimum mass of the smaller (secondary) black hole can change with the mass of the bigger (primary) black hole. We use binary black hole signals from the GWTC-3.0 with and without the relatively extreme mass-ratio GW190814. When excluding GW190814, models with a variable minimum mass are disfavoured compared to one with a constant minimum mass, with log Bayes factors of -2.49 to -0.98, indicating that the biggest black holes can merge with the smallest. When including GW190814, a parabola model that allows the minimum mass to decrease with increasing primary mass is favoured over a constant minimum-mass model with a log Bayes factor of 4.44. When allowing the minimum mass to decrease, the overall population distributions remain similar whether or not GW190814 is included. This shows that with added model flexibility, we can reconcile potential outlier observations within our population. These investigations motivate further explorations of correlations between mass ratio and component masses in order to understand how evolutionary processes may leave an imprint on these distributions.

2.2 | Introduction

Measuring the underlying distribution of BBH masses (primary mass m_1 and secondary mass m_2), mass ratios ($q = m_2/m_1 \leq 1$), spins, redshifts, and the correlations between these parameters from current GW observations can inform us of their formation mechanisms (Mapelli, 2021; Abbott et al., 2023b; Callister, 2024).

BBH formation processes are expected to leave signatures in the distribution of component masses and mass ratios of merging binaries. The mass ratios of binary

stars during their evolution depend on the initial fragmentation, accretion during star formation and evolution, and supernovae mechanisms (Bate & Bonnell, 1997; Belczynski et al., 2012; Moe & Di Stefano, 2017; Neijssel et al., 2019; Zevin et al., 2020; Broekgaarden et al., 2022; Briel et al., 2022b). Population-synthesis simulations indicate that field binaries that undergo a common-envelope stage can produce BBHs with mass-ratio distributions with support down to $q \sim 0.2$, though this depends on the treatment of the common-envelope process (Zevin et al., 2020; Olejak et al., 2021). Simulations of binaries that undergo stable mass transfer predict this channel cannot produce such unequal-mass binaries, but can produce preferentially more massive black holes than common envelope (van Son et al., 2022). However, other studies have found that mass ratio reversal, where the (initially less massive) second-formed black hole becomes the more massive object in the binary through stable mass transfer, can lead to lower mass ratios (Olejak et al., 2024). In contrast, BBHs that merge due to dynamical encounters may have more unequal masses and higher masses than isolated formation channels, especially if they are hierarchical mergers (Rodriguez et al., 2019; Sedda et al., 2023a, 2020; Bruel et al., 2024). The von Zeipel–Lidov–Kozai effect for field triples can produce BBHs with mass ratios down to 0.3 (Martinez et al., 2022). Similarly, active galactic nuclei can produce BBHs with more unequal masses than isolated binaries (Yang et al., 2019; Tagawa et al., 2020; Secunda et al., 2020). High-mass BBHs with unequal masses could therefore be a signature of formation via dynamical encounters or more exotic processes.

Investigations of the mass distributions of black holes have found structure in the primary masses, secondary masses, and the mass ratios of the population. Fishbach & Holz (2020) found that black holes do not randomly merge with each other regardless of mass; rather BBHs prefer equal masses. This is seen in the inferred mass-ratio distribution, which peaks at equal masses when considering a power-law model (Abbott et al., 2021b, 2023b; Abac et al., 2025a). Farah et al. (2024) modelled the secondary-mass distribution distinctly from the primary-mass distribution, finding that while the distributions of primary and secondary masses are consistent, secondary masses may follow a different distribution than primary masses. Different distributions would indicate that primary and secondary black holes undergo different evolutionary processes. Works considering more flexible models for the mass-ratio distribution find that the distribution peaks away from equal masses (Sadiq et al., 2024; Godfrey et al., 2023; Rinaldi et al., 2024; Callister & Farr, 2024; Rinaldi et al., 2025). Furthermore, there is evidence that the mass-ratio distribution may change with primary mass (Sadiq et al., 2024) and that different features in the primary-mass distribution have different mass-ratio distributions (Li et al., 2022; Godfrey et al., 2023; Li et al., 2024a; Galaudage & Lamberts, 2025; Roy et al., 2025). These results indicate that there is information in the mass distributions that could be used to uncover the uncertain astrophysics of BBH formation (e.g., Colloms et al., 2025a). Exploring parametric model variations for BBH masses allows us to test assumptions about the relationship between

component masses.

Callister (2024) highlights that the correlations between parameters, in addition to the parameters' marginal distributions, can be informative. When investigating correlations, it is necessary to check whether results are accurate descriptions of the population, or whether they are driven by the enforced model constraints. Adamcewicz & Thrane (2022) investigated the anti-correlation between mass ratio and effective inspiral spin. By fixing the shape of the marginal distributions, they avoid the case where the correlation results from an improved fit of the marginal distributions. The same can be true in the other direction: inferences of marginal distributions could be driven by assumed correlations (or the lack thereof) imposed by the chosen model. Potentially, individual source constraints that are not well accounted for by model correlations could influence the measured marginal distributions, especially if a correlation must be distorted to fit them. Updating population models so that outliers are well accounted for and correlations are appropriately incorporated helps us understand the entire population of BBHs.

Modelling BBH mass distributions requires a minimum black hole mass. This raises the question of the classification of merging objects with a mass that could correspond either to a black hole or a neutron star (Farah et al., 2022). It is unknown whether the maximum neutron star mass is the same as the minimum black hole mass. The maximum neutron star mass is uncertain, and depends on the neutron star equation of state (Rutherford et al., 2024; Koehn et al., 2025). Meanwhile, uncertainties in the minimum black hole mass come from the details of supernovae and core collapse (Fryer et al., 2012; Liu et al., 2021; Patton et al., 2022; Fryer et al., 2022). Following X-ray binary observations, it has been proposed that the maximum neutron star mass and the minimum black hole mass are not equal (Bailyn et al., 1998; Ozel et al., 2010; Farr et al., 2011; Siegel et al., 2023), and there is a mass gap with a dearth of black holes. Observations of objects with masses ~ 2.5–$5\,M_\odot$ indicate that at least some of the proposed gap is populated, though the rate of these objects is uncertain (Thompson et al., 2019; Wyrzykowski & Mandel, 2020; Abbott et al., 2020b; Jayasinghe et al., 2021; Abbott et al., 2023a; Lam et al., 2022; Abac et al., 2024; Fishbach et al., 2026). Including potential GW observations of low-mass black holes can allow us to measure the minimum black hole mass (Ray et al., 2026a).

Observations of (potential) low-m_2 BBHs include GW190814, which has $q = 0.112^{+0.008}_{-0.009}$ and $m_2 = 2.6^{+0.1}_{-0.1} M_\odot$ (Abbott et al., 2020b, 2024). Understanding the nature of GW190814's source as a BBH or a neutron star–black hole requires understanding if it is an outlier to the BBH population, and whether its outlier status is because of population misspecification. While Abbott et al. (2021b) and Essick et al. (2022) find GW190814 to be an outlier, Farah et al. (2022) consider models where it is not by including the full population of neutron stars and black holes, and considering a partially full mass gap. The outlier status of observations thus depends on the choice of population model. We seek to understand how to improve our models and incorporate observations that are currently challenging

to fit.

Current models for the BBHs' mass distribution assume that the minimum black hole mass is constant with the primary black hole mass. Given the preference for equal-mass mergers and the structure in the distribution of BBH masses, this may not be the case. We investigate if there is evidence for an evolving minimum secondary mass with primary mass: can black holes with higher masses merge with the lowest-mass black holes? We introduce models for the mass-ratio distribution of BBHs specifying a minimum secondary black hole mass that evolves with the primary mass (Section 2.3). Using these models on observations from the third Gravitational-Wave Transient Catalogue (GWTC-3.0; Abbott et al., 2023a) we infer the primary-mass and mass-ratio distributions (Section 2.4). Our conclusions for the inferred evolution of minimum secondary mass, and measured primary-mass and mass-ratio distributions are discussed in Section 2.5.

2.3 | Methods

We estimate a set of hyperparameters Λ that govern the shape of the mass, spin and redshift distributions of GW observations from GWTC-3.0 (Abbott et al., 2023a) using hierarchical Bayesian inference (Section 1.3.1, Mandel et al., 2019; Thrane & Talbot, 2019; Vitale et al., 2020). For each population distribution we choose a parametric model. We assume the distribution of the primary mass, spin and redshift distributions to be the default parametric models used in Abbott et al. (2023b), and introduce new models for the mass-ratio distribution as described in Section 2.3.1. Our inference framework is described in more detail in section Section 2.3.2 and the observations we use are described in Section 2.3.3.

2.3.1 | Variable minimum-mass models

To investigate the correlation between the minimum mass and the primary mass, we model the mass-ratio distribution of BBHs with a minimum secondary mass that varies with the primary mass. We include this as a modification to a power-law model of mass ratio (Fishbach & Holz, 2017; Kovetz et al., 2017), such that the mass-ratio distribution $p(q \mid m_1)$ with a power-law index β and a minimum secondary mass $m_2^{\min}(m_1)$ follows

$$\begin{aligned} p(q \mid m_1) &= (1+\beta)\left[1 - \frac{m_2^{\min}(m_1)}{m_1}\right]^{-(1+\beta)} q^{\beta} \\ &\propto q^{\beta} \qquad \left\{\frac{m_2^{\min}(m_1)}{m_1} \leq q \leq 1\right\}. \end{aligned} \tag{2.1}$$

This allows a varying lower bound on the secondary mass m_2 such that $m_2 \geq m_2^{\min}(m_1)$, the smallest possible secondary mass, and therefore the minimum mass ratio varies with the primary mass.

We consider two forms for how the minimum secondary mass varies with primary mass. Neither has significant astrophysical motivation, but both are simple progressions from a fixed minimum mass that enable us to explore whether the data support more complicated models. The two models complement each other in their analytic forms, and minimise additional model hyperparameters.

First, we consider a Power Law form that adds one additional hyperparameter:

$$m_2^{\rm min} = m_{\rm min} + (m_{\rm max} - m_{\rm min}) \left(\frac{m_1 - m_{\rm min}}{m_{\rm max} - m_{\rm min}} \right)^{\gamma}, \tag{2.2}$$

where $\gamma \geq 1$, $m_{\rm min}$ is the smallest primary black hole mass, and $m_{\rm max}$ is the maximum (stellar-mass) black hole mass. The Power Law model is a strictly increasing function that sets $m_{\rm min}$ as the minimum black hole mass for both the primary and secondary objects, and $m_2^{\rm min}(m_{\rm max}) = m_{\rm max}$. As γ increases, lower $m_2^{\rm min}$ is allowed for each m_1. Inspired by observations of binaries preferring more equal-mass components, this model prevents high-mass primaries from merging with low-mass secondaries.

Second, we consider a Parabola model for $m_2^{\rm min}$ that adds two additional hyperparameters:

$$m_2^{\rm min}(m_1) = m_{\rm min} + \xi(m_1 - m_{\rm min}) + \zeta(m_1 - m_{\rm min})^2. \tag{2.3}$$

This functional form automatically enforces that $m_2^{\rm min}(m_{\rm min}) = m_{\rm min}$. We apply additional constraints to ensure that the model remains physical. We require that $m_2^{\rm min}(m_1) < m_1$ for all $m_1 < m_{\rm max}$. Equivalently,

$$\begin{aligned} m_{\rm max} > m_{\rm min} &+ \xi(m_{\rm max} - m_{\rm min}) \\ &+ \zeta(m_{\rm max} - m_{\rm min})^2. \end{aligned} \tag{2.4}$$

This yields the bound

$$\zeta < \frac{1 - \xi}{m_{\rm max} - m_{\rm min}}. \tag{2.5}$$

We consider two variations of this model.

The Increasing Parabola model has $\xi > 0$ and $\zeta > 0$ such that $m_2^{\rm min}$ is always increasing with m_1.

The Relaxed Parabola model allows for a decreasing $m_2^{\rm min}$ with m_1 by allowing for negative ξ. This means that the minimum secondary mass can be less than $m_{\rm min}$. For this case, we require explicitly enforcing that $m_2^{\rm min}(m_1) > 0$ for all m_1. This necessitates that Eq. (2.3) has no real roots, and

$$\zeta > \frac{\xi^2}{4 m_{\rm min}}. \tag{2.6}$$

Combining Eq. (2.5) and Eq. (2.6) sets a limit on ξ such that

$$\frac{1 - \xi}{m_{\rm max} - m_{\rm min}} > \frac{\xi^2}{4 m_{\rm min}}, \tag{2.7}$$

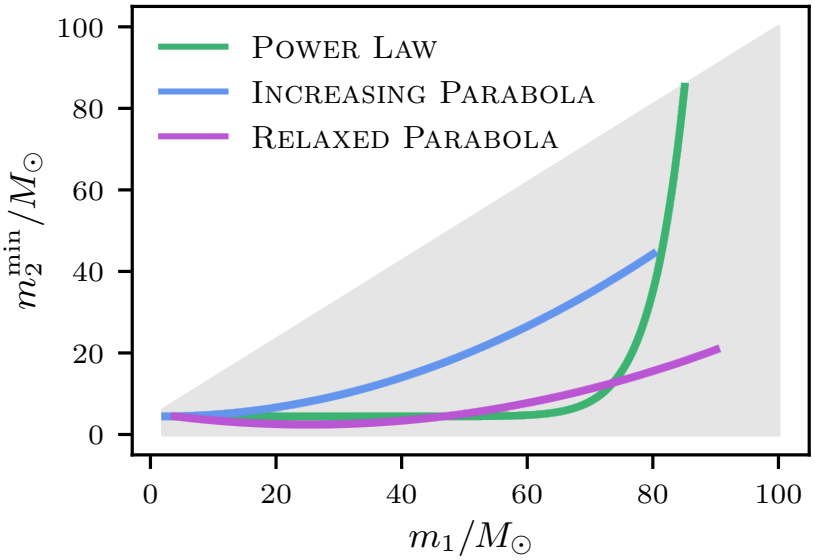


Figure 2.3.1: Example evolution of $m_2^{\min}$ with m_1 for the Power Law, Increasing Parabola, and Relaxed Parabola model variations we consider, for some fiducial hyperparameters within our prior range. The shaded region indicates where $m_1 > m_2^{\min} > 0$.

and hence,

$$-2\frac{\sqrt{m_{\min}}}{\sqrt{m_{\max}}-\sqrt{m_{\min}}} < \xi < 2\frac{\sqrt{m_{\min}}}{\sqrt{m_{\max}}+\sqrt{m_{\min}}}. \tag{2.8}$$

We consider these constraints in our priors on ξ and ζ, as detailed in Table 2.3.1.

Figure 2.3.1 shows graphical representations of each model for some fiducial hyperparameters.

2.3.2 | Statistical framework

To investigate the overall distribution of the BBH population, we jointly infer the distribution of BBH primary masses, mass ratios, spin magnitudes and tilts, and redshifts. We model the distribution of BBH primary masses, mass ratios, spin magnitudes $\chi_{1,2}$ and tilts $\theta_{1,2}$, and redshifts z as a separable probability distribution conditional on population hyperparameters Λ,

$$\begin{aligned} p(m_1, q, \chi_{1,2}, \cos\theta_{1,2}, z \mid \Lambda) = p(m_1 \mid \Lambda) \times p(q|m_1, \Lambda) \times \\ p(\chi_{1,2} \mid \Lambda) \times p(\cos\theta_{1,2} \mid \Lambda) \times p(z \mid \Lambda), \end{aligned} \tag{2.9}$$

where the mass-ratio distribution includes the Power Law, Increasing Parabola or Relaxed Parabola minimum secondary-mass model described above, and

Model	Hyperparameter priors	Minimum secondary mass
Power Law	$\gamma \sim \mathcal{U}(1, 35)$	Eq. (2.2)
Increasing Parabola	$\xi \sim \mathcal{U}(0, 1)$ $\zeta \sim \mathcal{U}\left(0, \frac{1-\xi}{m_{\rm max} - m_{\rm min}}\right)$	Eq. (2.3)
Relaxed Parabola	$\xi \sim \mathcal{U}\left(-2\frac{\sqrt{m_{\rm min}}}{\sqrt{m_{\rm max}} - \sqrt{m_{\rm min}}}, 2\frac{\sqrt{m_{\rm min}}}{\sqrt{m_{\rm max}} + \sqrt{m_{\rm min}}}\right)$ $\zeta \sim \mathcal{U}\left(\frac{\xi^2}{4m_{\rm min}}, \frac{1-\xi}{m_{\rm max} - m_{\rm min}}\right)$	Eq. (2.3)

Table 2.3.1: Prior ranges on γ, ξ and ζ hyperparameters for the three minimum secondary-mass models we consider, Power Law, Increasing Parabola, and Relaxed Parabola, and their corresponding forms for the minimum secondary mass.

the other observables have distributions that follow the default parametric models used in Abbott et al. (2023b). We use the Power Law+Peak model for $p(m_1)$ (Talbot & Thrane, 2018). This models the m_1 distribution as a two-component mixture of a power law of slope α and a Gaussian component. We choose similar priors on the default model parameters as Abbott et al. (2023b), with some refinements. We use primary-mass power-law index $\alpha \sim \mathcal{U}(-4, 12)$, mass-ratio power-law index $\beta \sim \mathcal{U}(-2, 7)$, primary minimum mass $m_{\rm min} \sim \mathcal{U}(2, 6)\,M_\odot$, and maximum mass $m_{\rm max} \sim \mathcal{U}(62, 100)\,M_\odot$. Both spin magnitudes are modelled with an identical beta distribution (Wysocki et al., 2019), and the spin tilts a mixture of isotropic and normally distributed spin tilts (Talbot & Thrane, 2017). The redshift distribution $p(z)$ is modelled with a power law (Fishbach et al., 2018) of index $\kappa \sim \mathcal{U}(-2, 8)$.

To infer hyperposteriors on Λ we use the population-inference package `gwpopulation_pipe` (Talbot et al., 2025; Talbot, 2021) with the nested sampler `dynesty` (Speagle, 2020). The inference imposes a cut on the likelihood variance of 1 during sampling (Talbot & Golomb, 2023).

2.3.3 | Gravitational-wave observations

We use the observations from GWTC-3.0 (Abbott et al., 2023a) selected with a false alarm rate threshold of less than 1 yr^{-1}. This is identical to the set of observations used in Abbott et al. (2023b). We consider only signals with BBH sources and the ambiguous GW190814 (Abbott et al., 2020b, 2024). As GW190814's secondary object may be above the expected maximum mass for a neutron star, we compare the results with and without GW190814, and investigate its potential as an outlier to the BBH population. GW200210_092254 (Abbott et al., 2023a), which has a similar secondary mass and mass ratio to GW190814, is not included in our analysis as it does not satisfy the false alarm rate threshold.

We use a combined injection set of semi-analytic injections at the sensitivity of the first and second observing runs, which assume a constant PSD across each observing run, and real injections at the third observing run sensitivity to account for selection effects (Abbott et al., 2023b,c). We consider the real injections with a false alarm rate of less than 1 yr^{-1} detectable, and the semi-analytic injections we cut in network signal-to-noise ratio of 10 (Abbott et al., 2018). These sensitivity estimates are the same as those used in GWTC-3.0 LVK population inference by Abbott et al. (2023b), where there did not exist real injections for the first and second observing runs (Abbott et al., 2023c).

2.4 | Results

2.4.1 | Evolution of minimum secondary mass

We show the inferred evolution of m_2^{min} with m_1 in Figure 2.4.1, with and without GW190814, for our three model variations. Uncertainties grow at larger m_1 where we have fewer observations.

Under the Power Law minimum secondary-mass model, the evolution of m_2^{min} is preferentially flat until high primary masses, where the uncertainty on m_2^{min} grows. At high m_1, m_2^{min} is forced to increase to m_{max} because of the functional form of the model. The hyperposteriors in Figure 2.4.2 show that high γ is preferred, meaning that the evolution of m_2^{min} is flatter with m_1, before sharply increasing as m_1 approaches m_{max}. The results are restricted by the prior on γ, as we allow a maximum γ of 35, while a flat evolution in m_2^{min} would require $\gamma \to \infty$. With our prior choice, we find that $\gamma > 4.4$ with GW190814 and $\gamma > 2.8$ without GW190814 at 90% credibility. The preference for a flatter evolution of m_2^{min} means that there is little support for a higher minimum secondary mass at higher primary masses. While this behaviour is consistent whether or not GW190814 is included, the overall minimum mass m_{min} and the primary-mass power-law index α are different when including GW190814. Without GW190814, the minimum mass is higher, with $m_{\mathrm{min}} = 5.8^{+0.2}_{-0.6}\,M_\odot$, while with GW190814, $m_{\mathrm{min}} = 2.2^{+0.4}_{-0.2}\,M_\odot$ at 90% credibility to accommodate GW190814's secondary mass.

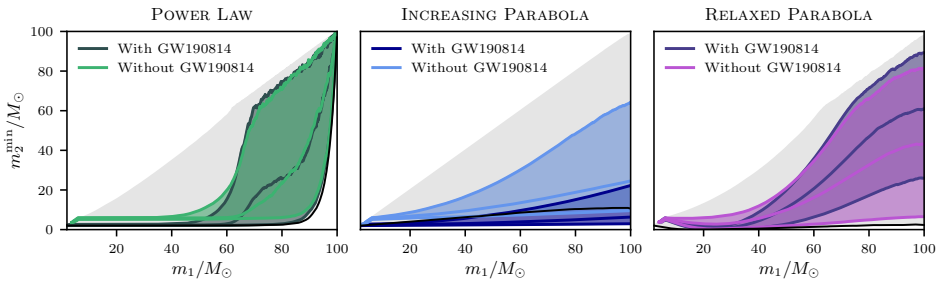


Figure 2.4.1: Inferred evolution of the minimum secondary mass with primary mass for the Power Law (left), Increasing Parabola (middle), and Relaxed Parabola (right) models. Each panel shows the resulting median and 90% credible interval on the minimum secondary mass, including and excluding GW190814. The grey shows the 99% credible interval for the prior for each model, with the lower edge in black. The interval is evaluated over $m_1 \in [2, 100]\,M_\odot$, where the plotted range extends to $m_1 > m_{\rm max}$ for some individual hyperposterior samples. With GW190814, the entire distribution of $m_2^{\rm min}$ is lower, apart from the Relaxed Parabola model, which decreases to allow for the lower secondary mass at $m_1 \sim 20\,M_\odot$. Due to the driving of the prior, $m_2^{\rm min}$ then increases at higher m_1 for this model. Otherwise, there is no strong evidence for an increasing $m_2^{\rm min}$ with m_1 when accounting for model limitations.

For the Increasing Parabola model, the inferred $m_2^{\rm min}$ behaviour is similar to the Power Law result, where $m_2^{\rm min}$ is consistent with being flat with m_1 for both sets of results. Hence, there is no evidence that the minimum secondary mass changes with primary mass when considering this model. This can be seen in the hyperposteriors on ξ and ζ, as plotted in Figure 2.4.3, with ξ and ζ both peaking at 0. Similarly to the Power Law model, when including GW190814, $m_{\rm min}$ is forced to be lower to include its secondary mass, and α is smaller. At large primary masses, the uncertainty on the minimum secondary mass is high, though this is reduced when including GW190814. This may result from the restriction that $m_2^{\rm min}$ cannot increase above $\sim 2.6\,M_\odot$ at $m_1 \sim 20\,M_\odot$, as can be seen by the more tightly constrained ξ and ζ when including GW190814.

In the Relaxed Parabola model, $m_2^{\rm min}$ remains largely unconstrained at high m_1, while $m_2^{\rm min}(m_{\rm max}) = m_{\rm min}$ is not ruled out when not including GW190814. With GW190814, $m_2^{\rm min}$ decreases below $m_{\rm min}$ until $m_1 \sim 20\,M_\odot$, then increases at high m_1, preferring $m_2^{\rm min} > m_{\rm min}$ at higher primary masses. The dip in $m_2^{\rm min}$ at $m_1 \sim 20\,M_\odot$ accommodates GW190814's secondary mass, while the increase of $m_2^{\rm min}$ at high m_1 is driven by the prior constraints when including negative ξ. For both sets of observations, $m_{\rm min}$ does not significantly change as $m_2^{\rm min}$ is instead allowed to decrease to allow for the secondary mass of GW190814. This is because $m_{\rm min}$ is no longer describing the lowest secondary mass of the

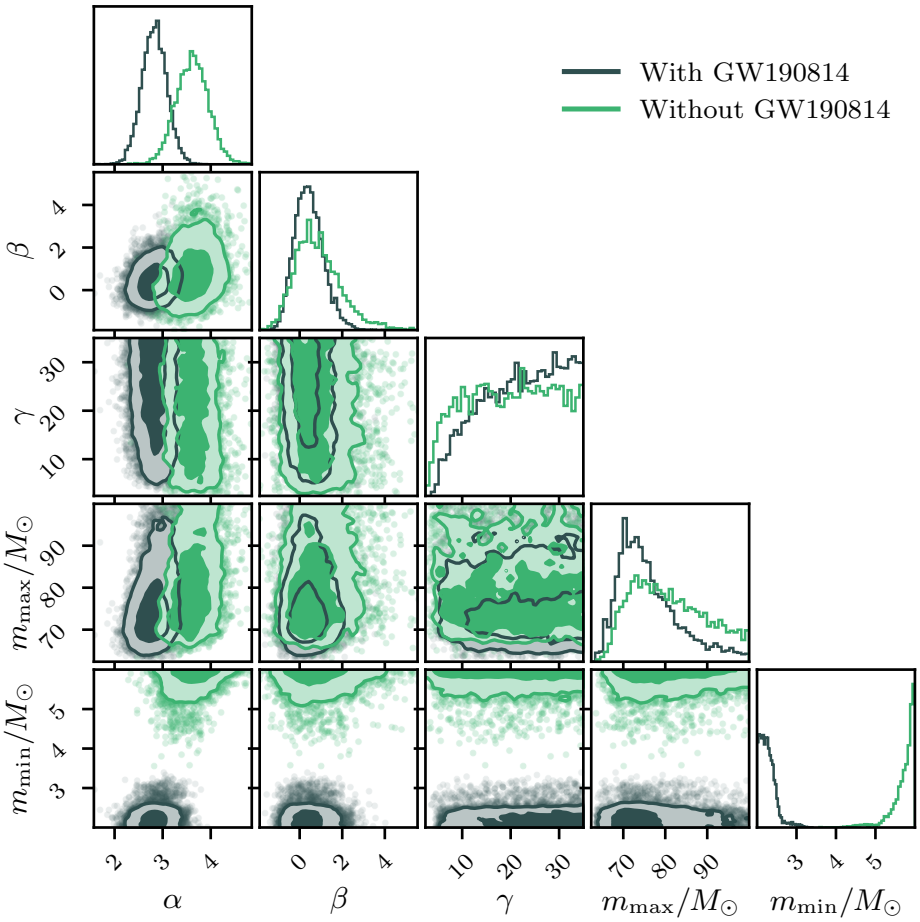


Figure 2.4.2: Hyperposteriors on the key mass and mass-ratio hyperparameters with Power Law model, with and without GW190814. The $m_2^{\rm min}$ hyperparameter γ prefers high values, supporting a non-increasing or shallowly increasing $m_2^{\rm min}$ with m_1. The overall minimum mass $m_{\rm min}$ shows strong disagreement between results including or excluding GW190814, while the primary-mass power-law index α also differs between the two sets of observations, though is consistent to within 90% credibility.

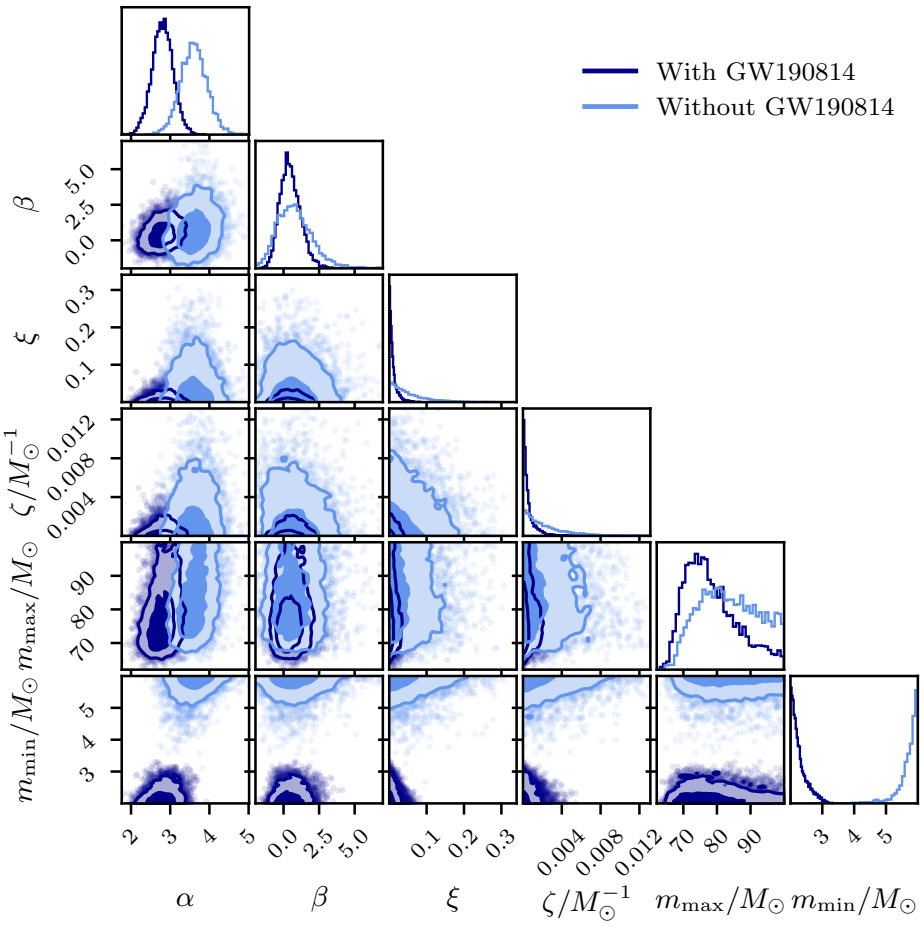


Figure 2.4.3: Hyperposteriors on the key mass and mass-ratio hyperparameters with the Increasing Parabola model, with and without GW190814. The $m_2^{\min}$ hyperparameters, ξ and ζ are railing at 0, showing a preference for a non-increasing $m_2^{\min}$ with m_1. The overall minimum mass $m_{\min}$ shows strong disagreement between results including or excluding GW190814, while the primary-mass power-law index α also differs between the two sets of observations, though is consistent to within 90% credibility.

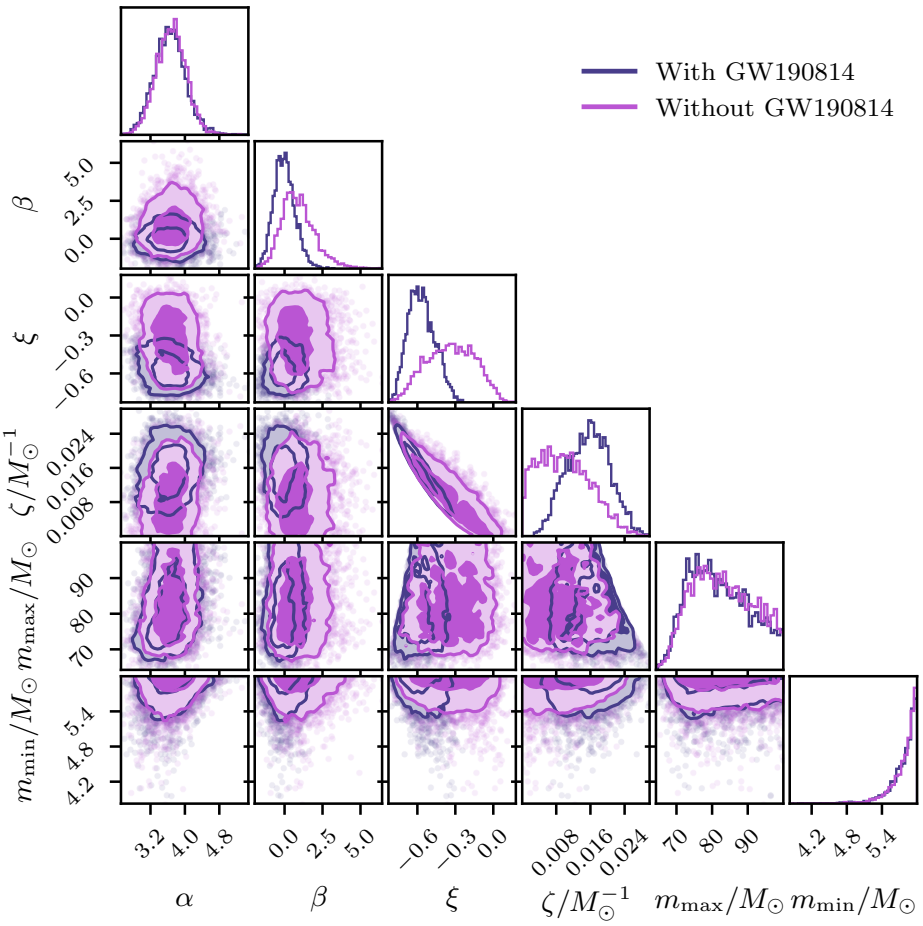


Figure 2.4.4: Hyperposteriors on the key mass and mass-ratio hyperparameters with the Relaxed Parabola model, with and without GW190814. The $m_2^{\min}$ hyperparameters, ξ and ζ are consistent with zero without GW190814, but negative ξ is preferred when including GW190814. The other hyperparameters are in agreement between the two choices of observations.

observations, but the lowest primary mass. As shown in Figure 2.4.4, while the hyperposteriors on ξ and ζ are different with the different sets of observations, the other hyperposteriors show less variation when including GW190814 than results with the strictly increasing $m_2^{\min}$ models. This hints that GW190814 is no longer an outlier under the model assumptions that allow $m_2^{\min}$ to decrease with m_1.

We compare each of our analyses to the results with the default Power Law + Peak model from Abbott et al. (2023b), with $m_2^{\min}$ fixed to be equal to $m_{\min}$. We find that our models are generally not preferred to the default model without a variable $m_2^{\min}$, as shown in Table 2.4.1. This reinforces the result that there is no

Model	$\log_{10} \mathcal{B}$	
	With GW190814	Without GW190814
Power Law	−0.54	−0.98
Increasing Parabola	−1.86	−2.49
Relaxed Parabola	4.44	−1.37

Table 2.4.1: Log Bayes factors between the Power Law, Increasing Parabola, and Relaxed Parabola minimum secondary-mass models and the default model without evolution of the minimum secondary mass from Abbott et al. (2023b). All models allowing for an evolving minimum secondary mass with primary mass are disfavoured compared to the default model, with the exception of the Relaxed Parabola model when including GW190814.

evidence that $m_2^{\min}$ (strictly) increases with m_1. The exception is the Relaxed Parabola model, which is preferred with a log Bayes factor of $\log_{10} \mathcal{B} = 4.44$ when including GW190814. Therefore, if GW190814's source is a BBH, this indicates that there is structure in the mass-ratio and component-mass distributions that should be accounted for by population models.

2.4.2 | Mass-ratio and primary-mass distributions

To understand the impact of a variable $m_2^{\min}$ on the mass-ratio distribution, we show $p(m_1)$ and $p(q)$ for our three model variations in Figure 2.4.5. For the Power Law and Increasing Parabola models, neither $p(m_1)$ nor $p(q)$ agree between the results with and without GW190814. Both of these models find a different $m_{\min}$ for the different sets of observations, resulting in a different $p(m_1)$ distribution at lower masses. Without GW190814, these models also are not required to go to as low mass ratios. With these strictly increasing $m_2^{\min}$ models, inclusion of GW190814 changes the overall primary-mass distribution despite its m_1 being unexceptional. For the Relaxed Parabola model, there is more agreement between both primary-mass and mass-ratio distributions. While $p(q)$ does extend to lower values with GW190814, this is expected due to its low mass ratio. The added flexibility in the mass-ratio model thus allows for GW190814-like systems to fit with the bulk BBHs population, and they do not need to be treated as population outliers.

Figure 2.4.6 shows $p(q)$ at slices of m_1 for the three model variations. This shows the strictly increasing $m_2^{\min}$ models' inconsistency between the mass-ratio distributions for the results with and without GW190814, especially for lower primary masses. At $m_1 = 10\,M_\odot$ for the Power Law and Increasing Parabola models, the mass-ratio distribution supports much lower mass ratios when in-

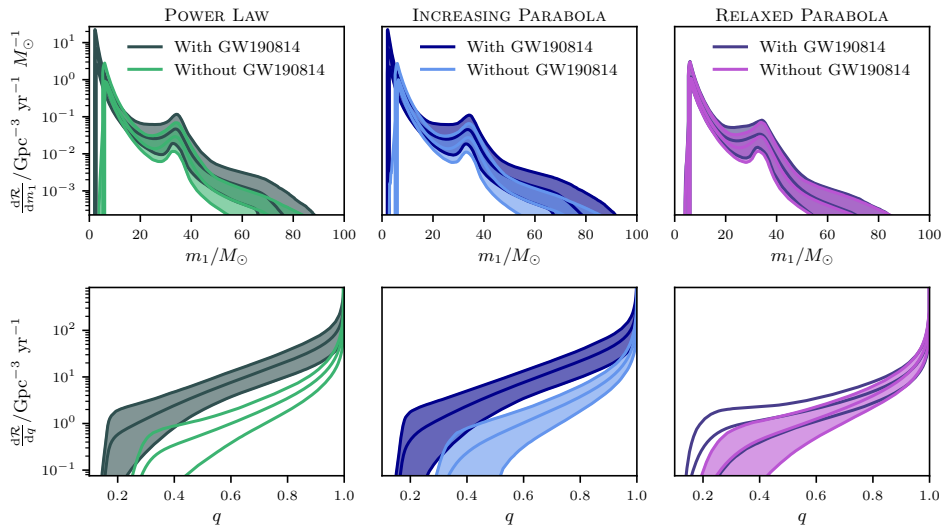


Figure 2.4.5: The inferred primary-mass (top) and mass-ratio (bottom) distributions for the Power Law (left), Increasing Parabola (middle), and Relaxed Parabola (right) models, each with and without GW190814. The solid line shows the posterior population distribution and the shaded region is the 90% credible interval for each distribution. For the Increasing Parabola and Power Law models, the distributions are different when including GW190814, while for the Relaxed Parabola model the intervals are consistent.

cluding GW190814 because of the overall lower $m_{\rm min}$. As $m_{\rm min} \sim 5\,M_\odot$ without GW190814, the lowest possible mass ratio is $q \sim 0.5$ at $m_1 = 10\,M_\odot$, while when including GW190814, $m_{\rm min} \sim 2\,M_\odot$, making the lowest possible mass ratio $q \sim 0.2$. Extreme mass ratios at $m_1 = 10\,M_\odot$ are not well supported by the prior, and GW190814 is highly informative for the inference. At all mass slices, $p(q)$ is flatter for inference with GW190814, driven by the available model space. The difference at $m_1 = 10\,M_\odot$ is reconciled for the Relaxed Parabola model, as $m_{\rm min}$ is in agreement whether or not GW190814 is included.

Compared to our analyses, Callister & Farr (2024) and Sadiq et al. (2024) find that $p(q)$ has more support for unequal-mass mergers when considering data-driven models. These models do not consider a parametrised $m_{\rm min}$ and therefore have more flexibility to extend support to lower q even at low m_1. Sadiq et al. (2024), when excluding GW190814, find that $p(q \mid m_1 = 10\,M_\odot)$ has support down to $q \sim 0.3$, while the distribution peaks away from equal masses at higher m_1. Using parametric models, Li et al. (2022) find that higher-mass BBHs are more likely to be equal mass than lower-mass binaries when excluding GW190814. Meanwhile, Galaudage & Lamberts (2025), Li et al. (2024a), and Roy et al. (2025) find that BBHs around $m_1 \sim 35\,M_\odot$ are more likely to prefer equal masses. As shown in Figure 2.4.6, our models induce that the mass-ratio distribution varies

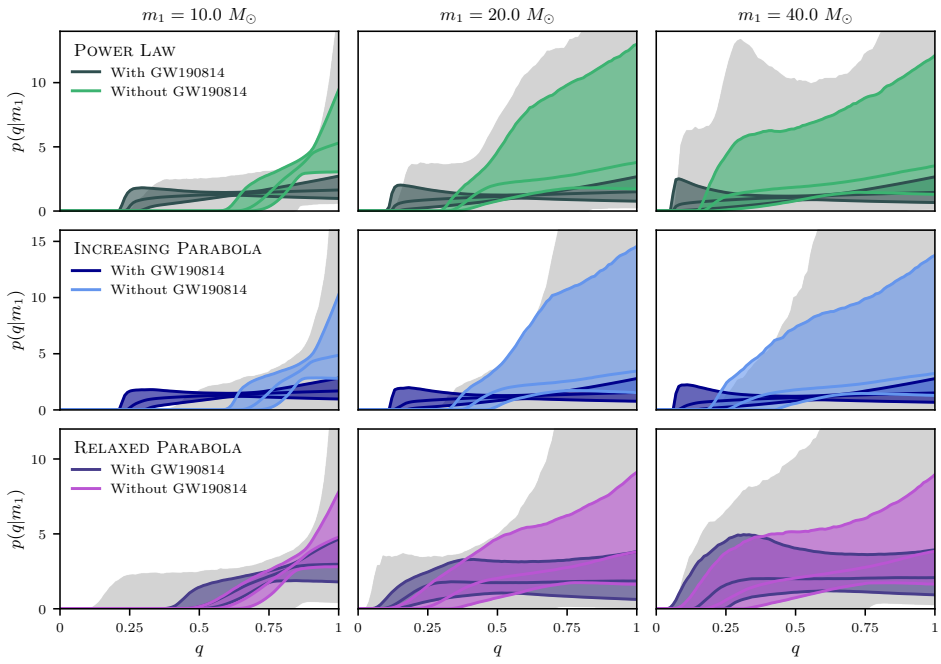


Figure 2.4.6: Slices of the median and 90% interval for the mass-ratio distribution at primary masses of 10 $M_\odot$ (left), 20 $M_\odot$ (middle), and 40 $M_\odot$ (right) for the Power Law (top), Increasing Parabola (middle), and Relaxed Parabola (bottom) models, each with and without GW190814. The 99% credible interval of the prior at each slice is shown in grey. For the Increasing Parabola and Power Law models, the distributions are different when including GW190814, while for the Relaxed Parabola model the intervals are more consistent.

with primary mass when excluding GW190814: at the lowest masses, the distribution looks steep because the range of potential mass ratios is tightly constrained by the minimum mass, while at larger masses the preferred distribution is comparatively flatter, though with broad uncertainties. The variation between results across the literature illustrates how different model constraints and prior assumptions can influence the interpretation of results, and highlights the importance of exploring alternative choices.

2.5 | Conclusions

We investigated the correlation between the mass ratio and primary mass of merging BBHs to understand whether the biggest stellar-mass black holes can merge with the smallest. We introduced population models that included a minimum secondary mass $m_2^{\rm min}(m_1)$ that varies with primary mass. We considered three model variations for $m_2^{\rm min}$, which we used with a power-law model for the mass-ratio

distribution and a Power Law + Peak distribution for primary mass (Talbot & Thrane, 2018; Abbott et al., 2023b). Using these models we inferred how the minimum secondary mass evolves for GWTC-3.0 observations (Abbott et al., 2023a).

Our results show that there is currently no evidence that $m_2^{\rm min}$ increases with m_1 when excluding GW190814. We find log Bayes factors of $\log_{10} \mathcal{B} = -2.49$ to -0.98 for our models compared to a constant minimum-mass model. These results imply that if GW190814 is not a BBH, then we do not yet have observational evidence to justify changes to fiducial LVK population models that assume that there is a constant minimum black hole mass across the entire primary-mass range. We repeated the same analysis including GW190814. If we enforce that $m_2^{\rm min}$ is strictly increasing with m_1, we find that the overall minimum black hole mass has to be lower to account for GW190814's secondary mass, while the primary-mass and mass-ratio distributions change with inclusion of GW190814. This suggests that GW190814 should be considered as an outlier for these population models. With a model allowing $m_2^{\rm min}$ to decrease with m_1, $m_2^{\rm min}$ dips to allow GW190814's secondary mass, and the minimum primary mass is in agreement with the inference excluding GW190814. The added flexibility in $m_2^{\rm min}$ means that under this model the differences in $p(m_1)$ and $p(q)$ are reconciled when doing a leave-one-out analysis, no longer causing GW190814 to be an outlier. This model is favoured when including GW190814 in the population compared to a constant minimum-mass model, with $\log_{10} \mathcal{B} = 4.44$. Hence, there is support for an evolving $m_2^{\rm min}$ when including GW190814's source in the BBH population.

The preference for a near-constant $m_2^{\rm min}$ with m_1 means that we do not rule out the possibility of high-primary mass, relatively extreme mass-ratio BBHs. Such sources could be indicators of BBH formation through dynamical interactions, in triples, or BBHs formed in AGN disks. Stronger constraints on the rate of high-primary mass, relatively extreme mass-ratio BBHs would motivate the development of models to account for these observations, thus help us understand the uncertainties in these formation channels. However, our model uncertainties are greatest at large primary masses due to a combination of a lack of observations and model limitations.

While our models provide one way to understand correlations between primary mass and mass ratio, improvements to our model could be made to ensure robustness and avoid model misspecification. Our model that allows for decreasing $m_2^{\rm min}$ enforces that the minimum secondary mass is always positive, but currently does not place other bounds on $m_2^{\rm min}$, e.g., it could go to subsolar masses. Future work could update this assumption by imposing a non-zero lower bound on $m_2^{\rm min}$, representing a minimum black hole mass, which may help us understand the classification of GW190814. Additional model flexibility also could be added, for example by relaxing the assumption that $m_2^{\rm min}(m_{\rm min}) = m_{\rm min}$; Farah et al. (2024) find that the inferred minimum primary and minimum secondary masses of the BBH population are in agreement within current uncertainties, but did

not include GW190814 in their analysis. Our parametric forms for the minimum secondary mass are also not motivated by particular astrophysics or trend in the data. A data-driven description of the primary-mass and mass-ratio distributions may give more insight into the underlying correlations, unrestricted by the choice of parametric form. Furthermore, we specify a power-law distribution in mass ratio, restricting the mass-ratio distribution to be strictly monotonic. However, models with more relaxed assumptions about the shape of the distribution show support for the mass-ratio distribution to peak away from 1 (Sadiq et al., 2024; Godfrey et al., 2023; Rinaldi et al., 2024; Callister & Farr, 2024; Rinaldi et al., 2025). Other works find that the mass-ratio distribution changes with primary mass, either continuously or for different subpopulations, motivating the exploration of a variable shape of the distribution in addition to a variable lower bound (Li et al., 2022; Baibhav et al., 2023; Galaudage & Lamberts, 2025; Roy et al., 2025). Considering a mass-ratio distribution that more accurately reflects the data would help avoid biases driven by an inaccurate models.

Other observable parameters such as the spin or eccentricity could provide more information about the astrophysical nature of such BBHs. Works looking at the correlations between mass and spin may identify subpopulations of sources from similar evolutionary pathways (e.g., Safarzadeh et al., 2020; Heinzel et al., 2024; Ray et al., 2025; Antonini et al., 2025). If the mass-ratio distribution is jointly correlated with component mass and spin distributions, neglecting to model these correlations could lead to biased inference (Alvarez-Lopez et al., 2025). Jointly investigating mass, mass-ratio, and spin distributions could help us understand how these parameters are correlated, and therefore give more insight into the astrophysical evolution of the population. However, adding additional parameters to our models and investigating the joint correlations between mass, mass ratio, and spin, requires having sufficient data to constrain the model parameters. The increased number of BBH observations coming with the fourth LVK observing run will allow us to better measure any correlations between the BBH observables, and explore the astrophysics encoded in these correlations. In Chapter 4, we show how our understanding of these correlations progresses with observations with the larger number of observations in GWTC-4.0.

3 Exploring the evolution of gravitational-wave emitters with efficient emulation: Constraining the origins of binary black holes using normalising flows

3.1 | Abstract

Binary population synthesis simulations allow detailed modelling of gravitational-wave sources from a variety of formation channels. These population models can be compared to the observed catalogue of merging binaries to infer the uncertain astrophysical input parameters describing binary formation and evolution, as well as the relative rates between various formation pathways. However, it is computationally infeasible to run population synthesis simulations for all variations of uncertain input physics. We demonstrate the use of normalising flows to emulate population synthesis results and interpolate between astrophysical input parameters. Using observations of BBHs from GWTC-3.0, we use our trained normalising flows to infer branching ratios between multiple formation channels, and simultaneously infer common-envelope efficiency and natal spins across a continuous parameter range. Given our set of formation channel models, we infer the natal spin to be $0.04^{+0.04}_{-0.01}$, and the common-envelope efficiency to be > 3.7 at 90% credibility, with the majority of underlying mergers coming from the common-envelope channel. Our framework allows us to measure population synthesis inputs where we do not have simulations, and better constrain the astrophysics underlying current gravitational-wave populations.

3.2 | Introduction

The first three LIGO–Virgo–KAGRA observing runs yielded 90 probable GW candidates (Abbott et al., 2024, 2023a). Analysis of the population properties of these events gives insight into how merging BBHs can form and how their progenitors evolve (Mapelli, 2021; Abbott et al., 2023b; Callister, 2024). Measuring the properties of the sub-population of BBHs formed through the evolution of field binaries can inform us about binary stellar evolution (Wysocki et al., 2018; Giacobbo et al., 2018; Bavera et al., 2021; Delfavero et al., 2023; Godfrey et al., 2023; Golomb et al., 2024; Galaudage & Lamberts, 2025); whereas the sub-population formed from dynamical interactions can inform us about the properties of their dense cluster environments (Romero-Shaw et al., 2021; Fishbach & Fragione, 2023; Ng et al., 2023; Kritos et al., 2025; Ginat et al., 2026). The specifics of these formation processes are still poorly understood, such as the star formation rates at given metallicities, the details of binary stellar evolution and interactions, and the relative rates of BBH mergers from different formation channels (Belczynski et al., 2022). The uncertainty in model predictions is currently large, and differs depending on the initial input physics (Mandel & Broekgaarden, 2022; Belczynski et al., 2022; Mandel & Farmer, 2022). Population analysis of GW observations offers a unique way to constrain the uncertainty from model predictions.

Modelling of compact binary merger populations has followed several paths, comprised of parametric models, data driven models, and models based on population synthesis (see Section 1.3.2). Fitting the population with parametric models risks forcing population features not present in the data, or missing features which are not allowed by the model. On the other hand, data-driven models can identify features in the data that are not captured by parametric models, as long as statistical fluctuations are accounted for. Both of these approaches do not directly use models of BBH progenitor evolution, making it difficult to link observed population features to astrophysical theory.

We aim to gain more insight into BBH astrophysics by using population synthesis simulations to predict the observed population. These simulations model the evolution of BBH progenitors, predicting the resulting observable parameters of sources for different input astrophysics and formation scenarios, which we will refer to as hyperparameters. Works comparing population synthesis to GW observations have constrained sets of these hyperparameters (Stevenson et al., 2017b; Zevin et al., 2017; Neijssel et al., 2019; Bouffanais et al., 2021; Zevin et al., 2021; Bavera et al., 2021; Mastrogiovanni et al., 2022; Stevenson & Clarke, 2022; Iwaya et al., 2023; Chen et al., 2025; Gayathri et al., 2025). However, simulating just a single population requires many simulated binaries, and investigating varying input physics requires simulating many populations. Current population synthesis codes take $O(10^2)$ CPU hours for a single population of 10^6 isolated binaries (Andrews et al., 2025). Therefore, a detailed exploration of uncertain physical processes that affect predicted BBH populations makes using population

synthesis much more computationally costly than using parametric or data-driven models.

Exploring more population hyperparameters in an analysis gives us more insight into the degeneracies and correlations between different model populations (Barrett et al., 2018; Stevenson & Clarke, 2022). Furthermore, there is evidence for multiple formation channels present in the GW data (Zevin et al., 2021; Cheng et al., 2023; Li et al., 2024b; Gayathri et al., 2025); each will have its own contribution to the total population. Joint inference of these parameters gives better constraints on the inferred population, and is necessary to avoid biased results coming from a neglected channel or variable that truly contributes to the astrophysical observations (Zevin et al., 2021; Cheng et al., 2023). To understand the correlations between different input astrophysical hyperparameters, we require a comprehensive set of simulations of BBH populations considering a broad range of simulation inputs.

Machine learning is a potential avenue for reducing the difficulties associated with these expensive simulations. Machine learning has been used in various aspects of GW data analysis in order to increase efficiency of data-quality improvement, waveform modelling, searches, and parameter estimation (Cuoco et al., 2024). In the context of BBH populations, we can employ machine learning to reduce the cost of performing population synthesis simulations: instead of performing numerous simulations across the input hyperparameter space, we may interpolate between simulations. The interpolation may either be done by reconstructing the probability distribution across different input hyperparameters given a set of observations (Delfavero et al., 2023) or by emulating how the astrophysical population varies across hyperparameter space. Different approaches have previously been explored for population emulation. Early studies used Gaussian process regression (GPR; Barrett et al., 2016; Taylor & Gerosa, 2018). While this emulation was effective on a small number of dimensions, GPR was shown to be limited at handling higher dimensional data with sufficient computational efficiency, the principal component analysis needed to reduce the input dimensionality to GPR giving an inevitable loss of information in the training data (Wong et al., 2020; Cheung et al., 2022). Deep neural networks (DNNs) have been used to emulate populations, training the DNNs on kernel density estimation (KDE) reconstructions of the population (Mould et al., 2022) or binned distributions (Riley & Mandel, 2023). The reliance on the evaluation of the population distributions once by the KDEs and then again by the DNNs adds uncertainty to the result, while using bins requires a trade-off between having sufficient training data per bin and having sufficient bins to resolve all the features in the population. Accurately emulating population synthesis models across many hyperparameters is therefore a challenging problem.

Normalising flows are a type of generative machine learning algorithm which can model complicated probability densities; they are therefore well suited to this problem. Normalising flows operate by learning a series of transformations from

a simple distribution to a target distribution (Papamakarios et al., 2021; Kobyzev et al., 2021). Applications of normalising flows to population inference problems have given fast and accurate results for population analyses of stellar mass BBHs (Wong et al., 2020, 2021b,a), supermassive BBHs (Laal et al., 2025), and studies of other stellar populations (Hon et al., 2024). Wong et al. (2020) demonstrated that normalising flows are capable of recovering astrophysical hyperparameters of simulated observations across a higher dimensional set of binary parameters than GPR. The application of normalising flows was extended by Wong et al. (2021a), inferring two astrophysical hyperparameters and the branching ratio between two formation channels with GWTC-2.0 observations (up to the first part of the third LIGO–Virgo–KAGRA observing run; Abbott et al., 2021a). These studies show the effectiveness of normalising flows when applied to problems with increased dimensionality, using the latest sets of GW observations. Using normalising flows as emulators of population synthesis models allows for astrophysically informed inference of BBH populations, while reducing the computational expense of simulating many populations.

We use normalising flows to emulate and interpolate population synthesis models conditional on the simulation input parameters, using this as our population model for hierarchical inference of two astrophysical hyperparameters and branching fractions between five different formation channels. We extend the Astrophysical Model Analysis and Evidence Evaluation (AMAZE) framework used in Zevin et al. (2021) and Cheng et al. (2023), to incorporate the normalising flows and move to a continuous hyperparameter space of our population models. AMAZE was originally developed to allow the comparison of observations to discrete population synthesis models using a hierarchical Bayesian inference framework. The interpolation allows for the inference of the astrophysical hyperparameters between the originally simulated models, giving us insight into the uncertainties and correlations of these hyperparameters. We illustrate our approach using GWTC-3.0 observations (Abbott et al., 2023a). In showing that our inference can reliably infer a continuous range of astrophysical hyperparameters, we open up possibilities of increasing the dimensionality of astrophysical hyperparameters studied in BBH population inference. Applying this method in cases of higher dimensional problems, together with our growing set of GW observations would give key insights into in the formation of BBHs and their formation environments.

In Section 3.3 we underline the methods for the analysis including the population models, inference framework, normalising flows, and the GW observations used. We show results with our trained normalising flows in Section 3.4; these include tests of the normalising flows' emulation, inference comparison to a non-interpolating method, continuous inference with interpolation, and the resulting fit to the observed data. In Section 3.5 we discuss the results and limitations of the current analysis, and avenues for future work.

3.3 | Methods

We infer astrophysical hyperparameters of the population of BBHs observed with GWs. To do this, we use hierarchical Bayesian inference to measure the population hyperparameters given our observed GW population. Our statistical framework is described in Section 3.3.2. This requires model BBH populations from population synthesis, on which we train normalising flows, and then compare to the observations, which are each described in Section 3.3.1, Section 3.3.3, and Section 3.3.4, respectively. We update the AMAχE framework developed by Zevin et al. (2021) to incorporate the addition of normalising flows. Using normalising flows to interpolate between the population models, we evaluate the model likelihood across population hyperparameter space *continuously*, therefore performing continuous inference of our astrophysical hyperparameters of interest.

3.3.1 | Population models

We consider five formation channels: three isolated channels (common envelope, chemically homogeneous evolution, stable mass transfer); and two dynamical cluster environments (globular clusters and nuclear star clusters). These are the same population models as used in Zevin et al. (2021), which contains further details of the simulations.

The common-envelope (CE) channel describes stellar binaries that, after one compact object has already formed, undergo unstable mass transfer leading to an envelope of the donor star's material surrounding the compact object and the donor's core. Drag forces from the envelope cause the system to harden, and the envelope is ejected with some efficiency, $\alpha_{\rm CE}$ (van den Heuvel, 1976; Paczynski, 1976; Belczynski et al., 2001; Dominik et al., 2012). This efficiency is the fraction of orbital energy used to eject the envelope, where a value of $\alpha_{\rm CE}$ greater than 1 requires energy other than orbital energy, such as radiative energy from the stellar objects, to be used in the envelope's ejection (Ivanova et al., 2013). A lower CE efficiency creates post-CE systems that are tighter, although at a risk of some systems merging in the envelope, while a higher efficiency creates systems that may still be too wide to merge within a Hubble time. For the stable mass transfer (SMT) channel, binaries undergo only mass transfer which is stable during their evolution (van den Heuvel et al., 2017; Neijssel et al., 2019; Gallegos-Garcia et al., 2021; van Son et al., 2022; Briel et al., 2022b). The CE and SMT channels were modelled with an early version of the `POSYDON` framework (Fragos et al., 2022; Andrews et al., 2025): `COSMIC` population synthesis (Breivik et al., 2020) up until the end of the second mass-transfer episode, and the detailed stellar evolution code `MESA` (Paxton et al., 2011, 2013, 2015, 2018, 2019) from the formation of a black hole–helium star binary until the formation of the second black hole.

The chemically homogeneous evolution (CHE) case accounts for BBHs that form from stars which are tidally synchronized during core hydrogen burning,

causing the stars to rapidly spin and their interiors to become chemically homogeneous, such that they do not expand during the red-giant phase (de Mink et al., 2008; Mandel & de Mink, 2016; du Buisson et al., 2020). The CHE channel models were adapted from du Buisson et al. (2020), who used a `MESA` grid of binary simulations.

In the globular cluster (GC) and nuclear star cluster (NSC) channels, binaries undergo dynamical scatterings and captures due to their dense formation environments (Heggie, 1975; Fitchett, 1983; Antonini & Rasio, 2016). NSCs can generally retain more black hole merger remnants, and have a significant fraction of hierarchical mergers, than GCs due to their higher masses and escape velocities. The GC models used the cluster Monte Carlo code `CMC` from Rodriguez et al. (2019), using the Binary Stellar Evolution (`BSE`) package of Hurley et al. (2000, 2002), with some updated evolutionary prescriptions from `COSMIC` (Breivik et al., 2020). The NSC channel models use the semi-analytical model from Antonini et al. (2019) to determine the merger rate of BBHs. This model assumes a regulation of the energy generated by BBHs by the process of two-body relaxation in the system.

The population models include associated weights for each simulated BBH. These weights account for the probability of binary formation at particular metallicities and the increasing co-moving volume probed at higher redshift. The populations are simulated over a range of different metallicity values. For all population models other than GCs (Zevin et al., 2021, appendix A.1.3), the distribution of metallicities at a given redshift follows a truncated log normal distribution around the empirical mean metallicity from Madau & Fragos (2017), with the birth redshifts of each binary following the star formation rate history from Madau & Fragos (2017). The associated metallicity weight of each sample is the density of the metallicity distribution given the birth redshift of each simulated binary. The total weight is the factor of the co-moving volume probed at the redshift of the merger multiplied by the metallicity weight (Zevin et al., 2021).

Each simulation predicts the set binary observables $\vec{\theta}$: the chirp mass $\mathcal{M}$, mass ratio $q = m_2/m_1$, effective inspiral spin $\chi_{\rm eff}$, and redshift z at which the BBH merges.

Two astrophysical hyperparameters across all formation channels, $\vec{\lambda}$, are varied: the birth spin of black holes born in isolation (i.e., born without prior spin-up of the stellar core due to tidal interactions) $\chi_{\rm b}$, and the CE efficiency $\alpha_{\rm CE}$. The value of $\alpha_{\rm CE}$ only influences the evolution of binaries from the CE channel, while $\chi_{\rm b}$ is varied consistently across all channels.

The birth spin of black holes $\chi_{\rm b}$ is a common hyperparameter between all formation channels. For the field formation channels, black holes are given a minimum spin of $\chi_{\rm b}$ upon formation. A field channel black hole forms with $\chi > \chi_{\rm b}$ if tidal effects or mass transfer sufficiently spin up the black holes. For black holes in the dynamical models, all first-generation black holes formed from

stellar collapse have a spin of $\chi_{\rm b}$, while merger products are born with a spin of $\chi \sim 0.7$, appropriate for nearly equal-mass mergers (Pretorius, 2005; Gonzalez et al., 2007; Buonanno et al., 2008). Varying the black hole birth spin is a proxy for the efficiency of angular momentum transfer in massive stars (Belczynski et al., 2020; Bavera et al., 2020). It is expected that efficient angular momentum transfer results in low birth spins of black holes, giving a narrow range of $\chi_{\rm b} \lesssim 0.01$ (Fuller & Ma, 2019). The models with larger $\chi_{\rm b}$ extend the assumption of a single birth spin value (when ignoring the effects of tidal spin-up or spin-up by mass transfer) to a case where there is some process during stellar evolution or collapse (potentially related to inefficient angular momentum transfer) that imparts a characteristic spin on newly formed black holes (Zevin et al., 2021).

The population synthesis simulations were generated at fixed values of $\vec{\lambda}$, $\chi_{\rm b}^{*} \in \{0, 0.1, 0.2, 0.5\}$ and $\alpha_{\rm CE}^{*} \in \{0.2, 0.5, 1, 2, 5\}$ (Zevin et al., 2021). These were the only points in parameter space considered in previous inferences, and we will use them as the training set for the normalising flows. Our inference parameter space is bound by the minimum and maximum values of the training models: for the continuous inference we explore $\chi_{\rm b}$ over the range $[0, 0.5]$ and $\alpha_{\rm CE}$ over the range $[0.2, 5]$.

We infer the branching ratios of the five formation channels $\vec{\beta} = \{\beta_{\rm CE}, \beta_{\rm CHE}, \beta_{\rm GC}, \beta_{\rm NSC}, \beta_{\rm SMT}\}$ as a mixture model between the channels. Therefore in total we determine 7 population hyperparameters (6 independent since $\vec{\beta}$ must be normalised) made up of $\vec{\Lambda} = \{\vec{\lambda}, \vec{\beta}\}$.

3.3.2 | Statistical framework

We infer a continuous posterior distribution over hyperparameters $\vec{\lambda}$ in addition to the branching fractions $\vec{\beta}$, following a hierarchical inference framework (Mandel et al., 2019; Thrane & Talbot, 2019; Vitale et al., 2020). To do this, we construct a hyperlikelihood for the population hyperparameters, using the posterior samples on the GW observables $\vec{\theta}$ from parameter estimation released in GWTC-2.1 and GWTC-3.0 (Abbott et al., 2023a, 2024). Altering the hyperlikelihood from Section 1.3.1 to include the contribution from each formation channel, we get

$$p(\{\vec{d}\}|\vec{\Lambda}) \propto \prod_{i=1}^{N_{\rm obs}} \frac{1}{S_i \tilde{\xi}(\vec{\lambda})} \sum_j \beta_j \sum_{k=1}^{S_i} \frac{p\left(\vec{\theta}_i^k \middle| \vec{\lambda}\right)}{\pi\left(\vec{\theta}_i^k \middle| \varnothing\right)}, \tag{3.1}$$

where $\{\vec{d}\}$ is the set of observed GW data, S_i is total number of posterior samples for the ith GW for $N_{\rm obs}$ detections, $\tilde{\xi}(\vec{\lambda}) = \sum_j \beta_j \int P_{\rm det}(\vec{\theta}) p(\vec{\theta}|\vec{\lambda}) {\rm d}\vec{\theta}$ is the detection efficiency of each population model, $\pi(\vec{\theta}_i^k|\varnothing)$ is the prior on each GW source for each posterior sample in the LVK analysis. We are solely interested in

the shape of the distributions of $\vec{\theta}$ and therefore implicitly marginalize over the expected number of detections (e.g., Fishbach et al., 2018).

The hyperposterior for $\vec{\Lambda}$ is sampled with `emcee` (Foreman-Mackey et al., 2013), using 250 walkers with 1000 steps and a 40% burn in. We inspect the sampling chains to check the chain convergence, and check the autocorrelation length and sample until we reach at least 50 times the autocorrelation length of the measured hyperparameters.

The previous AMAZE framework (Zevin et al., 2021) evaluated $p(\vec{\theta}_i^k \mid \vec{\lambda})$ and $\tilde{\xi}(\vec{\lambda})$ at discrete $\vec{\lambda} = \{\chi_b^*, \alpha_{CE}^*\}$, where there existed samples from the population synthesis simulations. This work expands the inference to be over any values of $\vec{\lambda}$ within the prior bounds, using normalising flows to evaluate $p(\vec{\theta}_i^k \mid \vec{\lambda})$. The detection efficiency, $\tilde{\xi}(\vec{\lambda})$ is evaluated at each training population point by using a signal-to-noise ratio cut on the population samples to find $P_{\rm det}$, then calculated across $\vec{\lambda}$ using a piecewise cubic Hermite interpolating polynomial (`pchip`) interpolation (Fritsch & Butland, 1984). This ensures a continuously differentiable detection efficiency across the populations in each channel, preserving monotonicity between the known detection efficiency points. This assumes that the values of the detection efficiency we calculate from the population synthesis are representative of the whole range of populations we explore. The interpolation of the detection efficiency allows the continuous inference framework to work together to calculate the hyperposterior for $\vec{\Lambda}$, using the interpolated population models by the normalising flows.

We assume an uninformative prior across $\vec{\beta}$, given by a Dirichlet distribution with equal concentration parameters of 1, such that $(0 < \beta_i < 1)$ for all β_i and $\sum_i \beta_i = 1$. We use a uniform prior on χ_b and log-uniform in α_{CE}, within the bounds of the training set.

As we have a finite number of samples for each population model, the tails of the population model distributions are extrapolated from a small number of samples. Evaluating the probability of the population models in the tails of the distribution therefore does not give a reliable estimate of the true probability of BBH formation in these regions. While the absolute error in reconstructing the distributions may be small, the relative error needed to compare different channels can be significant. To alleviate this problem, we add a regularisation term to the probability determined by the population models, such that a uniform value determined by constant N contributes to $p(\vec{\theta} \mid \vec{\lambda})$, giving a regularised distribution

$$p(\vec{\theta} \mid \vec{\lambda}) = \frac{N}{N+1} p_{\rm pop}(\vec{\theta} \mid \vec{\lambda}) + \frac{1}{N+1} U(\vec{\theta}). \tag{3.2}$$

Here, $U(\vec{\theta})$ is a uniform distribution over the population parameters, and $p_{\rm pop}(\vec{\theta} \mid \vec{\lambda})$ is the reconstructed population distribution for a channel. The additional regularisation assumes that $p(\vec{\theta} \mid \vec{\lambda})$ tends towards a uniform distribution in parts

of the parameter space where there is not enough information from the population synthesis simulations. This form would result from treating reconstruction of the distribution as an inference problem and applying Laplace's rule of succession (Jaynes, 2003, chapter 18). We start with a uniform prior where we have no samples to inform our result, and as we increase the number of population synthesis samples N, we place increasing weight in the population model probability $p_{\text{pop}}(\vec{\theta} \mid \vec{\lambda})$. In parts of $\vec{\theta}$ parameter space where there are many samples, the probability $p(\vec{\theta} \mid \vec{\lambda})$ is approximately equal to the population model probability $p_{\text{pop}}(\vec{\theta} \mid \vec{\lambda})$, but in the tails of the distribution, we tend towards an uninformative uniform distribution. Our analysis uses the same N for each channel, chosen to be the smallest number of training samples out of all of the training sub-populations. This means there is no preference between channels in regions of $\vec{\theta}$ parameter space where we have no samples for any channel.

The inference mirrors the AMAZE framework used in Zevin et al. (2021) and Cheng et al. (2023). This included a change to the calculation of the prior on each GW posterior sample from the work in Cheng et al. (2023), which ensures that there were no prior values with extremely small support for any of the event samples used.

3.3.3 | Normalising flows

Normalising flows are a class of generative neural networks that are trained to find the optimal mapping between a complicated distribution and a simple latent space through a series of transformations, allowing emulation of the more complex distribution (Papamakarios et al., 2021; Kobyzev et al., 2021). We use normalising flows to represent the distributions of GW observables for each formation channel given the population hyperparameters, $p_{\text{pop}}(\vec{\theta} \mid \vec{\lambda})$. As we have samples from these distributions, but not the probabilities over $\vec{\theta}$-space, normalising flows give us a way to directly evaluate these probability distributions.

During training, a normalising flow learns a set of transformations that can be used to reconstruct the distribution of interest from the latent space. For a single transformation f, where the target distribution x relates to the latent distribution $\mathcal{Z}$ as $\mathcal{Z} = f(x)$, the relationship between the target and latent distributions is

$$p_x(x \mid \phi) = p_{\mathcal{Z}}\left(f(x \mid \phi)\right) \left| \frac{\partial f(x \mid \phi)}{\partial x} \right|, \tag{3.3}$$

allowing the normalising flow to map from the target distribution $p_x(x|\phi)$ to the latent distribution $p_{\mathcal{Z}}(f(x|\phi))$, given some conditional parameter ϕ. In our case, x will be related to our source $\vec{\theta}$ parameters and ϕ will be our population hyperparameters $\vec{\lambda}$ for each channel. For a normalising flow the Jacobian of the transformation $|\partial f(x)/\partial x|$ is tractable, and the function f is invertible. Samples

are taken from this latent distribution, which has an analytic form and can be easily sampled from, before being transformed back through the inverse of the learned transformations for emulation of the target distributions. Our transformations use a multidimensional Gaussian for the latent distribution.

The optimised series of transformations are found by minimising the loss function of the network. We use the Kullback–Leibler (KL) divergence (Kullback & Leibler, 1951) between the normalising flow's approximated target distribution $p^*_x(x|\phi)$ and the original target distribution $p_x(x|\phi)$. As our training samples also have associated weights to account for the distribution of BBHs in cosmic time and metallicity, we train the normalising flows to learn the weighted distribution of these samples, $w(x|\phi)p_x(x|\phi)$, where $w(x|\phi)$ is the distribution of sample weights. The KL divergence between the weighted distribution and the emulated distribution produced by the flow in the target space can be written as

$$\begin{aligned} D_{\mathrm{KL}}\left[w_x(x|\phi)p^*_x(x|\phi)||p_x(x|\phi)\right] &= \int_{-\infty}^{+\infty} w_x(x|\phi)p^*_x(x|\phi)\ln\left(\frac{w_x(x|\phi)p^*_x(x|\phi)}{p_x(x|\phi)}\right)\mathrm{d}x \\ &= -\int_{-\infty}^{+\infty} w_x(x|\phi)p^*_x(x|\phi)\ln p_x(x|\phi)\mathrm{d}x + \text{constant} \\ &= -\mathbb{E}_{p^*_x(x|\phi)}\left[w_x(x|\phi)\ln p_x(x|\phi)\right] + \text{constant} \\ &= -\mathbb{E}_{p^*_x(x|\phi)}\left\{w_x(x|\phi)\left[\ln p_{\mathcal{Z}}\left(f(x|\phi)\right) + \ln\left|\det\frac{\partial f(x|\phi)}{\partial x}\right|\right]\right\} \\ &\quad + \text{constant}. \end{aligned} \tag{3.4}$$

In training, we find the gradient of the KL divergence, and thus ignore the constant terms, ultimately minimising the loss, $\mathcal{L}$,

$$\mathcal{L} = -\frac{1}{\mathcal{N}}\sum_{i=1}^{\mathcal{N}} w_x(x_i)\left[\ln p_{\mathcal{Z}}\left(f(x_i)\right) + \ln\left|\det\frac{\partial f(x_i)}{\partial x}\right|\right], \tag{3.5}$$

where $\mathcal{N}$ is the number of samples used in each batch of training.

We train one normalising flow for each of the five formation channels. Each normalising flow is used to reconstruct the population distribution $p_{\mathrm{pop}}(\vec{\theta}|\vec{\lambda})$ using samples of parameters $\vec{\theta}$ from population synthesis simulations. To improve the performance of the training, we use scaled and transformed versions of the source parameters, rather than using $\vec{\theta}$ directly. The training samples from populations synthesis are mapped with a set of logistic and hyperbolic transformations in order to improve the training and avoid hard boundaries in the parameter space of the GW observables. First, we use logistic or hyperbolic mappings such that

each dimension in the observable data space θ is in the range $(-\infty,\infty)$:

$$\begin{aligned}\tilde{\mathcal{M}} &= \text{logit}\left(\frac{\mathcal{M}}{\mathcal{M}^{\text{up}}}\right),\\ \tilde{q} &= \text{logit}\left(\frac{q}{q^{\text{up}}}\right),\\ \tilde{\chi}_{\text{eff}} &= \text{arctanh}(\chi_{\text{eff}}),\\ \tilde{z} &= \text{logit}\left(\frac{z}{z^{\text{up}}}\right),\end{aligned} \tag{3.6}$$

where the inverse logistic function on p is defined as

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right),\ 0 < p < 1, \tag{3.7}$$

and $\{\mathcal{M}^{\text{up}}, q^{\text{up}}, z^{\text{up}}\}$ are convenient upper bounds. The upper bounds are set to $\mathcal{M}^{\text{up}} = 100M_{\odot}$ $q^{\text{up}} = 1$ and $z^{\text{up}} = 10$, except for the GC channel, which contains samples where $q = 1$ and therefore $q^{\text{up}} = 1.001$ to ensure $0 < q/q^{\text{up}} < 1$. Finally, the transformed parameters are rescaled to give

$$\vec{\theta}' = \left\{\frac{\tilde{\mathcal{M}}}{\max(\tilde{\mathcal{M}})}, \frac{\tilde{q}}{\max(\tilde{q})}, \tilde{\chi}_{\text{eff}}, \frac{\tilde{z}}{\max(\tilde{z})}\right\}. \tag{3.8}$$

This rescaling is empirically chosen such that $\vec{\theta}'$ is defined on a similar domain across each dimension. The normalising flows are trained on samples of $\vec{\theta}'$ conditional on $\vec{\lambda}$, learning the distributions $p_{\text{pop}}(\vec{\theta}'|\vec{\lambda}, \text{channel})$.

Before training, 20% of the population synthesis samples were randomly chosen across different values of $\vec{\lambda}$ as the validation set used in the training. The normalising flows were implemented using CouplingNF spline normalising flows (Durkan et al., 2019) using the `glasflow` (Williams et al., 2022) package.

The specific parameters of the normalising flows are shown in Table 3.3.1. These parameters were chosen considering the results of network parameter optimisation tool Weights and Biases (Biewald, 2020). We used Weights and Biases to train 20 normalising flows for each formation channel, varying the number of transforms, neurons, and spline bins for each instance of training. We considered number of transforms in the range $[4, 10]$, number of neurons in the range $[10, 130]$, and number of spline bins in the range $[3, 8]$. Bayesian optimisation was used to find the network parameters that best minimised the validation loss (Dewancker et al., 2016). For the NSC and SMT channels, the choice of network parameters did not make a significant difference on the minimum validation loss. For the other channels, normalising flows with higher numbers of neurons resulted in lower validation losses. We chose the number of transforms and spline bins (given in Table 3.3.1) to reduce the computational expense while still retaining

Formation channel	Epochs	Transforms	Neurons	Blocks	Spline bins
CE	15000	6	128	2	5
NSC	10000	6	128	2	5
CHE, GC, SMT	10000	6	128	2	4

Table 3.3.1: The network parameters of the normalising flows, including the number of training epochs, number of transformations, number of neurons per layer, and number of spline bins in each transform. These parameters vary for the five normalising flows, which each learn the populations in a given formation channel.

the flexibility to emulate the training populations. When visually inspecting the emulated populations, we found that the CE and NSC channels better captured the target correlations with 5 spline bins than with 4 spline bins. The number of training epochs was chosen such that the validation loss had plateaued by the end of training when employing the chosen network configuration. In addition, the learning rate was annealed with cosine annealing throughout training (Loshchilov & Hutter, 2017), with an initial learning rate of 0.001. The training was performed on using a NVIDIA GeForce RTX 2080 Ti GPU and took $\sim$ 50 hr for five channels.

3.3.4 | Observational data

We use results from GWTC-3.0 (Abbott et al., 2023a), analysing candidates with a FAR of less than 1 yr^{-1}, following previous population analyses with this catalogue (Abbott et al., 2023b). The parameter estimation samples used are the mixed samples from GWTC-2.1 and GWTC-3.0 (Abbott et al., 2023a, 2024), containing an equal number of samples from `IMRPhenomXPHM` (Pratten et al., 2021) and `SEOBNRv4PHM` (Ossokine et al., 2020) waveform approximants. We additionally exclude events GW190521 (Abbott et al., 2020c) and GW190814 (Abbott et al., 2020b), as these are outliers within the context of our population models. The detection probabilities are calculated for the LIGO-Hanford, LIGO-Livingston, and Virgo network operating at `midhighlatelow` sensitivities, as a proxy for the sensitivity of the detectors in their third observing run (Abbott et al., 2020a). These choices are consistent with the analyses of Zevin et al. (2021) and Cheng et al. (2023), enabling us to compare our results.

As the code currently requires the same number of posterior samplers for each GW, we use 11,058 posterior samples per GW. For GWs that had fewer than 11,058 posterior samples, we repeated the total set of posterior samples for that GW, and then chose to repeat a subset of the posterior samples, chosen randomly.

Our choice of 11,058 total posterior samples minimised the number of samples in the repeated subset for all GWs, while allowing us to use more posterior samples from the GWs which had them.

3.4 | Results

Here we evaluate the emulation performance of our trained normalising flows with a series of tests to determine that they are trained sufficiently well for the following continuous inference. Our first tests look at the performance of the normalising flow in emulating the training data and interpolating across hyperparameter space (Section 3.4.1). We use the normalising flows for inference on a discrete hyperparameter space, and show that these results are in agreement with results using the previous (non-interpolating) representation of the population models (Section 3.4.2). We then present the results from our extension to inference across a continuous hyperparameter space (Section 3.4.3). Finally, we look at the inferred distribution of GW observables, given the continuous inference results (Section 3.4.4).

3.4.1 | Emulation and interpolation performance

Before using the normalising flows for continuous inference, we test their capabilities of accurately capturing the training data. We compare the trained normalising flows to the training samples from the population synthesis models as well as to KDE representations of the astrophysical models. The KDE representations of the astrophysical models give a smooth probability density across $\vec{\theta}$ space. They are only used to construct these distributions at the points in hyperparameter space where we have population synthesis results, i.e., $\{\chi_b^*, \alpha_{CE}^*\}$. Our KDE implementation is consistent with the KDE model representations used in Zevin et al. (2021) and Cheng et al. (2023). Comparing the two methods enables us to examine the robustness of our results to the approach used to represent the populations.

An example of the samples drawn from the trained normalising flows in $\vec{\theta}$-space is shown in Figure 3.4.1, alongside samples from KDE and the training samples. The trained normalising flow captures the primary features of these distributions, including both the marginal features and the correlations. Both the KDE and the normalising flow struggle to capture the density of sharp features, such as the height of the spike in $\chi_{\rm eff}$. However, the increased flexibility of the normalising flows compared to KDEs mean that this feature is in general better matched by the normalising flow than the KDEs across the training populations, as shown in the inset plot of Figure 3.4.1.

One difference observed between the normalising flow and the training samples is in the discrete bands observed at high redshifts. These are artefacts of the population synthesis models being generated at discrete lookback times before

being combined. This binning in lookback time is not an astrophysically expected feature in true populations. The normalising flow, however, smooths over the binning present in the models in the high redshift tails. This is a by-product of the normalising flow not having sufficient flexibility (given the network architecture) to pick up on more complicated features, and this feature only being present in the tails of the distribution with few training samples. This smoothing means that the normalising flows effectively interpolate over simulations at different lookback times at this GW observable level.

In order to numerically evaluate the difference between the model representations, we calculate the difference in the KL divergence between the normalising flows and training population models, and the KDE models and the training models. For all models across all formation channels, the KL divergence between the normalising flows and the training data is lower than that between the KDE and training data, with an average difference in KL divergence for each formation channel being -0.37 nat, -0.49 nat, -0.67 nat, -0.72 nat, -0.97 nat for the CE, CHE, GC, NSC and SMT channels, respectively. For the difficult $\chi_{\rm eff}$ distribution in the CE channel, we find that the Kolmogorov–Smirnov (K–S) test between the normalising flow and the training distribution is lower than the K–S test between the KDE and the training distribution by a factor of 10 averaged over all models. A lower KL divergence for the normalising flows than the KDEs indicates that the normalising flows are more similar to the training samples than the KDEs. Therefore, the normalising flows are on average a better representation of the target population synthesis models for each formation channel.

After testing the emulation of the flow at points where it has seen the training data, we test how the normalising flows perform at points in between the training data. Figure 3.4.2 shows the interpolation of the normalising flow between training points for the CE channel. The solid lines show the distributions at training points, with the darkest and lightest values at the two extremes of the training set. We plot with dashed lines the normalising-flow distribution between the simulated populations, showing how the normalising flow interpolates these distributions. Even though the normalising flow does not see data from the dashed distributions during training, it can reproduce the features of the nearby training data and evolve the distribution between the points informed by population synthesis.

We verify the performance of this normalising flow interpolation by removing one training population from the total training data, and testing how it is reproduced by the normalising flow. The reproduced unseen population is shown in Figure 3.4.3. This model was the $\chi_{\rm b} = 0.1, \alpha_{\rm CE} = 2.0$ training population from the CE channel, a population in the middle of the grid of the $\chi_{\rm b}^*$ and $\alpha_{\rm CE}^*$ values used in training. The remaining 19 training populations were used to train this normalising flow. The normalising flow emulates the test population data well even though they were not seen during training. As a figure of merit we compare the KL divergences between the underlying model and the normalising flow with the removed test data, the normalising flow trained on the whole

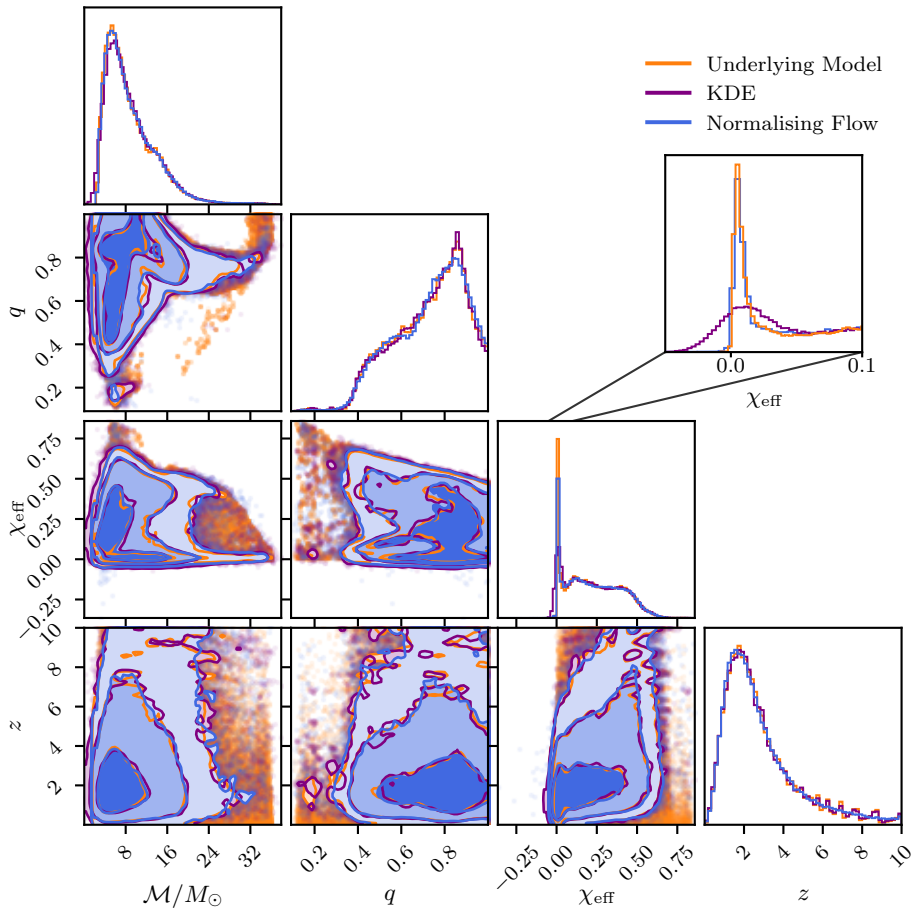


Figure 3.4.1: The distribution of GW observables from the training population synthesis samples (orange), the KDE of this population (purple), and the normalising flow (blue), for the CE channel with $\chi_{\rm b} = 0$ and $\alpha_{\rm CE} = 2$. The confidence levels represent the areas containing 50%, 90%, and 99% of the probability. The normalising flow matches the training distribution well both in the marginal distributions and in the correlations. The spike in the effective inspiral spin distribution, shown in more detail in the inset plot, is due to the model assumption that black holes are born with a common spin. The number of effective samples is the same for all three model representations.

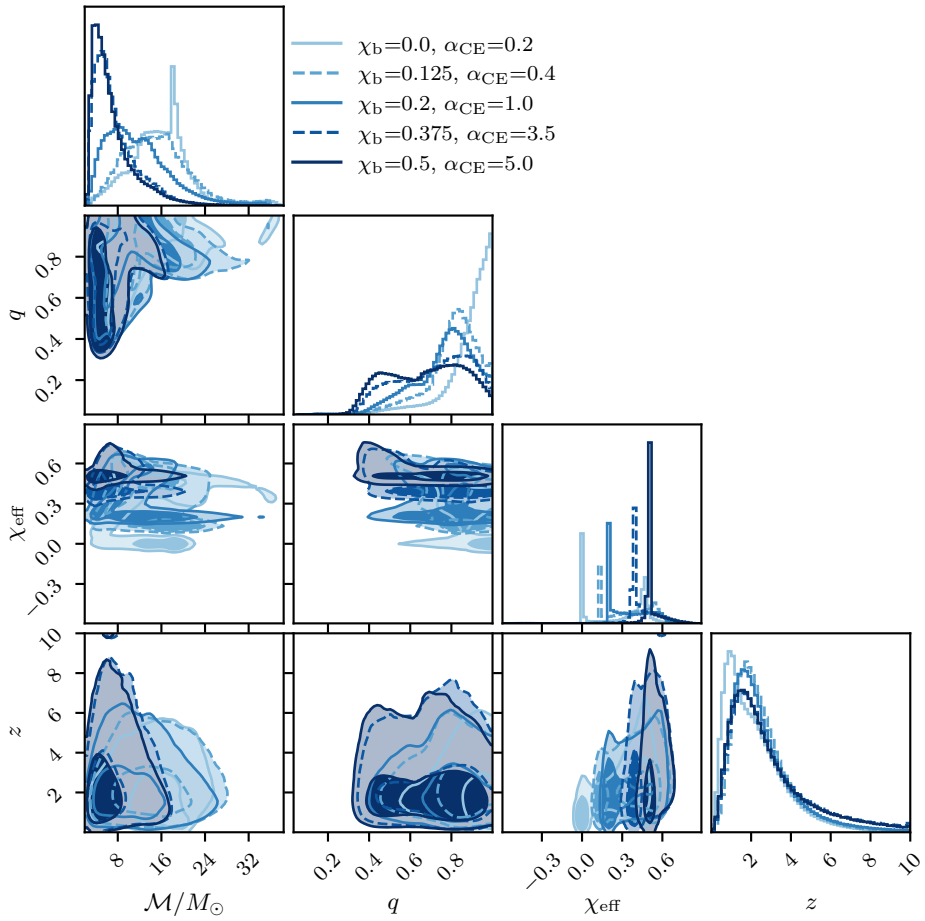


Figure 3.4.2: The normalising-flow emulated distributions of chirp mass, mass ratio, effective inspiral spin and redshift for the CE channel at five values of natal spin χ_b and common-envelope efficiency α_{CE}. The solid lines show the emulated distribution evaluated where there is training data from population synthesis. The dashed distributions are evaluated at $\{\chi_b, \alpha_{CE}\}$ where there is no training data, showing the interpolation of the normalising flow.

dataset, and the KDE for the population model with $\chi_b = 0.1, \alpha_{CE} = 2.0$. The KL divergence between the normalising flow at the removed population and the underlying model is comparable to the KL divergence between the normalising flow which saw all the training data and the underlying model, with a difference of -0.04 nat. The difference in KL divergences between the normalising flow at the removed population and the underlying model and between the KDE and the underlying model is -0.41 nat. There is a smaller difference in KL divergence when comparing the two normalising flows than when comparing the normalising flow with the unseen data to the KDE, meaning the two normalising flows have more similar distributions. This indicates that the normalising flow trained with the removed population emulates the unseen distribution better than the KDEs, and it has a similar performance to the normalising flows which saw the whole dataset. The normalising flows are therefore trained well enough, and with a powerful enough network architecture, to interpolate this set of training data. Conversely, the interpolation performance shows that there are sufficient training data, and that the training models are close enough together in parameter space to allow for good interpolation. By showing that the normalising flow can replicate an unseen test distribution, we infer that the normalising flow is reliable across the parameter space where we do not have training samples. Testing the reconstruction of omitted training data is an important verification step of the flow interpolation performance, and it motivates generating more training data with which to perform these tests.

While we have established that KDEs and normalising flows both do well at reproducing the bulk features of the distributions, we find that they do differ in the tails, where we must extrapolate the distributions due to the finite number of samples. Our uncertainties here can be ameliorated through our introduction of regularisation (Section 3.3.2). Figure 3.4.4 shows the difference in the likelihood surface $p(\{\vec{d}\}|\vec{\Lambda})$ of the normalising flows in comparison to that of KDEs for a slice through the two-dimensional $\mathcal{M}$–q plane with and without regularisation. Without regularisation, the biggest difference in the likelihood surface is present in the tails of the distribution where we have few population synthesis samples. Regularisation mitigates the differences in the tails by adding a uniform value for both the normalising flow and KDE. Hence, as desired, with the additional regularisation both representations tend to the same (uninformative) limit where we lack simulation points.

3.4.2 | Testing normalising flows with discrete inference

Having validated that the normalising flows can emulate the target population distributions, we now compare the flows to the KDEs in the discrete model selection framework. After training, we evaluate the probability of the observations given the normalising flows at the points $\{\chi_b^*, \alpha_{CE}^*\}$. The hyperposteriors on the

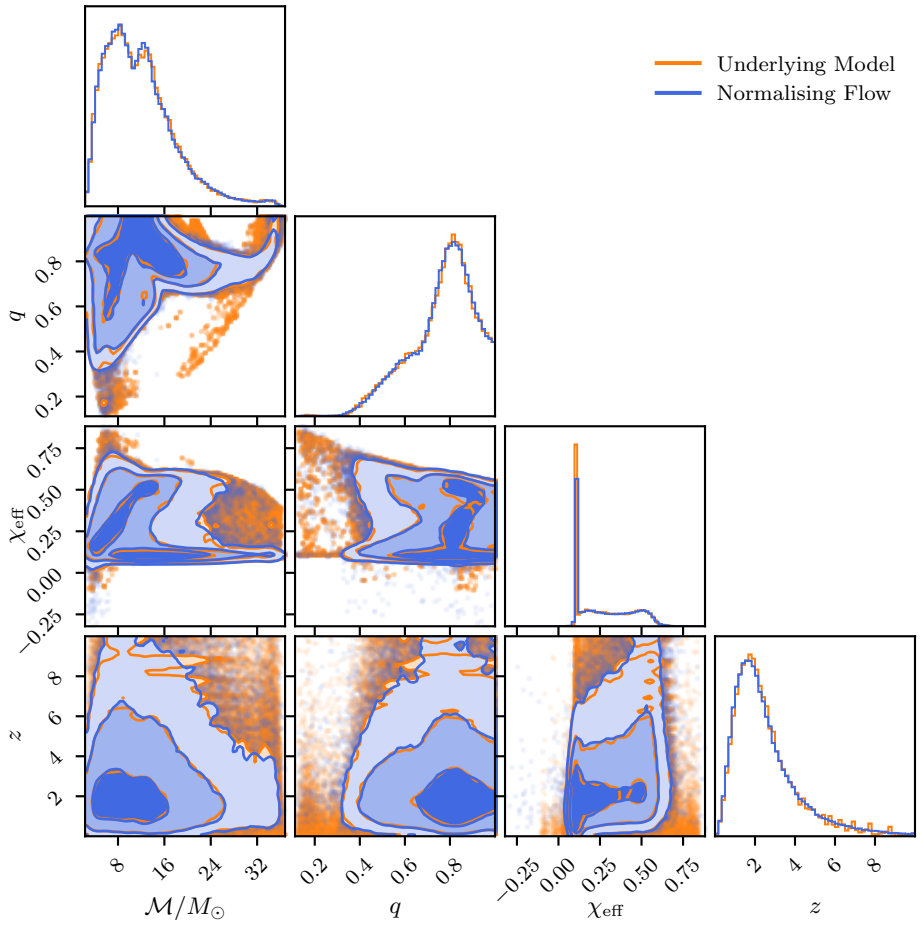


Figure 3.4.3: The distribution learned by the normalising flow (blue) for the CE channel for $\chi_b = 0.1$ and $\alpha_{CE} = 1$, without having seen the underlying training data (orange) for this particular population model. By seeing the training data for other values of χ_b and α_{CE} around this population, the normalising flow is able to interpolate and reproduce this unseen distribution. The number of effective samples is the same for both the underlying model and normalising flow samples.

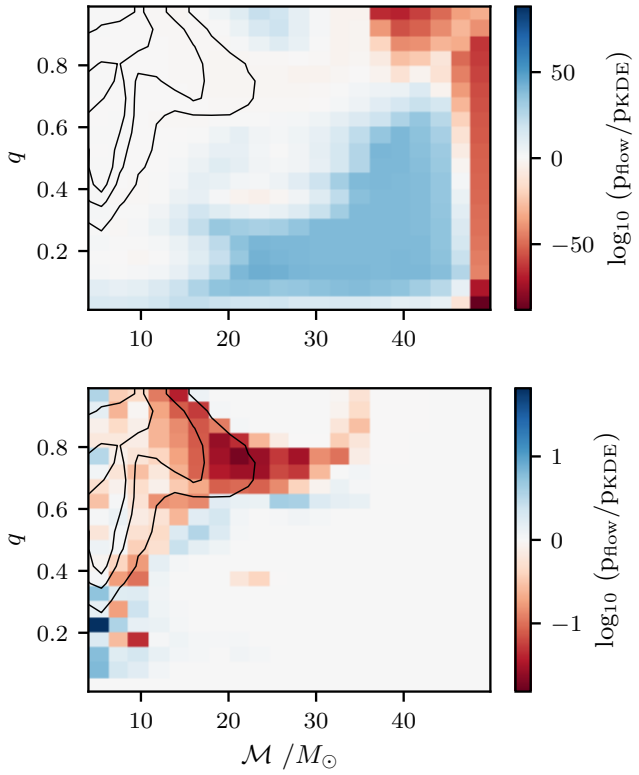


Figure 3.4.4: The log likelihood ratio between the normalising flow and the KDE representation of the CE population model with $\chi_b = 0$ and $\alpha_{CE} = 5$, at a slice through chirp mass and mass ratio at effective inspiral spin of 0.05 and redshift of 0.1. The black contours trace the areas containing 50%, 90%, and 99% of the training samples in a bin around this slice. The top panel shows no regularisation applied to the normalising flow or the KDE, with the likelihood ratio at the most extreme in the tails of the probability distribution. The bottom panel shows the log likelihood ratio when using regularisation for both representations, minimising the differences in the tails of the distribution. The scale bar in the lower panel covers a much smaller range of ratios.

	Normalising flows	KDE
β_{CE}	$0.897^{+0.049}_{-0.097}$	$0.913^{+0.042}_{-0.088}$
β_{CHE}	$0.006^{+0.010}_{-0.004}$	$0.005^{+0.009}_{-0.004}$
β_{GC}	$0.048^{+0.061}_{-0.034}$	$0.032^{+0.056}_{-0.029}$
β_{NSC}	$0.016^{+0.018}_{-0.010}$	$0.015^{+0.017}_{-0.009}$
β_{SMT}	$0.029^{+0.051}_{-0.024}$	$0.031^{+0.055}_{-0.026}$

Table 3.4.1: Inferred median and 90% symmetric credible interval for the underlying branching fractions from the discrete inference using normalising flows and using KDEs.

underlying branching fractions, and the model selection between the parameters χ_{b} and α_{CE}, comparing the two inference methods are shown in Figure 3.4.5.

The inferred branching fractions for the five formation channels for both the normalising flows and the KDEs are shown in Table 3.4.1. The branching fractions are consistent between both methods for all formation channels, with the CE channel having the highest branching fraction. Both the KDE and the normalising flows prefer the $\chi_{\mathrm{b}} = 0$ and $\chi_{\mathrm{b}} = 0.1$ models, with a slight preference for the $\chi_{\mathrm{b}} = 0.1$ model. There is a log Bayes factor between the models $\log_{10} \mathcal{B}^{\chi_{\mathrm{b}}=0.1}_{\chi_{\mathrm{b}}=0} = 1.3$ for the normalising flows, and $\log_{10} \mathcal{B}^{\chi_{\mathrm{b}}=0.1}_{\chi_{\mathrm{b}}=0} = 0.61$ for the KDEs. The normalising flow also prefers high α_{CE}, in line with the KDE results; both methods strongly preferring the $\alpha_{\mathrm{CE}} = 5$ model, with some support for $\alpha_{\mathrm{CE}} = 2$, and minimal support for the models with lower α_{CE}. The log Bayes factor between the two preferred α_{CE} values is $\log_{10} \mathcal{B}^{\alpha_{\mathrm{CE}}=5}_{\alpha_{\mathrm{CE}}=2} = 2.7$ for the normalising flows and $\log_{10} \mathcal{B}^{\alpha_{\mathrm{CE}}=5}_{\alpha_{\mathrm{CE}}=2} = 2.1$ for the KDEs. The general agreement of the normalising flows and the KDEs on the hyperposteriors of the branching fractions and the model selection of χ_{b} and α_{CE} indicates the robustness of the results, and shows that normalising flows can perform comparably to the established KDE approach in a hierarchical inference problem.

The branching fraction hyperposteriors for the CE and GC channels are wider for the normalising flow method, indicating that some of the degeneracies between these channels may be driven by the choice of model representation. The support for natal spin model differs somewhat between the normalising flows and the KDEs, with the KDEs giving a stronger preference for $\chi_{\mathrm{b}} = 0$. This may be due to the differences in each model representation's ability to capture the sharp spikes in the effective inspiral spin distributions. The KDEs also have a stronger preference for $\alpha_{\mathrm{CE}} = 2$. In general for this inference framework, α_{CE} is less well measured than χ_{b} due to only being constrained by one of the formation channels.

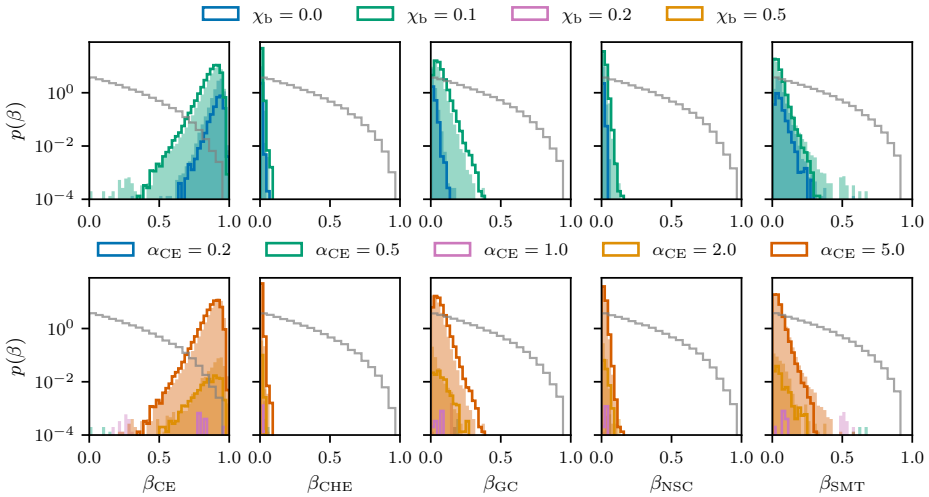


Figure 3.4.5: The hyperposteriors on the underlying branching fractions of each formation channel, using the normalising flows (solid lines) and KDEs (filled) for discrete inference. The top row shows the samples corresponding to each of the χ_b populations, and the bottom row shows those corresponding to the different α_{CE} models, with the prior in grey. We infer negligible contribution to the population from the $\chi_b = 0.2$ and $\chi_b = 0.5$ models, or the $\alpha_{CE} = 0.2$ model (both with normalising flows and KDE model representations), and hence these curves do not show up in the plot.

The minor updates to the AMAZE framework have not changed the discrete inference results significantly, and results are consistent with past analyses (Zevin et al., 2021; Cheng et al., 2023). As the model representations are different in some areas of the population parameter space, we expect some minor differences in the inference results, but the overall agreement indicates that the differences are not significant.

3.4.3 | Continuous inference with gravitational-wave observations

Having verified the emulation and checked agreement with discrete inference, we now show the unique feature of the normalising flows: the extension to continuous inference. We allow the astrophysical population parameters $\vec{\lambda}$ to vary anywhere within the prior, inferring a continuous hyperposterior on these parameters. Our recovered hyperposteriors on $\vec{\lambda}$ and the underlying and detectable branching fractions between formation channels are shown in Figure 3.4.6. The upper panels show the hyperposteriors (blue histograms) with the points in hyperparameter space where our population synthesis models lie (where the discrete

	β	β^{det}
CE	$0.908^{+0.045}_{-0.102}$	$0.22^{+0.13}_{-0.10}$
CHE	$0.005^{+0.009}_{-0.004}$	$0.06^{+0.08}_{-0.05}$
GC	$0.015^{+0.046}_{-0.014}$	$0.10^{+0.22}_{-0.09}$
NSC	$0.013^{+0.015}_{-0.008}$	$0.25^{+0.16}_{-0.13}$
SMT	$0.053^{+0.079}_{-0.036}$	$0.33^{+0.21}_{-0.21}$
χ_{b}	$0.04^{+0.04}_{-0.01}$	
α_{CE}	> 3.7	

Table 3.4.2: Inferred median and 90% symmetric credible intervals for the underlying and detectable branching fractions and the natal spin, and the 90% credible lower limit for CE efficiency from the continuous inference results using normalising flows.

inference was evaluated; grey vertical lines). The measured values on our hyperposteriors are quoted in Table 3.4.2.

We measure $\chi_{\mathrm{b}} = 0.04^{+0.04}_{-0.01}$, with most of the posterior support in between $\chi_{\mathrm{b}} = 0$ and $\chi_{\mathrm{b}} = 0.1$. Given the assumptions in our population models, we find that black holes in merging BBHs are born with low, but non-zero spin. We find that $\chi_{\mathrm{b}} > 0$ is preferred over $\chi_{\mathrm{b}} = 0$, with a log Bayes factor $\log_{10} \mathcal{B}^{\chi_{\mathrm{b}}>0}_{\chi_{\mathrm{b}}=0} = 2.5$.

The narrow hyperposterior on χ_{b} may be an artefact of the assumption in the population models that black holes share a common birth spin in the absence of spin-up effects (Zevin et al., 2021). This assumption causes the sharp feature at χ_{b} in the χ_{eff} distributions (Figure 3.4.1). Our models for $\chi_{\mathrm{b}} = 0$ follow the assumption of efficient angular momentum transfer, meaning that black holes uninfluenced by binary interactions (typically including the first-born black hole for the CE and SMT channels) are born with negligible spin (Fuller & Ma, 2019). In this case, natal spins are expected to be in a narrow range $\chi_{\mathrm{b}} \lesssim 0.01$. The assumption that black holes are born with low spin may not hold when considering different prescriptions for angular momentum transfer or tidal spin up (Belczynski et al., 2020; Bavera et al., 2020; Olejak & Belczynski, 2021; Fuller & Lu, 2022). For example, with less efficient angular momentum transfer, we expect the distribution of effective inspiral spin of BBHs to broaden (e.g., Périgois et al., 2023). Other influences on the natal spin of a black hole include formation kicks (Baibhav & Kalogera, 2024) and accretion during black hole formation (Issa et al., 2025), while super Eddington accretion could increase the spins of black holes post-formation (van Son et al., 2020; Bavera et al., 2021; Zevin & Bavera, 2022).

Larger spins are considered as a proxy for these influences within the population models we use by increasing χ_b; however, the models with $\chi_b > 0$ maintain the narrow range of natal spin values. Therefore our use of χ_b may not accurately reflect mechanisms that produce higher natal spins. Updating the assumptions pertaining to natal spin, angular momentum transfer, and spin up in our population synthesis training data could give us a more reliable spin measurement.

Our results for α_{CE} show preference for high values, with $\alpha_{CE} > 3.7$ at 90% credibility, and an extended tail towards lower values. These results motivate follow-up population synthesis simulations at points of interest, at high values of α_{CE} where we have the most support (and at higher values than in our training set), and around $\alpha_{CE} = 2$. These extra simulations could improve the emulation accuracy of the normalising flow around these points. The populations with highest α_{CE} correspond to populations with the lowest masses and greatest redshifts, hence the lowest detection efficiency. The detection efficiency decreases monotonically from $\alpha_{CE} = 1$ to $\alpha_{CE} = 5$, therefore the structure we see over this range is not driven by the detection efficiency.

These results reflect the findings from Zevin et al. (2021) and Cheng et al. (2023) that there is a preference for high α_{CE} when considering the same formation channels and hyperparameters. This is also in line with Bouffanais et al. (2021), who found the largest support at $\alpha_{CE} \sim 5$ and $\alpha_{CE} \sim 6$ when considering only BBHs from isolated evolution, and α_{CE} in the range $[1, 10]$. Wong et al. (2021a) found a weak preference for low α_{CE}, using mass and redshift observations to constrain a population made up of a mixture of binaries from isolated evolution and globular clusters. However, this was not well constrained as the CE channel was not found to be the majority channel in their results. Gayathri et al. (2025) found that α_{CE} was not well constrained when considering only `SEVN` models with varying metallicity, highlighting how the choice of physics varied within the included models can change the inferred result. The modelling of the CE phase using the α_{CE} prescription may not be best suited to these analyses, as there is growing evidence that α_{CE} is not a universal value for all systems (Politano, 2004; Iaconi & De Marco, 2019). Updated prescriptions for the CE phase (e.g., Hirai & Mandel, 2022) may benefit our constraints on the inference of CE astrophysics.

We find that the measurements of α_{CE} and χ_b have no strong correlations with each other or any of the branching fractions. This shows that despite overlap in the populations, there are sufficiently distinguishing features on a population level given the current observations.

The branching fraction results are consistent with our discrete inference, showing support for a mix of channels contributing to the observations, with the CE channel contributing the most to the underlying mergers. We infer the majority of astrophysical mergers to come from the CE channel, despite the prior preference against any single branching ratio being large. However, the CE channel has the lowest detection efficiency of the formation channels, and thus contributes significantly less to the detected population, as seen in the detectable

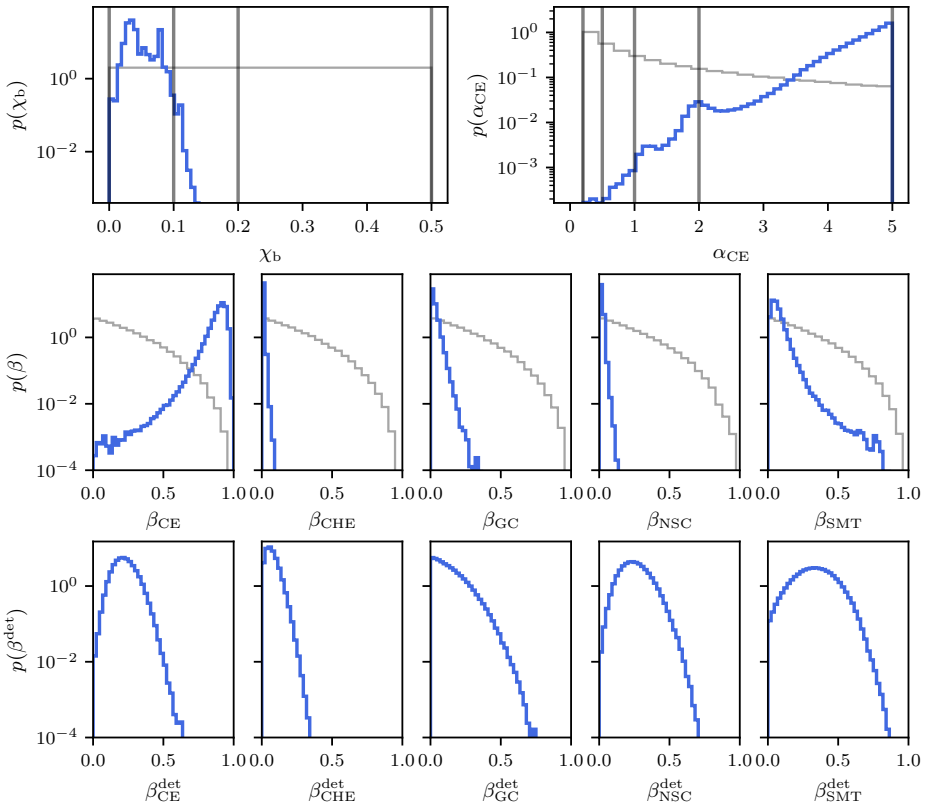


Figure 3.4.6: The hyperposteriors on astrophysical hyperparameters χ_b (top left) and α_{CE} (top right), the underlying branching fractions for the five formation channels (middle row), and the detectable branching fractions (bottom row), using normalising flows for continuous inference of χ_b and α_{CE}, and using data from GWTC-3.0 (Abbott et al., 2023a). The vertical lines on the top row show the location for the population synthesis models used to train the normalising flows. The grey histograms show the priors on χ_b, α_{CE}, and the underlying branching fractions.

branching fractions. The second largest contribution to the underlying channels is the SMT channel, for which we infer a higher branching fraction compared to the discrete inference. We find a strong correlation between β_{CE} and β_{SMT}. These populations contribute to the lower masses and effective inspiral spins close to χ_b, and therefore may be harder to distinguish than the other channels with our observations.

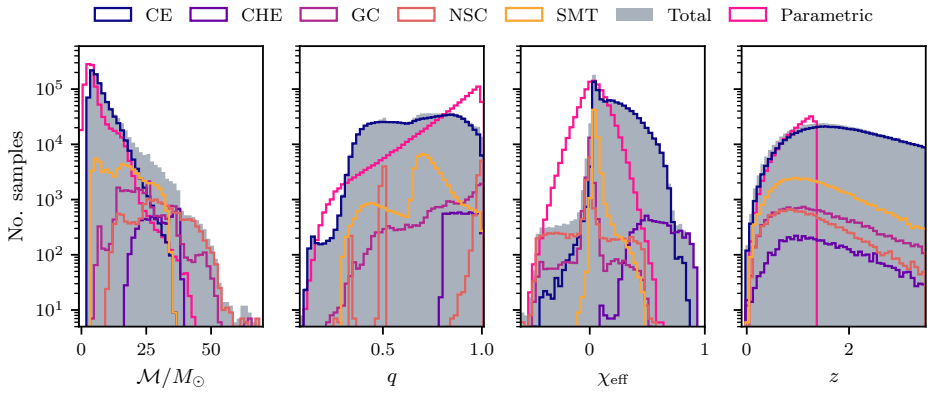


Figure 3.4.7: The distribution of GW observables given by the inferred hyperparameters using continuous inference, broken down into contributions from each formation channel, marginalised over the hyperposterior. The total contribution from all five channels is shown in dark grey. We compare this to inferred results using standard parametric models using GWTC-3.0 (Abbott et al., 2023b), shown in pink. Compared to the parametric models, our inferred population has a greater contribution at higher chirp masses, mass ratios around 0.5, and more positive effective inspiral spins.

3.4.4 | Comparing the inferred population to parametric models

Finally, we look at the inferred population in GW observable-space, given our results from the continuous inference using normalising flows. We compare these distributions to results from parametric models that are based on simple functional forms (and therefore are less reliant on specific astrophysical assumptions). We show the parametric inference results from the default models used in the GWTC-3.0 population analysis (Abbott et al., 2023b). This used a power-law-plus-peak model in primary mass (Talbot & Thrane, 2018), a power law in mass ratio, an identical beta distribution for both spin magnitudes, a mixture of isotropic and normally distributed spin tilts, and a power law in redshift. Figure 3.4.7 shows the contribution of the five formation channels, marginalised over the hyperposteriors from our continuous inference, in comparison with the parametric inference.

Our measured chirp mass distribution has support out to higher masses in comparison to the parametric model. While the CE and SMT channels are the primary contributors to the population between $\mathcal{M} \sim 2$–$30\,M_\odot$, the higher chirp mass tail is from hierarchical mergers in the dynamical formation channels, GC and NSC. Our results support the notion that the underlying BBH population has contributions to the high mass population from dynamical channels

(Kimball et al., 2021; Tagawa et al., 2021; Antonini et al., 2023, 2025). Both the parametric result and our result suggest that the data supports a large fraction of BBHs at low masses. The distribution from the parametric result peaks at smaller masses than our result; this is a consequence of a minimum chirp mass of $\sim 2.3\,M_\odot$ in the population synthesis models we consider. The CE channel is the only channel than can contribute to the population for $\mathcal{M} \sim 2.3$–$3.5\,M_\odot$, as the other channels do not have chirp masses this low. This necessitates that the majority of the contribution at low masses comes from the CE channel. As lower mass BBHs make up the majority of the underlying population, this drives the underlying branching fraction for the CE channel to be high. Exploring a variable minimum black hole birth mass as an input hyperparameter may help to accurately fit the mix of formation channels, and also uncover details of supernova mechanisms (Fryer et al., 2012; Zevin et al., 2020; Mandel et al., 2020; Farah et al., 2022; Olejak et al., 2022).

There is more structure in our inferred mass ratio distribution than the power-law mass ratio model. While the power-law model peaks at $q = 1$, we find a flatter distribution above $q \sim 0.5$, with an excess of mergers in the range $q \sim 0.3$–0.8 and a dearth of mergers at symmetric mass ratios compared to the power-law model. Our results are similar to those found by Godfrey et al. (2023), who use a data-driven model to infer the distributions of BBHs in GWTC-3.0, finding a flatter distribution of mass ratios between $q \sim 0.3$ and $q = 1$. Similarly, Rinaldi et al. (2024) find a mass ratio distribution which does not show support for equal mass binaries, instead the main contribution being centred around $q \sim 0.7$. The results from Callister & Farr (2024) also infer a mass ratio distribution that either peaks at $q = 1$ or is flat in mass ratio. While the parametric power-law model is limited to a monotonic function of mass ratio, these data-driven models identify alternative structure in the data. This raises the question whether BBHs do preferentially merge with equal masses (Fishbach & Holz, 2020), or if this is driven by assumptions in the power-law model. Our results add to the evidence that the mass ratio distribution of BBHs does not peak at $q = 1$, motivating the investigation of alternative parametric models for this distribution, as highlighted in Chapter 2.

For the effective inspiral spin distribution, our inferred population is significantly more asymmetric than the parametric distribution, with a skew towards positive values. Our population models assume that BBHs with large negative effective inspiral spin can only be produced by the two dynamical channels, and we infer only a small fraction of mergers come from dynamical channels. Furthermore, our constraints on χ_b to low values mean that there is a narrower spread of χ_eff for the dynamical channels, contributing to the lack of support for negative χ_eff in our measurements. There is a general preference for positive χ_eff in the data (e.g., Miller et al., 2020; Roulet et al., 2021; Callister et al., 2021; Abbott et al., 2023b; Edelman et al., 2023; Adamcewicz et al., 2024). Recent work by Banagiri et al. (2025a) find a positively skewed χ_eff distribution, although the distribution

is mostly contained to $\chi_{\text{eff}} < 0.4$. In comparison, we find the fraction of the population with $\chi_{\text{eff}} > 0.4$ to be 14.5%, compared to 1.6% from the `skewnormal` mixture model from Banagiri et al. (2025a), and 0.2% for the parametric results from the default models used in the GWTC-3.0 population analysis (Abbott et al., 2023b). Our results show a larger preference for positive χ_{eff}, in contrast to these previous works. The `skewnormal` mixture model from Banagiri et al. (2025a) allows for a subpopulation with high χ_{eff}, but the data do not strongly support such a feature. This suggests that the significant contribution from BBHs with $\chi_{\text{eff}} > 0.4$ that we infer could be driven by the strong astrophysical assumptions in our population models. However, our results are in agreement with these other works in that there is negligible support for a highly spinning subpopulation of BBHs, with our results finding that 99% of the inferred population has $\chi_{\text{eff}} < 0.63$. These comparisons between population models highlight the usefulness of models with different strengths of astrophysical assumptions in understanding what drives the shape of our inferred population in different regions of parameter space.

The redshift distribution varies the least between formation channels, and looks the most similar to the parametric result. The parametric distribution is cut off at $z = 1.35$ as our GW measurements are not strongly constrained beyond this range. Our result has a shallower redshift evolution than the parametric model. With measurements of GWs past $z > 2$, measuring the peak of the redshift distribution and the rate evolution at higher redshifts could help distinguish between different formation channels.

3.5 | Conclusion

We extended the AMAZE framework to continuous population inference over astrophysical hyperparameters, using normalising flows to interpolate simulations from population synthesis. We trained normalising flows to emulate the distributions of GW observables for five different formation channels, given two astrophysical input hyperparameters. Through this method we demonstrated how we can use astrophysically motivated population models of BBH populations to understand the origins of our observed population, without the computational expense of a dense grid of population synthesis simulations.

We validated our method with comparisons to a non-interpolating method which uses KDEs for the astrophysical model representation, ensuring that the normalising flows are fit for purpose. These tests show that the flexibility of the normalising flows can capture features in our training populations more accurately than the KDEs. Both methods produce consistent results when used for inference between the discrete training population models.

Using GWTC-3.0 observations (Abbott et al., 2023a), we applied the normalising flow interpolation to recover continuous hyperposteriors on our astrophysical

hyperparameters (natal spin and CE efficiency) alongside the branching fractions of the formation channels. Given our population model, the measurement of natal spin suggests that black holes are born with small spins of $\chi_b = 0.04^{+0.04}_{-0.01}$. We find a preference for high CE efficiency, with $\alpha_{CE} > 3.7$ at 90% credibility. This pushes to the edge of our prior on α_{CE}, motivating further population synthesis simulations around and above $\alpha_{CE} = 5$. While detailed simulations of CE evolution have indicated that α_{CE} may be greater than 1, it is not yet clear what the largest physically plausible description may be for BBH formation (Fragos et al., 2019; Law-Smith et al., 2020; Lau et al., 2022; Roepke & De Marco, 2023). In general, our framework can be used to find the region of highest (and lowest) probability for population synthesis input hyperparameters, and determine interesting locations for further simulations.

Our analysis indicates that the majority of mergers (in the underlying population) come from the CE channel, in line with our tests with discrete inference and the results from Cheng et al. (2023). Due to the much lower detection efficiency for this channel, we find that the detected fraction of CE binaries is 22^{+13}_{-10}% of the overall population, and no one channel dominates the observed population. Given this mix, it is important not to interpret all observations as coming from a single channel.

The continuous inference framework is crucial for future investigations of key aspects of BBH formation. Measurements of astrophysical hyperparameters vary when making different assumptions about the formation channels present (Zevin et al., 2021; Cheng et al., 2023), and their possible input parameters. For example, we measure higher CE efficiency than Wong et al. (2021a), partially driven by the choice of population inputs. Considering more formation channels in our analysis means that a single channel does not have to account for a wide range of observations. This motivates the inclusion of multiple channels and astrophysical hyperparameters in future analyses. Our framework employing normalising flows for continuous inference is well suited to population inference with multiple formation channels and varying astrophysical hyperparameters.

We find that our inferred population shows more structure in the distributions of chirp mass, mass ratio and effective inspiral spin than current results using parametric population models (Abbott et al., 2023b). Using population synthesis models provides a more astrophysically informed approach to interpreting observations. There are many potential uncertainties in these models, which may influence our conclusions. Using our continuous inference, it is possible to marginalise over the uncertainty in input hyperparameters, mitigating potential bias from uncertainty in these parameters. However, our results are only as reliable as the models used as input. Furthermore, we do not yet account for the absolute merger rate predicted by our population models, instead marginalising over this quantity. This means that the predicted rate from the population synthesis may not match the observed rate of BBH mergers; additionally, adding in information on the absolute merger rate could make results more precise (Barrett et al.,

2018; Mastrogiovanni et al., 2022). By comparing our results to parametric and data-driven population models, we may further check the consistency of results, and explore the limitations of population models. Our results illustrate the complementary nature of the different kinds of population analyses, and motivate including new structure to parametric models.

While we show that the normalising flows can effectively interpolate the population distributions over our hyperparameter space, for the detection efficiency we used a cubic polynomial interpolation. This encodes the assumption that the detection efficiency values we calculate from the population synthesis models are representative for the full model space we use in the continuous inference, which may not necessarily be true. The accuracy of this quantity is especially important with higher numbers of GW observations, as it scales to the power of the number of observations. Previous work has shown that neural networks can be used to emulate the selection function for GW detection (e.g., Mould et al., 2022; Chapman-Bird et al., 2023; Callister et al., 2024). A similar approach could be implemented in our framework, extending our implementation of neural networks to interpolate the detection efficiency, making our results more reliable. We explore this further in Chapter 4.

Since this work, hundreds more events detected over the fourth observing run (Abac et al., 2026b,c). These additional BBH observations will give us more constraining power to measure BBH population properties (Zevin et al., 2017; Barrett et al., 2018), and allow us to measure a larger subset of input hyperparameters to population synthesis. To fully leverage our population models with these observations, it is necessary that the models reflect our uncertainty on the astrophysical origins of BBHs and consider these larger sets of parameters. Our work illustrates that using normalising flows for astrophysical inference is a robust method that is capable of continuous inference over multiple population hyperparameters and multiple formation channels, with potential to be extended to higher dimensional inference. In the future, this method may be extended to infer additional parameters such as black hole natal kick magnitude, stellar cluster properties, and stellar wind parameters. With more population synthesis simulations varying these parameters, our continuous inference framework could be used to understand the degeneracies between these key variables in BBH progenitor evolution, and uncover the origins of these systems.

4 Inferring the origins of binary black holes with growing gravitational wave catalogues

4.1 | Abstract

The rapidly growing number of GW observations bring great potential for inferring detailed physics of BBH formation. While measurements of BBH population parameters become increasingly precise, attention is required to mitigate systematic biases in population inference. We demonstrate the increasing precision of measurements of the BBH population using parametric variable minimum secondary mass models, and normalising flow emulators of population synthesis, with GWTC-2.1, GWTC-3.0, and GWTC-4.0. We show that GWTC-4.0 data remains consistent with no strong evidence for a variable minimum mass. Comparing two methods for detection efficiency interpolation within the AMAZE framework, we find that small changes in the detection efficiency can significantly change measurements of astrophysical parameters in GWTC-4.0. This highlights the importance of having accurate population and detection efficiency models for robust population inference with current and future GW observations.

4.2 | Introduction

The LVK Collaboration's fourth observing run included more than two years of observing at the most sensitive detector configuration so far (Capote et al., 2025; Abac et al., 2025c). The first part of the fourth observing run (O4a) resulted in a total of 153 BBHs detected with FAR$< 1/\mathrm{yr}^{-1}$ included in GWTC-4.0 (Abac et al., 2026b). A further 104 confidently detected BBHs were recently published in GWTC-5.0 (Abac et al., 2026c). The sixth Gravitational-Wave Transient Catalogue (GWTC-6.0) will detail another $O(100)$ observations, which are to be published by the end of 2026 (Fig 1.1.4). After more than a decade's worth of observing GWs, we are approaching several hundred observations which can be

used to constrain BBH astrophysics.

Multiple studies have shown how increasing numbers of simulated observations allows us to more precisely infer the evolution of BBH progenitors (Stevenson et al., 2017a; Zevin et al., 2017; Mastrogiovanni et al., 2022; Zevin et al., 2021). Indeed, with growing catalogues we are now able to see the reduction of statistical uncertainties with real observations (e.g., Abac et al., 2025a, 2026a). With shrinking statistical uncertainties, potential systematic uncertainties become a growing concern (e.g., Gaebel et al., 2019; Talbot & Golomb, 2023; Cheng et al., 2023; Heinzel & Vitale, 2025; Wolfe et al., 2026). The growing number of observations allow us to explore the robustness of each element of population inference, and motivates their improvement.

One example of improvements to population inference frameworks is the development to BBH population models. These improvements include the identification of more favoured parametric model forms (e.g., Abac et al., 2025a), and an understanding of the regions where model systematics and priors dominate over data-driven regions (e.g., Abac et al., 2026a). With more observations, we are able to probe finer details in both marginal distributions and correlations, which helps to avoid model misspecification.

Additionally, producing accurate, computationally efficient descriptions of selection effects for the BBH population is extremely important for population inference. The selection effect contribution is proportional to the power of the number of observations, which can lead to fast-growing systematic biases with expanding catalogues (Talbot & Golomb, 2023). Of the different methods to estimate selection effects, injection campaigns are the most robust, though they require significant computational effort and coordination of person power (e.g., Essick et al., 2025). Additionally, injections also only cover a fixed range of intrinsic distributions; quickly reproducing injections with higher-density coverage for certain ranges of intrinsic parameters is challenging. Alternative methods have been developed to calculate the detection efficiency with machine-learning based emulators (Gerosa et al., 2020; Chapman-Bird et al., 2023; Callister et al., 2024), or analytic approximants (Lorenzo-Medina & Dent, 2025). These methods achieve order-of-magnitudes speed up, and can offer the ability to evaluate detection efficiency where injections are not well suited, for example over areas where injections gave sparse coverage.

In the case where we are dominated by statistical uncertainty, the statistical uncertainties on the hyperparameters are expected to scale inversely with $\sqrt{N}$, where N is the number of observations (Mandel, 2010; Barrett et al., 2018; Gair et al., 2022). Barrett et al. (2018) suggests that this approximation requires ~ 1000 observations to be valid. With hundreds of observations, we are now in a regime to test this scaling with real events, to get a better sense of the current impact of statistical versus systematic uncertainties on population inference.

Here, we update results from Chapters 2 and 3 including the new BBH obser-

vations from GWTC-4.0. We compare these results to those with the preceding two catalogues, GWTC-2.1 and GWTC-3.0, exploring how the additional observations impact our inferred population within these modelling frameworks. We investigate both how the population hyperparameter uncertainties scale with number of observations, and how the inference results change with developments to the modelling choices in both the population distributions and the selection effects.

In Chapter 2, we asked whether big black holes can merge with small black holes, finding that GWTC-3.0 observations were consistent with high mass, unequal mass-ratio BBHs, though with large uncertainties at high masses due to lack of observations. Now, we ask whether this is still the case considering GWTC-4.0 observations. We update the marginal primary mass and spin models, informed by LVK model comparison studies (Abac et al., 2025a), and investigate the outlier status of GW190814 under these new models. Comparing these results to inference with GWTC-2.1 and GWTC-3.0 using the same models, we explore how increasing numbers of observations impact the measurement of correlations between primary masses and mass ratios.

In Chapter 3, we introduced an inference framework to compare population synthesis models to observations, inferring inputs to simulations over a continuous range using normalising flows. We show how the results of this inference change using GWTC-2.1, GWTC-3.0 and GWTC-4.0 observations. While this work implemented sophisticated interpolation techniques for the population distributions, to calculate the detection efficiency we assumed a polynomial interpolation based on the population synthesis data rather than the population normalising flows. Motivated by the growing number of GW observations, we develop a new detection efficiency interpolation to minimise systematic uncertainties through misspecification of the selection effects. We investigate the use of machine learning to emulate the detection efficiency in order to ensure consistency between our population model and the detection efficiency interpolation. We show how these new assumptions for the detection efficiency interpolation effects inference results, and how this changes with increasing numbers of observations.

This chapter is structured as follows. In Section 4.3, we detail the updates to the variable minimum mass ratio analysis, and show the results with GWTC-4.0 as well as the comparison between the results using successive GW catalogues. Section 4.4 shows the extension of the population synthesis emulation analysis to successive catalogues, with the same detection efficiency emulation as is in Chapter 3. In Section 4.5, we detail the methods underlying the machine learning detection efficiency interpolation, and show the inference results on successive catalogues, comparing to the original interpolation method. Section 4.6 presents the conclusion to this chapter, discussing the impact of increasing GW observations on our population inference both for parametric and population synthesis-based models, and the necessary work to reduce systematic biases in BBH catalogues with future data.

4.3 | Measuring a variable minimum mass ratio in successive catalogues

In this section, we apply variable minimum secondary mass models from Chapter 2 to successive catalogues of observations, GWTC-2.1 (Abbott et al., 2024), GWTC-3.0 (Abbott et al., 2023a), and GWTC-4.0 (Abac et al., 2026b), using hierarchical Bayesian inference. While Chapter 2 considers a Power Law + Peak model in primary mass, we update this model considering GWTC-4.0 observations. We first show our results with GWTC-4.0 observations, and then with successive catalogues, both to probe changes in evidence for a variable minimum secondary mass, and to understand the effect of marginal distributions modelling choices on BBH correlations.

4.3.1 | Methods

We investigate the evidence for a variable minimum secondary mass with three catalogues of observations: GWTC-2.1 (Abbott et al., 2024), GWTC-3.0 (Abbott et al., 2023a), and GWTC-4.0 (Abac et al., 2026b). GW190814 remains the most confidently measured unequal mass-ratio observation in GWTC-4.0 data, hence we continue to investigate its status as a potential population outlier with leave-one-out tests. We use a FAR threshold of less than 1 yr^{-1}, analysing a total of 47 BBH observations from GWTC-2.1, 69 from GWTC-3.0, and 152 from GWTC-4.0 when not including GW190814, consistent with previous LVK population studies apart from our exclusion of GW231123 (Abbott et al., 2021b, 2023b; Abac et al., 2025a).

For all sets of observations, we consider identical variable minimum secondary mass models to Chapter 2: Power Law, Increasing Parabola, and Relaxed Parabola models. We assume the same priors on γ, ζ, and ξ for all models, as detailed in Table 2.3.1. However, we update the primary mass and spin models to the default models from Abac et al. (2025a), as detailed below.

For the primary mass model, $p(m_1)$, we use the Broken Power Law + 2 Peaks model, the default primary mass model used in GWTC-4.0 analyses (Abac et al., 2025a). We use a uniform prior on the maximum mass, m_{max} from 100–300 $M_\odot$, which gives the minimum secondary mass models more flexibility in their evolution. We also extend the prior on the minimum primary mass, m_1^{low} down to 2 $M_\odot$, as we find that our Increasing Parabola did not well describe the set of observations including GW190814 when we required $m_1^{\mathrm{low}} > 3\,M_\odot$. Otherwise, we use the default priors on the mass hyperparameters as described in Abac et al. (2025a). We update the spin models to the Gaussian Component Spins model, which describes the spin magnitudes as a truncated Gaussian, and the spin tilts as a mixture between isotropic and truncated Gaussian sub-populations, as was used in Abac et al. (2025a). We continue to model the

redshift distribution as a power law (Fishbach et al., 2018); we consider this an appropriate model given that 99% of detectable BBHs fall below $z = 1.5^{+0.2}_{-0.2}$ in GWTC-4.0 (Abac et al., 2025a), and there is no evidence for a turnover in redshift even with GWTC-5.0 (Abac et al., 2026a).

We perform population inference using `gwpopulation_pipe` (Talbot, 2021) with nested sampler `dynesty` (Speagle, 2020), requiring the likelihood variance to be less than 1 (Talbot & Golomb, 2023). Our injection sets consists of semi-analytic injections for O1 and the second observing run (O2), and additional real injections through the first part of the third observing run (O3a), all of the third observing run (O3), and O4a for GWTC-2.1, GWTC-3.0 and GWTC-4.0, respectively, as originally produced for LVK population analyses (Abbott et al., 2021b, 2023b,c; Essick et al., 2025). We maintain a SNR threshold of 10 for the semi-analytic injections, and a FAR threshold of less than 1 yr^{-1} for the real injections (Essick, 2023; Essick et al., 2025).

4.3.2 | Results with GWTC-4.0

Figure 4.3.1 shows the evolution of m_2^{min} with the Power Law, Increasing Parabola, and Relaxed Parabola models using GWTC-4.0 observations, with and without GW190814. Overall, we continue to find that there is a preference for flat evolution of m_2^{min} with m_1, i.e. no evidence for a variable secondary minimum mass with any model.

The top rows of Table 4.3.1 show the Bayes factors of each model in GWTC-4.0, compared to the default LVK model which assumes a constant minimum secondary mass. An evolving m_2^{min} is disfavoured compared to the default model when excluding GW190814 with all three model variations, with $\log_{10}$ Bayes factors in the range -0.1 to -1.1. With GW190814, the Power Law and Increasing Parabola models are disfavoured, while the Relaxed Parabola model is favoured compared to the default model, consistent with results from Chapter 2.

With the Power Law minimum secondary mass model, we find strong preference for a flat evolution in m_2^{min}, with $\gamma > 6.6$ with GW190814 and $\gamma > 4.3$ without GW190814 at 90% credibility. When including GW190814, we require a higher γ and a lower minimum primary mass, m_{min}, with $m_{\mathrm{min}} = 2.24^{+0.30}_{-0.22}\,M_\odot$ including GW190814 and $m_{\mathrm{min}} = 3.45^{+2.37}_{-1.30}\,M_\odot$ excluding GW190814. While including GW190814 does require lower m_{min}, these measurements are consistent within 90% credible intervals, and show less discrepancy than measurements of m_{min} from Chapter 2. This indicates that GW190814 is less of an outlier to the Power Law m_2^{min} model and the updated $p(m_1)$ model in GWTC-4.0 than with the Power Law + Peak $p(m_1)$ model and GWTC-3.0 observations. This may be due to the change in model choices, or the dilution of GW190814 with respect to the other observations. Above $\sim 100\,M_\odot$ (the lower end of our prior on m_{max}), we find that uncertainties on m_2^{min} grow larger. This is partially due to the requirement that $m_2^{\mathrm{min}}(m_{\mathrm{max}}) = m_{\mathrm{max}}$ under this model's assumptions.

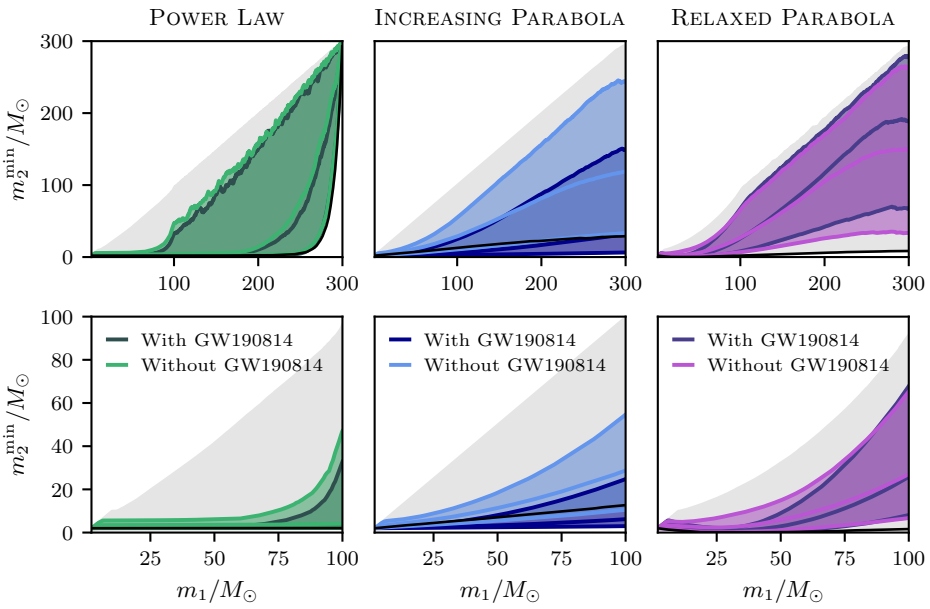


Figure 4.3.1: The evolution of $m_2^{\rm min}$ with m_1 with and without GW190814 for our three variable minimum secondary mass models and GWTC-4.0 observations. The top row shows the whole prior space up to $m_{\rm max} = 300\,M_\odot$, while the bottom row shows a zoom in on $m_1 \in [0, 100]\,M_\odot$.

With GWTC-4.0 observations, we find that the Power Law models finds little support for an increasing $m_2^{\rm min}$ with increasing m_1, until larger primary masses where there are larger uncertainties on $m_2^{\rm min}$ and we have few observations.

Under the Increasing Parabola model, we find that flat $m_2^{\rm min}$ is preferred with GWTC-4.0 observations. No evolution in $m_2^{\rm min}$ is favoured over increasing $m_2^{\rm min}$ with a $\log_{10}$ Bayes factor of 3.2 with GW190814, and 1.4 without GW190814. This indicates that GW190814 is highly informative for the evolution of $m_2^{\rm min}$, consistent with results from Chapter 2. Results with the Relaxed Parabola model also demonstrate similar behaviour to Chapter 2: there is a preference for flatter evolution of $m_2^{\rm min}$, with GW190814 causing a dipping in $m_2^{\rm min}$ around $m_1 \sim 20\,M_\odot$ to accommodate GW190814's secondary mass. Overall, there are no significant changes to the evidence for $m_2^{\rm min}$ with the Parabola $m_2^{\rm min}$ models when including GWTC-4.0 observations and default $p(m_1)$ model.

The marginal primary mass and mass ratio distributions for each model are shown in Figure 4.3.2. Compared to Chapter 2, we find more agreement between the distributions with the Power Law and Increasing Parabola models with and without GW190814. Both models find similar constraints on $m_{\rm min}$ with and without GW190814, meaning that the low-mass end of the $p(m_1)$ distributions are more similar than results from Chapter 2. Furthermore, the updated form of $p(m_1)$ models the peak at $m_1 \sim 10\,M_\odot$ as a Gaussian rather than the truncation

		$\log_{10} \mathcal{B}$	
Catalogue	Model	With GW190814	Without GW190814
GWTC-4.0	Power Law	−1.2	−0.1
	Increasing Parabola	−3.3	−1.1
	Relaxed Parabola	0.9	−0.1
GWTC-3.0	Power Law	−0.6	−0.5
	Increasing Parabola	−1.4	−1.4
	Relaxed Parabola	1.1	−0.6
GWTC-2.1	Power Law	−0.7	−0.4
	Increasing Parabola	−2.2	−1.0
	Relaxed Parabola	1.0	−0.4

Table 4.3.1: Log Bayes factors between the Power Law, Increasing Parabola, and Relaxed Parabola minimum secondary-mass models and the default model without evolution of the minimum secondary mass from Abac et al. (2025a) in three successive catalogues. All models allowing for an evolving minimum secondary mass with primary mass are disfavoured compared to the default model, with the exception of the Relaxed Parabola model when including GW190814. The uncertainty on each log Bayes factor is ±0.4, indicating no statistically significant difference in results between catalogues, apart from the Increasing Parabola result which shows some changes when including GW190814.

of a power law. The constraints on the mean and width of the peak are similar with and without GW190814 for all models, suggesting that GW190814 lies below the peak, allowing the low-mass regime to be in more agreement with and without this observation. While the mass ratio distributions when including GW190814 have increased support at lower mass ratios, they are consistent with the results excluding GW190814 to within 90% bounds. This suggests that the outlier status of GW190814 can depend in the choice of chosen form for $p(m_1)$. With more observations, not only to we get better constraints on population parameters, but we can update our models to new forms that are able to encompass observations that were previously thought of as outliers.

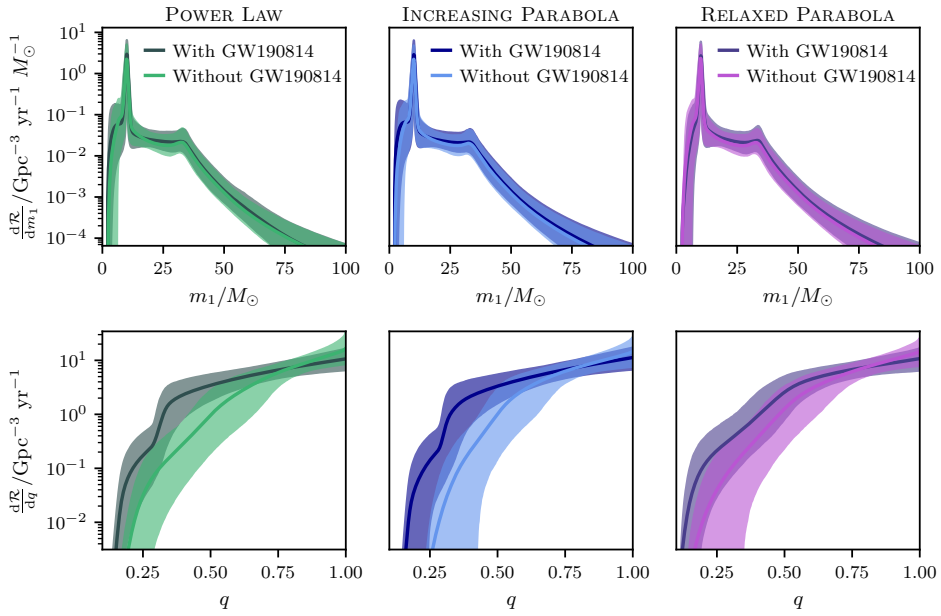


Figure 4.3.2: Inferred primary mass and mass ratio distributions with and without GW190814 for our three variable minimum secondary mass models and GWTC-4.0 observations.

4.3.3 | Comparison across catalogues

To investigate the impact of GWTC-4.0 on measurements of a variable minimum secondary mass, we compare to inference with GWTC-2.1 and GWTC-3.0 observations, using the same models. The inferred m_2^{min} for all models with and without GW190814 is shown in Figure 4.3.3 for the three catalogues. It is not clear from these results that there is any growing evidence for a m_2^{min} that varies with m_1. We find that the measured intervals on m_2^{min} are consistent from GWTC-2.1 to GWTC-4.0, with small changes in the Increasing Parabola model without GW190814 from GWTC-2.1 to GWTC-3.0. The lack of evolution between catalogues reflects the lack of observations of high-mass, unequal mass-ratio sources in GWTC-3.0 and GWTC-4.0, which would be smoking gun evidence that big black holes can merge with small black holes. It is likely that the observations which drive the evolution in m_2^{min} the most are those with confidentially-measured, unequal mass ratios, namely GW190412 and GW190814 (Abbott et al., 2020d,b). Just as Mancarella & Gerosa (2025) find that the highest and lowest mass observations are the most informative for the minimum and maximum mass of $p(m_1)$, respectively, we find that the observations with the lowest secondary masses and most unequal mass ratios are likely the most informative for inference of $m_2^{\mathrm{min}}(m_1)$. Future observations of sources with secondary masses close to our inferred $m_2^{\mathrm{min}}(m_1)$ could provide more constraints on this correlation.

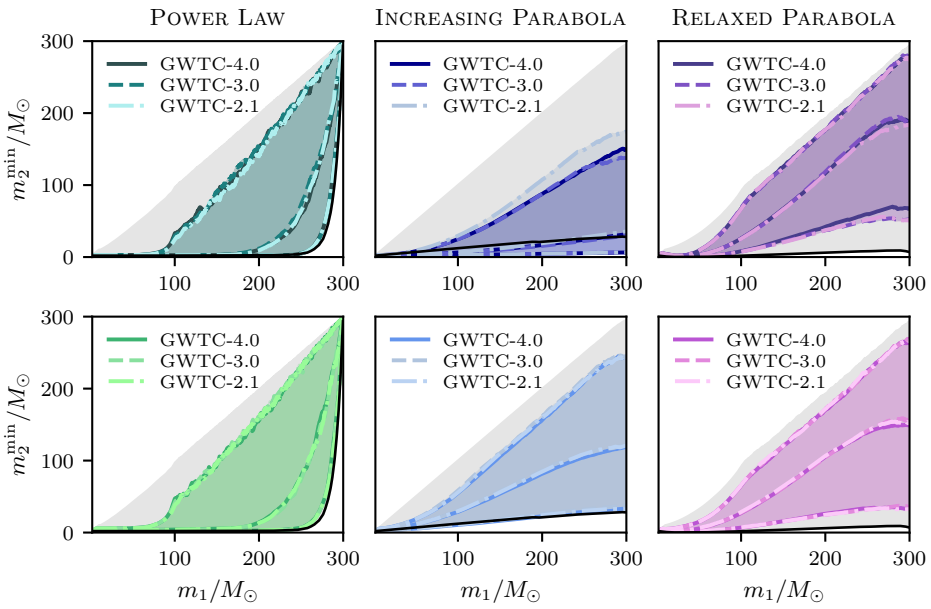


Figure 4.3.3: The evolution of $m_2^{\min}$ with m_1 for our three variable minimum secondary mass models with GWTC-2.1, GWTC-3.0, and GWTC-4.0 observations. The top row shows inference including GW190814, and the bottom row excluding GW190814. We find little difference between the constraints on $m_2^{\min}$ with successive catalogues, except the Increasing Parabola model from GWTC-2.1 to GWTC-3.0.

Directly comparing the results from Chapter 2 to those with GWTC-3.0 and the updated $p(m_1)$ model lets us understand the impact of different marginal distribution modelling choices, as all else between the results is held the same. We find no evidence for a variable $m_2^{\min}$ under the new $p(m_1)$. The uncertainties on $m_2^{\min}$ are smaller over the $m_1 \in [m_{\min}, 100]\,M_\odot$ range under the updated model, due to the increased prior on $m_{\max}$, especially for the Power Law model. This demonstrates how strong priors can impact our measurements of correlations where we lack observations. Additionally, we find that all our $p(m_1)$ and $p(q)$ distributions are consistent with and without GW190814 in GWTC-3.0 under the new marginal models, as shown in Figure 4.3.4. This confirms that even without the additional observations from GWTC-4.0, different choices in modelling assumptions can change our interpretation of outlier observations. While in Chapter 2 we showed that extra flexibility was required within our correlations to reconcile GW190814 with the bulk population, these results indicate that additional flexibility within the marginal distributions can achieve the same aim.

The Bayes factors for all model variations with and without GW190814 for GWTC-2.1 and GWTC-3.0 are included in Table 4.3.1. Without GW190814, we

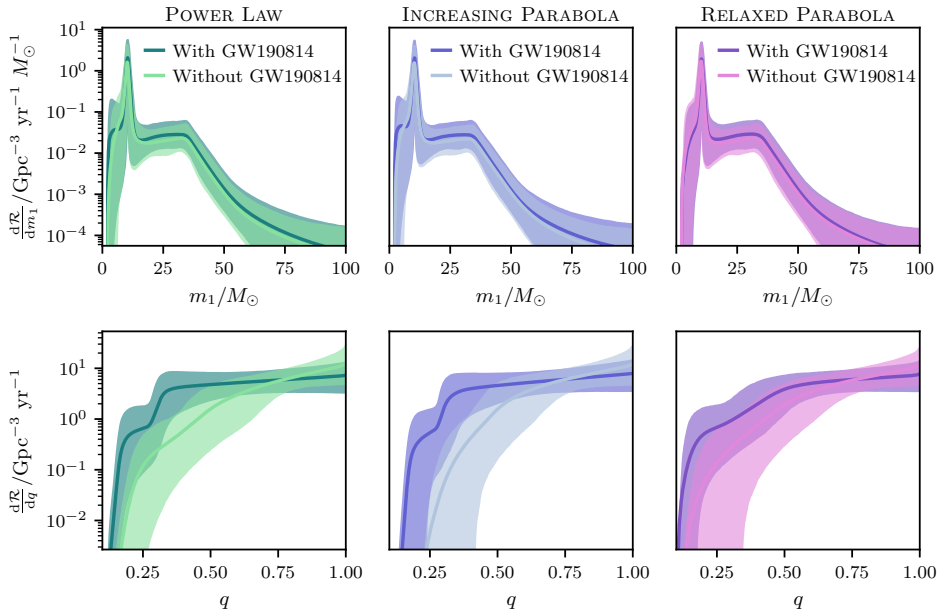


Figure 4.3.4: Inferred primary mass and mass ratio distributions found with our three variable minimum secondary mass models with and without GW190814, with GWTC-3.0 observations, and a Broken Power Law + Two Peaks model in $p(m_1)$.

find that the variable minimum secondary mass models remain more disfavoured compared to a model with no evolution of m_2^{min}, with $\log_{10}$ Bayes factors ranging from -0.5 to -1.4 in GWTC-3.0, and -0.4 to -1.0 in GWTC-2.1. The Bayes factors between catalogues are consistent to within statistical uncertainties when excluding GW190814, reinforcing that there is no change in evidence for a variable minimum secondary mass between GWTC-2.1 and GWTC-4.0. The Relaxed Parabola model remains moderately favoured across catalogues when including GW190814, suggesting that increased flexibility between m_1 and m_2^{min} is favoured across catalogues to encompass GW190814. However, the Bayes factor in favour of the Relaxed Parabola model with GWTC-3.0 observations including GW190814 is smaller than that from Chapter 2. This suggests that there is less need for additional flexibility to encompass GW190814 within the bulk population when using the Broken Power Law + Two Peaks model than using the Power Law + Peak model, reinforcing our findings from the marginal distributions themselves.

In contrast to constraints on the minimum secondary mass distributions, we find that the measurement of the marginal mass and mass ratio distributions show shrinking uncertainties with successive catalogues, as shown in Figure 4.3.5. The 90% credible intervals on $\log_{10} p(m_1)$ at the $m_1 \sim 10\,M_\odot$ peak shrink by

36% (39%), and at the $m_1 \sim 35\,M_\odot$ peak by 12% (13%) between GWTC-2.1 and GWTC-3.0 when including (excluding) GW190814, averaged over our three model variations. From GWTC-3.0 to GWTC-4.0, the 90% credible intervals on $\log_{10} p(m_1)$ at the $m_1 \sim 10\,M_\odot$ peak shrink by 26% (25%), and at the $m_1 \sim 35\,M_\odot$ peak by 4% (4%) including (excluding) GW190814, averaged over our three model variations. These results imply that we are generally making better improvements to our measurement of the $m_1 \sim 10\,M_\odot$ peak with successive catalogues than the $m_1 \sim 35\,M_\odot$ peak.

Figure 4.3.6 shows how the change in uncertainties compare to an inverse scaling with the number of observations. For $m_1 \sim 35\,M_\odot$, the uncertainties do not yet follow a scaling with $1/\sqrt{N}$, suggesting that the results are not dominated by statistical uncertainties, with the prior potentially having a larger influence in our understanding of the higher-mass BBH population. However, the evolution of the uncertainties at $m_1 \sim 10\,M_\odot$, where there is the highest rate of underlying BBHs, may be starting to approach the scaling expected with dominant statistical uncertainties. Seeing if this scaling persists with GWTC-5.0 will help us confirm the role of systematics around the low-mass peak in m_1. Furthermore, the $p(m_1)$ rates at high masses and $p(q)$ rates at unequal mass ratios decrease from GWTC-2.1 to GWTC-3.0. This indicates that more observations do not only improve the precisions of our population inference, but also the accuracy considering these population models.

4.4 | Inferring binary black hole astrophysics from population synthesis models in successive catalogues

In this section, we show the impact of the increasing number of observations on the inference of BBH astrophysics from population synthesis simulations, using the AMAZE framework as detailed in Chapter 3. We use normalising flows to infer the properties of BBHs over a continuous range of population hyperparameters for three consecutive catalogues of GWs: GWTC-2.1, GWTC-3.0 and GWTC-4.0. We show how our constraints on hyperparameters describing BBH evolution improve with the growing number of observations, leveraging continuous inference with normalising flows.

4.4.1 | Method

Our population model and statistical framework are consistent with those described in Chapter 3. We infer the black hole natal spin χ_b, common envelope efficiency α_CE, and branching ratios between five formation channels (CE, CHE, GC, NSC and SMT) using hierarchical Bayesian inference. We use the already-trained

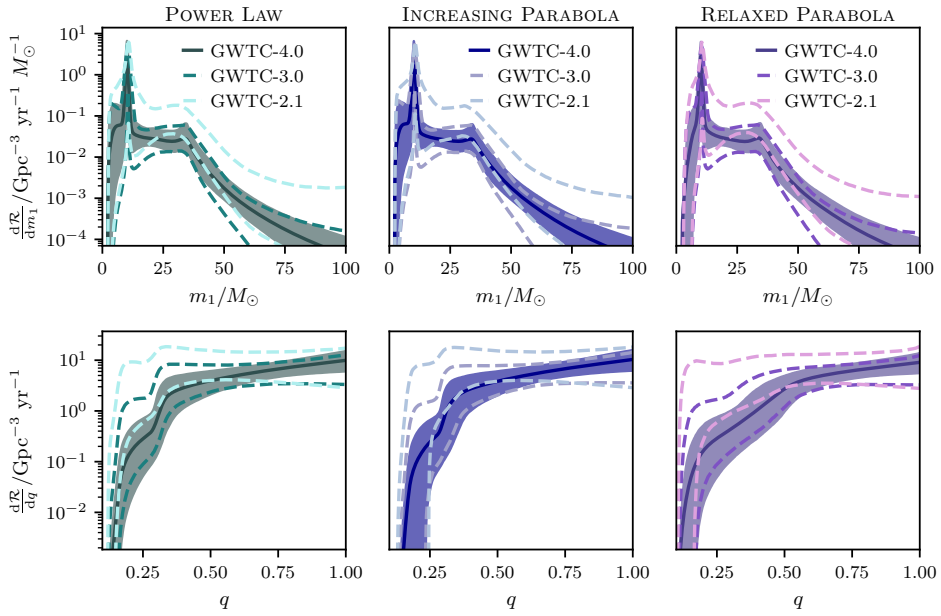


Figure 4.3.5: The change in the primary mass and mass ratio distributions found with our three variable minimum secondary mass models including GW190814, with GWTC-2.1, GWTC-3.0, and GWTC-4.0 observations. The solid band shows the PPD and 90% credible interval from GWTC-4.0, with the dashed lines show the 90% credible intervals from GWTC-2.1 and GWTC-3.0.

normalising flows to emulate the population synthesis simulations (henceforth population flows) as detailed in Section 3.3.3. These population flows are trained to emulate $p(\vec{\theta}|\vec{\lambda})$, where $\vec{\theta}$ are the BBH observables, $\{\mathcal{M}, q, \chi_{\rm eff}, z\}$, and $\vec{\lambda}$ are the astrophysical parameters varied within the formation channels, $\{\chi_{\rm b}, \alpha_{\rm CE}\}$. The hyperpriors on our population hyperparameters and the hyperlikelihood regularisation are identical to those described in Section 3.3.2. Our hyperlikelihood is therefore consistent with Equation 3.1, aside from the detection efficiency $\xi(\vec{\lambda})$.

We make several changes to the calculation of detection efficiency for our population models. These changes assume more realistic observing scenarios for our population models, taking into account actual detector sensitivities and the probability of observing across multiple observing runs.

We continue to calculate the detection efficiency via a SNR cut on the population synthesis samples to find the detection probabilities, $p_{\rm det}$. The choice of a SNR cut to find $p_{\rm det}$ rather than using injections as done in Section 4.3 allows us to directly calculate the sensitivity from the population synthesis samples rather than rely on the intermediary distribution in GW observables used to generate the injections. While the distribution of injections in GW-observable space may have a good match to the data, they may have larger discrepancies with our strongly

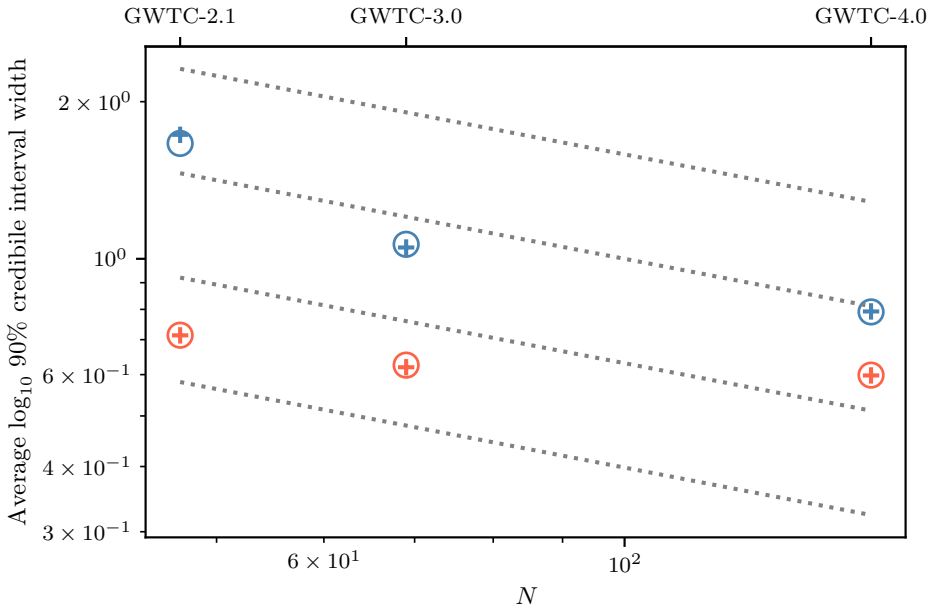


Figure 4.3.6: The number of observations N compared to the width of the 90% credible interval of $p(m_1)$ at $m_1 \sim 10\,M_\odot$ (blue) and $m_1 \sim 35\,M_\odot$ (red), averaged over our three model variations with GW190814 (circles) and without GW190814 (crosses). Comparing the measured change in uncertainties to the expected evolution for dominating statistical uncertainties ($\propto 1/\sqrt{N}$), the distribution at $m_1 \sim 35\,M_\odot$ does not appear to follow the expected proportionality, while at $m_1 \sim 10\,M_\odot$ the uncertainties may be dominated by statistical uncertainties after GWTC-3.0.

parametrised models. Using an SNR cut ensures that we have sufficient coverage in our sensitivity estimates over the full range of our models, and avoids Monte-Carlo error arising from sharp features in our population models and a lack of injections in that area of parameter space (e.g., Talbot & Golomb, 2023). Previously, the optimal SNRs were calculated assuming `midhighlatelow` detector sensitivity as a proxy for the sensitivity of the detectors in their third observing run (Zevin et al., 2021; Abbott et al., 2020a). We update the calculation of the SNR of each population synthesis sample to use representative detector ASDs for sensitivities from O1 to O4a (Abbott et al., 2020a, 2019; Acernese et al., 2018; Abbott et al., 2023a,d; Abac et al., 2025c). For each population synthesis sample, we draw a set of random uninformative extrinsic parameters, and evaluate the optimal SNR using `bilby` with a IMRPhenomXPHM waveform (Ashton et al., 2019; Pratten et al., 2021). The network SNR is calculated as the root sum of squared SNRs in each detector that contributed to the corresponding observing run.

We approximate the $p_{\rm det}$ of each sample by first finding the noise-realised SNR from a non-central χ^2 distribution with degrees of freedom equal to the number of detectors, and non-centrality parameter equal to the optimal SNR (Maggiore, 2007b; Chapman-Bird et al., 2023). The mean detection probability in a single observing run k, $p_{\rm det}(\vec{\theta}|{\rm k})$, is found by evaluating the survival function of a noise-realised SNR greater than our chosen SNR threshold, given the calculated optimal SNR (Chapman-Bird et al., 2023). We then find the detection probability averaged over each observing sensitivity weighted by the relative observing time. To find the detection efficiency at each point in $\{\chi_{\rm b}^*, \alpha_{\rm CE}^*\}$ where we have samples from population synthesis, we take a Monte Carlo sum over the average detection probabilities. Combined, we find the detection efficiencies $\tilde{\xi}(\vec{\lambda})$ from the $p_{\rm det}(\vec{\theta}|{\rm k})$ of each population synthesis sample as

$$\tilde{\xi}(\vec{\lambda}) = \frac{1}{\sum_i^N w_i} \sum_i^N w_i \sum_{\rm k} \frac{p_{\rm det}(\vec{\theta}|{\rm k})}{T_k}, \tag{4.1}$$

where N is the total number of population synthesis samples in a given simulation, w_i is the associated weight of each sample, and T_k is the observing time of each observing run k. We continue to interpolate $\tilde{\xi}(\vec{\lambda})$ with a `pchip` interpolation for continuous inference over $\vec{\lambda}$ (Fritsch & Butland, 1984).

For consistency, we choose one SNR threshold for both event selection and the detection efficiency selection threshold. We choose an SNR threshold of 10.6, requiring the network matched-filter SNR of observed signals to be greater than this in at least one matched-filter search pipeline. This threshold approximates the standard threshold of FAR$< 1\,{\rm yr}^{-1}$, while minimising the inclusion of events with FAR$> 1\,{\rm yr}^{-1}$. We exclude GW240105_151143 and GW230728_083628, which met our SNR threshold but were both coincident with excess repeating noise in the Hanford detector, and GW231231_120147 which only met our SNR threshold with a non-matched-filtered search pipeline. For GWTC-4.0, we additionally exclude GW231123_135430, GW231028_153006 and GW241230_233618 which have PE samples with $\mathcal{M} > 100\,M_\odot$, and are therefore outwith the domain of our population models. We assume that these high-mass events would return the hyperprior, with no support in any channel. Our analyses consider 32, 47, and 84 total observations from inference with GWTC-2.1, GWTC-3.0, and GWTC-4.0 catalogues, respectively, lower than the usual count for LVK analyses due to the different choice of selection cut.

To sample from the hyperposterior we continue to use `emcee` (Foreman-Mackey et al., 2013), using 250 walkers with 1000 steps and 40% burn in. We analyse a maximum of 11000 samples from each event; for events with less than 11000 samples, we instead use the total number of samples.

4.4.2 | Results

Figure 4.4.1 shows the hyperposteriors on χ_b, α_{CE}, and the underlying and detectable branching fractions of each formation channel for GWTC-2.1, GWTC-3.0, and GWTC-4.0 observations. We show the 90% confidence intervals on each hyperparameter with each catalogue in Table 4.4.1.

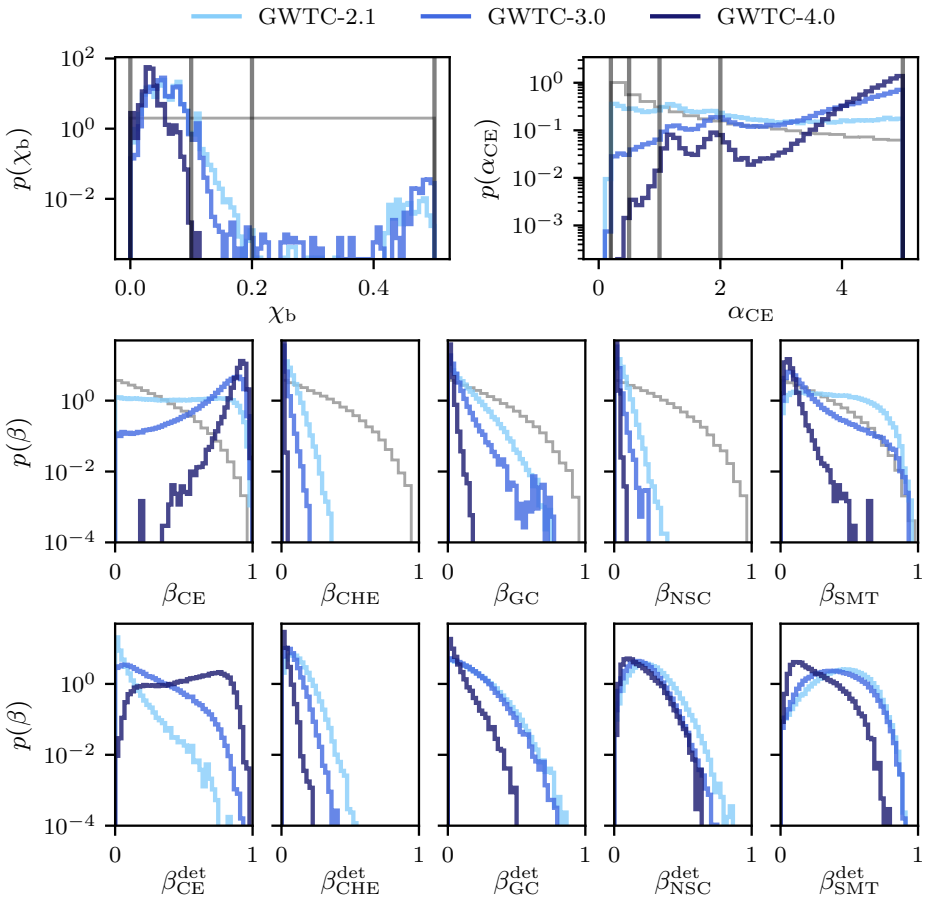


Figure 4.4.1: The inferred natal spin, common envelope efficiency, underlying and detectable branching fractions from GWTC-2.1 (light blue), GWTC-3.0 (medium blue) and GWTC-4.0 (dark blue) using continuous inference with normalising flows, and `pchip` interpolation of the detection efficiency. Grey histograms indicate the priors on χ_b, α_{CE}, and the five branching fractions, while vertical lines in the top row show the placement of the population synthesis training data. With growing catalogues, our measurement of the underlying population gets more precise, showing a preference for low natal spins, high common envelope efficiencies, and a large underlying proportion of CE mergers.

With GWTC-2.1, we favour small spins, with some support up to $\chi_b \sim 0.5$.

There is some preference away from the prior towards larger $\alpha_{\rm CE}$, though the hyperposterior is largely uninformative, with $\alpha_{\rm CE} > 0.51$ at 90% credibility. The branching fractions of the CHE, GC, and NSC channels prefer small values, with $\beta_{\rm CHE} = 0.04^{+0.09}_{-0.03}$, $\beta_{\rm GC} = 0.07^{+0.22}_{-0.06}$ and $\beta_{\rm NSC} = 0.04^{+0.07}_{-0.03}$ respectively. Meanwhile, the branching fractions of the CE and SMT channels have support away from the prior up to larger values, though are largely unconstrained.

Results with GWTC-3.0 are qualitatively similar to those with GWTC-2.1, though show more preference for larger $\alpha_{\rm CE}$ and $\beta_{\rm CE}$, and smaller $\beta_{\rm SMT}$. These measurements have larger uncertainties than those presented in Chapter 3 due to the inclusion of fewer total observations, as a consequence of the stricter selection.

We find that the GWTC-4.0 results are similar to those in Chapter 3, with observations favouring small natal spins, high common envelope efficiencies, and the majority of underlying mergers originating from the CE channel. Furthermore, $\chi_{\rm b} \gtrsim 0.1$ and $\beta_{\rm CE} \lesssim 0.5$ are ruled out at much higher confidence than with GWTC-2.1 and GWTC-3.0 catalogues. The structure around $\alpha_{\rm CE} \sim 1$–2 is somewhat more pronounced with GWTC-4.0 than results shown in Chapter 3. It is unclear where this comes from: it could be a true feature of the data, with observations favouring features at lower $\alpha_{\rm CE}$ whilst low detection efficiencies boost the high $\alpha_{\rm CE}$ contribution. Else, it could be fluctuations in the normalising flow interpolation performance around this region of hyperparameter space. However, this structure is not the data clinging to points in hyperparameter space where we have population synthesis samples, as when we consider the result in both $\chi_{\rm b}$, $\alpha_{\rm CE}$, the hyperposterior on $\chi_{\rm b}$ lies largely in between the points with training data. Understanding where this structure comes from motivates follow-up population synthesis simulations around $\alpha_{\rm CE} \sim 2$. This result reveals how the increasing number of GW observations can be used to focus work both in developing population synthesis simulations that reflect observations, and motivates the development of accurate population synthesis emulators.

We also place much more stringent constraints on CHE, GC, and NSC channels, with $\beta_{\rm CHE} = 0.003^{+0.006}_{-0.002}$, $\beta_{\rm GC} = 0.01^{+0.03}_{-0.01}$ and $\beta_{\rm NSC} = 0.01^{+0.01}_{-0.01}$. With $O(100)$ observations, we can start to put upper limits of order a few percent on specific formation channels within this modelling framework. We find that we can produce similar constraints on astrophysical parameters compared to inference using 100 simulated observations as shown in Zevin et al. (2021).

Table 4.4.1 additionally shows the percent point change in width of the credible interval between GWTC-2.1 and GWTC-4.0. The branching fraction of the CE channel shows the largest change, with the credible interval width changing by 70%. The hyperparameter $\alpha_{\rm CE}$ associated solely with the CE channel also sees a large change in measurement uncertainty, with a 54% improvement in measurement. It is likely that this change absorbed largely from SMT channel, which sees a 55% change in width. The CE and SMT channels show similar features in binary parameter space, with common features in chirp mass and spin distributions (Zevin et al., 2021; Colloms et al., 2025a; Alvarez-Lopez et al., 2025). Our

measured change in β measurements shows that catalogues with $O(100)$ allow us to distinguish between formation channels with similar features.

	GWTC-2.1	GWTC-3.0	GWTC-4.0	CI % change
χ_b	$0.05^{+0.04}_{-0.03}$	$0.05^{+0.03}_{-0.03}$	$0.03^{+0.02}_{-0.01}$	7
α_{CE}	> 0.51	> 1.61	> 3.23	54
β_{CE}	$0.45^{+0.41}_{-0.41}$	$0.83^{+0.10}_{-0.47}$	$0.92^{+0.04}_{-0.09}$	70
β_{CHE}	$0.04^{+0.09}_{-0.03}$	$0.01^{+0.03}_{-0.01}$	$0.003^{+0.006}_{-0.002}$	11
β_{GC}	$0.07^{+0.22}_{-0.06}$	$0.03^{+0.11}_{-0.03}$	$0.01^{+0.03}_{-0.01}$	25
β_{NSC}	$0.04^{+0.07}_{-0.03}$	$0.01^{+0.03}_{-0.01}$	$0.01^{+0.01}_{-0.01}$	8
β_{SMT}	$0.34^{+0.38}_{-0.28}$	$0.11^{+0.36}_{-0.08}$	$0.06^{+0.07}_{-0.03}$	55

Table 4.4.1: The measured natal spin (χ_b), common envelope efficiency (α_{CE}), and branching fractions for our five formation channels from GWTC-2.1, GWTC-3.0, and GWTC-4.0 observations. We show the median and 90% credible intervals for χ_b and branching ratios, and the 90% lower limit for α_{CE}. The final column shows the percentage change (in percent points) in the measured confidence interval between GWTC-2.1 and GWTC-4.0.

On the other hand, χ_b shows the least change over catalogues, indicating that distinguishing the location of broad features in the spin distribution is possible even with few observations (e.g., Abbott et al., 2021b; Miller et al., 2020), though not finer features within the spin distribution. For the channels which supply the most underlying mergers to the inferred population, CE and SMT, χ_b sets a lower bound on the resulting χ_{eff} distribution, with the possibility that a fraction of BBHs gain additional spin through binary interactions. Therefore, our constraints on χ_b are largely set by the lowest upper bound on χ_{eff} from individual events, if we neglect the support for negative χ_{eff} from dynamical channels. Like the inference on m_2^{min}, we expect the hyperposterior on χ_b to vary mostly only as we detect well-constrained, low spinning observations. This is consistent with studies finding that we cannot identify a sharp excess of BBHs with extremely small spins, which would indicate a subset of observations that have well-constrained small spins (e.g., Tong et al., 2022; Callister et al., 2022; Prasad & Sathyaprakash, 2026). Our constraints on a low underlying fraction of dynamically formed BBHs indicate that we are predicting a low fraction of mergers with negative χ_{eff}, as these are the only channels with support in this region. This is inconsistent with other studies which have found a non-negligible fraction of negative χ_{eff} BBHs (e.g., Vitale et al., 2022; Callister & Farr, 2024; Banagiri et al., 2025a; Abac et al., 2025a). This suggests that we may be dominated by modelling priors for negative χ_{eff}

BBHs, motivating updates to the astrophysical assumptions in the underlying population synthesis models.

4.5 | Updating detection efficiency interpolation

As the number of GW observations increases, the detection efficiency plays an increasingly important role in the hyperlikelihood used for hierarchical Bayesian inference. The detection efficiency term goes to the power of $-N_{\rm obs}$, where $N_{\rm obs}$ is the number of events, as shown in Equation 1.28.

In Chapter 3 and Section 4.4, we employed a `pchip` interpolation to evaluate the detection efficiency at points in hyperparameter space where there were no population synthesis simulations (Fritsch & Butland, 1984). At points where we have population synthesis simulations, we calculate the detection efficiency from the population synthesis samples (see Subsection 4.4.1). The `pchip` interpolation uses monotonic cubic splines to evaluate the detection efficiency in between these points, preserving monotonicity and avoiding overshooting. This assumes that the values of the detection efficiency we calculate from the population synthesis are representative of the entire range populations we consider, and that the detection efficiencies of the population flows are well represented by the population synthesis samples.

With the increasing importance of detection efficiency in GWTC-4.0, we seek to test alternative interpolation techniques and assess their relative accuracy for population inference. We test a new implementation of detection efficiency interpolation that is (a) consistent with the population flow interpolation of the binary observable distributions, $p(\vec{\theta}|\vec{\lambda})$, and (b) can capture evolution of the detection efficiency that differs from the analytic form described by `pchip` interpolation. In this section, we introduce a machine-learning interpolation of the detection efficiency that employs the population flows' interpolation of $p(\vec{\theta}|\vec{\lambda})$. We apply this method to the same three catalogues of GWs, showing how the new detection efficiency interpolation effects the inference results.

4.5.1 | Outline of methods

Our aim is to create an interpolator of the detection efficiency, $\tilde{\xi}(\vec{\lambda})$ that follows the population flow interpolation of $p(\vec{\theta}|\vec{\lambda})$. We can do this by finding the distribution of SNRs, and hence $p_{\rm det}$, from samples from the population flow at any location in $\vec{\lambda}$-space. However, calculating the network SNRs for $O(10^6)$ samples drawn from the population flow for each evaluation of the hyperlikelihood would be computationally infeasible. We therefore require a faster method to interpolate $\tilde{\xi}(\vec{\lambda})$, for which we choose to use neural networks.

Neural networks have been shown to be accurate emulators of the probability of detection from intrinsic binary parameters based on SNR cuts (Gerosa et al., 2020) and injection studies (Callister et al., 2024). Emulating p_{det} via SNRs requires fixing the SNR threshold prior to training. Emulation via injections presents challenges in generating additional training data if necessary, due to the computational expense of generating injections. Chapman-Bird et al. (2023) demonstrate a two stage emulation process, implementing a SNR emulator before analytically calculating p_{det}, and then achieving additional computational efficiency by training a separate $\bar{\xi}(\vec{\lambda})$ emulator. We base our implementation on this two stage framework, giving us the ability to target the SNR training data to areas of BBH parameter space that need more coverage considering our population models, and to avoid Monte Carlo integrals to estimate $\bar{\xi}(\vec{\lambda})$ for each likelihood evaluation.

First, we require a method to evaluate the SNRs from the outputs of the population flows, $\vec{\theta}$. There is a distribution of SNRs that can occur from one combination of $\vec{\theta}$, due to our chosen random configuration of a source's extrinsic parameters. To learn this distribution, we train a normalising flow that takes $\vec{\theta}$ as inputs and gives the SNR distribution as output. Sampling from these normalising flows can give the SNRs associated with $p(\vec{\theta}|\vec{\lambda})$ from the population flows anywhere in population hyperparameter space. We then follow the same steps used to calculate the detection efficiency as done for the SNRs calculated directly from the population synthesis samples (Subsection 4.4.1). Last, to speed up the interpolation during inference, we emulate the detection efficiency with a neural network, based on the framework from Chapman-Bird et al. (2023).

We describe the method and resulting network performance from training the normalising flows for SNR emulation (henceforth SNR flows) in section 4.5.2, and those for the detection efficiency neural network in section 4.5.3.

4.5.2 | Training normalising flows for SNR emulation

Our SNR flows take $\vec{\theta}$ as inputs and output the SNR distribution at a single detector sensitivity. We train one normalising flow for each detector sensitivity from O1 to O4a. Each network is implemented as a CouplingNF spline normalising flow (Durkan et al., 2019) using the `glasflow` (Williams et al., 2022) package, as we found these were suitably flexible for the more complex, higher-dimensional population flows (Chapter 3). We optimise our network hyperparameters with Weights and Biases (Biewald, 2020), and choose architectures with 64 neurons, 6 transforms, and 6 spline bins.

Our population models have support over a wide range of $\vec{\theta}$-space including at high masses, unequal mass ratios, and extreme spins. Furthermore, our population models have support up to redshifts of 10, further than the detector horizon for

any BBH at current detector sensitivities (Zevin et al., 2021). Therefore to evaluate the sensitivity to our population models, we require accurate SNR emulation over a large range of intrinsic BBH parameters. Following these requirements, we choose the distributions of the SNR flow training data to be able to emulate SNRs at high masses, unequal mass ratios, and high spins.

We generate 10^7 training data samples with randomly drawn intrinsic and extrinsic binary parameters, and calculate the SNRs for each sample at each observing sensitivity from O1 to O4a. We draw uniformly in chirp mass between $2\,M_\odot$ and $100\,M_\odot$. Half of the mass ratio samples are drawn from log uniform distribution between 0.01 and 1, and half from power law distribution with index of 2, in order to generate additional training data down to low mass ratios. We make a cut on the secondary masses, requiring that $m_2 > 3\,M_\odot$, to ensure our training data are consistent with the secondary black hole masses in the target population models. For spins, we draw randomly in spin magnitudes and tilts, and use this to calculate $\chi_{\rm eff}$. We draw half of the spin magnitudes uniformly between 0 and 1, and half from a power law distribution with an index of 3 to give more samples with high spins. The cosine tilts are drawn as a beta distribution which has a support across all tilt values, but peaks at -1 and 1. This allows our distribution in $\chi_{\rm eff}$ to cover a broad range of values. These choices for the distributions of mass ratio and spin parameters were empirically motivated to improve the network's training in particular areas of parameter space.

For redshifts, we drawn from a power law distribution with a power-law index of 1.2, up to a maximum redshift which differs for each detector sensitivity. We choose maximum redshifts equal to 1.5 in O1 and O2, 1.9 in O3, and 3 in O4a, as conservative estimates for the maximum redshift detectable given our SNR threshold (Abac et al., 2025a). Any samples from our population models which have redshift greater than these values are assigned zero SNR at that sensitivity, as is done with a hopeless cut for sensitivity injections (e.g., Essick et al., 2025). This redshift cut allows us to have more densely concentrated training data across the area of parameter space where we realistically expect detections.

We follow several pre-processing steps to modify the training data to improve network training. Our input parameter space is modified such that the SNR flows are trained on the detector-frame chirp mass $\mathcal{M}_{\rm det} = \mathcal{M}(1+z)$, $\log(q)$, and $\chi_{\rm eff}$, all scaled such that they lie on a unit cube. Combining these inputs with randomised extrinsic parameters, we calculate the SNR again using `bilby` and an IMRPhenomXPHM waveform, consistent with Subsection 4.4.1. The training data outputs are scaled such that the normalising flows output the distribution of $\log(\mathrm{SNR} \times D_{\rm L})$, where we calculate the luminosity distance $D_{\rm L}$ assuming cosmological parameters from Aghanim et al. (2020). Rescaling by $D_{\rm L}$ simplifies the relationship between the input and the target data, exploiting the proportionality between SNR and $1/D_{\rm L}$ (Finn & Chernoff, 1993). These steps ensure the input and output training data have mean and spread of unity, making the data easier to learn by the normalising flows.

We test the SNR flows against 300 test populations, each made up of 10000 samples generated holding the intrinsic parameters the same, while varying the extrinsic parameters as done for the training data. Three examples of the $\log_{10}$ SNR distribution predicted by the SNR flow at O4a sensitivity compared to the corresponding test populations are plotted in Figure 4.5.1. The SNR flow appears to learn the distribution of these three test cases well, and is able to identify the location and spread of the target distribution for multiple combinations of intrinsic parameters.

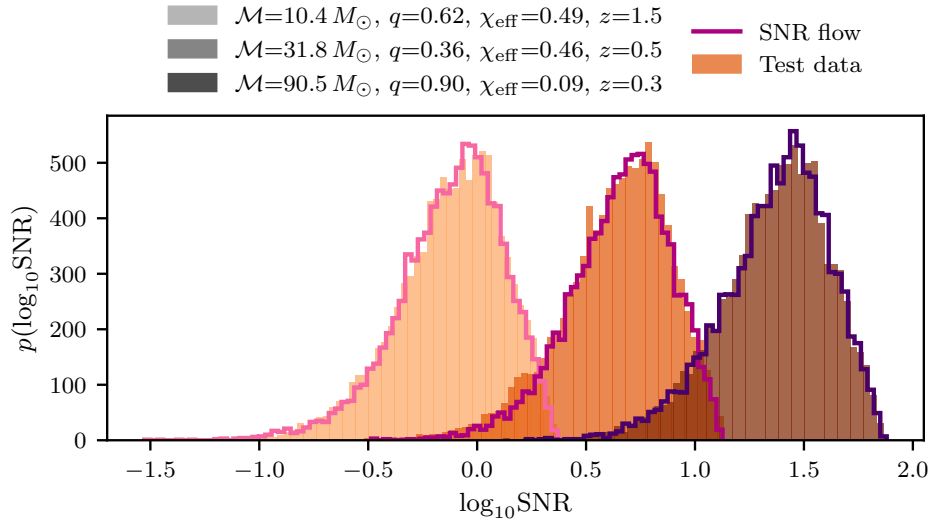


Figure 4.5.1: Distribution of $\log_{10}$ SNR for three different values of intrinsic parameters (source frame chirp mass, mass ratio, effective inspiral spin, and redshift) at O4a sensitivity. Orange filled histograms show the test data, while pink solid lines show the performance of the SNR normalising flow at the same intrinsic parameter values. The emulator can learn the distribution of $\log_{10}$ SNR effectively, closely matching the three test populations.

We quantify the performance of the trained SNR flows by finding the percentage difference between the locations of the 5%, 50% and 90% quantiles found with the normalising flows and the test populations. The distribution of these errors and their variation with mass ratio of the test populations for the O3b and O4a networks is shown in Figure 4.5.2. Above $q \sim 0.3$, most of the errors are below 10% (dashed line), showing that the SNR flows can capture the location of the peak and the tails of the target distribution to reasonable accuracy. At mass ratios closer to zero, the errors are larger, with some up to and above 100%. We do not find a correlation between the performance of the SNR flows and any other intrinsic BBH parameter. Training samples with low q exhibit a much larger range of SNRs, and test populations with low q show SNR distributions with different

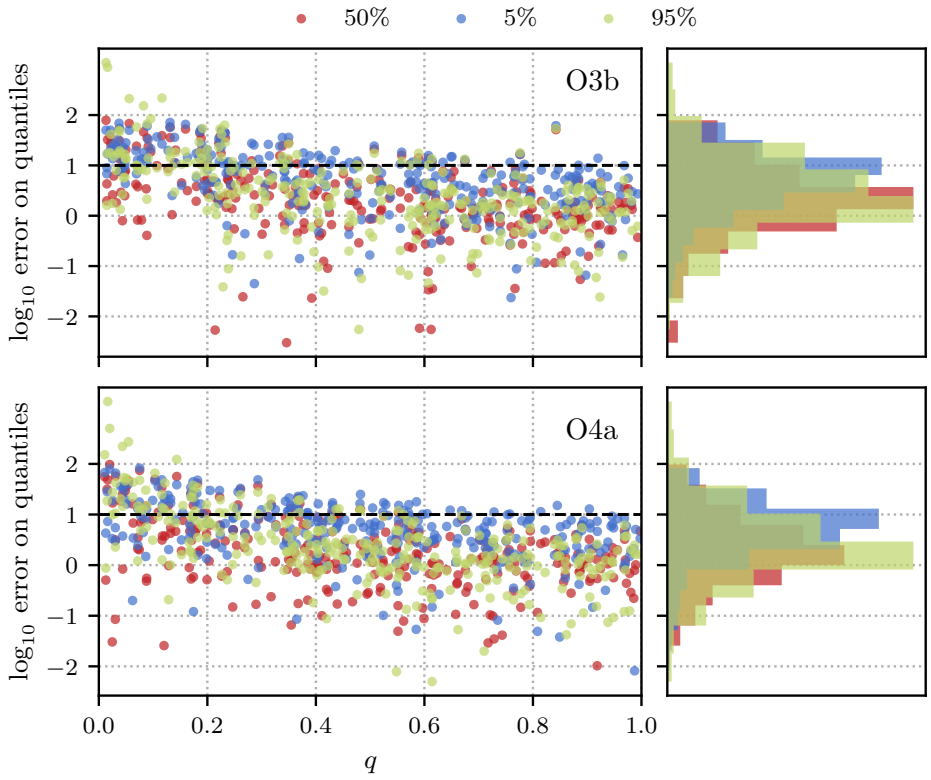


Figure 4.5.2: The $\log_{10}$ percentage errors on the median (red), 5% (blue), and 95% (green) quantiles of the SNR distributions for the SNR flows compared to a set of test populations, as a function of mass ratio (left) and the total distribution (right). Top shows the performance of the SNR flow for O3b sensitivity, and bottom for O4a sensitivity. Both normalising flows are able to capture the distribution's quantile's to within $\sim 10\%$ accuracy (dashed line) for the majority of test populations, though the errors are larger at lower mass ratios.

morphologies, which could be the reason that the networks struggle in this regime despite the increased training data. The relationship between mass ratios and SNR may be a consequence of the difficulties of simulating waveforms at unequal mass ratios (e.g., Mahapatra et al., 2026), and the heightened impact of higher order multipole moments in the unequal mass-ratio signals. Most of our population model support lies above $q > 0.3$, therefore we expect that the errors on the SNR emulation for unequal mass ratio samples will not have a significant impact on the final estimates for the detection efficiency. Below we implement these SNR flows for detection efficiency interpolation, and show tests of the impact of the SNR emulation compared to direct calculations of SNR.

4.5.3 | Detection efficiency emulation

Once we have trained normalising flows to evaluate the SNR from the population flows, we can use these to calculate the detection efficiency, and train neural networks to interpolate the detection efficiency for each formation channel. A schematic diagram of our full emulation method is shown in Figure 4.5.3. We train five neural networks, one for each formation channel, each taking the channel's population hyperparameters as input and outputting the detection efficiency for the corresponding population. Each neural network is defined as a multi-layer perceptron (MLP; Hastie et al., 2009), which learns the direct relationship between the inputs and the outputs, using 3 layers of 128 neurons. We find that we can achieve neural networks which can predict the detection efficiency to percent level accuracy without optimising the network hyperparameters.

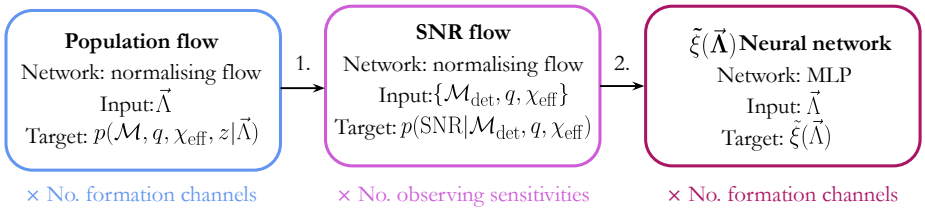


Figure 4.5.3: A schematic diagram of the different steps involved in training our machine learning detection efficiency interpolation, including the network type, input, and target of each emulator. First, we draw binary parameters from our population flows at some $\vec{\Lambda}$, and evaluate $\mathcal{M}_{\rm det}$ at step 1. We use $\mathcal{M}_{\rm det}$, q, and $\chi_{\rm eff}$ to evaluate the SNRs of each sample at each observing sensitivity using the SNR flows. At step 2, we use the SNRs to evaluate the $p_{\rm det}$ of each sample, and hence the detection efficiency $\tilde{\xi}(\vec{\Lambda})$. Once trained, this results in a network which can evaluate $\tilde{\xi}(\vec{\Lambda})$ for any input $\vec{\Lambda}$, for each formation channel.

To generate the training data for each network, we draw 10^6 samples of $\vec{\theta}$ from the channel's population flow for 10^4 combinations of population hyperparameters $\vec{\lambda}$. Evaluating the SNRs of the population flow samples from the SNR flows at each sensitivity, we calculate the $p_{\rm det}$ in each observing run and subsequent detection efficiency as done in Section 4.4.1. We train the networks for 5000 epochs, where we observe flattening of the training loss.

Samples of the detection efficiency from the trained networks across $\chi_{\rm b}$ for each channel are shown in Figure 4.5.4. We compare these samples to a test dataset, generated identically to our training data, and calculate the percent absolute error between the test data and samples from the neural networks. These errors are shown in the right panels of Figure 4.5.4. The errors are below 10% for all channels, with the majority of samples at a few percent error for all channels, and slightly larger errors in the CE channel. This indicates that the networks are effectively

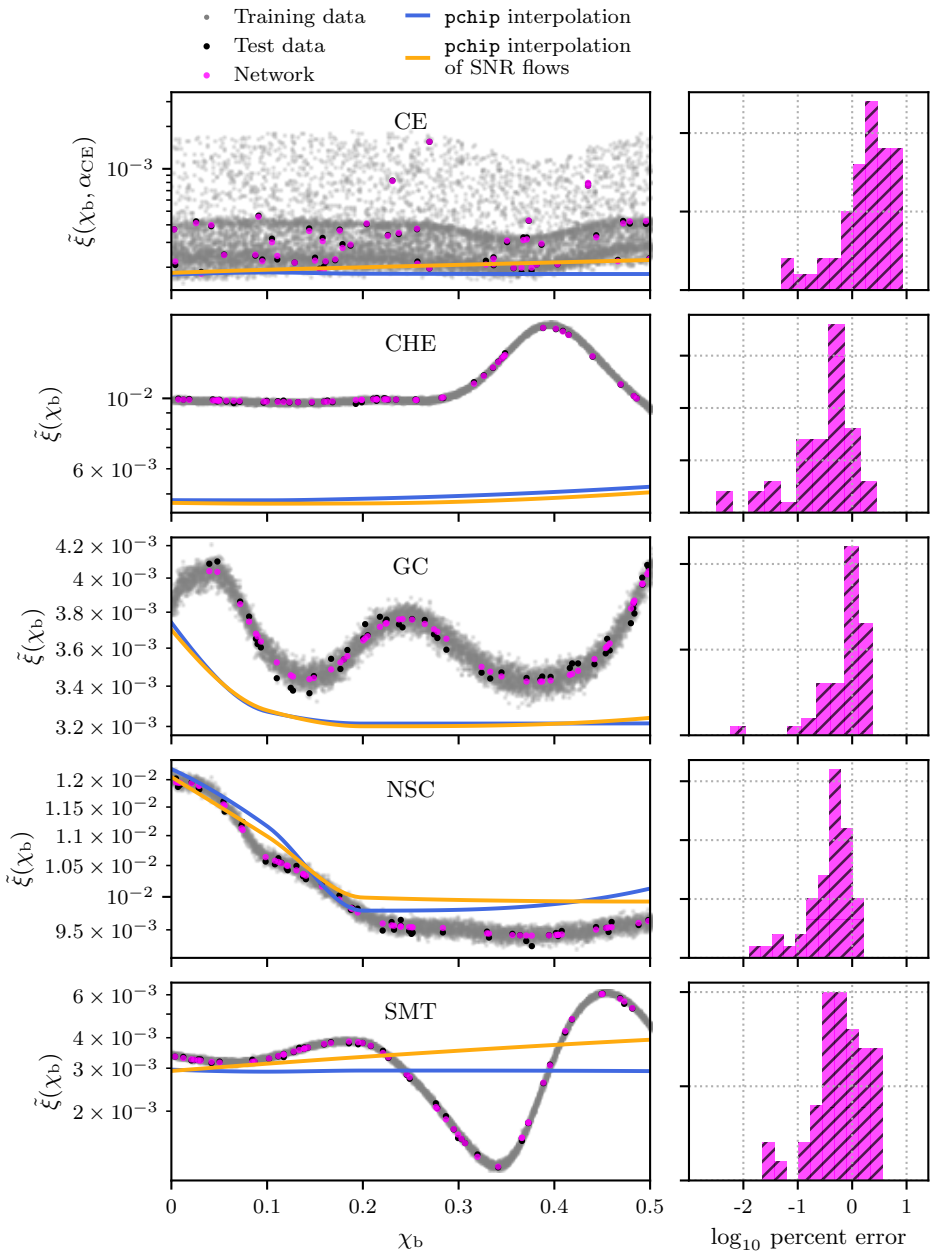


Figure 4.5.4: Left: The detection efficiency interpolation with `pchip` interpolation (blue), `pchip` interpolation of the detection efficiency calculated using the SNR flows (orange), compared to the neural network interpolation (pink) at the location of several test points (black), which trace the training data set for the neural networks (grey). Each row shows a different formation channel at O4a sensitivity. Right: The $\log_{10}$ percent error on the detection efficiency between the neural network and the test data set. While the neural network errors are low, they predict a different trend to the `pchip` interpolation for most channels.

emulating the training data. The CE channel network has to interpolate across two input hyperparameters (χ_b and α_{CE}), while for the other channels the detection efficiency only depends on χ_b. With the same number of training points and network architecture, it is expected that the CE network targetting a higher-dimensional space performs worse. With a larger training set, it is likely that the CE network errors could be reduced further.

Comparing the network performance to the whole training dataset (grey), the networks' predicted detection efficiencies lie within the spread of the training data. For the CE channel, this width is largely a result of the projection of the training data onto a single dimension. Additional spread in the training data may arise from using a finite number of samples from the population flow to estimate the detection efficiencies, or the stochastic draws from the SNR flow, as we do not average over SNR flow predictions.

Despite the accurate predictions of the neural networks with respect to the training data, when we compare the network outputs to the detection efficiencies predicted via the `pchip` interpolation (blue), it is clear that the two interpolation methods give very disparate predictions. This includes some large overall offsets for some channels, as well as additional fluctuations that are not present in the `pchip` interpolation. For example, despite the minimal change the the detection efficiency across χ_b predicted by the `pchip` interpolation for the CHE channel, the training data and network predict a rise in the detection efficiency around $\chi_b \sim 0.4$. Similar structures are apparent in the neural networks' interpolation of the GC and SMT channels' detection efficiencies. Some of these fluctuations occur in between the points where we have samples from population synthesis (e.g., SMT in between $\chi_b = 0.2$ and $\chi_b = 0.5$), while others show fluctuations towards population synthesis sample locations (e.g., GC at $\chi_b = 0.5$).

Figure 4.5.5 shows the detection efficiency interpolation in the CE channel with α_{CE}, comparing the network performance and training data to the `pchip` interpolation at a fixed $\chi_b = 0.1$. Overall, the machine learning method predicts a slightly larger detection efficiency than the `pchip` method. Furthermore, the network interpolation does not decrease as fast towards $\alpha_{CE} = 5$ compared to the `pchip` method, resulting in a difference in the detection efficiency by approximately a factor of two at $\alpha_{CE} = 5$.

To understand where these discrepancies come from between the different methods, we additionally show a hybrid interpolation method for all channels: a `pchip` interpolation of the detection efficiencies found from calculating the SNRs for the population synthesis samples using the SNR flows. This is shown in yellow for each channel in Figures 4.5.4 and 4.5.5. We find that the hybrid interpolation largely follows the original `pchip` interpolation. This indicates that the additional structure in the network detection efficiency interpolation is not due to differences between the direct calculation of the population synthesis sample SNRs and the SNR flow predictions. If the discrepancies are not due to inaccuracies in the SNR flow nor due to the neural network interpolator (the

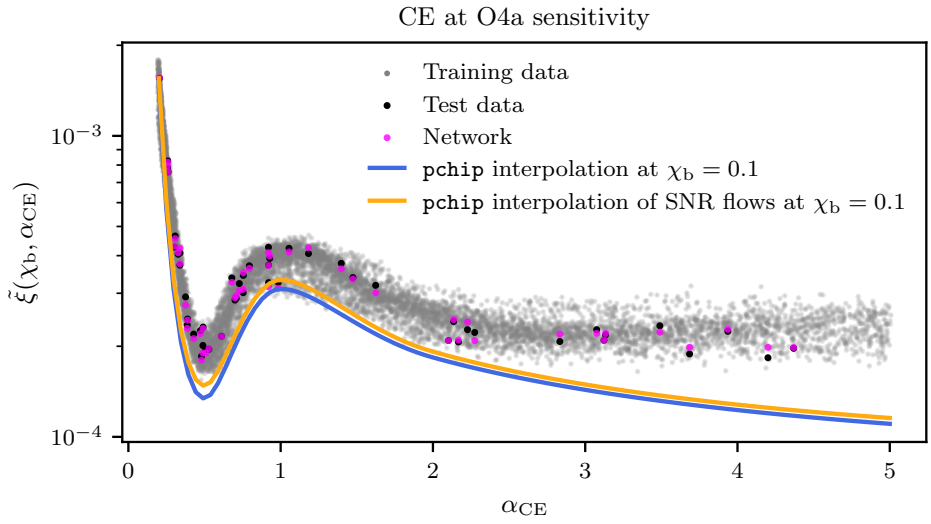


Figure 4.5.5: The common envelope efficiency versus detection efficiency for the common envelope formation channel at O4a sensitivity, found using three different interpolation methods: machine learning based interpolation (network samples in pink, test data in black, training data in grey), `pchip` interpolation of detection efficiency found with SNRs calculated directly from population synthesis samples (blue), and `pchip` interpolation of detection efficiency found with SNRs from the SNR flows. While the machine learning interpolation follows the general trend of `pchip` methods, it gives a heightened detection efficiency at high α_{CE}.

second and third components in Figure 4.5.3), this suggests that the differences arise from the use of the population flow samples rather than the population synthesis samples (the first component in Figure 4.5.3) as a basis for calculating the detection efficiency.

It is not obvious whether network or `pchip` interpolation is the correct method to use for inference. While the network interpolation is consistent with the population flow interpolation of $p(\vec{\theta}|\vec{\lambda})$, it is not always consistent with the 'true' detection efficiency calculated from the population synthesis samples. On the other hand, the `pchip` interpolation always passes through the 'true' detection efficiency by construction, though is less self-consistent in interpolation. For the CHE channel (the channel with the most discrepancies) we find that the population flows predict a higher probability of samples with much closer redshifts, and therefore higher SNRs. This suggests that differences between the tails of the population flow distributions and the population synthesis samples can lead to large differences in the detection efficiencies.

Astrophysically it is expected that changing χ_b should have little effect on

the CHE channel's BBH population, and hence minimal effect on the detection efficiency (Zevin et al., 2021). This is reflected in the `pchip` detection efficiency, which remains fairly constant across χ_b, but not the machine learning interpolation. For the other channels, it is less clear whether the structure in the detection efficiency from the machine learning interpolation is physically motivated or not. It may be that the population flows are adding additional variation to the population distributions across the population hyperparameters. We showed the effective emulation and interpolation abilities of the population flows in Chapter 3, demonstrating that any variation in the population flows must be small, or just in the tails of the distribution (and hence does not have a large effect on the population distribution's contribution to the hyperlikelihood). However, these variations may cause large changes in the detection efficiency, hence motivating future work to consider the detection efficiency during the population flow training process.

4.5.4 | Inference with machine learning detection efficiency

We implement the machine learning detection efficiency interpolation into our hierarchical inference framework, and show results using GWTC-2.1, GWTC-3.0, and GWTC-4.0 observations in Figure 4.5.6.

We find that the posterior on χ_b is similar to the results with `pchip` interpolation, with $\chi_b = 0.04^{+0.02}_{-0.02}$ in GWTC-4.0, and little change between catalogues. This is due to the measurement on χ_b mostly being driven by the lowest spin observations, rather than by the selection effects. However, the hyperposteriors on α_{CE} and the branching fractions become noticeably different to those with `pchip` detection efficiency interpolation in GWTC-4.0, with $\alpha_{CE} = 1.86^{+1.49}_{-1.24}$, $\beta_{CE} = 0.78^{+0.12}_{-0.33}$, and $\beta_{SMT} = 0.16^{+0.28}_{-0.10}$. The relative disfavouring of $\alpha_{CE} = 5$ comes from the increased detection efficiency towards higher α_{CE} in the machine-learning emulation. This decreases the hyperlikelihood in this region when we hold the observations and population models the same. The change in branching fractions could arise from the either the relative change of detection efficiencies between channels, or could be compensation for the shift in α_{CE}. Regardless, this shows that (relatively small) changes in the detection efficiency can cause significant changes to inferred astrophysical parameters. Accurately modelling the detection efficiency is therefore of crucial importance with $O(100)$ observations to avoid biases in BBH population inference.

In general, the α_{CE} parameter remains highly uncertain for BBH progenitors, due to an interplay between degeneracies with other evolutionary parameters, lack of observational constraints, and dependencies on modelling assumptions (Broekgaarden et al., 2026). Other works constraining α_{CE} with GWTC-3.0 show preference for higher values (e.g., Bouffanais et al., 2021), although different mod-

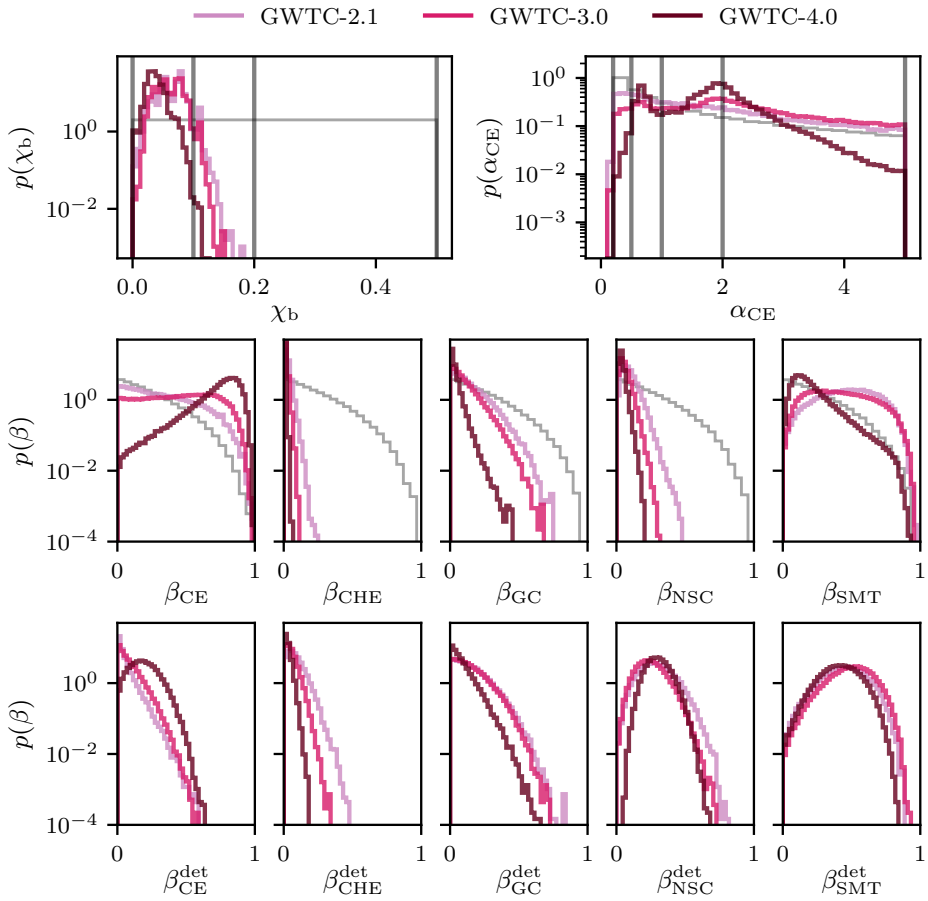


Figure 4.5.6: The inferred natal spin, common envelope efficiency, underlying and detectable branching fractions with GWTC-2.1, GWTC-3.0, and GWTC-4.0 observations, using continuous inference with normalising flows, and a machine learning interpolation of the detection efficiency. Grey histograms indicate the priors on χ_b, α_{CE}, and the five branching fractions, while vertical lines in the top row show the placement of the population synthesis training data.

elling assumptions result in different measurements (e.g., Wong et al., 2021a; Bavera et al., 2021; Gayathri et al., 2025). The constraints that we find on α_{CE} with a machine learning detection efficiency interpolation therefore are not in tension with these other works. Instead, our results emphasise the importance of exploring a range of α_{CE} values alongside multiple other varying astrophysical parameters in future comparisons of GW observations to population synthesis, due to the sensitivity of α_{CE} measurements to modelling assumptions.

Figure 4.5.7 compares the inference results using GWTC-4.0 observations between the original `pchip` method, the fully machine learning interpolation, and the hybrid `pchip` interpolation using population synthesis samples and the

SNR flows. The original and hybrid `pchip` methods predict very similar hyper-posteriors, as expected from the similar detection efficiencies. This reassures us that the SNR flows are sufficiently accurate emulators for use in calculating the detection efficiencies of these population models. However, the machine learning interpolation shows significantly different results to both `pchip` interpolation methods. This confirms that using the population flow, rather than the population synthesis samples, to interpolate the detection efficiency can make a large difference to the inference results. Evaluating the detection efficiency using accurately trained population flows is necessary for estimating detection sensitivities comparable to the true sensitivities estimated via the population synthesis samples, and therefore avoiding biases in the inference results.

4.6 | Conclusions

We extended our inference from Chapters 2 and 3 to measure the population properties of BBHs in GWTC-2.1, GWTC-3.0, and GWTC-4.0, in order to investigate how population inference results change with growing catalogues. We showed the impact of the increasing numbers of observations on the inferred population, as well as the importance of improving the accuracy of population and detection efficiency models.

Using parametric models from Chapter 2, we showed that there is still no significant evidence for a variable minimum secondary in GWTC-4.0, with no significant change in constraints on $m_2^{\min}$ from GWTC-2.1. This can explained by the correlation between m_1 and $m_2^{\min}$ being driven by observations with the most extreme secondary masses and mass ratios, rather than scaling inversely with the square root of the number of observations as expected for statistical uncertainties. We expect that our constraints on $m_2^{\min}$ will update if further extreme observations are observed in future catalogues. On the other hand, with more observations we are able to measure the marginal primary mass and mass ratio distributions with increasing precision. This scaling does not yet follow an inverse scaling with the square root of the number of observations across all masses, indicating that we do not yet have enough observations to be dominated solely by statistical uncertainty.

Furthermore, with the updated parametric model in $p(m_1)$ from Abac et al. (2025a), GW190814 lies within the bulk primary mass and mass ratio distributions for all of our model variations. More BBH observations allows for the development of better fitting parametric models for marginal distributions, allowing for previously considered outlier observations to be reconciled with the BBH population. This demonstrates how we can reduce our systematic biases arising from modelling choices, bringing us closer to the statistical uncertainty-dominated regime.

We infer the natal spin, common envelope efficiency, and branching ratios between five formation channels to GWTC-2.1, GWTC-3.0, and GWTC-4.0, using

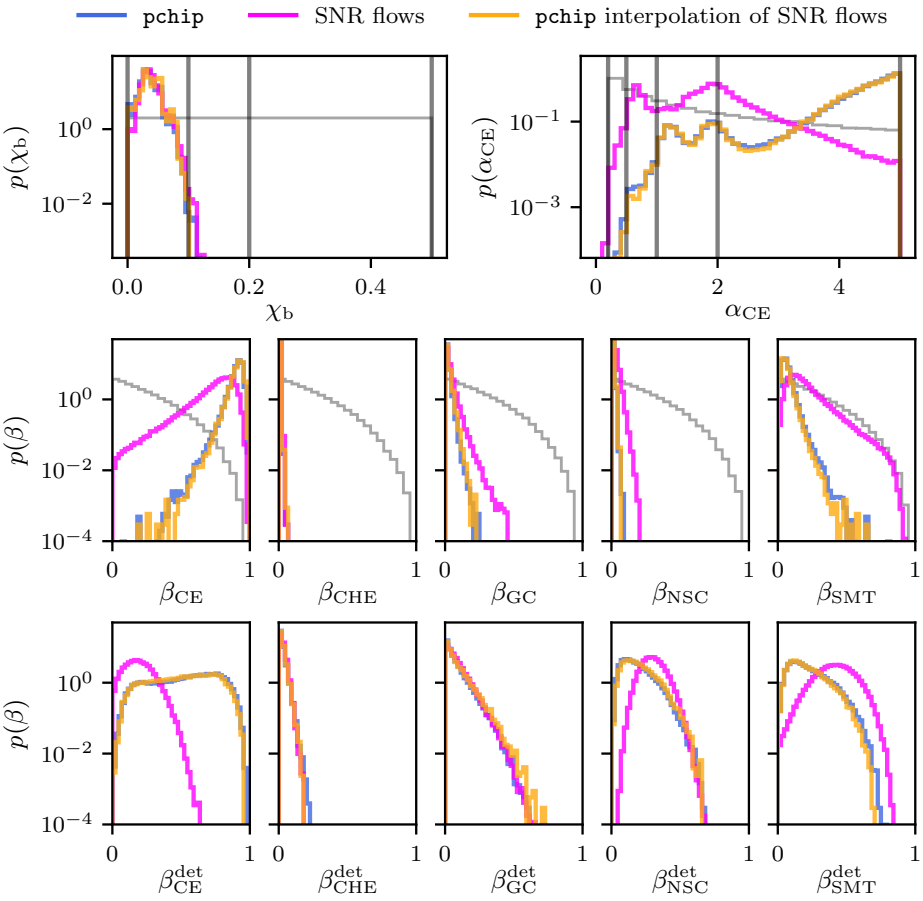


Figure 4.5.7: The inferred natal spin, common envelope efficiency, underlying and detectable branching fractions from GWTC-4.0 using continuous inference with normalising flows, comparing three different detection efficiency interpolation methods. Blue shows the original `pchip` interpolation with SNRs calculated directly from the population synthesis samples; pink shows the machine-learning method using SNR emulators to evaluate the sensitivity to population flow samples, interpolated using neural networks; and yellow show the hybrid interpolation using SNR flows to evaluate the sensitivity to the population synthesis samples, and interpolating the detection efficiencies with a `pchip` method.

the AMAZE framework from Chapter 3. We find that results with GWTC-4.0 observations are similar to those found in Chapter 3, with preference for low natal spins, high common envelope efficiencies, and majority of underlying BBHs produced by the CE channel. We find that increasing the number of observations significantly improves our measurements of formation channel branching ratios, but that the width of the uncertainty on the natal spin hyperparameter is driven by measurement uncertainties on individual observations.

We developed a new method to interpolate the detection efficiency within the AMAZE framework, aiming to improve robustness by ensuring consistency between the interpolation of the population models and the detection efficiency. We developed machine learning emulators which offered accurate predictions of SNR and detection efficiency. However, we found that small differences between the population synthesis samples and the population emulators led to discrepancies in the resulting detection efficiencies and the inference results.

Our results imply that the accuracy of normalising flows trained to emulate population synthesis simulations is critical for robust inference: even if high accuracy is not strictly required for comparable emulation to other model representations, e.g., KDEs, it is highly important for evaluating the detection efficiency. Efforts to improve the training of the population flows may improve this. It may be that there is not sufficient training data for the population flows in key areas of the parameter space. Adding follow up population synthesis simulations in regions favoured by data, or sensitive to the detection efficiency (e.g., at high $\alpha_{\rm CE}$) would likely improve the population flow training, and hence make the inference results more robust. Otherwise, accounting for training inaccuracies in the tails of the population distributions where there are few training samples, for example via regularisation or Bayesian normalising flows (e.g., Ray, 2025), could alleviate differences, especially at low redshifts where the tails strongly effect the detection efficiency. Reducing the impact of the differences between population flows and population synthesis on detection efficiencies is necessary to create a fully robust inference framework for hundreds of observations.

Regardless of the detection efficiency interpolation, evaluating the sensitivity using a FAR threshold rather than an SNR threshold would allow the use of an observational dataset that is consistent with LVK population analyses, and almost double the number of observations we can include within this analysis for GWTC-4.0. Neural network emulators trained on injections or analytic fits to the detector sensitivity may be promising emulators to estimate the sensitivity to population synthesis models based on FAR (e.g., Callister et al., 2024; Lorenzo-Medina & Dent, 2025). Implementing a FAR-based detection efficiency emulator would let us take full advantage of the growing sets of observations.

With data from GWTC-5.0 (Abac et al., 2026c,a) and soon approaching GWTC-6.0, we expect population inference to be increasingly precise, enabling more detailed insights into astrophysical constraints. With increasing precision, it is essential that we develop robust inference frameworks for accurate constraints.

We have shown how improvements to population models with multiple population inference techniques are necessary to leverage the growing number of GW observations, and gain complementary insights into BBH astrophysics. With our parametric models, more observations would let us introduce even more model flexibility to understand the underlying correlations between masses and mass ratios. Updating the physical assumptions going into the population synthesis models is a key next step to ensure trustworthy constraints for inference with strongly astrophysically informed models. This work shows that jointly with improving the population synthesis models, we need to ensure robustly trained population flows for detection efficiency interpolation to maximise insights with future GW catalogues.

5 Conclusions

5.1 | Summary of results

This thesis details new developments in BBH population inference from GW observations. Studying the population properties of BBHs from now hundreds of observations allows us to map out the unknown astrophysical evolution of the systems' progenitors. I have presented new population modelling frameworks to investigate correlations in the observed population, and to enable comparisons of observations to detailed population synthesis simulations. These frameworks enable previously intractable insights into the the BBH population, allowing us to take full advantage of the increasing number of GW observations. Applying these analyses to growing GW catalogues, I have shown how this requires careful consideration of systematics in our analyses. Below, I recap findings from each chapter in this thesis, summarising results with both modelling frameworks, as well as investigations into systematic uncertainties arising from modelling choices of the population and the selection effects.

5.1.1 | Probing binary black hole correlations with parametric models

Chapter 2 details my development of new parametric models to explore potential correlations between the black hole primary mass and mass ratios of BBHs, motivated by a desire to understand whether the largest black holes can merge with the smallest black holes. With these models, I examine the relationship between the primary mass and the minimum secondary mass of BBHs, using GW observations from GWTC-3.0 (Abbott et al., 2023a), with and without the relatively extreme mass-ratio GW190814 (Abbott et al., 2020b). These initial investigations show that a minimum secondary mass that evolves with primary mass is not preferred to a constant minimum mass, meaning that big black holes *can* merge with the smallest black holes. GWTC-3.0 observations are therefore consistent with the presence of high primary-mass, relatively extreme mass-ratio BBHs, contrary to conventional expectations that BBHs prefer to merge with similar masses (e.g., Fishbach & Holz, 2020; Abbott et al., 2021b, 2023b). These systems could arise from hierarchical mergers in dense environments or other exotic processes such as

mergers from stellar triples or formation in the disks of AGN. While uncertainties are large at high masses where we have fewer observations, these models enable novel insights into the correlation between black hole masses and mass ratios.

While GW190814 is found to be an outlier to other population models (Abbott et al., 2021b; Essick et al., 2022), our model which allows the minimum secondary mass to decrease with primary mass reconciles this observation with the bulk population. This result brings into question the established wisdom that GW190814 is an outlier to the BBH population, supporting the notion that the lowest mass black holes observed with GWs are not consistent with the presence of a lower mass gap (Ray et al., 2026a). Understanding outlier observations in light of new population models helps us understand the nature of such compact objects, and in this case, the physics of neutron star matter and supernova processes that create the smallest black holes (Fryer et al., 2012; Koehn et al., 2025). The work presented in this thesis shows how the development of new population models can shed light onto the nature of compact objects with unclear classifications, and the formation physics that leads to these objects.

5.1.2 | Inferring the origins of binary black holes with normalising flows

In Chapter 3, I detail the development of a new analysis framework to efficiently infer input parameters to population synthesis and reduce the required number of computationally expensive simulations. This method harnesses the interpolation abilities of normalising flows, a type of machine-learning model especially suitable for emulating distributions (Papamakarios et al., 2021; Kobyzev et al., 2021; Wong et al., 2020). Inference of inputs to detailed astrophysical models using a hierarchical Bayesian framework strictly requires a method to evaluate the probability distributions of population synthesis samples. The development of accurate emulators is therefore crucial for obtaining statistical constraints on uncertainties in BBH evolution. By training normalising flows to emulate a range of population synthesis models, we demonstrate the effectiveness of normalising flows for this emulation task, with the additional bonus of enabling measurements of input parameters over a continuous range, without the computational expense of a dense grid of simulations. This framework makes it possible to investigate the correlations and degeneracies between multiple physical assumptions in population synthesis simulations, which is essential to minimise biases in our understanding of BBH astrophysics.

Training the emulators on population synthesis models for five formation channels, we demonstrate inference over a continuous range of black hole natal spins, common envelope efficiencies, and the branching ratios between channels using GWTC-3.0 data. We verify the normalising flows' emulation and interpolation performance, and inference results compared to a non-interpolating method.

The normalising flows outperform the non-interpolating emulator in their match to the original population synthesis data, even on distributions which present challenging features, such as narrow spikes and multi-modalities. Furthermore, the normalising flows allow for a speed up in the total inference time of ~ 3 times compared to the non-interpolating emulator. This work establishes normalising flows as accurate and effective emulators of population synthesis models, enabling previously intractable insights into the impact of multiple astrophysical processes on the BBH population.

The resulting framework provides the most sophisticated and statistically robust method currently available to infer correlations between parameters governing BBH evolution. This analysis may be applied to other population synthesis models, formation channels, and astrophysical parameters using this framework, provided that there are sufficient training data for the normalising flows. Measurements from our method are the best diagnostic for necessary improvements to BBH and binary stellar evolutionary theories. Inference results can be used to identify interesting regions of population parameter space to run follow-up population synthesis simulations, strengthening our understanding of the population in regions most favoured by observations.

5.1.3 | The effect of a growing catalogue on binary black hole population inference

In Chapter 4, I apply each model framework to GWTC-2.1 (Abbott et al., 2024), GWTC-3.0 (Abbott et al., 2023a), and GWTC-4.0 (Abac et al., 2026b) observations, to demonstrate how astrophysical insights and systematics are impacted with the rapidly growing numbers of observations.

With our parametric minimum mass models, we find that GWTC-4.0 observations continue to show no evidence for an evolving $m_2^{\rm min}$ with m_1. This arises from the most confidently-measured unequal mass-ratio observations being the most informative for constraints on $m_2^{\rm min}$, with the majority of these events included in GWTC-2.1. With future observations, we may observe more informative, unequal mass ratio events that could lead to better constraints on $m_2^{\rm min}$. Further observations of high primary mass sources that do not exhibit unequal mass ratios should reveal shrinking uncertainties towards an evolving $m_2^{\rm min}$ at higher primary masses. While at lower primary masses ($< 100\,M_\odot$), our inferred $m_2^{\rm min}$ indicates that BBH are allowed to pair with any combination of masses with as few observations as is in GWTC-2.1, we are still largely prior-dominated at higher masses even in GWTC-4.0.

We update our models to use new marginal distributions more favoured by GWTC-4.0 observations, and show the implications for our inference results when testing the outlier status of GW190814. With the updated marginal distribution in m_1, there is less flexibility required to reconcile GW190814 with BBH

population; all of our results show more agreement in $p(m_1)$ and $p(q)$ when testing with and without GW190814 using the updated marginal model than in Chapter 2. This demonstrates how additional observations allow improvements to the features in parametric marginal models, in turn impacting the measurements of correlations and our understanding of outlier observations.

With AMAZE, I show the results of continuous inference in successive catalogues, using the trained normalising flows from Chapter 3. Our results with GWTC-4.0 show preference for low but non-zero natal spin and high common envelope efficiency, given our population synthesis assumptions. We find that our measurements of natal spin are the least affected by the growing catalogue size, indicating that there have been few well-constrained, low-spin measurements on individual observations since GWTC-2.1, which are the most informative for this hyperparameter. In contrast, the growing number of observations allows us to place increasingly precise constraints on the branching ratios between formation channels, which are informed by the combination of masses, spins, and redshifts of our observations. This highlights how this inference framework developed in this thesis enables us to pinpoint the underlying astrophysics of BBH formation by enabling inference over a continuous range of inputs to population synthesis simulations.

The growing number of observations requires an accurate description of the population models' detection efficiency to avoid systematic biases in our results. Simplifying assumptions in detection efficiency interpolation as implemented in Chapter 3 meant that it lacked consistency with the interpolation of the population distributions. I investigate a new detection efficiency interpolation based on three layers of machine learning emulation, ensuring consistency between the population distributions and sensitivity estimates. I find significant differences in the resulting detection efficiency depending on the choice of interpolation method, despite showing that our SNR and detection efficiency emulation is accurate across the majority of our target space. This indicates that small difference between the population flows and population synthesis samples can create substantial differences in detection efficiency, even though the population flows proved to be robust emulators of the population distributions as presented in Chapter 3. This follows from the compound effect of the detection efficiency in the hyperlikelihood — for a given value of population hyperparameters, we take the product over the same contribution to the detection efficiency, inflating inaccuracies faster than the population model term in the numerator of the hyperlikelihood, which has a different contribution for different observations. Implementing the new detection efficiency method into the inference framework, I find that the different detection efficiency estimates lead to significant differences in the inference results on our hyperposteriors which are more sensitive to selection effects. These differences are concerning, as they show that systematic errors in the inference framework can lead to potentially large misinterpretations of BBH astrophysics.

Overcoming systematics from detection efficiency emulation is necessary to achieve an robust inference framework allowing the comparison of population synthesis models to GW observations. The differences between the detection efficiency evaluated using the population synthesis samples versus the emulated samples motivates us to factor in the detection efficiency in training our population flows for inference with GWTC-4.0 observations and beyond. Achieving an accurate estimation of the detection efficiency will allow us to explore systematics arising from the inputs to population synthesis, and probe the degeneracies between multiple astrophysical parameters.

5.2 | Looking forward

The future of population inference with the frameworks presented in this thesis requires further developments to the population models and detection efficiency in order to achieve robust astrophysical constraints. Overcoming the impacts of these systematics will allow us to take full advantage of the abundance of GW observations, and deliver meaningful constraints on the underlying astrophysics of BBH observations. Below, I explore future work to be done to mitigate these systematic uncertainties, and potential avenues to decipher the evolution of BBH building on this work.

The results presented in Chapter 4 demonstrate how the choice of parametric population models change our understanding of BBH observations and astrophysics. Incorporating GW observations from GWTC-5.0 and beyond will allow us to probe finer features in both marginal distributions and in correlations. Considering evidence for more structures in the correlations between m_1 and q (e.g., Galaudage & Lamberts, 2025; Banagiri et al., 2025b; Ray et al., 2026b; Abac et al., 2026a), as well as in marginal distributions (e.g., Abac et al., 2025a, 2026a), future work systematically exploring the impact of different choices of population models within our $m_2^{\rm min}$ analyses will help clarify the underlying astrophysics leading to the observed BBH masses. In addition, investigating the joint mass, mass ratio and spin correlations has allowed the identification of the properties of multiple subpopulations of black holes (Banagiri et al., 2025b; Ray et al., 2026b; Cheng et al., 2026). The results presented in this thesis show how measurements of correlations may couple to modelling choices, cautioning the choice of a single default model for population correlations, and motivating the systematic investigation of parametric modelling choices for population correlations, extending beyond the correlation between masses and mass ratios , e.g., as was done for the marginal mass distribution in Abac et al. (2025a). By exploring the joint correlations between multiple BBH properties, and introducing more model flexibility in both marginal and correlated distributions, we can gain more clarity into the formation astrophysics that leads to the observed distributions of BBHs.

For robust inference using the AMAZE framework, Chapter 4 shows the

necessity of improving the fit of the population flows to the population synthesis data to deliver accurate detection efficiency estimates. Follow up population synthesis simulations could improve the training of the normalising flows in regions favoured by the data, and where there are the most discrepancies between detection efficiency interpolation methods. Without additional training data, we need to find ways to improve the training of the population flows to account for to selection effect estimates, or account for the uncertainties in the fit to the training data. Bayesian neural networks can be used to probe the uncertainties in training flow-based models and marginalise over them (Goan & Fookes, 2020; Ray, 2025). Accounting for the population flow uncertainties in the detection efficiency interpolation is a potential way forward towards implementing a robust inference framework.

Beyond the developments to the AMAZE inference framework to mitigate the systematic uncertainties from selection effects, the assumptions in the population synthesis models used to train the normalising flows are ultimately our strongest systematic. The proof-of-concept inference framework presented in this thesis allows us to vary these assumptions, both in the parameters varied within individual formation channels, and the choice of formation channels included in the analysis. Applying this work to population synthesis models that consider the joint variation between a larger set of astrophysical parameters will let us better understand how channel contributions and population parameters are correlated. This will be crucial for disentangling the astrophysical processes that lead to merging BBHs with growing GW catalogues.

BBH formation has a strong dependence on single stellar evolution, such as supernovae processes, stellar winds, and interior mixing; binary physics such as mass transfer; and the underlying metallicity and redshift evolution, but the physical details which lead to merging BBHs are uncertain due to the correlated impacts of these dependencies (e.g., Giacobbo & Mapelli, 2018; Tang et al., 2020; Iorio et al., 2022; van Son et al., 2025; Borchers et al., 2025). The current differences in prescriptions of single stellar evolution and implementation of binary physics causes large discrepancies between population synthesis predictions (Gallegos-Garcia et al., 2021; Bavera et al., 2023; Broekgaarden et al., 2026; Guerrero et al., 2026); these choices in turn impact the predictions from other formation channels which depend on the underlying binary stellar evolution processes (e.g., Rodriguez et al., 2016a; Chatterjee et al., 2017). Studies varying individual or a subset of prescriptions have revealed how the choices of binary physics changes the resulting distribution of BBH properties (e.g., Bavera et al., 2021; Levina et al., 2026; Maclean et al., 2026), but how these effects compound is less clear. Incorporating the effects of multiple free hyperparameters consistently across formation channel models within our inference framework is the necessary next step, and is made possible by the inference framework presented in this thesis. This will allow us to probe the combined effects multiple uncertain astrophysical processes on BBHs.

Analysis using AMAZE on an expanded hyperparameter space will allow insights into which astrophysical parameters are well constrained with current GW observations, and which are highly correlated with other parameters. Degeneracies between hyperparameters will help population synthesis developers understand which input parameters need to be jointly varied, while well constrained parameters may be fixed to certain values for predicting BBH astrophysics. Allowing certain input parameters to be fixed could allow a deeper exploration of more uncertain parameters, and in the longer term, help unify predictions between different population synthesis codes.

Despite progress in exploring key stages in binary stellar evolution in detailed simulations of individual systems, assumptions are still necessary to feasibly simulate populations of BBH progenitors. For example, the common envelope stage remains one of the most uncertain stages in binary stellar evolution, and future work is required to bridge the gap between approximations such as the $\alpha_{\rm CE}$ formalism and computationally expensive hydrodynamical simulations for population synthesis studies (Broekgaarden et al., 2026). Comparison of population synthesis models to GWs with AMAZE will help bridge the gap by offering constraints on what is required from common envelope evolution to produce the observed population. Furthermore, the population synthesis models considered for our proof-of-concept work make simplifying assumptions about the nature of natal spin, resulting in a narrow feature in $\chi_{\rm eff}$ in a subset of our population models. Updating the formalism for birth spins in our population synthesis models to more accurately represent inefficient angular momentum transfer (e.g., Périgois et al., 2023), supernovae kicks (e.g., Baibhav & Kalogera, 2024) and accretion during black hole formation (e.g., Issa et al., 2025) will result in more robust constraints on black hole natal spins, with the potential added bonus of easier to emulate features in $\chi_{\rm eff}$. While these assumptions serve as important caveats to measurements of astrophysical parameters, complementary developments in both theory and observations will be critical to clarify where we can place stringent constraints on BBH evolution, and where there are gaps in our knowledge.

Since the models we use to train our population flows were developed for Zevin et al. (2021), population synthesis simulations have seen significant developments (e.g., Andrews et al., 2025; Mapelli et al., 2021; Kritos et al., 2024). Applying the AMAZE framework to cutting-edge population synthesis simulations will maximise our insights into the uncertain astrophysics of BBH formation. Additionally, population synthesis models have been recently developed to account for BBHs from channels other than those considered in our proof-of-concept models (e.g. BBH formed in AGN or young clusters; McKernan et al., 2025; Ye et al., 2026). These additional channels may contribute to current BBH observations, therefore including them in inference is essential to avoid biases in astrophysical constraints. Generating new population synthesis simulations gives us the opportunity to incorporate new physical assumptions, formation channels, and calculate additional training data, which is key to improve the population flow

emulation for detection efficiency interpolation.

The extension to new population synthesis models provides the opportunity to overcome systematics in emulation accuracy and expand the hyperparameter space, though at the same time it will present new challenges. Increasing the population parameter space may result in sparser training data, and new models may present challenging features for a normalising flow to learn. Training normalising flows on the most detailed population synthesis models may serve the largest payoff, as they consider some effects of binary evolution on each object in the binary. However, this training data would be the most expensive to produce, meaning that there may be trade-off between detailed simulations and cost of training data.

Further developments in machine-learning techniques provide another avenue to ensure sufficient emulation of new target data. Flow matching has been shown to be highly suitable for learning multi-dimensional data with complex data modalities, with increased flexibility compared to discrete normalising flows (Lipman et al., 2023; Wildberger et al., 2023; Raymond et al., 2025). Combining flow matching and Bayesian neural networks will allow future work to account for emulation uncertainty in the interpolation of state-of-the-art population synthesis models. This will ensure that uncertainties arising from machine learning can be well understood, allowing us to robustly probe the uncertain astrophysics of BBHs.

Iteratively comparing population inference results across parametric, data-driven, and population synthesis-based models will be necessary to fully exploit the growing number of observations. Data-driven models will allow us to discover new features and recognise misspecified astrophysical assumptions in the inferred population. Comparisons of results using population synthesis-based models to more astrophysics-agnostic population models will indicate which features in the observed population current population synthesis theory struggles to fit. This will direct priorities of model developers of both population synthesis and parametric models, informing the next updates to physical assumptions used for cutting-edge BBH models. Delivering robust constraints on BBH astrophysics with future GW observations therefore extends beyond any one analysis framework delivering astrophysical constraints, instead requiring collaborative analysis of the observed population using a variety of novel methods.

Combining information from GWs with multi-messenger observations of massive stars which may be BBH progenitors will deliver even stronger constraints on their evolution. These observations could include x-ray binaries, stripped stars, LGRBs, and dormant black holes (e.g., Perley et al., 2016; Drout et al., 2023; Panuzzo et al., 2024; van Son et al., 2026). Data from astrometric surveys, for example using *Gaia* (Prusti et al., 2016), promise further observations of black holes within our galaxy. Spectroscopic surveys of stars, for example with the Binarity at LOw Metallicity campaign (Shenar et al., 2024; Villaseñor et al., 2025), or the future Habitable Worlds Observatory (Senchyna et al., 2025), will expand

our understanding of the physics of massive, low-metallicity stars. Studying these observations in the context of BBHs can let us understand the differences or commonalities between the evolution leading to these observations (e.g., Gallegos-Garcia et al., 2022; Kotko & Belczynski, 2024; Wang et al., 2025; Fishbach et al., 2026; Shah et al., 2026), or constrain astrophysical implications for BBHs formation (Vigna-Gómez et al., 2024). Incorporating multi-phase observations to analyses of the BBH population could be used to provide additional constraints on evolutionary parameters (e.g. Bavera et al., 2022; Wu & Fishbach, 2024). Furthermore, incorporating transient observations from high redshifts, such as LGRBs, allows us to probe the astrophysics of BBH progenitors at at redshifts beyond the current GW horizon. Each category of observations may trace different subpopulations of massive stars, at different evolutionary stages, and with different selection biases. Understanding the differences in the selection effects is a key challenge in deciphering how different sources have evolved, and therefore the physics of different massive stellar evolution pathways (Fishbach & Kalogera, 2022; Liotine et al., 2023). Even if these sources are not necessarily the direct progenitors of GW sources, they give us crucial insight into the outcomes of binary interactions, and therefore into the astrophysics determining the evolution of BBHs.

Looking forward, there are an abundance of pathways to investigate the unknown astrophysics of BBH formation. With now hundreds GW observations (Abac et al., 2026c), and soon thousands with future LVK observing runs, we will have a rich dataset with which to study in detail the lives and deaths of BBHs. Using these data for inference with increasingly detailed models of BBH progenitor populations has the potential to inform our theory of massive stellar binaries, triples, primordial black holes, stellar clusters, and AGN. Incorporating these theoretical improvements into population synthesis should in turn improve constraints from population inference. Together with complementary datasets from surveys of stellar mass black holes and their progenitors, we will be able to constrain how the uncertain astrophysics of stellar evolution creates the sources we observe with GWs. Combining multiple observational sources, different inference methods, and simulation techniques does not just provide a promising opportunity, but will be necessary to be able to look back from the stellar graveyard and understand with certainty the origins of BBHs.

Bibliography

Aarseth, S. J., & Heggie, D. C. 1976, Astronomy and Astrophysics, 53, 259

Aasi, J., et al. 2015, Classical and Quantum Gravity, 32, 074001, doi: 10.1088/0264-9381/32/7/074001

Abac, A. G., et al. 2024, Astrophysical Journal Letters, 970, L34, doi: 10.3847/2041-8213/ad5beb

—. 2025a. https://arxiv.org/abs/2508.18083

—. 2025b. https://arxiv.org/abs/2509.04348

—. 2025c, Astrophysical Journal Letters, 995, L18, doi: 10.3847/2041-8213/ae0c06

—. 2026a. https://arxiv.org/abs/2605.27226

—. 2026b, Astrophysical Journal Letters, 1004, L22, doi: 10.3847/2041-8213/ae2c74

—. 2026c. https://arxiv.org/abs/2605.27225

—. 2026d, Astrophysical Journal Letters, 1004, L21, doi: 10.3847/2041-8213/ae447b

—. 2026e. https://arxiv.org/abs/2603.19019

—. 2026f. https://arxiv.org/abs/2605.27227

Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2018, Living Reviews in Relativity, 21, 3, doi: 10.1007/s41114-018-0012-9

Abbott, B. P., et al. 2016a, Physical Review Letters, 116, 061102, doi: 10.1103/PhysRevLett.116.061102

—. 2016b, Physical Review Letters, 116, 241102, doi: 10.1103/PhysRevLett.116.241102

—. 2016c, Physical Review Letters, 116, 131103, doi: 10.1103/PhysRevLett.116.131103

—. 2016d, Physical Review D, 93, 112004, doi: 10.1103/PhysRevD.93.112004

—. 2017, Astrophysical Journal Letters, 848, L12, doi: 10.3847/2041-8213/aa91c9

—. 2019, Physical Review X, 9, 031040, doi: 10.1103/PhysRevX.9.031040

—. 2020a, Living Reviews in Relativity, 23, 3, doi: 10.1007/s41114-020-00026-9

Abbott, R., et al. 2020b, Astrophysical Journal Letters, 896, L44, doi: 10.3847/2041-8213/ab960f

—. 2020c, Physical Review Letters, 125, 101102, doi: 10.1103/PhysRevLett.125.101102

—. 2020d, Physical Review D, 102, 043015, doi: 10.1103/PhysRevD.102.043015

—. 2021a, Physical Review X, 11, 021053, doi: 10.1103/PhysRevX.11.021053

—. 2021b, Astrophysical Journal Letters, 913, L7, doi: 10.3847/2041-8213/abe949

—. 2023a, Physical Review X, 13, 041039, doi: 10.1103/PhysRevX.13.041039

—. 2023b, Physical Review X, 13, 011048, doi: 10.1103/PhysRevX.13.011048

—. 2023c, 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, Zenodo, doi: 10.5281/zenodo.5636815

—. 2023d, GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run — Data behind the figures, Zenodo, doi: 10.5281/zenodo.7997424

—. 2024, Physical Review D, 109, 022001, doi: 10.1103/PhysRevD.109.022001

Abe, H., et al. 2023, Progress of Theoretical and Experimental Physics, 2023, 10A101, doi: 10.1093/ptep/ptac093

Acernese, F., et al. 2015, Classical and Quantum Gravity, 32, 024001, doi: 10.1088/0264-9381/32/2/024001

—. 2018, Classical and Quantum Gravity, 35, 205004, doi: 10.1088/1361-6382/aadf1a

—. 2026. https://arxiv.org/abs/2603.20342

Adamcewicz, C., Galaudage, S., Lasky, P. D., & Thrane, E. 2024, Astrophysical Journal Letters, 964, L6, doi: 10.3847/2041-8213/ad2df2

Adamcewicz, C., Guttman, N., Lasky, P. D., & Thrane, E. 2025, Astrophysical Journal, 994, 261, doi: 10.3847/1538-4357/ae1370

Adamcewicz, C., & Thrane, E. 2022, Monthly Notices of the Royal Astronomical Society, 517, 3928, doi: 10.1093/mnras/stac2961

Adams, S. M., Kochanek, C. S., Gerke, J. R., Stanek, K. Z., & Dai, X. 2017, Monthly Notices of the Royal Astronomical Society, 468, 4968, doi: 10.1093/mnras/stx816

Aghanim, N., et al. 2020, Astronomy and Astrophysics, 641, A6, doi: 10.1051/0004-6361/201833910

Agrawal, P., Hurley, J., Stevenson, S., Szécsi, D., & Flynn, C. 2020, Monthly Notices of the Royal Astronomical Society, 497, 4549, doi: 10.1093/mnras/staa2264

Ajith, P., Hannam, M., Husa, S., et al. 2011, Physical Review Letters, 106, 241101, doi: 10.1103/PhysRevLett.106.241101

Allen, B. 2005, Physical Review D, 71, 062001, doi: 10.1103/PhysRevD.71.062001

Allen, B., Anderson, W. G., Brady, P. R., Brown, D. A., & Creighton, J. D. E. 2012, Physical Review D, 85, 122006, doi: 10.1103/PhysRevD.85.122006

Alléné, C., et al. 2025, Classical and Quantum Gravity, 42, 105009, doi: 10.1088/1361-6382/add234

Alvarez-Lopez, S., Heinzel, J., Mould, M., & Vitale, S. 2025. https://arxiv.org/abs/2506.20731

Andrés-Carcasona, M., Iovino, A. J., Vallejo-Pagès, E., et al. 2026. https://arxiv.org/abs/2605.15749

Andrews, J. J., Bavera, S. S., Briel, M., et al. 2025, Astrophysical Journal Supplement Series, 281, 3, doi: 10.3847/1538-4365/adfb78

Antonini, F. 2014, Astrophysical Journal, 794, 106, doi: 10.1088/0004-637X/794/2/106

Antonini, F., Gieles, M., Dosopoulou, F., & Chattopadhyay, D. 2023, Monthly Notices of the Royal Astronomical Society, 522, 466, doi: 10.1093/mnras/stad972

Antonini, F., Gieles, M., & Gualandris, A. 2019, Monthly Notices of the Royal Astronomical Society, 486, 5008, doi: 10.1093/mnras/stz1149

Antonini, F., & Rasio, F. A. 2016, Astrophysical Journal, 831, 187, doi: 10.3847/0004-637X/831/2/187

Antonini, F., Romero-Shaw, I. M., & Callister, T. 2025, Physical Review Letters, 134, 011401, doi: 10.1103/PhysRevLett.134.011401

Apostolatos, T. A., Cutler, C., Sussman, G. J., & Thorne, K. S. 1994, Physical Review D, 49, 6274, doi: 10.1103/PhysRevD.49.6274

Arun, K. G., Iyer, B. R., Sathyaprakash, B. S., & Sundararajan, P. A. 2005, Physical Review D, 71, 084008, doi: 10.1103/PhysRevD.71.084008

Ashton, G., Hübner, M., Lasky, P. D., et al. 2019, Astrophysical Journal Supplement Series, 241, 27, doi: 10.3847/1538-4365/ab06fc

Aso, Y., Michimura, Y., Somiya, K., et al. 2013, Physical Review D, 88, 043007, doi: 10.1103/PhysRevD.88.043007

Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, Annual Review of Astronomy and Astrophysics, 47, 481–522, doi: 10.1146/annurev.astro.46.060407.145222

Atallah, D., Weatherford, N. C., Trani, A. A., & Rasio, F. 2024, Astrophysical Journal, 970, 112, doi: 10.3847/1538-4357/ad5185

Baade, W., & Zwicky, F. 1934, Proceedings of the National Academy of Sciences of the United States of America, 20, 254, doi: 10.1073/pnas.20.5.254

Bagui, E., et al. 2025, Living Reviews in Relativity, 28, 1, doi: 10.1007/s41114-024-00053-w

Baibhav, V., Doctor, Z., & Kalogera, V. 2023, Astrophysical Journal, 946, 50, doi: 10.3847/1538-4357/acbf4c

Baibhav, V., & Kalogera, V. 2024. https://arxiv.org/abs/2412.03461

Bailyn, C. D., Jain, R. K., Coppi, P., & Orosz, J. A. 1998, Astrophysical Journal, 499, 367, doi: 10.1086/305614

Baker, J. G., Centrella, J., Choi, D.-I., et al. 2006, Astrophysical Journal Letters, 653, L93, doi: 10.1086/510448

Banagiri, S., Berry, C. P. L., Cabourn Davies, G. S., Tsukada, L., & Doctor, Z. 2023, Physical Review D, 108, 083043, doi: 10.1103/PhysRevD.108.083043

Banagiri, S., Callister, T. A., Adamcewicz, C., Doctor, Z., & Kalogera, V. 2025a, Astrophysical Journal, 990, 147, doi: 10.3847/1538-4357/adf4c6

Banagiri, S., Thrane, E., & Lasky, P. D. 2025b. https://arxiv.org/abs/2509.15646

Banerjee, S. 2022, Physical Review D, 105, 023004, doi: 10.1103/PhysRevD.105.023004

Barrett, J. W., Gaebel, S. M., Neijssel, C. J., et al. 2018, Monthly Notices of the Royal Astronomical Society, 477, 4685, doi: 10.1093/mnras/sty908

Barrett, J. W., Mandel, I., Neijssel, C. J., Stevenson, S., & Vigna-Gomez, A. 2016, Proceedings of the International Astronomical Union, 325, 46, doi: 10.1017/S1743921317000059

Bartos, I., Kocsis, B., Haiman, Z., & Márka, S. 2017, Astrophysical Journal, 835, 165, doi: 10.3847/1538-4357/835/2/165

Bate, M. R., & Bonnell, I. A. 1997, Monthly Notices of the Royal Astronomical Society, 285, 33, doi: 10.1093/mnras/285.1.33

Bavera, S. S., Fragos, T., Qin, Y., et al. 2020, Astronomy and Astrophysics, 635, A97, doi: 10.1051/0004-6361/201936204

Bavera, S. S., et al. 2021, Astronomy and Astrophysics, 647, A153, doi: 10.1051/0004-6361/202039804

—. 2022, Astronomy and Astrophysics, 657, L8, doi: 10.1051/0004-6361/202141979

—. 2023, Nature Astronomy, 7, 1090, doi: 10.1038/s41550-023-02018-5

Bayes, T., & Price, R. 1763, Philosophical Transactions (1683-1775), 53, 370. http://www.jstor.org/stable/105741

Belczynski, K., Bulik, T., Fryer, C. L., et al. 2010, Astrophysical Journal, 714, 1217, doi: 10.1088/0004-637X/714/2/1217

Belczynski, K., Kalogera, V., & Bulik, T. 2001, Astrophysical Journal, 572, 407, doi: 10.1086/340304

Belczynski, K., Kalogera, V., Rasio, F. A., et al. 2008, Astrophysical Journal Supplement Series, 174, 223, doi: 10.1086/521026

Belczynski, K., Wiktorowicz, G., Fryer, C., Holz, D., & Kalogera, V. 2012, Astrophysical Journal, 757, 91, doi: 10.1088/0004-637X/757/1/91

Belczynski, K., et al. 2020, Astronomy and Astrophysics, 636, A104, doi: 10.1051/0004-6361/201936528

Belczynski, K., Romagnolo, A., Olejak, A., et al. 2022, Astrophysical Journal, 925, 69, doi: 10.3847/1538-4357/ac375a

Biewald, L. 2020, Experiment Tracking with Weights and Biases. https://www.wandb.com/

Biscoveanu, S., Callister, T. A., Haster, C.-J., et al. 2022, Astrophysical Journal Letters, 932, L19, doi: 10.3847/2041-8213/ac71a8

Biscoveanu, S., Isi, M., Varma, V., & Vitale, S. 2021, Physical Review D, 104, 103018, doi: 10.1103/PhysRevD.104.103018

Blanchet, L. 2002, Living Reviews in Relativity, 5, 3, doi: 10.12942/lrr-2002-3

Blanchet, L., Damour, T., Iyer, B. R., Will, C. M., & Wiseman, A. G. 1995, Physical Review Letters, 74, 3515, doi: 10.1103/PhysRevLett.74.3515

Boccioli, L., Vartanyan, D., O'Connor, E. P., & Kasen, D. 2025, Monthly Notices of the Royal Astronomical Society, 540, 3885, doi: 10.1093/mnras/staf963

Borchers, A., Ye, C. S., & Fishbach, M. 2025, Astrophysical Journal, 987, doi: 10.3847/1538-4357/addec6

Bouffanais, Y., Mapelli, M., Santoliquido, F., et al. 2021, Monthly Notices of the Royal Astronomical Society, 505, 3873, doi: 10.1093/mnras/stab1589

Breen, P. G., & Heggie, D. C. 2013, Monthly Notices of the Royal Astronomical Society, 432, 2779–2797, doi: 10.1093/mnras/stt628

Breivik, K. 2026, in Encyclopedia of Astrophysics (First Edition), ed. I. Mandel (Oxford: Elsevier), 570–591, doi: 10.1016/B978-0-443-21439-4.00115-2

Breivik, K., Coughlin, S., Zevin, M., et al. 2020, Astrophysical Journal, 898, 71, doi: 10.3847/1538-4357/ab9d85

Briel, M. M., Eldridge, J. J., Stanway, E. R., Stevance, H. F., & Chrimes, A. A. 2022a, Monthly Notices of the Royal Astronomical Society, 514, 1315, doi: 10.1093/mnras/stac1100

Briel, M. M., Stevance, H. F., & Eldridge, J. J. 2022b, Monthly Notices of the Royal Astronomical Society, 520, 5724, doi: 10.1093/mnras/stad399

Broekgaarden, F. S., Stevenson, S., & Thrane, E. 2022, Astrophysical Journal, 938, 45, doi: 10.3847/1538-4357/ac8879

Broekgaarden, F. S., et al. 2026. https://arxiv.org/abs/2606.05322

Bruegmann, B., Gonzalez, J. A., Hannam, M., Husa, S., & Sperhake, U. 2008, Physical Review D, 77, 124047, doi: 10.1103/PhysRevD.77.124047

Bruel, T., Rodriguez, C. L., Lamberts, A., et al. 2024, Astronomy and Astrophysics, 686, A106, doi: 10.1051/0004-6361/202348716

Buonanno, A., Kidder, L. E., & Lehner, L. 2008, Physical Review D, 77, 026004, doi: 10.1103/PhysRevD.77.026004

Burrows, A., Radice, D., Vartanyan, D., et al. 2020, Monthly Notices of the Royal Astronomical Society, 491, 2715, doi: 10.1093/mnras/stz3223

Burrows, A., & Vartanyan, D. 2021, Nature, 589, 29, doi: 10.1038/s41586-020-03059-w

Callister, T. A. 2024. https://arxiv.org/abs/2410.19145

Callister, T. A., Essick, R., & Holz, D. E. 2024, Physical Review D, 110, 123041, doi: 10.1103/PhysRevD.110.123041

Callister, T. A., & Farr, W. M. 2024, Physical Review X, 14, 021005, doi: 10.1103/PhysRevX.14.021005

Callister, T. A., Haster, C.-J., Ng, K. K. Y., Vitale, S., & Farr, W. M. 2021, Astrophysical Journal Letters, 922, L5, doi: 10.3847/2041-8213/ac2ccc

Callister, T. A., Miller, S. J., Chatziioannou, K., & Farr, W. M. 2022, Astrophysical Journal Letters, 937, L13, doi: 10.3847/2041-8213/ac847e

Campanelli, M., Lousto, C. O., Zlochower, Y., & Merritt, D. 2007, Astrophysical Journal Letters, 659, L5, doi: 10.1086/516712

Capote, E., et al. 2025, Physical Review D, 111, 062002, doi: 10.1103/PhysRevD.111.062002

Carr, B., & Kuhnel, F. 2020, Annual Review of Nuclear and Particle Science, 70, 355, doi: 10.1146/annurev-nucl-050520-125911

Chan, C., Müller, B., & Heger, A. 2020, Monthly Notices of the Royal Astronomical Society, 495, 3751, doi: 10.1093/mnras/staa1431

Chan, C., Müller, B., Heger, A., Pakmor, R., & Springel, V. 2018, Astrophysical Journal Letters, 852, L19, doi: 10.3847/2041-8213/aaa28c

Chapman-Bird, C. E. A., Berry, C. P. L., & Woan, G. 2023, Monthly Notices of the Royal Astronomical Society, 522, 6043, doi: 10.1093/mnras/stad1397

Chatterjee, S., Rodriguez, C. L., & Rasio, F. A. 2017, Astrophys. J., 834, 68, doi: 10.3847/1538-4357/834/1/68

Chen, Z., Zhang, J., Lu, Y., Liu, J., & Zeng, C. 2025, Monthly Notices of the Royal Astronomical Society: Letters, 544, L89, doi: 10.1093/mnrasl/slaf105

Cheng, A. Q., Toubiana, A., Biscoveanu, S., & Gair, J. 2026. https://arxiv.org/abs/2605.25980

Cheng, A. Q., Zevin, M., & Vitale, S. 2023, Astrophysical Journal, 955, 127, doi: 10.3847/1538-4357/aced98

Cheung, D. H. T., Wong, K. W. K., Hannuksela, O. A., Li, T. G. F., & Ho, S. 2022, Physical Review D, 106, 083014, doi: 10.1103/PhysRevD.106.083014

Chruslinska, M., Nelemans, G., & Belczynski, K. 2019, Monthly Notices of the Royal Astronomical Society, 482, 5012, doi: 10.1093/mnras/sty3087

Colloms, S., Berry, C. P. L., Veitch, J., & Zevin, M. 2025a, Astrophysical Journal, 988, 189, doi: 10.3847/1538-4357/ade546

—. 2026, Journal of Physics: Conference Series, 3177, 012077, doi: 10.1088/1742-6596/3177/1/012077

Colloms, S., Doctor, Z., & Berry, C. P. L. 2025b, Astrophysical Journal, 995, 123, doi: 10.3847/1538-4357/ae1f09

Costa, G., Bressan, A., Mapelli, M., et al. 2020, Monthly Notices of the Royal Astronomical Society, 501, 4514–4533, doi: 10.1093/mnras/staa3916

Creighton, J., & Anderson, W. 2011, Gravitational-Wave Detectors (John Wiley & Sons, Ltd), 197–267, doi: 10.1002/9783527636037.ch6

Cuoco, E., Cavaglià, M., Heng, I. S., Keitel, D., & Messenger, C. 2024, Living Reviews in Relativity, 28, 2, doi: 10.1007/s41114-024-00055-8

Cutler, C., & Flanagan, E. E. 1994, Physical Review D, 49, 2658, doi: 10.1103/PhysRevD.49.2658

Dal Canton, T., Nitz, A. H., Gadre, B., et al. 2021, Astrophysical Journal, 923, 254, doi: 10.3847/1538-4357/ac2f9a

Davis, D., et al. 2021, Classical and Quantum Gravity, 38, 135014, doi: 10.1088/1361-6382/abfd85

Davis, P. J., Kolb, U., & Knigge, C. 2011, Monthly Notices of the Royal Astronomical Society, 419, 287–303, doi: 10.1111/j.1365-2966.2011.19690.x

De Donder, E., & Vanbeveren, D. 2004, New Astronomy Reviews, 48, 861, doi: 10.1016/j.newar.2004.07.001

De Marco, O., & Izzard, R. G. 2017, Publications of the Astronomical Society of Australia, 34, doi: 10.1017/pasa.2016.52

de Mink, S. E., Cantiello, M., Langer, N., et al. 2008, Proceedings of the International Astronomical Union, 252, 365, doi: 10.1017/S1743921308023223

de Mink, S. E., Langer, N., Izzard, R. G., Sana, H., & de Koter, A. 2013, Astrophysical Journal, 764, 166, doi: 10.1088/0004-637X/764/2/166

Delfavero, V., O'Shaughnessy, R., Belczynski, K., Drozda, P., & Wysocki, D. 2023, Physical Review D, 108, 043023, doi: 10.1103/PhysRevD.108.043023

Dewancker, I., McCourt, M., & Clark, S. 2016. https://arxiv.org/abs/1612.04858

Dittmann, A. J., Dempsey, A. M., & Li, H. 2024, Astrophysical Journal, 964, 61, doi: 10.3847/1538-4357/ad23ce

Dominik, M., Belczynski, K., Fryer, C., et al. 2012, Astrophysical Journal, 759, 52, doi: 10.1088/0004-637X/759/1/52

Dorozsmai, A., & Toonen, S. 2024, Monthly Notices of the Royal Astronomical Society, 530, 3706, doi: 10.1093/mnras/stae152

Drout, M. R., Götberg, Y., Ludwig, B. A., et al. 2023, Science, 382, 1287, doi: 10.1126/science.ade4970

du Buisson, L., Marchant, P., Podsiadlowski, P., et al. 2020, Monthly Notices of the Royal Astronomical Society, 499, 5941, doi: 10.1093/mnras/staa3225

Duez, M. D., & Zlochower, Y. 2019, Reports on Progress in Physics, 82, 016902, doi: 10.1088/1361-6633/aadb16

Durkan, C., Bekasov, A., Murray, I., & Papamakarios, G. 2019, in Advances in Neural Information Processing Systems, ed. H. Wallach, H. Larochelle, A. Beygelzimer, F. dAlché-Buc, E. Fox, & R. Garnett, Vol. 32 (Curran Associates, Inc.). https://arxiv.org/abs/1906.04032

Edelman, B., Farr, B., & Doctor, Z. 2023, Astrophysical Journal, 946, 16, doi: 10.3847/1538-4357/acb5ed

Einstein, A. 1918, Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften, 154

Ekström, S. 2026, in Encyclopedia of Astrophysics (First Edition), ed. I. Mandel (Elsevier), 252–263, doi: 10.1016/B978-0-443-21439-4.00089-4

Eldridge, J. J. 2020, Astronomy & Geophysics, 61, 2.24, doi: 10.1093/astrogeo/ataa029

Eldridge, J. J., & Stanway, E. R. 2022, Annual Review of Astronomy and Astrophysics, 60, 455, doi: 10.1146/annurev-astro-052920-100646

Eldridge, J. J., Stanway, E. R., Xiao, L., et al. 2017, Publications of the Astronomical Society of Australia, 34, e058, doi: 10.1017/pasa.2017.51

Essick, R. 2023, Physical Review D, 108, 043011, doi: 10.1103/PhysRevD.108.043011

Essick, R., Farah, A., Galaudage, S., et al. 2022, Astrophysical Journal, 926, 34, doi: 10.3847/1538-4357/ac3978

Essick, R., et al. 2025, Physical Review D, 112, 102001, doi: 10.1103/44x3-hv3y

Evans, M., et al. 2021. https://arxiv.org/abs/2109.09882

Fairhurst, S. 2009, New Journal of Physics, 11, 123006, doi: 10.1088/1367-2630/11/12/123006

Farag, E., Renzo, M., Farmer, R., Chidester, M. T., & Timmes, F. X. 2022, Astrophysical Journal, 937, 112, doi: 10.3847/1538-4357/ac8b83

Farah, A. M., Fishbach, M., Essick, R., Holz, D. E., & Galaudage, S. 2022, Astrophysical Journal, 931, 108, doi: 10.3847/1538-4357/ac5f03

Farah, A. M., Fishbach, M., & Holz, D. E. 2024, Astrophysical Journal, 962, 69, doi: 10.3847/1538-4357/ad0558

Farah, A. M., Vijaykumar, A., & Fishbach, M. 2026, Astrophysical Journal Letters, 1001, L40, doi: 10.3847/2041-8213/ae4e19

Farmer, R., Renzo, M., de Mink, S. E., Marchant, P., & Justham, S. 2019, Astrophysical Journal, 887, 53, doi: 10.3847/1538-4357/ab518b

Farr, W. M., Gair, J. R., Mandel, I., & Cutler, C. 2015, Physical Review D, 91, 023005, doi: 10.1103/PhysRevD.91.023005

Farr, W. M., Sravan, N., Cantrell, A., et al. 2011, Astrophysical Journal, 741, 103, doi: 10.1088/0004-637X/741/2/103

Fichtner, Y. A., Grassitelli, L., Romano-Díaz, E., & Porciani, C. 2022, Monthly Notices of the Royal Astronomical Society, 512, 4573–4592, doi: 10.1093/mnras/stac785

Finn, L. S., & Chernoff, D. F. 1993, Physical Review D, 47, 2198, doi: 10.1103/PhysRevD.47.2198

Fishbach, M., Breivik, K., Willcox, R., & van Son, L. A. C. 2026, Astrophysics and Space Science, 371, 36, doi: 10.1007/s10509-026-04554-8

Fishbach, M., & Fragione, G. 2023, Monthly Notices of the Royal Astronomical Society, 522, 5546, doi: 10.1093/mnras/stad1364

Fishbach, M., & Holz, D. E. 2017, Astrophysical Journal Letters, 851, L25, doi: 10.3847/2041-8213/aa9bf6

—. 2020, Astrophysical Journal Letters, 891, L27, doi: 10.3847/2041-8213/ab7247

Fishbach, M., Holz, D. E., & Farr, W. M. 2018, Astrophysical Journal Letters, 863, L41, doi: 10.3847/2041-8213/aad800

Fishbach, M., & Kalogera, V. 2022, Astrophysical Journal Letters, 929, L26, doi: 10.3847/2041-8213/ac64a5

Fitchett, M. J. 1983, Monthly Notices of the Royal Astronomical Society, 203:3, doi: 10.1093/mnras/203.4.1049

Flanagan, E. E., & Hughes, S. A. 2005, New Journal of Physics, 7, 204, doi: 10.1088/1367-2630/7/1/204

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, Publications of the Astronomical Society of Australia, 125, 306, doi: 10.1086/670067

Fowler, W. A., & Hoyle, F. 1964, Astrophysical Journal Supplement Series, 9, 201, doi: 10.1086/190103

Fragos, T., Andrews, J. J., Ramirez-Ruiz, E., et al. 2019, Astrophysical Journal Letters, 883, L45, doi: 10.3847/2041-8213/ab40d1

Fragos, T., Tremmel, M., Rantsiou, E., & Belczynski, K. 2010, Astrophysical Journal Letters, 719, L79, doi: 10.1088/2041-8205/719/1/L79

Fragos, T., Andrews, J. J., Bavera, S. S., et al. 2022, Astrophysical Journal Supplement Series, doi: 10.3847/1538-4365/ac90c1

Franciolini, G., & Pani, P. 2022, Physical Review D, 105, 123024, doi: 10.1103/PhysRevD.105.123024

Fritsch, F. N., & Butland, J. 1984, SIAM Journal on Scientific and Statistical Computing, 5, 300, doi: 10.1137/0905021

Fryer, C. L. 1999, Astrophysical Journal, 522, 413, doi: 10.1086/307647

Fryer, C. L., Belczynski, K., Wiktorowicz, G., et al. 2012, Astrophysical Journal, 749, 91, doi: 10.1088/0004-637X/749/1/91

Fryer, C. L., Olejak, A., & Belczynski, K. 2022, Astrophysical Journal, 931, 94, doi: 10.3847/1538-4357/ac6ac9

Fuller, J., & Lu, W. 2022, Monthly Notices of the Royal Astronomical Society, 511, 3951, doi: 10.1093/mnras/stac317

Fuller, J., & Ma, L. 2019, Astrophysical Journal Letters, 881, L1, doi: 10.3847/2041-8213/ab339b

Gaebel, S. M., Veitch, J., Dent, T., & Farr, W. M. 2019, Monthly Notices of the Royal Astronomical Society, 484, 4008, doi: 10.1093/mnras/stz225

Gair, J. R., Antonelli, A., & Barbieri, R. 2022, Monthly Notices of the Royal Astronomical Society, 519, 2736, doi: 10.1093/mnras/stac3560

Galaudage, S., & Lamberts, A. 2025, Astronomy and Astrophysics, 694, A186, doi: 10.1051/0004-6361/202451654

Gallegos-Garcia, M., Berry, C. P. L., Marchant, P., & Kalogera, V. 2021, Astrophysical Journal, 922, 110, doi: 10.3847/1538-4357/ac2610

Gallegos-Garcia, M., Fishbach, M., Kalogera, V., Berry, C. P. L., & Doctor, Z. 2022, Astrophysical Journal Letters, 938, L19, doi: 10.3847/2041-8213/ac96ef

Gallegos-Garcia, M., Jacquemin-Ide, J., & Kalogera, V. 2024, Astrophysical Journal, 973, 168, doi: 10.3847/1538-4357/ad6769

Gayathri, V., et al. 2025. https://arxiv.org/abs/2509.09647

Gerosa, D., Berti, E., O'Shaughnessy, R., et al. 2018, Physical Review D, 98, 084036, doi: 10.1103/PhysRevD.98.084036

Gerosa, D., Pratten, G., & Vecchio, A. 2020, Physical Review D, 102, 103020, doi: 10.1103/PhysRevD.102.103020

Giacobbo, N., & Mapelli, M. 2018, Monthly Notices of the Royal Astronomical Society, 480, 2011, doi: 10.1093/mnras/sty1999

Giacobbo, N., Mapelli, M., & Spera, M. 2018, Monthly Notices of the Royal Astronomical Society, 474, 2959, doi: 10.1093/mnras/stx2933

Ginat, Y. B., Antonini, F., Flanagan, E., & Gieles, M. 2026. https://arxiv.org/abs/2604.07456

Goan, E., & Fookes, C. 2020, Bayesian Neural Networks: An Introduction and Survey (Springer International Publishing), 45–87, doi: 10.1007/978-3-030-42553-1_3

Godfrey, J., Edelman, B., & Farr, B. 2023. https://arxiv.org/abs/2304.01288

Golomb, J., Isi, M., & Farr, W. M. 2024, Astrophysical Journal, 976, 121, doi: 10.3847/1538-4357/ad8572

Golomb, J., & Talbot, C. 2022, Physical Review D, 108, 103009, doi: 10.1103/PhysRevD.108.103009

Gonzalez, J. A., Sperhake, U., Bruegmann, B., Hannam, M., & Husa, S. 2007, Physical Review Letters, 98, 091101, doi: 10.1103/PhysRevLett.98.091101

Goodman, J., & Tan, J. C. 2004, Astrophysical Journal, 608, 108, doi: 10.1086/386360

Gottlieb, O., Jacquemin-Ide, J., Lowell, B., Tchekhovskoy, A., & Ramirez-Ruiz, E. 2023, Astrophysical Journal Letters, 952, L32, doi: 10.3847/2041-8213/ace779

Graham, M. J., et al. 2020, Physical Review Letters, 124, 251102, doi: 10.1103/PhysRevLett.124.251102

Green, A. M., & Kavanagh, B. J. 2021, Journal of Physics G, 48, 043001, doi: 10.1088/1361-6471/abc534

Guerrero, A. G., Zevin, M., Maclean, D. B., et al. 2026. https://arxiv.org/abs/2608.21609

Guglielmetti, F., Fischer, R., & Dose, V. 2009, Monthly Notices of the Royal Astronomical Society, 396, 165, doi: 10.1111/j.1365-2966.2009.14739.x

Hamers, A. S., Rantala, A., Neunteufel, P., Preece, H., & Vynatheya, P. 2021, Monthly Notices of the Royal Astronomical Society, 502, 4479, doi: 10.1093/mnras/stab287

Hannam, M. 2014, General Relativity and Gravitation, 46, 1767, doi: 10.1007/s10714-014-1767-2

Harris, W. E. 1996, Astronomical Journal, 112, 1487, doi: 10.1086/118116

Hastie, T., Tibshirani, R., & Friedman, J. 2009, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer Series in Statistics (Springer), doi: 10.1007/978-0-387-84858-7

Hastings, W. K. 1970, Biometrika, 57, 97. http://www.jstor.org/stable/2334940

Heger, A., Langer, N., & Woosley, S. E. 2000, Astrophysical Journal, 528, 368, doi: 10.1086/308158

Heggie, D. C. 1975, Monthly Notices of the Royal Astronomical Society, 173, 729, doi: 10.1093/mnras/173.3.729

Heinzel, J., Biscoveanu, S., & Vitale, S. 2024, Physical Review D, 109, 103006, doi: 10.1103/PhysRevD.109.103006

Heinzel, J., & Vitale, S. 2025. https://arxiv.org/abs/2509.07221

Hénon, M. 1971a, Astrophysics and Space Science, 13, 284. https://api.semanticscholar.org/CorpusID:119724558

—. 1971b, Astrophysics and Space Science, 14, 151. https://api.semanticscholar.org/CorpusID:115854730

Hirai, R., & Mandel, I. 2022, Astrophysical Journal Letters, 937, L42, doi: 10.3847/2041-8213/ac9519

Hirschi, R., et al. 2025, doi: 10.1093/mnras/staf1470

Hon, M., Li, Y., & Ong, J. 2024, Astrophysical Journal, 973, 154, doi: 10.3847/1538-4357/ad6320

Hooper, S., Chung, S. K., Luan, J., et al. 2012, Physical Review D, 86, 024012, doi: 10.1103/PhysRevD.86.024012

Hopkins, P. F., Keres, D., Onorbe, J., et al. 2014, Monthly Notices of the Royal Astronomical Society, 445, 581, doi: 10.1093/mnras/stu1738

Hurley, J. R., Pols, O. R., & Tout, C. A. 2000, Monthly Notices of the Royal Astronomical Society, 315, 543, doi: 10.1046/j.1365-8711.2000.03426.x

Hurley, J. R., Tout, C. A., & Pols, O. R. 2002, Monthly Notices of the Royal Astronomical Society, 329, 897, doi: 10.1046/j.1365-8711.2002.05038.x

Husa, S., Khan, S., Hannam, M., et al. 2016, Physical Review D, 93, 044006, doi: 10.1103/PhysRevD.93.044006

Hussain, A., Isi, M., & Zimmerman, A. 2026, Astrophysical Journal, 996, 71, doi: 10.3847/1538-4357/ae1574

Hypki, A., & Giersz, M. 2013, Monthly Notices of the Royal Astronomical Society, 429, 1221, doi: 10.1093/mnras/sts415

Iaconi, R., & De Marco, O. 2019, Monthly Notices of the Royal Astronomical Society, 490, 2550, doi: 10.1093/mnras/stz2756

Iorio, G. 2025, The chart of rapid population synthesis codes, Zenodo, doi: 10.5281/zenodo.15775292

Iorio, G., Costa, G., Mapelli, M., et al. 2022, doi: 10.48550/arXiv.2211.11774

Isi, M. 2022, Parametrizing gravitational-wave polarizations, v1.0.0, Zenodo, doi: 10.5281/zenodo.7016058

—. 2023, Classical and Quantum Gravity, 40, 203001, doi: 10.1088/1361-6382/acf28c

Islam, T., & Wadekar, D. 2026, Physical Review D, 113, 104017, doi: 10.1103/4jvv-qg4h

Issa, D., Lowell, B., Jacquemin-Ide, J., Liska, M., & Tchekhovskoy, A. 2025. https://arxiv.org/abs/2502.08732

Ivanova, N., et al. 2013, Astronomy and Astrophysics Review, 21, 59, doi: 10.1007/s00159-013-0059-2

Iwaya, M., Kinugawa, T., & Tagoshi, H. 2023. https://arxiv.org/abs/2312.17491

Izzard, R. G., Ramirez-Ruiz, E., & Tout, C. A. 2004a, Monthly Notices of the Royal Astronomical Society, 348, 1215, doi: 10.1111/j.1365-2966.2004.07436.x

Izzard, R. G., Tout, C. A., Karakas, A. I., & Pols, O. R. 2004b, Monthly Notices of the Royal Astronomical Society, 350, 407, doi: 10.1111/j.1365-2966.2004.07446.x

Janka, H.-T. 2012, Annual Review of Nuclear and Particle Science, 62, 407, doi: 10.1146/annurev-nucl-102711-094901

Jayasinghe, T., et al. 2021, Monthly Notices of the Royal Astronomical Society, 504, 2577, doi: 10.1093/mnras/stab907

Jaynes, E. T. 2003, Probability Theory: The Logic of Science, ed. G. L. Bretthorst (Cambridge: Cambridge University Press), doi: 10.1017/CBO9780511790423

Jermyn, A. S., et al. 2023, Astrophysical Journal Supplement Series, 265, 15, doi: 10.3847/1538-4365/acae8d

Joyce, M., & Tayar, J. 2023, Galaxies, 11, 75, doi: 10.3390/galaxies11030075

Kalogera, V. 2000, Astrophysical Journal, 541, 319, doi: 10.1086/309400

Kalogera, V., & Baym, G. 1996, Astrophysical Journal Letters, 470, L61, doi: 10.1086/310296

Kalogera, V., et al. 2021. https://arxiv.org/abs/2111.06990

Kidder, L. E. 1995, Physical Review D, 52, 821, doi: 10.1103/PhysRevD.52.821

Kimball, C., et al. 2021, Astrophysical Journal Letters, 915, L35, doi: 10.3847/2041-8213/ac0aef

Kıroğlu, F., Lombardi, J. C., Kremer, K., Vanderzyden, H. D., & Rasio, F. A. 2025, Astrophysical Journal Letters, 983, L9, doi: 10.3847/2041-8213/adc263

Klimenko, S., et al. 2016, Physical Review D, 93, 042004, doi: 10.1103/PhysRevD.93.042004

Kobyzev, I., Prince, S. J. D., & Brubaker, M. A. 2021, IEEE Transactions on Pattern Analysis and Machine Intelligence, 43, 3964, doi: 10.1109/TPAMI.2020.2992934

Kochanek, C. S. 2015, Monthly Notices of the Royal Astronomical Society, 446, 1213, doi: 10.1093/mnras/stu2056

Kochanek, C. S., Beacom, J. F., Kistler, M. D., et al. 2008, Astrophysical Journal, 684, 1336, doi: 10.1086/590053

Koehn, H., et al. 2025, Physical Review X, 15, 021014, doi: 10.1103/PhysRevX.15.021014

Kotko, I., & Belczynski, K. 2024, Astronomy and Astrophysics, 683, A85, doi: 10.1051/0004-6361/202346880

Kovetz, E. D., Cholis, I., Breysse, P. C., & Kamionkowski, M. 2017, Physical Review D, 95, 103010, doi: 10.1103/PhysRevD.95.103010

Kremer, K. 2026, in Encyclopedia of Astrophysics (First Edition), ed. I. Mandel (Elsevier), 458–472, doi: 10.1016/B978-0-443-21439-4.00103-6

Kremer, K., Ye, C. S., Rui, N. Z., et al. 2020, Astrophysical Journal Supplement Series, 247, 48, doi: 10.3847/1538-4365/ab7919

Kritos, K., Reali, L., Ng, K. K. Y., Antonini, F., & Berti, E. 2025, Physical Review D, 111, 063056, doi: 10.1103/PhysRevD.111.063056

Kritos, K., Strokov, V., Baibhav, V., & Berti, E. 2024, Physical Review D, 110, 043023, doi: 10.1103/PhysRevD.110.043023

Kroupa, P. 2002, Science, 295, 82, doi: 10.1126/science.1067524

Kruckow, M. U., Tauris, T. M., Langer, N., Kramer, M., & Izzard, R. G. 2018, Monthly Notices of the Royal Astronomical Society, 481, 1908, doi: 10.1093/mnras/sty2190

Kullback, S., & Leibler, R. A. 1951, The Annals of Mathematical Statistics, 22, 79, doi: 10.1214/aoms/1177729694

Kummer, F., Toonen, S., & de Koter, A. 2023, Astronomy and Astrophysics, 678, A60, doi: 10.1051/0004-6361/202347179

Laal, N., et al. 2025, Astrophysical Journal, 982, 55, doi: 10.3847/1538-4357/adb4ef

Lam, C. Y., et al. 2022, Astrophys. J. Lett., 933, L23, doi: 10.3847/2041-8213/ac7442

Lamers, H. J., & M. Levesque, E. 2017, in Understanding Stellar Evolution, 2514-3433 (IOP Publishing), 15–1 to 15–12, doi: 10.1088/978-0-7503-1278-3ch15

Langer, N. 1992, Astronomy and Astrophysics, 265, L17

Langer, N. 2012, Annual Review of Astronomy and Astrophysics, 50, 107–164, doi: 10.1146/annurev-astro-081811-125534

Lau, M. Y. M., Hirai, R., González-Bolívar, M., et al. 2022, Monthly Notices of the Royal Astronomical Society, 512, 5462, doi: 10.1093/mnras/stac049

Law-Smith, J. A. P., et al. 2020. https://arxiv.org/abs/2011.06630

Legred, I., Golomb, J., & Chatziioannou, K. 2026. https://arxiv.org/abs/2604.01420

Levin, Y. 2007, Monthly Notices of the Royal Astronomical Society, 374, 515, doi: 10.1111/j.1365-2966.2006.11155.x

Levina, S., Broekgaarden, F., van Son, L., et al. 2026. https://arxiv.org/abs/2601.20202

Li, L., Lü, G., Zhu, C., et al. 2025, Physical Review D, 112, 103005, doi: 10.1103/drq9-dpy4

Li, Y.-J., Tang, S.-P., Gao, S.-J., Wu, D.-C., & Wang, Y.-Z. 2024a, Astrophysical Journal, 977, 67, doi: 10.3847/1538-4357/ad83b5

Li, Y.-J., Wang, Y.-Z., Tang, S.-P., & Fan, Y.-Z. 2024b, Physical Review Letters, 133, 051401, doi: 10.1103/PhysRevLett.133.051401

Li, Y.-J., Wang, Y.-Z., Tang, S.-P., et al. 2022, Astrophysical Journal Letters, 933, L14, doi: 10.3847/2041-8213/ac78dd

Liotine, C., Zevin, M., Berry, C. P. L., Doctor, Z., & Kalogera, V. 2023, Astrophysical Journal, 946, 4, doi: 10.3847/1538-4357/acb8b2

Lipman, Y., et al. 2023. https://arxiv.org/abs/2210.02747

Lipunov, V. M., Postnov, K. A., Prokhorov, M. E., & Bogomazov, A. I. 2009, Astronomy Reports, 53, 915, doi: 10.1134/S1063772909100047

Liu, T., Wei, Y.-F., Xue, L., & Sun, M.-Y. 2021, Astrophysical Journal, 908, 106, doi: 10.3847/1538-4357/abd24e

Lorenzo-Medina, A., & Dent, T. 2025, Classical and Quantum Gravity, 42, 045008, doi: 10.1088/1361-6382/ad9c0e

Loshchilov, I., & Hutter, F. 2017. https://arxiv.org/abs/1608.03983

Luo, R., Zhu, C., Lü, G., et al. 2025, Astronomy and Astrophysics, 704, A46, doi: 10.1051/0004-6361/202555081

MacFadyen, A., & Woosley, S. E. 1999, Astrophysical Journal, 524, 262, doi: 10.1086/307790

Mackey, A. D., Wilkinson, M. I., Davies, M. B., & Gilmore, G. F. 2008, Monthly Notices of the Royal Astronomical Society, 386, 65, doi: 10.1111/j.1365-2966.2008.13052.x

Maclean, D. B., Agrawal, P., Breivik, K., et al. 2026. https://arxiv.org/abs/2606.28506

Madau, P., & Dickinson, M. 2014, Annual Review of Astronomy and Astrophysics, 52, 415, doi: 10.1146/annurev-astro-081811-125615

Madau, P., & Fragos, T. 2017, Astrophysical Journal, 840, 39, doi: 10.3847/1538-4357/aa6af9

Maeder, A. 1975, Astronomy and Astrophysics, 40, 303

Maeder, A. 1989, in Physics of Luminous Blue Variables, ed. K. Davidson, A. F. J. Moffat, & H. J. G. L. M. Lamers (Dordrecht: Springer Netherlands), 15–26

Maggiore, M. 2007a, in Gravitational Waves: Volume 1: Theory and Experiments (Oxford University Press), doi: 10.1093/acprof:oso/9780198570745.003.0009

—. 2007b, in Gravitational Waves: Volume 1: Theory and Experiments (Oxford University Press), doi: 10.1093/acprof:oso/9780198570745.003.0004

Maggiore, M., et al. 2020, JCAP, 03, 050, doi: 10.1088/1475-7516/2020/03/050

Mahapatra, P., Thompson, J. E., Fauchon-Jones, E., & Hannam, M. 2026. https://arxiv.org/abs/2603.26521

Mancarella, M., & Gerosa, D. 2025, Physical Review D, 111, 103012, doi: 10.1103/PhysRevD.111.103012

Mandel, I. 2010, Physical Review D, 81, 084029, doi: 10.1103/PhysRevD.81.084029

Mandel, I., & Broekgaarden, F. S. 2022, Living Reviews in Relativity, 25, 1, doi: 10.1007/s41114-021-00034-3

Mandel, I., & de Mink, S. E. 2016, Monthly Notices of the Royal Astronomical Society, 458, 2634, doi: 10.1093/mnras/stw379

Mandel, I., & Farmer, A. 2022, Physics Reports, 955, 1, doi: 10.1016/j.physrep.2022.01.003

Mandel, I., Farr, W. M., Colonna, A., et al. 2017, Monthly Notices of the Royal Astronomical Society, 465, 3254, doi: 10.1093/mnras/stw2883

Mandel, I., Farr, W. M., & Gair, J. R. 2019, Monthly Notices of the Royal Astronomical Society, 486, 1086, doi: 10.1093/mnras/stz896

Mandel, I., Müller, B., Riley, J., et al. 2020, Monthly Notices of the Royal Astronomical Society, 500, 1380, doi: 10.1093/mnras/staa3390

Mapelli, M. 2021, in Handbook of Gravitational Wave Astronomy, doi: 10.1007/978-981-15-4702-7_16-1

Mapelli, M., et al. 2021, Monthly Notices of the Royal Astronomical Society, 505, 339, doi: 10.1093/mnras/stab1334

Marchant, P., & Bodensteiner, J. 2024, Annual Review of Astronomy and Astrophysics, 62, 21, doi: 10.1146/annurev-astro-052722-105936

Marchant, P., Langer, N., Podsiadlowski, P., Tauris, T. M., & Moriya, T. J. 2016, Astronomy and Astrophysics, 588, A50, doi: 10.1051/0004-6361/201628133

Martinez, M. A. S., Rodriguez, C. L., & Fragione, G. 2022, Astrophysical Journal, 937, 78, doi: 10.3847/1538-4357/ac8d55

Martini, A., et al. 2026, Classical and Quantum Gravity, 43, 055016, doi: 10.1088/1361-6382/ae4717

Mastrogiovanni, S., Lamberts, A., Srinivasan, R., Bruel, T., & Christensen, N. 2022, Monthly Notices of the Royal Astronomical Society, 517, 3432, doi: 10.1093/mnras/stac2850

McKernan, B., & Ford, K. E. S. 2024, Monthly Notices of the Royal Astronomical Society, 531, 3479, doi: 10.1093/mnras/stae1351

McKernan, B., Ford, K. E. S., Kocsis, B., Lyra, W., & Winter, L. M. 2014, Monthly Notices of the Royal Astronomical Society, 441, 900, doi: 10.1093/mnras/stu553

McKernan, B., Ford, K. E. S., Lyra, W., & Perets, H. B. 2012, Monthly Notices of the Royal Astronomical Society, 425, 460, doi: 10.1111/j.1365-2966.2012.21486.x

McKernan, B., Ford, K. E. S., Cook, H. E., et al. 2025, Astrophysical Journal, 990, 217, doi: 10.3847/1538-4357/adf114

Merritt, J., Stevenson, S., Sander, A., et al. 2025. https://arxiv.org/abs/2507.17052

Messick, C., et al. 2017, Physical Review D, 95, 042001, doi: 10.1103/PhysRevD.95.042001

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. 1953, The Journal of Chemical Physics, 21, 1087, doi: 10.1063/1.1699114

Miller, M. C., & Lauburg, V. M. 2009, Astrophysical Journal, 692, 917, doi: 10.1088/0004-637X/692/1/917

Miller, S., Callister, T. A., & Farr, W. 2020, Astrophysical Journal, 895, 128, doi: 10.3847/1538-4357/ab80c0

Miller, S. J., Ko, Z., Callister, T., & Chatziioannou, K. 2024, Physical Review D, 109, 104036, doi: 10.1103/PhysRevD.109.104036

Moe, M., & Di Stefano, R. 2017, Astrophysical Journal Supplement Series, 230, 15, doi: 10.3847/1538-4365/aa6fb6

Moreno, M. M., Schneider, F. R. N., Roepke, F. K., et al. 2022, Astronomy and Astrophysics, 667, A72, doi: 10.1051/0004-6361/202142731

Morris, M. 1993, Astrophysical Journal, 408, 496, doi: 10.1086/172607

Mould, M., Gerosa, D., & Taylor, S. R. 2022, Physical Review D, 106, 103013, doi: 10.1103/PhysRevD.106.103013

Muccillo, L., Patricelli, B., & Razzano, M. 2026, Journal of High Energy Astrophysics, 100618, doi: 10.1016/j.jheap.2026.100618

Nagarajan, N., & Messenger, C. 2025, Physical Review D, 112, 103002, doi: 10.1103/zwj9-ycyz

Negu, S. H., & Tessema, S. B. 2015, International Journal of Astronomy and Astrophysics, 5, 222, doi: 10.4236/ijaa.2015.53026

Neijssel, C. J., Vigna-Gómez, A., Stevenson, S., et al. 2019, Monthly Notices of the Royal Astronomical Society, 490, 3740, doi: 10.1093/mnras/stz2840

Neumayer, N., Seth, A., & Boeker, T. 2020, Astronomy and Astrophysics Review, 28, 4, doi: 10.1007/s00159-020-00125-0

Ng, K. K. Y., Kritos, K., Antonelli, A., Cotesta, R., & Berti, E. 2023, Physical Review D, 108, 083015, doi: 10.1103/PhysRevD.108.083015

Nicholl, M., & Andreoni, I. 2025, Philosophical Transactions of the Royal Society A, 383, 20240126, doi: 10.1098/rsta.2024.0126

Nomoto, K., Kobayashi, C., & Tominaga, N. 2013, Annual Review of Astronomy and Astrophysics, 51, 457, doi: 10.1146/annurev-astro-082812-140956

Nuttall, L. K. 2018, Philosophical Transactions of the Royal Society A, 376, 20170286, doi: 10.1098/rsta.2017.0286

O'Connor, E., & Ott, C. D. 2011, The Astrophysical Journal, 730, 70, doi: 10.1088/0004-637X/730/2/70

Offner, S. S. R., Moe, M., Kratter, K. M., et al. 2023, in Astronomical Society of the Pacific Conference Series, Vol. 534, Protostars and Planets VII, ed. S. Inutsuka, Y. Aikawa, T. Muto, K. Tomida, & M. Tamura, 275, doi: 10.48550/arXiv.2203.10066

O'Leary, R. M., Kocsis, B., & Loeb, A. 2009, Monthly Notices of the Royal Astronomical Society, 395, 2127, doi: 10.1111/j.1365-2966.2009.14653.x

Olejak, A. 2026. https://arxiv.org/abs/2604.18676

Olejak, A., & Belczynski, K. 2021, Astrophysical Journal Letters, 921, L2, doi: 10.3847/2041-8213/ac2f48

Olejak, A., Belczynski, K., & Ivanova, N. 2021, Astronomy and Astrophysics, 651, A100, doi: 10.1051/0004-6361/202140520

Olejak, A., Fryer, C. L., Belczynski, K., & Baibhav, V. 2022, Monthly Notices of the Royal Astronomical Society, 516, 2252, doi: 10.1093/mnras/stac2359

Olejak, A., Klencki, J., Xu, X.-T., et al. 2024, Astronomy and Astrophysics, 689, A305, doi: 10.1051/0004-6361/202450480

Ossokine, S., et al. 2020, Physical Review D, 102, 044055, doi: 10.1103/PhysRevD.102.044055

Ozel, F., Psaltis, D., Narayan, R., & McClintock, J. E. 2010, Astrophysical Journal, 725, 1918, doi: 10.1088/0004-637X/725/2/1918

Paczyński, B. 1971, Annual Review of Astronomy and Astrophysics, 9, 183, doi: 10.1146/annurev.aa.09.090171.001151

Paczynski, B. 1976, Symposium - International Astronomical Union, 73, 75, doi: 10.1017/S0074180900011864

Panuzzo, P., et al. 2024, Astronomy and Astrophysics, 686, L2, doi: 10.1051/0004-6361/202449763

Papamakarios, G., Nalisnick, E., Rezende, D. J., Mohamed, S., & Lakshminarayanan, B. 2021, Journal of Machine Learning Research, 22, 1, doi: 10.48550/arXiv.1912.02762

Patton, R. A., & Sukhbold, T. 2020, Monthly Notices of the Royal Astronomical Society, 499, 2803, doi: 10.1093/mnras/staa3029

Patton, R. A., Sukhbold, T., & Eldridge, J. J. 2022, Monthly Notices of the Royal Astronomical Society, 511, 903, doi: 10.1093/mnras/stab3797

Paxton, B., Bildsten, L., Dotter, A., et al. 2011, Astrophysical Journal Supplement Series, 192, 3, doi: 10.1088/0067-0049/192/1/3

Paxton, B., Cantiello, M., Arras, P., et al. 2013, Astrophysical Journal Supplement Series, 208, 4, doi: 10.1088/0067-0049/208/1/4

Paxton, B., Marchant, P., Schwab, J., et al. 2015, Astrophysical Journal Supplement Series, 220, 15, doi: 10.1088/0067-0049/220/1/15

Paxton, B., Schwab, J., Bauer, E. B., et al. 2018, Astrophysical Journal Supplement Series, 234, 34, doi: 10.3847/1538-4365/aaa5a8

Paxton, B., Smolec, R., Schwab, J., et al. 2019, Astrophysical Journal Supplement Series, 243, 10, doi: 10.3847/1538-4365/ab2241

Périgois, C., Mapelli, M., Santoliquido, F., Bouffanais, Y., & Rufolo, R. 2023, Universe, 9, 507, doi: 10.3390/universe9120507

Perley, D. A., et al. 2016, Astrophysical Journal, 817, 7, doi: 10.3847/0004-637X/817/1/7

Peters, P. C. 1964, Physical Review, 136, B1224, doi: 10.1103/PhysRev.136.B1224

Peters, P. C., & Mathews, J. 1963, Physical Review, 131, 435, doi: 10.1103/PhysRev.131.435

Picco, A., Marchant, P., Sana, H., & Nelemans, G. 2024, Astronomy and Astrophysics, 681, A31, doi: 10.1051/0004-6361/202347090

Pierra, G., Mastrogiovanni, S., & Perriès, S. 2024, Astronomy and Astrophysics, 692, A80, doi: 10.1051/0004-6361/202452545

Pierra, G., & Papadopoulos, A. 2026. https://arxiv.org/abs/2601.03257

Plunkett, C., Callister, T., Zevin, M., & Vitale, S. 2026, Physical Review Letters, doi: 10.1103/n6p4-ftgq

Podsiadlowski, Ph., Joss, P. C., & Hsu, J. J. L. 1992, Astrophysical Journal, 391, 246, doi: 10.1086/171341

Poelarends, A. J. T., Herwig, F., Langer, N., & Heger, A. 2008, Astrophysical Journal, 675, 614, doi: 10.1086/520872

Politano, M. 2004, Astrophysical Journal, 604, 817, doi: 10.1086/381958

Popov, S., Müller, B., & Mandel, I. 2025, New Astronomy Reviews, 101, 101734, doi: 10.1016/j.newar.2025.101734

Portegies Zwart, S. F., & Verbunt, F. 1996, Astronomy and Astrophysics, 309, 179

Prandtl, L. 1925, Zeitschrift Angewandte Mathematik und Mechanik, 5, 136, doi: 10.1002/zamm.19250050212

Prasad, V., & Sathyaprakash, B. S. 2026. https://arxiv.org/abs/2602.12859

Pratten, G., et al. 2021, Physical Review D, 103, 104056, doi: 10.1103/PhysRevD.103.104056

Pretorius, F. 2005, Physical Review Letters, 95, 121101, doi: 10.1103/PhysRevLett.95.121101

Prusti, T., et al. 2016, Astronomy and Astrophysics, 595, A1, doi: 10.1051/0004-6361/201629272

Pürrer, M., Hannam, M., Ajith, P., & Husa, S. 2013, Physical Review D, 88, 064007, doi: 10.1103/PhysRevD.88.064007

Pürrer, M., Hannam, M., & Ohme, F. 2017, in 14th Marcel Grossmann Meeting on Recent Developments in Theoretical and Experimental General Relativity, Astrophysics, and Relativistic Field Theories, Vol. 3, 3144–3148, doi: 10.1142/9789813226609_0399

Quinlan, G. D., & Shapiro, S. L. 1987, Astrophysical Journal, 321, 199, doi: 10.1086/165624

Rakavy, G., & Shaviv, G. 1967, Astrophysical Journal, 148, 803, doi: 10.1086/149204

Ramírez-Agudelo, O. H., et al. 2013, Astronomy and Astrophysics, 560, A29, doi: 10.1051/0004-6361/201321986

Rasio, F. A., Freitag, M., & Atakan Gurkan, M. 2003, in Carnegie Observatories Centennial Symposium. 1. Coevolution of Black Holes and Galaxies. https://arxiv.org/abs/astro-ph/0304038

Ray, A. 2025. https://arxiv.org/abs/2506.05657

Ray, A., Farr, W., & Kalogera, V. 2026a, Astrophysical Journal, 1002, 105, doi: 10.3847/1538-4357/ae581e

Ray, A., & Kalogera, V. 2026, Astrophysical Journal Letters, 998, L20, doi: 10.3847/2041-8213/ae374d

Ray, A., Magaña Hernandez, I., Breivik, K., & Creighton, J. 2025, Astrophysical Journal, 991, 17, doi: 10.3847/1538-4357/adf22a

Ray, A., Magaña Hernandez, I., Mohite, S., Creighton, J., & Kapadia, S. 2023, Astrophysical Journal, 957, 37, doi: 10.3847/1538-4357/acf452

Ray, A., Mukherjee, S., Zevin, M., & Kalogera, V. 2026b. https://arxiv.org/abs/2603.17987

Raymond, V., Al-Shammari, S., & Göttel, A. 2025, Monthly Notices of the Royal Astronomical Society, 1103, 1108, doi: 10.1093/mnras/staf1282

Riley, J., & Mandel, I. 2023, Astrophysical Journal, 950, 80, doi: 10.3847/1538-4357/accf90

Riley, J., Mandel, I., Marchant, P., et al. 2021, Monthly Notices of the Royal Astronomical Society, 505, 663, doi: 10.1093/mnras/stab1291

Rinaldi, S., & Del Pozzo, W. 2021, Monthly Notices of the Royal Astronomical Society, 509, 5454, doi: 10.1093/mnras/stab3224

Rinaldi, S., Del Pozzo, W., Mapelli, M., Lorenzo-Medina, A., & Dent, T. 2024, Astronomy and Astrophysics, 684, A204, doi: 10.1051/0004-6361/202348161

Rinaldi, S., Liang, Y., Demasi, G., Mapelli, M., & Del Pozzo, W. 2025, Astronomy and Astrophysics, 702, A52, doi: 10.1051/0004-6361/202555870

Rodriguez, C. L., Amaro-Seoane, P., Chatterjee, S., & Rasio, F. A. 2018, Physical Review Letters, 120, 151101, doi: 10.1103/PhysRevLett.120.151101

Rodriguez, C. L., Chatterjee, S., & Rasio, F. A. 2016a, Physical Review D, 93, 084029, doi: 10.1103/PhysRevD.93.084029

Rodriguez, C. L., & Loeb, A. 2018, Astrophysical Journal Letters, 866, L5, doi: 10.3847/2041-8213/aae377

Rodriguez, C. L., Zevin, M., Amaro-Seoane, P., et al. 2019, Physical Review D, 100, 043027, doi: 10.1103/PhysRevD.100.043027

Rodriguez, C. L., Zevin, M., Pankow, C., Kalogera, V., & Rasio, F. A. 2016b, Astrophysical Journal Letters, 832, L2, doi: 10.3847/2041-8205/832/1/L2

Rodriguez, C. L., et al. 2022, Astrophysical Journal Supplement, 258, 22, doi: 10.3847/1538-4365/ac2edf

Roepke, F. K., & De Marco, O. 2023, Living Reviews in Computational Astrophysics, 9, 2, doi: 10.1007/s41115-023-00017-x

Romagnolo, A., Broekgaarden, F. S., Antoniadis, K., & Gormaz-Matamala, A. C. 2026. https://arxiv.org/abs/2601.02263

Romero-Shaw, I. M., Kremer, K., Lasky, P. D., Thrane, E., & Samsing, J. 2021, Monthly Notices of the Royal Astronomical Society, 506, 2362, doi: 10.1093/mnras/stab1815

Rose, S. C., Naoz, S., Sari, R., & Linial, I. 2022, Astrophysical Journal Letters, 929, L22, doi: 10.3847/2041-8213/ac6426

Roulet, J., Chia, H. S., Olsen, S., et al. 2021, Physical Review D, 104, 083010, doi: 10.1103/PhysRevD.104.083010

Roulet, J., & Zaldarriaga, M. 2019, Monthly Notices of the Royal Astronomical Society, 484, 4216, doi: 10.1093/mnras/stz226

Roy, S. K., van Son, L. A. C., & Farr, W. M. 2025, Classical and Quantum Gravity, 42, 225008, doi: 10.1088/1361-6382/ae1921

Rutherford, N., et al. 2024, Astrophysical Journal Letters, 971, L19, doi: 10.3847/2041-8213/ad5f02

Sadiq, J., Dent, T., & Gieles, M. 2024, Astrophysical Journal, 960, 65, doi: 10.3847/1538-4357/ad0ce6

Safarzadeh, M., Farr, W. M., & Ramirez-Ruiz, E. 2020, Astrophysical Journal, 894, 129, doi: 10.3847/1538-4357/ab80be

Saleem, M., et al. 2022, Classical and Quantum Gravity, 39, 025004, doi: 10.1088/1361-6382/ac3b99

Samsing, J., D'Orazio, D. J., Kremer, K., Rodriguez, C. L., & Askar, A. 2020, Physical Review D, 101, 123010, doi: 10.1103/PhysRevD.101.123010

Samsing, J., MacLeod, M., & Ramirez-Ruiz, E. 2014, Astrophysical Journal, 784, 71, doi: 10.1088/0004-637X/784/1/71

Sana, H., Shenar, T., Bodensteiner, J., et al. 2025, Nature Astronomy, 9, 1337–1346, doi: 10.1038/s41550-025-02610-x

Sana, H., de Mink, S. E., de Koter, A., et al. 2012, Science, 337, 444, doi: 10.1126/science.1223344

Sana, H., et al. 2014, Astrophysical Journal Supplement Series, 215, 15, doi: 10.1088/0067-0049/215/1/15

Sasaki, M. 2025, General Relativity and Gravitation, 57, 82, doi: 10.1007/s10714-025-03412-2

Sathyaprakash, B. S., & Schutz, B. F. 2009, Living Reviews in Relativity, 12, 2, doi: 10.12942/lrr-2009-2

Schmidt, P., Ohme, F., & Hannam, M. 2015, Physical Review D, 91, 024043, doi: 10.1103/PhysRevD.91.024043

Schneider, F. R. N., Podsiadlowski, P., & Laplace, E. 2023, Astrophysical Journal Letters, 950, L9, doi: 10.3847/2041-8213/acd77a

Schneider, F. R. N., Podsiadlowski, P., & Müller, B. 2021, Astronomy and Astrophysics, 645, A5, doi: 10.1051/0004-6361/202039219

Schroder, K.-P., Pols, O. R., & Eggleton, P. P. 1997, Monthly Notices of the Royal Astronomical Society, 285, 696, doi: 10.1093/mnras/285.4.696

Schutz, B. F. 2011, Classical and Quantum Gravity, 28, 125023, doi: 10.1088/0264-9381/28/12/125023

Sciarini, L., Rosu, S., Ekström, S., et al. 2026, Astronomy and Astrophysics, 707, A22, doi: 10.1051/0004-6361/202556992

Secunda, A., Bellovary, J., Mac Low, M.-M., et al. 2020, Astrophysical Journal, 903, 133, doi: 10.3847/1538-4357/abbc1d

Sedda, M. A., Mapelli, M., Benacquista, M., & Spera, M. 2023a, Monthly Notices of the Royal Astronomical Society, 520, 5259, doi: 10.1093/mnras/stad331

Sedda, M. A., Mapelli, M., Spera, M., Benacquista, M., & Giacobbo, N. 2020, Astrophysical Journal, 894, 133, doi: 10.3847/1538-4357/ab88b2

Sedda, M. A., Naoz, S., & Kocsis, B. 2023b, Universe, 9, 138, doi: 10.3390/universe9030138

Senchyna, P., Hawcroft, C., Garcia, M., et al. 2025. https://arxiv.org/abs/2511.03715

Shah, N., Renzo, M., Sen, K., Grichener, A., & Breivik, K. 2026. https://arxiv.org/abs/2602.12327

Shenar, T., Bodensteiner, J., Sana, H., et al. 2024, Astronomy and Astrophysics, 690, A289, doi: 10.1051/0004-6361/202451586

Siegel, J. C., et al. 2023, Astrophysical Journal, 954, 212, doi: 10.3847/1538-4357/ace9d9

Silver, E., Krastev, P., & Berger, E. 2025. https://arxiv.org/abs/2512.17204

Skilling, J. 2004, AIP Conference Proceedings, 735, 395, doi: 10.1063/1.1835238

Smartt, S. J. 2009, Annual Review of Astronomy and Astrophysics, 47, 63–106, doi: 10.1146/annurev-astro-082708-101737

Smith, N. 2014, Annual Review of Astronomy and Astrophysics, 52, 487, doi: 10.1146/annurev-astro-081913-040025

Speagle, J. S. 2020, Monthly Notices of the Royal Astronomical Society, 493, 3132, doi: 10.1093/mnras/staa278

Spitzer, Jr., L. 1969, Astrophysical Journal Letters, 158, L139, doi: 10.1086/180451

Stanway, E. R., & Eldridge, J. J. 2018, Monthly Notices of the Royal Astronomical Society, 479, 75, doi: 10.1093/mnras/sty1353

Stegmann, J., Antonini, F., & Moe, M. 2022, Monthly Notices of the Royal Astronomical Society, 516, 1406, doi: 10.1093/mnras/stac2192

Stevenson, S., Berry, C. P. L., & Mandel, I. 2017a, Monthly Notices of the Royal Astronomical Society, 471, 2801, doi: 10.1093/mnras/stx1764

Stevenson, S., & Clarke, T. 2022, Monthly Notices of the Royal Astronomical Society, doi: 10.1093/mnras/stac2936

Stevenson, S., Vigna-Gómez, A., Mandel, I., et al. 2017b, Nature Communications, 8, 14906, doi: 10.1038/ncomms14906

Stone, N. C., Metzger, B. D., & Haiman, Z. 2017, Monthly Notices of the Royal Astronomical Society, 464, 946, doi: 10.1093/mnras/stw2260

Sukhbold, T., Ertl, T., Woosley, S. E., Brown, J. M., & Janka, H. T. 2016, Astrophys. J., 821, 38, doi: 10.3847/0004-637X/821/1/38

Sukhbold, T., Woosley, S., & Heger, A. 2018, Astrophysical Journal, 860, 93, doi: 10.3847/1538-4357/aac2da

Tagawa, H., Haiman, Z., & Kocsis, B. 2020, Astrophysical Journal, 898, 25, doi: 10.3847/1538-4357/ab9b8c

Tagawa, H., Kocsis, B., Haiman, Z., et al. 2021, Astrophysical Journal, 908, 194, doi: 10.3847/1538-4357/abd555

Talbot, C. 2021, doi: 10.5281/zenodo.5654673

Talbot, C., Farah, A., Galaudage, S., Golomb, J., & Tong, H. 2025, Journal of Open Source Software, 10, 7753, doi: 10.21105/joss.07753

Talbot, C., & Golomb, J. 2023, Monthly Notices of the Royal Astronomical Society, 526, 3495, doi: 10.1093/mnras/stad2968

Talbot, C., & Thrane, E. 2017, Physical Review D, 96, 023012, doi: 10.1103/PhysRevD.96.023012

—. 2018, Astrophysical Journal, 856, 173, doi: 10.3847/1538-4357/aab34c

Tang, P. N., Eldridge, J. J., Stanway, E. R., & Bray, J. C. 2020, Monthly Notices of the Royal Astronomical Society, 493, L6, doi: 10.1093/mnrasl/slz183

Taracchini, A., et al. 2014, Physical Review D, 89, 061502, doi: 10.1103/PhysRevD.89.061502

Tauris, T. M. 2022, Astrophysical Journal, 938, 66, doi: 10.3847/1538-4357/ac86c8

Taylor, S. R., & Gerosa, D. 2018, Physical Review D, 98, 083017, doi: 10.1103/PhysRevD.98.083017

Temaj, D., Schneider, F. R. N., Laplace, E., Wei, D., & Podsiadlowski, P. 2024, Astronomy and Astrophysics, 682, A123, doi: 10.1051/0004-6361/202347434

Thompson, T. A., et al. 2019, Science, 366, 637, doi: 10.1126/science.aau4005

Thrane, E., & Talbot, C. 2019, Publications of the Astronomical Society of Australia, 36, e010, doi: 10.1017/pasa.2019.2

Tiwari, V. 2018, Classical and Quantum Gravity, 35, 145009, doi: 10.1088/1361-6382/aac89d

—. 2021, Classical and Quantum Gravity, 38, 155007, doi: 10.1088/1361-6382/ac0b54

—. 2025. https://arxiv.org/abs/2510.25579

Tong, H., Callister, T. A., Fishbach, M., et al. 2025. https://arxiv.org/abs/2511.05316

Tong, H., Galaudage, S., & Thrane, E. 2022, Physical Review D, 106, 103019, doi: 10.1103/PhysRevD.106.103019

Toonen, S., Boekholt, T. C. N., & Zwart, S. P. 2022, Astronomy and Astrophysics, 661, A61, doi: 10.1051/0004-6361/202141991

Toonen, S., Claeys, J. S. W., Mennekens, N., & Ruiter, A. J. 2014, Astronomy and Astrophysics, 562, A14, doi: 10.1051/0004-6361/201321576

Toonen, S., Hamers, A., & Portegies Zwart, S. 2016, Computational Astrophysics and Cosmology, 3, 6, doi: 10.1186/s40668-016-0019-0

Toonen, S., Nelemans, G., & Portegies Zwart, S. 2012, Astronomy and Astrophysics, 546, A70, doi: 10.1051/0004-6361/201218966

Toonen, S., Portegies Zwart, S., Hamers, A. S., & Bandopadhyay, D. 2020, Astronomy and Astrophysics, 640, A16, doi: 10.1051/0004-6361/201936835

Toubiana, A., Katz, M. L., & Gair, J. R. 2023, Monthly Notices of the Royal Astronomical Society, 524, 5844, doi: 10.1093/mnras/stad2215

Tremaine, S. D., Ostriker, J. P., & Spitzer, Jr., L. 1975, Astrophysical Journal, 196, 407, doi: 10.1086/153422

Vaccaro, M. P., Mapelli, M., Périgois, C., et al. 2024, Astronomy and Astrophysics, 685, A51, doi: 10.1051/0004-6361/202348509

van den Heuvel, E. P. J. 1976, in Structure and Evolution of Close Binary Systems, ed. P. Eggleton, S. Mitton, & J. Whelan, Vol. 73, 35, doi: 10.1017/S0074180900011839

van den Heuvel, E. P. J., Portegies Zwart, S. F., & de Mink, S. E. 2017, Monthly Notices of the Royal Astronomical Society, 471, 4256, doi: 10.1093/mnras/stx1430

van den Heuvel, E. P. J., Portegies Zwart, S. F., & de Mink, S. E. 2017, Monthly Notices of the Royal Astronomical Society, 471, 4256, doi: 10.1093/mnras/stx1430

van Son, L. A. C., de Mink, S. E., Chruslinska, M., et al. 2022a, Astrophysical Journal, 948, 105, doi: 10.3847/1538-4357/acbf51

van Son, L. A. C., de Mink, S. E., Broekgaarden, F. S., et al. 2020, Astrophysical Journal, 897, 100, doi: 10.3847/1538-4357/ab9809

van Son, L. A. C., de Mink, S. E., Renzo, M., et al. 2022b, Astrophysical Journal, 940, 184, doi: 10.3847/1538-4357/ac9b0a

van Son, L. A. C., de Mink, S. E., Callister, T., et al. 2022, Astrophysical Journal, 931, 17, doi: 10.3847/1538-4357/ac64a3

van Son, L. A. C., Roy, S. K., Mandel, I., et al. 2025, Astrophysical Journal, 979, 209, doi: 10.3847/1538-4357/ada14a

van Son, L. A. C., Yamaguchi, N., Nagarajan, P., et al. 2026. https://arxiv.org/abs/2605.31290

Veitch, J., & Vecchio, A. 2010, Physical Review D, 81, 062003, doi: 10.1103/PhysRevD.81.062003

Vigna-Gómez, A. 2025, Astronomy and Astrophysics, 701, L3, doi: 10.1051/0004-6361/202556051

Vigna-Gómez, A., et al. 2024, Physical Review Letters, 132, 191403, doi: 10.1103/PhysRevLett.132.191403

Vijaykumar, A., Farah, A. M., & Fishbach, M. 2026, Astrophysical Journal Letters, 999, L30, doi: 10.3847/2041-8213/ae4878

Villaseñor, J. I., Sana, H., Bodensteiner, J., et al. 2025. https://arxiv.org/abs/2511.08675

Vink, J. S., de Koter, A., & Lamers, H. J. G. L. M. 2001, Astronomy and Astrophysics, 369, 574, doi: 10.1051/0004-6361:20010127

Vink, J. S., Sabhahit, G. N., & Winch, E. R. J. 2025. https://arxiv.org/abs/2511.00199

Vitale, S., Biscoveanu, S., & Talbot, C. 2022, Astronomy and Astrophysics, 668, L2, doi: 10.1051/0004-6361/202245084

Vitale, S., Gerosa, D., Farr, W. M., & Taylor, S. R. 2020, in Handbook of Gravitational Wave Astronomy, ed. C. Bambi, S. Katsanevas, & K. D. Kokkotas (Singapore: Springer Singapore), 1–60, doi: 10.1007/978-981-15-4702-7_45-1

Vitale, S., Lynch, R., Raymond, V., et al. 2017, Physical Review D, 95, 064053, doi: 10.1103/PhysRevD.95.064053

Vitale, S., Lynch, R., Veitch, J., Raymond, V., & Sturani, R. 2014, Physical Review Letters, 112, 251101, doi: 10.1103/PhysRevLett.112.251101

Vitale, S., & Mould, M. 2025, Physical Review D, 112, 083015, doi: 10.1103/drsl-n3wz

Wang, L., Spurzem, R., Aarseth, S., et al. 2015, Monthly Notices of the Royal Astronomical Society, 450, 4070–4080, doi: 10.1093/mnras/stv817

Wang, M., & Zhang, F. 2023, Association of Asia Pacific Physical Societies Buletin, 33, 5, doi: 10.1007/s43673-023-00076-5

Wang, Z.-Y., Qin, Y., Hu, R.-C., et al. 2026, Astronomy and Astrophysics, 708, A62, doi: 10.1051/0004-6361/202557224

Wang, Z.-Y., Qin, Y., Meynet, G., et al. 2025, Monthly Notices of the Royal Astronomical Society, 548, doi: 10.1093/mnras/stag585

Wei, D., Schneider, F. R. N., Podsiadlowski, P., et al. 2024, Astronomy and Astrophysics, 688, A87, doi: 10.1051/0004-6361/202348560

Wildberger, J., Dax, M., Buchholz, S., et al. 2023, in Advances in Neural Information Processing Systems, ed. A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, & S. Levine, Vol. 36 (Curran Associates, Inc.), 16837–16864. https://arxiv.org/abs/2305.17161

Willcox, R., MacLeod, M., Mandel, I., & Hirai, R. 2023, Astrophysical Journal, 958, 138, doi: 10.3847/1538-4357/acffb1

Willcox, R., Schneider, F. R. N., Laplace, E., et al. 2025. https://arxiv.org/abs/2510.07573

Williams, M. J., Mcginn, J., Stachurski, F., & Veitch, J. 2022, Uofgravity/Glasflow, 0.1.2. https://github.com/uofgravity/glasflow

Wolfe, N. E., Mould, M., Heinzel, J., & Vitale, S. 2026, Physical Review D, 113, 103025, doi: 10.1103/15w1-n9jx

Wong, K. W. K., Breivik, K., Kremer, K., & Callister, T. 2021a, Physical Review D, 103, 083021, doi: 10.1103/PhysRevD.103.083021

Wong, K. W. K., Contardo, G., & Ho, S. 2020, Physical Review D, 101, 123005, doi: 10.1103/PhysRevD.101.123005

Wong, K. W. K., Franciolini, G., De Luca, V., et al. 2021b, Physical Review D, 103, 023026, doi: 10.1103/PhysRevD.103.023026

Woosley, S. E. 1993, Astrophysical Journal, 405, 273, doi: 10.1086/172359

Woosley, S. E., Heger, A., & Weaver, T. A. 2002, Reviews of Modern Physics, 74, 1015, doi: 10.1103/RevModPhys.74.1015

Wu, T. Y., & Fishbach, M. 2024, Astrophysical Journal, 977, 239, doi: 10.3847/1538-4357/ad98ed

Wyrzykowski, Ł., & Mandel, I. 2020, Astronomy and Astrophysics, 636, A20, doi: 10.1051/0004-6361/201935842

Wysocki, D., Gerosa, D., O'Shaughnessy, R., et al. 2018, Physical Review D, 97, 043014, doi: 10.1103/PhysRevD.97.043014

Wysocki, D., Lange, J., & O'Shaughnessy, R. 2019, Physical Review D, 100, 043012, doi: 10.1103/PhysRevD.100.043012

Yang, Y., Bartos, I., Haiman, Z., et al. 2019, Astrophysical Journal, 876, 122, doi: 10.3847/1538-4357/ab16e3

Ye, C. S., Fishbach, M., Kremer, K., & Reina-Campos, M. 2026, Astrophysical Journal, 997, 267, doi: 10.3847/1538-4357/ae317f

Yoon, S.-C., Langer, N., & Norman, C. 2006, Astronomy and Astrophysics, 460, 199, doi: 10.1051/0004-6361:20065912

Zevin, M., & Bavera, S. S. 2022, Astrophysical Journal, 933, 86, doi: 10.3847/1538-4357/ac6f5d

Zevin, M., Pankow, C., Rodriguez, C. L., et al. 2017, Astrophysical Journal, 846, 82, doi: 10.3847/1538-4357/aa8408

Zevin, M., Spera, M., Berry, C. P. L., & Kalogera, V. 2020, Astrophysical Journal Letters, 899, L1, doi: 10.3847/2041-8213/aba74e

Zevin, M., Bavera, S. S., Berry, C. P. L., et al. 2021, Astrophysical Journal, 910, 152, doi: 10.3847/1538-4357/abe40e