

THE UNIVERSITY OF CHICAGO

ASTROPHYSICS AND COSMOLOGY WITH THE MASS DISTRIBUTION OF
GRAVITATIONAL WAVE SOURCES

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHYSICS

BY
AMANDA MIRNA FARAH

CHICAGO, ILLINOIS
AUGUST 2025

Copyright © 2025 by Amanda Mirna Farah
All Rights Reserved

To the survivors and martyrs of the Gaza genocide.

May our people soon look to the sky and find stars instead of missiles.

TABLE OF CONTENTS

LIST OF FIGURES	vii
LIST OF TABLES	xviii
ACKNOWLEDGMENTS	xix
ABSTRACT	xxi
1 INTRODUCTION	1
1.1 Astrophysics with the GW source population	2
1.2 Spectral sirens: cosmology with the GW source mass distribution	4
1.3 Outline of this work	6
2 HIERARCHICAL BAYESIAN INFERENCE FOR GW POPULATION ANALYSIS	8
2.1 Introduction	8
2.2 Inferring the population	9
2.3 Sampling	12
2.4 Directed acyclic graphs	13
2.4.1 Example: spectral sirens	15
2.4.2 Example: constraining population synthesis parameters	17
3 BRIDGING THE GAP: THE TRANSITION BETWEEN NEUTRON STAR AND BLACK HOLE MASSES	21
3.1 Introduction	22
3.2 Methods	25
3.2.1 Event list	25
3.2.2 Population models	26
3.2.3 Characterizing the prominence of features in the mass distribution	32
3.2.4 Statistical framework	33
3.3 Population-level insights	34
3.3.1 Deviation from a single power law describing BH and NS masses	34
3.3.2 Steep decline after NS masses	35
3.3.3 Searching for a lower mass gap	38
3.4 Event-level insights	41
3.4.1 Event classification	44
3.4.2 GW190814	47
3.4.3 GW200105 and GW200115	48
3.5 Discussion	50
3.6 Additional analysis details	53
3.6.1 Hyperposteriors inferred with and without NSBH events	53
3.6.2 Most flexible model fit	53
3.6.3 Model comparison	56

4	THINGS THAT MIGHT GO BUMP IN THE NIGHT: SIGNIFICANCE OF STRUCTURE IN THE BINARY BLACK HOLE MASS DISTRIBUTION	61
4.1	Introduction	62
4.2	Motivation	65
4.3	Hierarchical analysis and results	70
4.3.1	POWER LAW + SPLINE mass model	70
4.3.2	Metrics of feature significance	72
4.3.3	Simulation study	73
4.4	Discussion	81
4.5	Additional analysis details	86
4.5.1	Generation of mock observations in GWMockCat	86
4.5.2	Validation of mock catalogs	92
4.5.3	Accuracy of mock catalogs when used in a population analysis	98
5	NO NEED TO KNOW: ASTROPHYSICS-AGNOSTIC COSMOLOGY WITH THE MASS SPECTRUM OF GRAVITATIONAL-WAVE SOURCES	102
5.1	Introduction	103
5.2	Cosmology with an astrophysically-agnostic mass model	106
5.2.1	Simulated data	108
5.2.2	The spectral siren method	109
5.2.3	Gaussian process-based mass distribution	110
5.3	Results	112
5.3.1	Projections for future measurements	114
5.4	Discussion	115
5.5	Additional analysis details	120
5.5.1	Details of data simulation	120
5.5.2	Effects of fitting for secondary mass	122
5.5.3	Gaussian process-based mass distribution	123
5.5.4	Biases induced by the BROKEN POWER LAW model	126
6	TWO OF A KIND: THE TWO-DIMENSIONAL BINARY BLACK HOLE MASS DISTRIBUTION	128
6.1	Introduction	129
6.2	Population models	132
6.3	Results	135
6.3.1	Primary and secondary masses are consistent with having the same underlying distribution	139
6.3.2	Improved constraints on peak location	140
6.3.3	Evidence for a peak in the secondary mass distribution	143
6.3.4	BBHs are picky	146
6.4	Looking to the future	148
6.5	Summary and implications for astrophysics of merging BBHs	149
6.6	Additional Analysis Details	152
6.6.1	Comparison of Common Mass Distribution Parametrizations	152

6.6.2	General Population Model	158
7	DON'T JUST LEAVE ONE OUT: PROBING EXTREMAL GRAVITATIONAL- WAVE EVENTS WITH COARSE-GRAINED LIKELIHOODS	166
7.1	Introduction	167
7.2	Coarse-grained leave-one-out likelihoods	170
7.2.1	Derivation of coarse-grained inference	171
7.2.2	Implications from the choice of $\mathcal{S}$	176
7.2.3	Quantifying inconsistencies with inferred populations	178
7.3	Astrophysical results	179
7.3.1	GW190814 is an outlier in secondary mass	180
7.3.2	GW190412 is not an outlier	183
7.3.3	GW190521 is not an outlier	185
7.4	Discussion	188
7.5	Additional details on reweighing existing hyperposteriors	190
	REFERENCES	192

LIST OF FIGURES

2.1	Two examples of directed acyclic graphs (DAGs). The left panel describes the data generation process for a generic astrophysical catalog. The right panel describes a specific example in gravitational wave (GW) astronomy, outlined in Section 2.4.2. In both panels, the joint likelihood between all nodes in the directed acyclic graph (DAG) is written at the bottom. Each arrow in the directed acyclic graph (DAG) represents a conditional distribution, i.e. an arrow from X to Y means that the conditional distribution of Y depends on X	15
3.1	90% posterior credible intervals for the component masses for all CBCs included in this population study assuming uniform priors in detector-frame masses. Events classified by the LVK as BBHs, BNSs and NSBHs are shown in black, green, and blue, respectively. The ambiguously classified event GW190814 is shown in orange. The grey band indicates the approximate location of the purported “lower mass gap.” GW190814 is the only event within this region at more than 90% credibility.	27
3.2	Illustration of the primary mass distribution for the three mass models used in this work. All models are depicted with the same values for m_{break} , α_1 , and α_2 .	31
3.3	Draws from the inferred mass distributions between $1 - 10 M_{\odot}$ using the BROKEN POWER LAW (<i>top</i>), BROKEN POWER LAW + DROP (<i>middle</i>), and BROKEN POWER LAW + DIP (<i>bottom</i>) models. Each panel has 200 random draws from the hyperposterior (<i>thin black lines</i>). As an illustration, the draw that maximizes the overall posterior is highlighted in color. 90% credible intervals on the locations of features in the mass distribution are indicated by colored bands. All three models find a steep falloff in the mass distribution between $\sim 2.2 - 3 M_{\odot}$ and BROKEN POWER LAW + DIP finds a subsequent rise at $\sim 6 M_{\odot}$	36
3.4	Values of the spectral index governing the low mass part of the spectrum α_1 and drop depth A inferred by the BROKEN POWER LAW + DROP model. Contours indicate 1- σ , 2- σ , and 3- σ credible regions. The two parameters are correlated and the lower right region is excluded: mild power law slopes are only allowed when A is large. Thus, the inference finds a dramatic drop-off in the merger rate, either through a steep negative initial power-law slope (α_1), or a discontinuity (large A). The two scenarios are both allowed by the data due to the inability to resolve the shape of the NS distribution.	39
3.5	Posterior on the gap depth parameter A under the BROKEN POWER LAW + DIP model including all events. $A = 1$ indicates a maximal gap, and $A = 0$ indicates no mass gap. A peaks at 0.90, with 90% highest density interval between 0.55 and 1.0 (<i>vertical dashed lines</i>), indicating support for both a partially-filled and totally empty mass gap.	42

3.6	Corner plot of the parameters governing the location (γ_{high}) and steepness (η_{high}) of the upper edge of the mass gap in the case where a steep decline in the mass distribution is enforced after NS masses ($A = 0.99$). The 90% highest probability density interval for η_{high} is indicated by vertical dashed lines. η_{high} is found to be positive, indicating a rise in the mass distribution near γ_{high} after the drop enforced at γ_{low}	43
3.7	Effects of leaving out event GW190814 under the BROKEN POWER LAW + DIP model. Green contours correspond to the inference when GW190814 is included, and grey contours correspond to the inference when it is excluded. Almost all hyperparameters remain unchanged with the inclusion vs exclusion of GW190814, indicating that this event is consistent with the inferred population using this model. The only parameter that exhibits a noticeable change is γ_{low} , where the bimodality of the posterior distribution disappears with the exclusion of GW190814.	49
3.8	A subset of hyperposteriors for the BROKEN POWER LAW + DIP model inferred with (<i>green</i>) and without (<i>brown</i>) the events GW200105 and GW200115. The mass distribution does not change significantly with the exclusion of these events, indicating that the NSBHs are consistent with the rest of the detected population. When the NSBHs are included, the upper edge of the mass gap, γ_{high} , is slightly better constrained and support for a wider gap (low γ_{low} and high γ_{high}) diminishes. This suggests that these events probe the gap location. Hyperparameters on which the exclusion of the NSBH events has no effect are not included in this plot.	54
3.9	Corner plot for an analysis where all hyperparameters are simultaneously fit. On the diagonal are the one-dimensional marginal distributions for each hyperposterior. Above each column are written the median and 68% credible intervals for the corresponding hyperposterior.	57
3.10	Corner plot for the BROKEN POWER LAW model. On the diagonal are the one-dimensional marginal distributions for each hyperposterior. Above each column are written the median and 68% credible intervals for the corresponding hyperposterior.	58
3.11	Same as Fig. 3.10 but for BROKEN POWER LAW + DROP.	59
3.12	Same as Fig. 3.10 but for BROKEN POWER LAW + DIP.	60

- 4.1 Distribution of primary BBH masses inferred using the third gravitational-wave transient catalog (GWTC-3) and three different population models. The smoothed power law model (*grey*) consists of a single power law slope between a minimum and maximum mass, with the merger rate set to exactly zero outside of those bounds. It also includes a smoothing parameter at the low-mass end that allows for an offset between the minimum BH mass and the global maximum of the distribution. The POWER LAW + PEAK model is similar to the smoothed power law, but also includes a Gaussian component. The POWER LAW + SPLINE model adds a cubic spline modulation to a smoothed power law to allow for additional substructure. We seek to determine if the perturbations beyond a power law found by POWER LAW + SPLINE and other semi-parametric models can be explained by random associations in the data due to a finite number of observations, or if they are features of the true underlying distribution. 65
- 4.2 Observed source-frame primary mass distributions. Black solid lines contain the median *a posteriori* values for the binary black holes in GWTC-3. Pink and blue bands indicate the 90% credible interval on the *observed* distributions predicted from astrophysical distributions that are power laws in primary mass with spectral index $\alpha = 3.25$ and $\alpha = 2.7$, respectively. The top panel shows a histogram of observed primary masses. For GWTC-3's distribution to be consistent with the null distributions, we expect its bin heights in the top panel to be within the 90% credible intervals in 18 out of the 20 bins. The uncertainties in these predicted distributions are due only to Poisson noise resulting from a finite number of observations, rather than modeling uncertainty or uncertainty in parameter estimation. Therefore, the cumulative distribution functions in the bottom panel are similar to a conventional posterior predictive check, but with only one source of uncertainty. The large deviations of the black curve from the shaded bands in some regions indicate the difficulty that a single power law with Poisson shot noise has in fully explaining the observations. However, many of the apparent excursions from a power law are well-contained within the predicted bands. . . 66
- 4.3 Median (*top panel*) and 90% credible interval (*bottom panel*) of the perturbation function resulting from the POWER LAW + SPLINE fit to the primary masses in GWTC-3 (*orange*) and in 10 mock catalogs (*grey*). The perturbation function multiplies a smoothed power law in primary mass to add modulations to an otherwise monotonic distribution, making it a direct measure of deviations from a power law. It is a cubic spline with knots fixed at the locations indicated by the black vertical tick marks. The prior on the perturbation heights is the unit normal distribution, as can be seen below $\sim 5 M_{\odot}$ where there are no detections to constrain the likelihood and the posterior reverts to the prior. The perturbation function corresponding to GWTC-3 events appears large in amplitude in three locations: $\sim 10 M_{\odot}$, $\sim 14 M_{\odot}$, and $\sim 35 M_{\odot}$. While the medians of the perturbation function at these distributions are comparable in amplitude, the posterior distribution at $\sim 35 M_{\odot}$ ($\sim 14 M_{\odot}$) is the most (least) tightly constrained. . . 74

4.4	Three largest deviations from a power law observed in GWTC-3 compared to mock catalogs. <i>Left column:</i> The posterior distribution of perturbation function heights at the location where the posterior distribution is maximal for mock catalogs (<i>light grey</i>) and GWTC-3 (<i>solid orange</i>). <i>Middle column:</i> Null distribution (<i>black</i>) and GWTC-3 distribution (<i>orange</i>) of Kolmogorov–Smirnov (KS) divergences between the individual distributions in the left column. Smaller values of the KS divergence indicate more similar distributions. <i>Right column:</i> Null distribution (<i>black</i>) and GWTC-3 distribution (<i>orange</i>) of percentiles. Large deviations from the diagonal indicate a more significant rightward shift of the GWTC-3 distribution relative to the mock catalogs. Each row corresponds to a different local extremum for GWTC-3: $m_1 = 13.8 M_\odot$ (<i>top</i>), $m_1 = 10.3 M_\odot$ (<i>middle</i>), and $m_1 = 35.7 M_\odot$ (<i>bottom</i>), while the global extrema for each mock catalog are shown in all rows, along with the aggregated distribution across all mock catalogs (<i>solid black</i>). The $\sim 35 M_\odot$ peak is an outlier with respect to both the KS and percentile statistics, but the other two features are more ambiguous.	76
4.5	Percentile at which the posterior distribution of the perturbation function excludes zero for GWTC-3 (<i>orange vertical lines</i>) and catalogs drawn from a featureless distribution (<i>black histogram</i>). For GWTC-3, we evaluate the perturbation function’s posterior distribution at primary mass (m_1) values of $13.8 M_\odot$ (<i>dotted</i>), $10.3 M_\odot$ (<i>dashed</i>), and $35.7 M_\odot$ (<i>solid</i>). For mock catalogs, we find the primary mass value at which the perturbation function is maximal and evaluate its posterior distribution there. The values reported here are the percentage of the posterior distribution that is greater than zero at those values in m_1 . The $13.5 M_\odot$ feature excludes zero to a level comparable to some of the mock catalogs, but the other two features exclude zero to a level not reproducible by any mock catalogs.	79
4.6	Injected (<i>Solid black line</i>) and recovered (<i>colored shaded bands</i>) distributions for 10 mock catalogs. <i>Top:</i> probability density function of primary masses. <i>Bottom left:</i> hyperposterior distribution for β , the power law spectral index of the mass ratio distribution. <i>Bottom Right:</i> hyperposterior distribution for κ , the spectral index of the power law factor in the redshift distribution.	94
4.7	A corner plot of the inferred hyperposterior from a fit to a mock catalog with 345 events. The injected values are shown in orange. The recovered hyperposterior is consistent with the injected population.	95
4.8	Mock posteriors simulated with GWMockCat (<i>red</i>) compared to posteriors made with full parameter estimation as released by the LVK (<i>blue</i>) for the event GW191215_223052. The two sets of posterior samples appear consistent to the level needed for a population analysis.	96

4.9	Perturbation functions of a population fit to GWTC-3 when using LVK-released posterior samples for each event (<i>orange</i>), and when using posterior samples simulated by GWMockCat (<i>blue and green</i>). The two sets of figures correspond to two selection criteria, one which is analogous to the one used for mock catalogs in this work ($\text{SNR} > 8$, <i>top two panels</i>), and one which is identical to that used for the LVK analysis ($\text{FAR} > 1$, <i>bottom two panels</i>). The main difference between the selection criteria is that they result in catalog sizes that are different by nearly a factor of two, and therefore the statistical error on the perturbation function is noticeably different between them. Using the same number of events in the simulated catalog as the LVK-released catalog results in a perturbation function that is similar in amplitude to the LVK-released population analysis. We conclude that the prescriptions used in GWMockCat are sufficient for the purposes of population analyses. We use the green curves in the bottom two panels for the analysis presented in the body of this chapter, and the orange curves in the bottom two panels for the analysis presented in this Section.	97
5.1	Spectral siren measurement for a simulated catalog with the correct parametric model (POWER LAW + PEAK, green dashed), a deliberately incorrect parametric model (BROKEN POWER LAW, pink solid) and the non-parametric model presented in this work (Gaussian process, blue filled). The left panel shows the recovered source frame primary mass distribution for each model, and the simultaneously-inferred posteriors on H_0 are shown in the right panel. The bands in the left panel represent the 90% credible intervals. The mass distribution and H_0 value used to generate the data are shown by a solid black line in each panel. The deliberately incorrect parametric model fails to recover the true mass distribution and therefore produces an estimate of H_0 that is offset from the true value, whereas both the correct and non-parametric models recover the mass distribution and H_0 . As the true mass distribution is unknown for real observations, using a non-parametric model mitigates systematic uncertainty that would otherwise arise from mismodeling the compact binary coalescence (CBC) population.	106
5.2	Draws from the Gaussian process (GP) used to model the mass distribution. The left panel shows prior draws from the Gaussian process (GP) and the right panel shows posterior draws once the population inference is performed on the simulated data. The posterior draws in the right panel are a subset of those used to create the 90% credible intervals in Fig. 5.1.	111

5.3	Projected constraints on multiple cosmological parameters after one year of observing at the LIGO-Virgo-KAGRA collaboration (LVK)'s design sensitivity, using the Gaussian process-based spectral siren method. The right panel shows the inferred expansion history of the universe, $H(z)$. It will be measured most precisely at $z = 0.9$, as can be seen by the narrowing of the inferred $H(z)$ curves there. The inset shows the posterior on $H(z = 0.9)$. Black solid lines indicate the true value of $H(z)$ in both the inset and main panel. The left panel shows the two-dimensional posterior on H_0 and Ω_M , with the true value indicated by a black “+”. The two parameters are strongly degenerate because of the multiple ways of measuring $z = 0.9$. Spectral sirens are particularly well suited for measuring cosmological parameters that affect the Universe at $z \gtrsim 0.2$	116
5.4	Comparison of a parametric spectral siren analysis performed while fitting for the distributions of primary and secondary masses (magenta solid) and primary masses only (green dashed). The posterior on H_0 is relatively unchanged between the two, and only minor differences are observed in the mass spectra.	123
5.5	Recovered posteriors on H_0 from spectral siren analyses on 50 mock catalogs with the POWER LAW + PEAK (green) and BROKEN POWER LAW (pink) parametric models. The true, injected value is shown by a vertical black line.	127
6.1	Illustration of some possible two-dimensional mass distributions under the models considered in this work. These distributions are not indicative of specific astrophysical predictions, but are instead meant to be illustrative of the morphologies accessible by current models. Overdensities/peaks in the mass distribution appear as darker filled contours in these figures. The different columns correspond to different power law spectral indices for the mass ratio-dependent pairing function, β_q . In the case where components strongly prefer to pair with nearly equal-mass partners, it becomes difficult to determine whether a feature appears only in one component mass distribution (as in the PAIRING:NO PEAK IN $p_2(m_2)$ model, <i>bottom row</i>) or in both (PAIRING:SYMMETRIC, <i>top row</i>). The diagonal contours in the middle and right columns are caused by a preference for equal-mass binaries and follow lines of constant mass ratio. The goal of this chapter is to distinguish between the different scenarios illustrated in this figure.	136

- 6.2 Two-dimensional posterior predictive distributions (PPDs) for PAIRING:SYMMETRIC (*left, orange*), PAIRING:GENERIC (*middle, purple*), and PAIRING:NO PEAK IN $p_2(m_2)$ (*right, pink*), all fit to the BBHs in GWTC-3. Darker colors indicate a higher rate of sources, and each panel is individually normalized to its maximum value. All three models find a higher rate of events near the $m_1 = m_2$ diagonal, as well as peaks at $m_1 \sim 10 M_\odot$ and $m_1 \sim 35 M_\odot$. PAIRING:SYMMETRIC and PAIRING:GENERIC both find peaks in m_2 as well, as indicated by horizontal bands in the two leftmost panels, whereas PAIRING:NO PEAK IN $p_2(m_2)$ is unable to model features in the m_2 direction. Additionally, PAIRING:SYMMETRIC and PAIRING:GENERIC display peaks in similar locations despite the fact that PAIRING:SYMMETRIC requires that both m_1 and m_2 follow the same underlying distribution but PAIRING:GENERIC is able to model each component separately. This suggests that both components may be drawn from the same underlying distribution, up to a pairing function. 137
- 6.3 Underlying distribution of primary and secondary masses for the PAIRING:SYMMETRIC (*orange*) and PAIRING:GENERIC (*purple*) models. Under the PAIRING:SYMMETRIC model, $p_1(m_1) = p_2(m_2) \equiv p(m)$, so only $p(m)$ is shown. $p(m)$ under PAIRING:SYMMETRIC is more tightly constrained than $p_1(m_1)$ (*purple filled band*) or $p_2(m_2)$ (*dot-dashed lines*) under PAIRING:GENERIC as it has twice as many observations per free parameter. These distributions are constructed to describe the population of black holes (BHs) *before* the function that pairs components into merging binaries is applied. All three underlying distributions are consistent with one another, though $p_2(m_2)$ appears to have a large peak at $\sim 35 M_\odot$ while $p_1(m_1)$ has some support for no peak in that region. $p(m)$ does not have support for no peak, but its peak is constrained to be small, while the peak in $p_2(m_2)$ is less well-constrained and may be larger in amplitude. This hints at the possibility that the peak identified in the primary mass distribution by the LVK 2023 formalism may have been driven by a peak in $p_2(m_2)$ rather than $p_1(m_1)$, though hyper-parameter uncertainties within PAIRING:GENERIC are too large to definitively determine this, and the data is consistent with $p_1(m_1) = p_2(m_2)$. . . 138

6.4	Conditional posterior predictive distributions (PPDs) for the models considered in this work. We show the secondary mass distributions conditioned on $m_1 = 20 M_\odot$ (<i>dashed</i>), $m_1 = 35 M_\odot$ (<i>solid</i>) and $m_1 = 70 M_\odot$ (<i>dotted</i>) to exemplify when the primary mass is below, inside, and above the Gaussian peak, respectively. Orange lines correspond to PAIRING:SYMMETRIC, purple lines to PAIRING:GENERIC, and green lines to LVK 2023. Lines denote the mean posterior probability, marginalized over hyper-parameter uncertainty, and credible intervals are omitted for clarity. When m_1 is below the peak, all models agree. When m_1 is above or within the peak, PAIRING:SYMMETRIC and PAIRING:GENERIC exhibit a larger preference for m_2 to be in the peak than the LVK 2023 model does because the latter is constructed to behave like a power law in the range $[m_{\min}, m_1]$. The fact that PAIRING:GENERIC and PAIRING:SYMMETRIC approximately agree on the peak location and height indicates that the primary and secondary masses may follow the same underlying distribution up to a pairing function.	141
6.5	Marginal posteriors on μ , the location of the Gaussian peak, for the LVK 2023 and PAIRING:SYMMETRIC models. This parameter is more precisely measured under the PAIRING:SYMMETRIC formalism by $\sim 1/\sqrt{2}$, indicating that m_1 and m_2 independently contribute to this measurement. μ is the most correlated parameter with the Hubble constant when using “spectral sirens” [Ezquiaga and Holz, 2022, Abbott et al., 2023b], so using the PAIRING:SYMMETRIC model will presumably improve cosmological measurements. The central values of the two distributions differ because the hyper-parameter has slightly different effects on the resulting 2D mass distribution each model.	144

- 6.6 Hyperposteriors under pairing function models. *Left:* Corner plot of hyperparameters governing the height of the Gaussian peak for the primary (λ_1) and secondary (λ_2) mass distributions under the PAIRING:GENERIC model. The medians and 90% credible intervals are shown above their respective marginal distributions. The medians indicate that roughly 17% of BBHs have secondary masses in the Gaussian peak while 4% have primary masses in the peak, though this preference for a larger peak in $p_2(m_2)$ may be due to the relatively poor constraint on λ_2 rather than a true preference in the data. The lower left portion of the two-dimensional posterior is excluded, indicating that there must be a peak in either $p_1(m_1)$ or $p_2(m_2)$, and a slight preference exists for the peak to be in $p_2(m_2)$ (upper left portion of plot) or both distributions (central portion of plot), though the peak being in $p_1(m_1)$ only is not completely ruled out. For reference, the 1D posterior on λ under the LVK 2023 model is shown in green, with $\lambda = 0.04^{+0.03}_{-0.02}$ [Abbott et al., 2023a]. The dashed grey line indicates where $\lambda_1 = \lambda_2$. *Right:* Marginal posteriors on β_q , the hyper-parameter controlling the steepness of the pairing function under the PAIRING:SYMMETRIC, PAIRING:GENERIC, and PAIRING:NO PEAK IN $p_2(m_2)$ models. When we set $\lambda_2 = 0$, β_q increases, indicating a preference for equal-mass components. This behavior is likely caused to accommodate an excess of events with $m_2 \sim 35 M_\odot$, which can either be caused by a peak in $p_2(m_2)$ or with a peak in $p_1(m_1)$ and a strong preference for equal-mass binaries (see discussion in Section 6.6.1). 147
- 6.7 Illustration of some possible two-dimensional mass distributions under the commonly used “CONDITIONED-Q” formalism described by Equation 6.3 (*top row*) and the pairing function formalism described by Equation 6.5 (*bottom row*). Overdensities/peaks in the mass distribution appear as darker filled contours in these figures. The CONDITIONED-Q formalism is only able to produce models with features in the m_1 distribution, as shown by the vertical bands in the top row, whereas the pairing function formalism can model features in either m_1 or m_2 , or both. The different columns correspond to different power law spectral indices for the mass ratio distribution (*top row*, β) and mass ratio-dependent pairing function (*bottom row*, β_q). In the case of a steeply rising mass ratio distribution, or if components strongly prefer to pair with nearly equal-mass partners, the CONDITIONED-Q model and the pairing function model become difficult to tell apart and likely explain the data equally well, as shown by the two panels in the leftmost column. The diagonal contours in the middle and right columns are caused by a preference for equal-mass binaries and follow lines of constant mass ratio. 157
- 6.8 Corner plot of all mass-related hyper-posteriors in the most general form of the base model. Power law spectral indecies governing the slope of the primary and secondary mass distributions are degenerate with one another and with the pairing function power law, so only two of the three parameters can be meaningfully constrained at a time. All other parameters that perform the same function in $p_1(m_1)$ and $p_2(m_2)$ are consistent between the two underlying distributions. . . 161

6.9	Underlying distributions (i.e. before a pairing function is applied) of the primary (<i>shaded band</i>) and secondary (<i>dot-dashed lines</i>) masses under the most general form of the base model. Despite fitting all parameters separately between the two distributions, they appear consistent with one another. However, both are relatively poorly constrained. Interestingly, the support for a peak in $p_1(m_1)$ lessens in this more generalized model, while the support for a peak in $p_2(m_2)$ remains the same as in PAIRING:GENERIC.	162
6.10	Marginal posterior predictive distributions (PPDs) of primary mass (<i>top</i>), secondary mass (<i>middle</i>) and mass ratio (<i>bottom</i>) for CONDITIONED-Q (<i>green</i>), PAIRING:SYMMETRIC (<i>orange</i>) and PAIRING:GENERIC (<i>purple</i>). Solid lines are the mean value and shaded regions represent the 90% credible interval. These are obtained by integrating the 2D posterior predictive distributions (PPDs) shown in Fig. 6.2 along each dimension in turn. Since the CONDITIONED-Q model is parameterized in terms of m_1 and q , we reconstruct its marginal m_2 distribution by sampling from the joint distribution and creating a kernel density estimate, causing some artificial “wiggles” in these plots. The same is true for the marginal q distributions of the pairing function models. Despite the 2D posterior predictive distributions (PPDs) appearing different between CONDITIONED-Q and the other two models, the marginal posterior predictive distributions (PPDs) all appear similar. This is because features in the primary mass distribution induce features in the marginal secondary mass distribution if equal-mass binaries are preferred. Therefore, marginal posterior predictive distributions (PPDs) are not sensitive to the features in the 2D mass distribution in which we are interested.	164
6.11	Two-dimensional (<i>bottom</i>) and conditional m_2 (<i>top</i>) posterior predictive distributions (PPDs) for the CONDITIONED-Q and PAIRING:NO PEAK IN $p_2(m_2)$ models.	165
7.1	(<i>top left</i>) Depiction of $\mathcal{S}_{\text{GW190814}}$ (<i>shaded region</i> , Eq. 7.14), chosen to be the region ($m_2 \leq 5.0 M_\odot$) or ($q \leq 0.28$) based on the approximate median m_2 and q of GW190924 and GW190412, respectively. Colored shaded regions show 50% and 90% highest-probability-density credible regions from each event assuming flat priors in component masses and uniform priors in comoving volume for (<i>black</i> ellipse, circled in lower left corner) GW190814, (<i>blue</i>) GW190924 and GW190412, as well as (<i>red</i>) all other BBH systems considered in Abbott et al. [2021d]. (<i>bottom left</i>) Distribution of (<i>black</i>) GW190814’s m_2 and (<i>blue</i>) the smallest expected m_2 in catalogs of 45 events based on the $(N - 1)$ -event coarse-grained hyperposterior for the POWERLAW+PEAK mass model. (<i>right</i>) Hyperposteriors inferred with different amounts of information about GW190814. Contours in the joint distributions denote 50% and 90% credible regions.	182
7.2	(<i>top</i>) Depiction of (<i>shaded grey</i>) $\mathcal{S}_{\text{GW190412}}$ (Eq. 7.15) with (<i>shaded black</i>) GW190412 and (<i>red</i>) all other events considered in Abbott et al. [2021d] (except GW190814) with flat priors on component masses as well as (<i>blue</i>) the prediction for the smallest q out of 44 events based on the $(N - 1)$ -event coarse-grained hyperposterior with the POWERLAW+PEAK mass model. (<i>bottom</i>) Hyperposteriors inferred with different amounts of information about GW190412.	184

7.3	(<i>top</i>) Depiction of (<i>shaded grey</i>) $\mathcal{S}_{\text{GW190521}}$ (Eq. 7.16) with (<i>black shaded</i>) GW190521 and (<i>red</i>) all other events considered in [Abbott et al., 2021d] informed by the POWERLAW+PEAK mass model as inferred under the $(N - 1)$ -event coarse-grained hyperposterior as well as the distribution of the largest predicted m_1 out of 44 events based on the $(N - 1)$ -event coarse-grained hyperposterior. The hard boundary in GW190521's distribution at $m_1 \sim 60 M_\odot$ is due to our labeling convention ($m_1 \geq m_2$) and approximately corresponds to equal-mass systems. (<i>bottom</i>) Hyperposteriors inferred with different amounts of information about GW190521. Note that m_{max} extends below $50 M_\odot$ because it only limits the power law part of POWERLAW+PEAK; the peak still has support at higher masses.	186
7.4	Comparison between reweighed samples and direct sampling for GW190814's $(N - 1)$ +coarse-grained hyperposterior (Sec. 7.3.1). Contours in the joint distributions denote 50% and 90% credible regions. Reweighed samples were obtained by applying Eq. 7.18 to public hyperposterior samples from an $(N - 1)$ -event analysis that excluded GW190814	191

LIST OF TABLES

3.1	Events classified in the BROKEN POWER LAW formalism. Values are the probabilities that each event’s component masses lie in a given region of parameter space, considering uncertainty in the component masses as well as in the inferred values of m_{break} . If m_{break} is interpreted as the delimiter between NS and BH masses, the probabilities can be understood as labeled in the second row. The uncertainties in these probabilities are estimated to be less than 1%. Only events with $> 5\%$ probability of containing a NS according to this classification are included in this table.	46
3.2	Events classified in the BROKEN POWER LAW + DIP formalism. Values are the probabilities that each event’s component masses lie in a given region of parameter space, considering uncertainty in the component masses as well as in the inferred values of γ_{low} and γ_{high} . If γ_{low} and γ_{high} are interpreted as the lower and upper edges of the mass gap and γ_{low} is therefore interpreted as the maximum observed NS mass in merging binaries, the probabilities can be understood as labeled in the second row. The uncertainties in these probabilities are estimated to be less than 1%. Only events with $> 5\%$ probability of being in one of these categories are included in this table.	46
4.1	Hyperparameter values for the underlying population of mock catalogs described by SMOOTHED POWER LAW (Equations 4.12–4.15).	93

ACKNOWLEDGMENTS

This dissertation is the result of seven years of work by many dedicated people. For scientific advice, professional mentorship, and academic support, I would like to thank my advisor, Daniel Holz, and my closest collaborators, Maya Fishbach, Jose María Ezquiaga, Mike Zevin, Reed Essick, Tom Callister, Colm Talbot, Bruce Edelman, and Zoheyr Doctor. I am also grateful for the patience, technical training, and invaluable scientific network provided by members of the LIGO-Virgo-KAGRA collaboration, especially in the Rates and Populations subgroup.

For professional advice, emotional labor, and material support (in the form of coffee, meals, hand-me-downs, and rides), I thank my shadow advisor and chosen sister, Eman Abdelhadi. Without my friends, siblings, community members, and fellow organizers, I would not have survived these last seven years, and certainly would not have completed this work. I therefore thank Dania Hallak, Frederick Farah, Daniel Farah, Hazem Hallak, Maddi McKay, Rama Wahbe, Shireen Hamza, Alaa Mezanin, Sarah Abboud, Fareaba Moradiahani, Leila Elaquad, Kell Oliver, Sarah Daoud, Anwaar Nabulsi, Julia Hage, Julie Rayan, Eugenia Huang, Annalisa Perez, Steph Barkan, Marcus Banks, Jordan Kemp, Samar Alqatari, Mandy Medley, Mack, Rosie, Charlotte, Erik Wallenberg, Storm Colloms, Aditya Vijaykumar, Graeme Eddols, Sophie Hourihane, Simona Miller, Naji Hattar, Osama El-Quqa, Azora Jalili, Yasmin Zacaria-Mikhail, Shana, Jad, Tariq, and, of course, Sultan.

I also acknowledge the students and TA in the "Introduction to Observational Astronomy" class of the Prison+Neighborhood Arts/Education Project, who taught me the importance of knowledge for knowledge's sake and pushed me to accept an entirely different understanding of astronomy: Daniel Perkins, Reginald BoClair, Robert Curry, Tylon Hudson, Jerel Matthews, Michael Sullivan, David Wales, Ed Bell, Maurice Darrell Brown, Cameron Matlock, Antwon Tyler, Robert Washington, Cameron White, Charles Hill.

Most importantly, I gratefully acknowledge my mother, Mirna M'farrej, for her uncondi-

tional support throughout the writing of this dissertation and the 22 years leading up to it. In addition to teaching me the power of critical thinking, instilling in me a love for problem solving, and being my most inspiring role model, her open-mindedness, dedication to breaking cycles of trauma, inhuman persistence, endless reservoir of care, and raw intelligence enabled me to pursue all of my personal and academic desires without hesitation.

ABSTRACT

Since the first direct detection of gravitational waves (GWs) just under 10 years ago, astronomers and physicists have gained immense physical insight by analyzing the individual compact binary coalescences (CBCs) from which GWs arise. Now, more than 90 GW detections have been published, and the most sensitive network of GW detectors is in observing mode and has produced over 200 additional public GW detection candidates so far. The rapidly-growing catalog of GW detections means that GW astronomy is firmly in the population era, where we stand to gain more physical insight by analyzing the ensemble of GW detections than any one source alone. This dissertation focuses on extracting physical information from one dimension of the GW source population: the mass distribution of CBCs.

The first part of the dissertation introduces GW cosmology and GW astrophysics and describes the statistical methods used to infer the population of CBCs from GW data. The second part of the dissertation applies this method to identify and characterize features in the mass distribution of CBC. We search for imprints of nuclear physics and the supernova explosion mechanism in the low-mass region of the mass distribution, finding evidence of the former and hints of the latter. We then assess the robustness of features in the high-mass region that were originally found using a flexible, rather than astrophysically-informed, search. We find that only two of the previously-identified features are statistically robust. The third part of the dissertation applies the features found in the second part to open problems in astrophysics: the stellar origin of CBCs and the Hubble tension. We demonstrate that comparing features in the distributions of primary and secondary black hole (BH) masses can aid in determining whether CBCs are mostly formed through dynamical encounters in dense stellar environments, or if they originate as binary star systems in the comparably diffuse galactic field. However, we show that such knowledge of the processes that produce CBCs is not necessary in order to use mass distribution features to measure the expansion

rate of the Universe. The final part of the dissertation discusses the scenario in which population models do not describe all of the data, or when they are significantly altered by a single GW event. We demonstrate how to identify such outliers to the inferred CBC population, and apply this to current GW catalogs.

CHAPTER 1

INTRODUCTION

Since the first direct detection of GWs just under 10 years ago, astronomers and physicists have used these ripples in spacetime to test Einstein’s theory of general relativity (GR) [Abbott et al., 2021h], measure the expansion rate of the Universe [Abbott et al., 2023b], determine the origin of heavy elements [Tanvir et al., 2017], characterize the properties of dense matter [Landry and Essick, 2019], constrain the rate of strong gravitational lensing [Abbott et al., 2021e], and learn about the varied astrophysical processes governing the massive stars from which GW sources arise [Abbott et al., 2023a]. Now, the LIGO detector [Aasi et al., 2015] has been joined by Virgo [Acernese et al., 2014] and KAGRA [Akutsu et al., 2021], and more than 90 GW detections have been published in the form of the first, second, and third Gravitational-Wave Transient Catalogs [GWTC-1, 2, and 3, respectively Abbott et al., 2019b, 2021b,g]. At the time of this writing, the LIGO-Virgo-KAGRA collaboration (LVK)’s fourth and most sensitive observing run is ongoing, and over 200 additional public GW detection candidates have been identified. The rapid growth of GW catalogs means GW astronomy is firmly in the population era, where we stand to gain more physical insight by analyzing the ensemble of GW detections than any one source alone.

Each GW event corresponds to a CBC, a merger of two neutron stars (NSs), two BHs, or one of each. Since GWs carry information about their source properties, we can use them to reconstruct the distribution of CBC masses, spins, and distances in the Universe. This population is influenced by the life cycles of massive stars, the dynamics of dense stellar systems, the expansion history of the Universe, and the size and abundance of dark matter halos. It is therefore a powerful probe of a wide range of physical phenomena.

All parameters that govern CBCs can provide insight into formation mechanisms of merging binaries. However, the masses of the objects in the system are the best measured and span the largest dynamic range. This dissertation therefore focuses on extracting physical

information from the mass distribution of GW sources.

1.1 Astrophysics with the GW source population

The formation mechanism of CBCs is still largely uncertain. While it is understood that the stellar-mass NSs and BHs detectable by the LVK are likely created from the collapse of massive stars, how these stars became compact objects that merged within the age of the Universe is still an open question. These massive stars may be born as a binary system in the galactic field that then evolves into a compact binary system, or they may be born in a dense stellar environment, such as a star cluster or an active galactic nucleus, in which stellar remnants dynamically assemble into tightly-bound binaries.

While it is not possible to know the formation history of any one CBC with certainty, the population of all merging CBCs encodes information about which astrophysical processes give rise to the bulk of these systems [e.g. Stevenson et al., 2015, Zevin et al., 2017, Mapelli, 2020, Mandel and Farmer, 2022]. This information includes the initial conditions of their progenitor stellar systems – such as the initial mass function of either binary or single stars in different environments, the birth metallicities of the progenitor stars – as well as the evolutionary processes that occur between star formation and compact object merger.

For example, the presence or dearth of black holes with masses between $\sim 2\text{--}5\ M_{\odot}$ [Özel et al., 2010, Farr et al., 2011, Fishbach et al., 2020a, Farah et al., 2022a] may unveil the maximum possible NS mass, the stability of mass transfer, and the timescales relevant for the engines that drive supernova explosions [e.g., Fryer et al., 2012, Zevin et al., 2020, Mandel and Müller, 2020, Li et al., 2021a, van Son et al., 2022b, Patton et al., 2022, Siegel et al., 2022]. On the high mass end, a sharp decrease in the mass spectrum for BHs with masses $\gtrsim 45 - 65\ M_{\odot}$ [Fishbach et al., 2017, Edelman et al., 2021] would be a strong indication that the pair-instability supernova (PISN) process is at play, as it is predicted to limit stellar core masses and therefore create a maximum-allowed BH mass above this limit [Fowler and

Hoyle, 1964, Barkat et al., 1967, Heger and Woosley, 2002, Heger et al., 2003, Woosley and Heger, 2015, Belczynski et al., 2016, Woosley, 2017, 2019, Marchant et al., 2019, Renzo et al., 2020]. Although PISN has been widely accepted as a necessary process for high-mass single stars, binary interactions may change this picture.

Other overdensities and underdensities in the observed mass distribution [Edelman et al., 2022, Tiwari, 2021, 2022, Edelman et al., 2023], as well as the evolution of the mass distribution with redshift [Fishbach et al., 2021, van Son et al., 2022a, Karathanasis et al., 2022a, van Son et al., 2022b], will further inform the dominant BBH formation channels, binary evolution physics, and the metallicity evolution of the Universe.

There are currently many more detected BBH mergers than binary neutron star (BNS) or neutron star-black hole (NSBH) mergers, so much of the activity on GW source populations has been on population properties of the BBH distribution, though the mass distribution of BNSs and NSBHs has also been considered [Fishbach et al., 2020a, Landry and Read, 2021, Farah et al., 2022a, Ye and Fishbach, 2022, Biscoveanu et al., 2022b].

The mass distribution of merging BBHs is typically parameterized by the primary mass m_1 , the larger of the two component masses in the binary, and the mass ratio $q = m_2/m_1$, the ratio of the less massive object’s mass to the primary mass, though other parameterizations are possible and valid [e.g., Farah et al., 2022a, Fishbach and Holz, 2020b, Tiwari, 2021]. The community has thus far gained a robust understanding of the large-scale features of the BBH mass distribution, and is just beginning to resolve its finer details. After the release of the first gravitational-wave transient catalog (GWTC-1) [Abbott et al., 2019b, GWTC-1], minimum and maximum masses at $\sim 5 M_\odot$ and $\sim 40 M_\odot$ were identified in the BBH primary mass distribution, but it was not yet possible to distinguish between a uniform distribution and a power law between those two bounds [Fishbach and Holz, 2017, Talbot and Thrane, 2018, Abbott et al., 2019a]. The second gravitational-wave transient catalog (GWTC-2) [Abbott et al., 2021b] brought dozens of additional events, and the BBH mass

distribution was found to have a global maximum at $\sim 8 M_\odot$ and an excess of BHs between $\sim 30 M_\odot$ – $40 M_\odot$ followed by a steep, although not infinitely sharp, drop off in the rate at higher masses extending to $\sim 80 M_\odot$ (instead of sharp cutoff at $\sim 40 M_\odot$). At the time, there were not enough observations to determine whether the mass distribution had a local maximum at $\sim 35 M_\odot$, represented by a Gaussian peak on top of a power law, or whether the steepening towards higher masses was better described as a break in the power law [Abbott et al., 2021d].

At the end of the third LVK observing run (O3), the same two features at $\sim 8 M_\odot$ and $\sim 35 M_\odot$ remained, and the feature at $35 M_\odot$ was classified as a peak rather than a break in the power law [Abbott et al., 2023a]. Additionally, non-parametric [Mandel et al., 2017, Rinaldi and Del Pozzo, 2022, Sadiq et al., 2022, Edelman et al., 2023] and semi-parametric [Edelman et al., 2022] analyses found robust evidence for an additional peak at $\sim 10 M_\odot$, the same peak at $\sim 35 M_\odot$, as well as modest evidence for a paucity of events near $\sim 14 M_\odot$ [Abbott et al., 2023a]. These features in the primary mass distribution correspond to similar ones in the chirp mass distribution, occurring at $\sim 9 M_\odot$, $\sim 11 M_\odot$, and $\sim 26 M_\odot$, respectively [Tiwari, 2021, 2022]. The current picture of the BBH mass distribution is therefore a decreasing power law from low to high masses, with a global maximum at $m_1 \sim 10 M_\odot$, a potential underdensity at $m_1 \sim 14 M_\odot$, and an overdensity at $m_1 \sim 35 M_\odot$.

1.2 Spectral sirens: cosmology with the GW source mass distribution

Like light, GWs are redshifted as they propagate across the Universe, thereby bearing imprints of the Universe’s cosmic expansion history. Unlike light, however, the form of GW signals are known from first principles, directly from the theory of GR. Furthermore, because GWs propagate across the Universe without attenuation from intervening matter, and

because the properties of GW detectors are well characterized, GW selection effects are extremely well understood. This allows for a precise estimate of each GW catalogs’ completeness and an unbiased measurement of the true GW source population [Abbott et al., 2021g, 2023a]. This is to be contrasted with many electromagnetic observations, where effects such as reddening, selection bias, and k -corrections can lead to significant complications and systematic uncertainty. Additionally, detected GW signals provide direct measurements of the distance to their sources. This makes them “standard sirens”: direct probes of cosmological parameters that circumvent the need for a cosmological distance ladder [Schutz, 1986, Holz and Hughes, 2005].

A well-known demonstration of standard siren cosmology was the multi-messenger event GW170817 [Abbott et al., 2017c, Coulter et al., 2017, Valenti et al., 2017, Tanvir et al., 2017], whose clear association with a host galaxy provided a precise redshift measurement and allowed for a direct “bright siren” measurement of the Hubble constant, H_0 [Abbott et al., 2017a]. External redshift information can also come from galaxy catalogs, which provide an ensemble of possible redshifts for each GW signal, allowing for a probabilistic “dark siren” measurement of H_0 when multiple GW detections are combined [Del Pozzo, 2012, Chen et al., 2018, Fishbach et al., 2019, Soares-Santos et al., 2019, Gray et al., 2020, Abbott et al., 2021a, 2023b, Gray et al., 2022, 2023, Mastrogiovanni et al., 2023a, Gair et al., 2023].

Electromagnetic information about GW source redshifts need not be available in order to use them as standard sirens, however. GW signals provide direct measurements of each source’s luminosity distance, D_L , and redshifted (detector frame) masses, $m_{\text{det}} = m_{\text{source}}(1 + z)$ [e.g. Chen et al., 2019]. Therefore, if the source frame mass is known, each GW signal provides a direct mapping between luminosity distance and redshift, allowing for a measurement of the expansion of the Universe at the time the GW signal was emitted, $H(z)$.

In practice, the source-frame masses of individual GW signals are not known (unless tidal information is available, e.g. Messenger and Read [2012], Chatterjee et al. [2021]). It is, however, possible to consider the population of CBCs at large and use known features in their source-frame mass distribution to obtain self-calibrated redshift estimates. The full mass distribution therefore acts analogously to an electromagnetic spectrum, in which the apparent locations of spectral features relative to their rest-frame locations provide a redshift measurement. Multiple features must be present in the mass distribution to disentangle the effects of an expanding Universe from astrophysically-induced redshift evolution of the mass spectrum, and more features further aid in breaking this degeneracy [Ezquiaga and Holz, 2022]. The method of using the mass distribution of GW sources to measure cosmological parameters has therefore been coined “spectral sirens” [Ezquiaga and Holz, 2022]. Spectral sirens were first demonstrated to be a feasible method to measure the Hubble constant by Chernoff and Finn [1993] and Taylor and Gair [2012] using the BNS mass distribution, and extended to the binary black hole (BBH) mass distribution by Farr et al. [2019]. Spectral siren analyses have since been implemented by the LVK using the latest GW catalog [Abbott et al., 2023b].

1.3 Outline of this work

In Chapter 2, we describe the most common statistical analysis technique in GW astronomy, hierarchical Bayesian inference. We also give practical considerations for how GW data analysts use this method. Chapter 3 explores the region between the most massive NSs and the least massive BHs, and Chapter 4 focuses on the robustness of features identified in the BBH mass distribution. We then use these features to measure the Universe’s expansion in Chapter 5, pointing out that prior knowledge of their physical origin is not necessary for their use in cosmological measurements. Chapter 6 takes the opposite approach, assuming the Universe’s expansion rate and attempting to determine the astrophysical origin of a

specific feature in the mass distribution by comparing the distribution of more and less massive BHs in the observed BBH population. Finally, Chapter 7 discusses GW events that do not seem to come from the same underlying population as the rest of the catalog, outlining a rigorous method for identifying outliers in the presence of large statistical uncertainties.

CHAPTER 2

HIERARCHICAL BAYESIAN INFERENCE FOR GW POPULATION ANALYSIS

2.1 Introduction

The most common statistical tool used to infer the parameters of the GW source population is hierarchical Bayesian inference. This framework is useful because it self-consistently incorporates the substantial and non-Gaussian measurement uncertainty in each individual event’s parameters, informs the each event’s true parameters simultaneously with the population-level parameters, and accounts for GW detectors’ selection biases [Loredo, 2004, Thrane and Talbot, 2019, Mandel et al., 2019, Vitale et al., 2020].

“Hierarchical” refers to the structure inherent in the method, whereby multiple levels of inference are constructed based on the data generation process. In GW astronomy, hierarchical models typically have three levels: observed data, true individual-event parameters, and population¹ parameters. These are typically denoted by $\{d\}$, $\{\theta\}$, and Λ , respectively. Examples of observed data include observed strain, maximum likelihood parameters, and the best-fit matched-filter template. Examples of true individual-event parameters include the primary mass, secondary spin, right ascension, and redshift of a CBC. Examples of population parameters include cosmological parameters such as H_0 , the maximum allowed NS spin, the slope of the BBH mass distribution, and the efficiency of mass transfer between binary stars of a given metallicity and mass ratio.

Inherent in the structure of a hierarchical Bayesian analysis are our assumptions for how GW data is generated: a population of CBCs exists in the Universe (top level), from which the true parameters of all individual CBC are drawn (middle level). A subset of these individual CBC are observed in our detectors, and their data (bottom level) is influenced

1. The terms “population parameters” and “hyper-parameters” are used interchangeably

both by their true parameters and detector noise. The population parameters, Λ are dictated entirely by astrophysical and cosmological processes (described in Sections 1.1 and 1.2), whereas individual-event parameters are only influenced by these processes through the fact that they are drawn from the population described by Λ . Population analyses are therefore a necessary component of GW astronomy.

2.2 Inferring the population

Using either parametric or non-parametric population models, a hierarchical Bayesian inference can be constructed to determine the appropriate population-level parameters, Λ given the observed set of data $\{d\}$ for N_{ev} observed events.

Observation of GWs is a Poisson² process: it is characterized by the random arrival of discrete events, for which we wish to infer the average expected number, N_{exp} . However, the process is *inhomogeneous*: the number of observations depends on the parameters θ of each event rather than being the same for all CBCs. This means that the likelihood of observing data $\{d\}$ that contains N_{ev} detected GW signals, each with parameters θ_i , is not the homogeneous Poisson likelihood (i.e. $p(N_{\text{ev}}|N_{\text{exp}}) = N_{\text{exp}}^{N_{\text{ev}}} e^{-N_{\text{exp}}} / N_{\text{ev}}!$) and is instead given by [Loredo, 2009, Taylor and Gair, 2012, Mandel et al., 2019, Vitale et al., 2020]

$$p(\{d\}|\{\theta\}, \Lambda) \propto e^{-N_{\text{exp}}(\Lambda)} \prod_i^{N_{\text{ev}}} p(d_i|\theta_i) \frac{dN}{dt_{\text{det}} d\theta}(\theta_i; \Lambda). \quad (2.1)$$

Here, $N_{\text{exp}}(\Lambda)$ is the expected number of detections given the population model and the GW detector sensitivity, and $p(d_i|\theta_i)$ is the likelihood of observing an event's data, d_i , given some true event-level parameters θ_i . Additionally, $\frac{dN}{dt_{\text{det}} d\theta}(\theta; \Lambda)$ is the expected number of

2. While we *observe* a finite number of events, there are an unknown number of CBCs in the Universe. We therefore do not know how many "trials" were made to observe a CBC, so we cannot model CBCs observation as a binomial distribution. We instead make the reasonable assumption that the Universe is extremely large and the number of trials is therefore well-approximated by infinity. This leads to a Poisson distribution, because of the Poisson limit theorem.

events per unit detector-frame time per single-event-parameter hypervolume, conditioned on hyper-parameters Λ . This is more colloquially called the merger rate density of CBCs. Following Callister and Farr [2024], we use a semicolon to explicitly indicate that it is a function of Λ , not a density over Λ . We also assume that GW signals arrive uniformly in detector-frame time, and hence $\frac{dN}{dt_{\text{det}}d\theta}$ only varies with θ . It is common to model the merger rate density as

$$\frac{dN}{dt_{\text{det}}d\theta}(\theta; \Lambda) = \mathcal{R}p(\theta|\Lambda), \quad (2.2)$$

where $\mathcal{R}$ is the total merger rate and $p(\theta|\Lambda)$ is the population model normalized over θ .

We can calculate each expression in Eq. 2.1 with an assumed form for our population model, knowledge of GW detectors and detection pipelines, and pre-existing inference on single-event parameters.

We first discuss $N_{\text{exp}}(\Lambda)$. The expected rate of detected events with parameters θ is the rate of those events in the Universe ($dN/dt_{\text{det}}d\theta$) multiplied by the probability of detecting an event with those parameters and given a noise realization n , $P(\text{det}|\theta, n)$. The total number of expected detections is then obtained simply by summing the contributions from each value of every source parameter, i.e. marginalizing over θ and t_{det} :

$$N_{\text{exp}}(\Lambda) = \int d\theta dt_{\text{det}} \frac{dN}{dt_{\text{det}}d\theta}(\theta; \Lambda) P(\text{det}|\theta, n). \quad (2.3)$$

GW experiments are unique because $P(\text{det}|\theta, n)$ is precisely knowable for GWs from CBCs, which is nontrivial for other astrophysical messengers, which typically have large theoretical uncertainties in the models that connect their source parameters to their observability. Because GR provides a complete mapping from source properties θ to the GW signal, we are able to “undo” GW selection biases via Eq. 2.3. Often in the GW literature, all possible noise realizations are marginalized over to write $P(\text{det}|\theta)$. In practice, though, this marginalization is not precise as a complete detector noise model does not exist. Thus, analyses either

work under the assumption of stationary Gaussian detector noise, or the observed strain data is used as the detector noise realization and no marginalization is performed.

Eq. 2.3 is often calculated using “software injections.” Software injections are a large set of N_{inj} samples drawn from a known distribution $p_{\text{draw}}(\theta^{\text{inj}}, t_{\text{det}}^{\text{inj}})$. These are then “injected” in the data stream by calculating the GR waveforms that would be made by each draw’s source parameters and adding this to the strain measured by the GW detectors. Finally, the search pipelines that are used to identify real GW events are run on the data stream that has been injected with the simulated signals. This results in N_{found} found injections with known true parameters θ^{inj} . Eq. 2.3 can then be approximated with a Monte-Carlo sum over *found* injections, as $P(\text{det}|\theta^{\text{inj}}, n)$ is zero when the injection is not found:

$$N_{\text{exp}}(\Lambda) \approx \frac{1}{N_{\text{inj}}} \sum_j^{N_{\text{found}}} \frac{dN}{dt_{\text{det}} d\theta}(\theta_j^{\text{inj}}; \Lambda) \frac{1}{p_{\text{draw}}(\theta_j^{\text{inj}}, t_{\text{det}}^{\text{inj}})}. \quad (2.4)$$

We will now discuss the products in Eq. 2.1. Since parameters of individual events are not perfectly measured, we marginalize over the possible properties of each event:

$$p(\{d\}|\Lambda) \propto e^{-N_{\text{exp}}(\Lambda)} \prod_i^{N_{\text{ev}}} \int d\theta_i p(d_i|\theta_i) \frac{dN}{dt_{\text{det}} d\theta}(\theta_i; \Lambda) \quad (2.5)$$

If we wished to simultaneously infer the properties of individual events along with the population hyper-parameters, we simply would not marginalize over those events’ parameters, and instead perform a high-dimensional inference, as in Mancarella and Gerosa [2025]. However, this can be impractical, so it is more typical to investigate the different levels of our hierarchical inference separately, examining the hyper-posterior for Λ in the population-level inference after marginalizing over uncertainty in individual-event parameters or examining the event-level population-informed posteriors for individual event parameters after marginalizing over the uncertainty in the inferred population. This allows us to factorize the inference problem, first inferring θ_i for each event, and then using the resulting posteriors in a population

inference, thereby reducing computational cost.

The expression $p(d_i|\theta_i)$ is the individual-event likelihood, which is related to the posterior on θ_i through Bayes' theorem:

$$p(d_i|\theta_i) \propto \frac{p(\theta_i|d_i)}{\pi_{\text{PE}}(\theta_i)}, \quad (2.6)$$

where π_{PE} is the prior probability of the event's θ_i . Individual-event posteriors are obtained via Bayesian parameter estimation on the strain associated with each event, using GR waveforms as the model within which to constrain parameters. A review on this process can be found in Thrane and Talbot [2019], and conventional prior choices can be found in Appendix C of Abbott et al. [2019b]. For our purposes, the parameter estimation process returns samples of θ_i drawn from the posterior in Eq. 2.6, which we can then use to evaluate the integral in Eq. 2.5 via a Monte Carlo average over each event's posterior samples, k :

$$p(\{d\}|\Lambda) \approx e^{-N_{\text{exp}}(\Lambda)} \prod_i^{N_{\text{ev}}} \frac{1}{N_{\text{samps}}} \sum_{k=1}^{N_{\text{samps}}} \frac{\frac{dN}{dt_{\text{det}} d\theta}(\theta_{k,i}; \Lambda)}{\pi_{\text{PE}}(\theta_{k,i})}. \quad (2.7)$$

Having calculated each term in the population likelihood, we can use Bayes' rule to turn it into a posterior on Λ by multiplying Eq. 2.1 by a prior on the hyper-parameters, $p(\Lambda)$.

2.3 Sampling

Sampling Eq. 2.1 to obtain a posterior on Λ can be done with standard sampling algorithms. However, some packages also exist expressly for this purpose, such as `gwpopulation` and the accompanying `gwpopulation-pipe` [Talbot et al., 2024].

A current trend in GW data analysis is to use `jax` to write population models and likelihoods, so as to take advantage of the package's auto-differentiation features and GPU integration. This enables the use of fast sampling algorithms such as No-U-Turn Hamiltonian Monte Carlo [Hoffman and Gelman, 2011], as implemented in codes such as `PyMC3` and `numpyro` [Phan et al., 2019]. It also significantly speeds up more traditional nested sampling

algorithms, such as `dynesty` [Speagle, 2020].

2.4 Directed acyclic graphs

DAGs are conceptual tools that can be helpful in constructing a self-consistent hierarchical model [e.g. Essick and Fishbach, 2024]. They are physical pictures of data generation and provide an algorithmic way of relating this picture to a joint likelihood between all parameters and data in a hierarchical Bayesian inference problem. They are therefore particularly useful in writing and interpreting population likelihoods, such as Eq. 2.1.

The left panel of Fig. 2.1 is an example of a DAG that generically describes the process of constructing an astrophysical catalog. Each level is a step in the data-generation process, and each node is an observation or parameter. At the top level are parameters that describe the population of astrophysical sources and the detectors. These hyper-parameters are considered properties of the Universe. In our case, this would be the form of the CBC population, Λ , and the noise properties of GW detectors, Λ_n . Individual events, described by individual-event parameters, $\{\theta_i\}$, are drawn from this population. This is represented by the second level. Finally, individual event parameters are combined with detector noise, n , to produce the observed data, $\{d_i\}$, as represented by the bottom level. This bottom level could be presented with one node symbolizing the complete dataset. However, in GW data analysis, the data stream is typically broken into segments that are believed to contain only one signal, so we represent the bottom level as discrete observations via discrete nodes. Future experiments where overlapping signals are present, or simultaneous analyses of the stochastic background and resolved signals may benefit from representing the full dataset as a single node. The direction of the arrows enforces physicality: parameters on lower levels do not impact parameters on higher levels, and no process is circular (i.e. $X \rightarrow Y \rightarrow X$ is prohibited) [Reich, 2020]. In other words, the the graph is “directed” and “acyclic.”

We can use this picture to write down a joint distribution of every node in the graph,

$p(\{d\}, \{\theta\}, \Lambda, \Lambda_n)$, using the rule that an arrow from X to Y means that the conditional distribution of Y depends on X and only on X . Thus,

$$p(\{d\}, \{\theta\}, \Lambda, \Lambda_n) = p(\Lambda, \Lambda_n)p(\{\theta\}|\Lambda)p(n|\Lambda_n)p(\{d\}|\{\theta\}, n). \quad (2.8)$$

Detector noise properties and astrophysical processes are typically assumed to be independent, so that $p(\Lambda, \Lambda_n) = p(\Lambda)p(\Lambda_n)$. Another typical, but perhaps less reasonable, assumption is that the detector power spectral densities are known and fixed, so $p(\Lambda_n)$ becomes a Kronecker delta function. Thus, population inference almost always excludes inference of Λ_n .

To confirm that the rules governing DAGs lead to valid joint probability distributions, we now show that Eq. 2.8 follows from Bayes' law and our assumptions about the data-generation process. Bayes' law would yield

$$p(\{d\}, \{\theta\}, \Lambda, \Lambda_n, n) \propto p(\Lambda, \Lambda_n)p(\{\theta\}, n|\Lambda, \Lambda_n)p(\{d\}|\{\theta\}, n, \Lambda, \Lambda_n). \quad (2.9)$$

Then, if we assume (as we did when constructing the DAG) that the observed data does not depend on the astrophysical properties of the Universe except for through the true event parameters, and make the analogous assumption for detector noise, $p(\{d\}|\{\theta\}, n, \Lambda, \Lambda_n)$ becomes $p(\{d\}|\{\theta\}, n)$. We conceptualize detector characteristics as completely independent from the properties of CBCs, which implies that the detector noise realization is drawn independently from individual CBC parameters, and that hyperparameters describing noise do not describe CBC parameters and vice versa, which yields

$$p(\{\theta\}, n|\Lambda, \Lambda_n) = p(\{\theta\}|\Lambda, \Lambda_n)p(n|\Lambda, \Lambda_n) \quad (2.10)$$

$$= p(\{\theta\}|\Lambda)p(n|\Lambda_n). \quad (2.11)$$

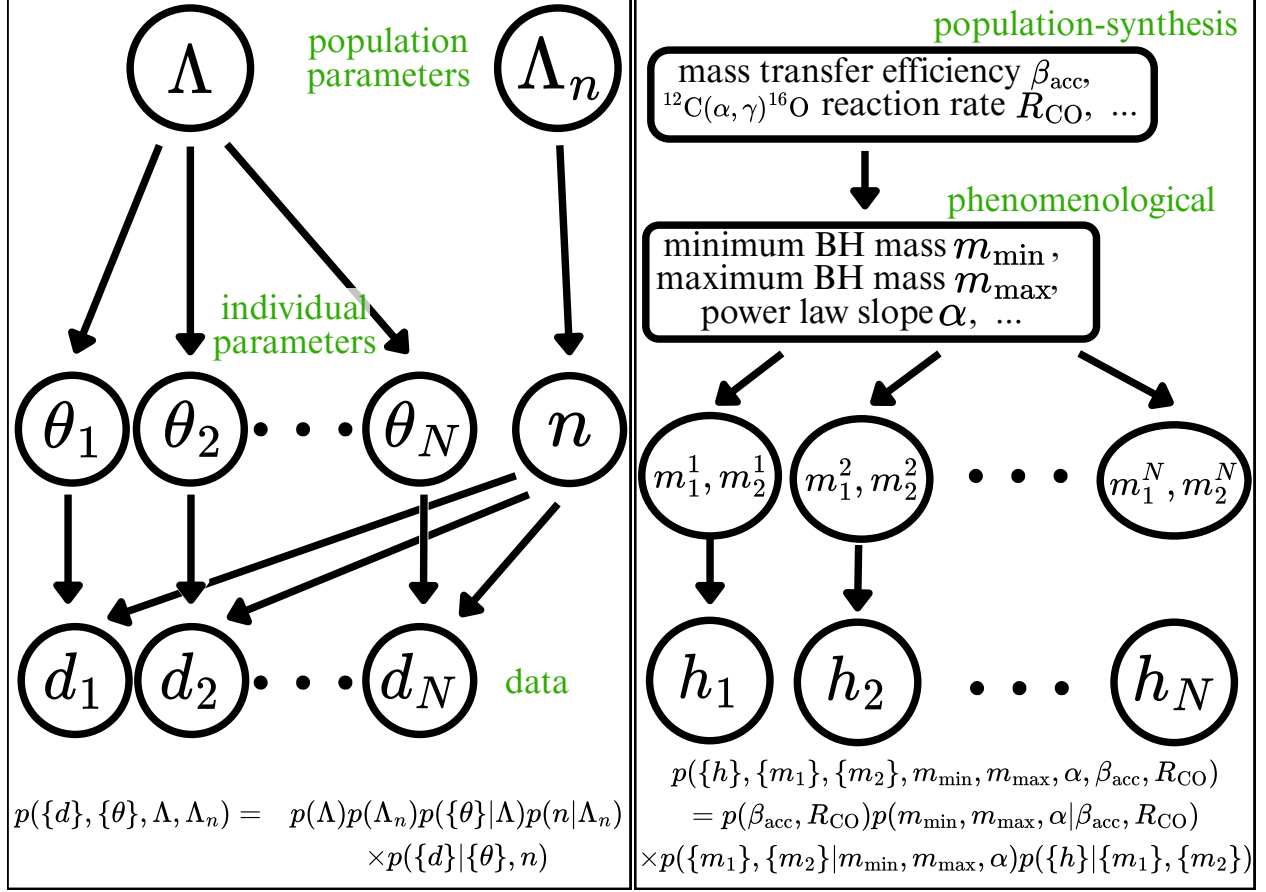


Figure 2.1: Two examples of DAGs. The left panel describes the data generation process for a generic astrophysical catalog. The right panel describes a specific example in GW astronomy, outlined in Section 2.4.2. In both panels, the joint likelihood between all nodes in the DAG is written at the bottom. Each arrow in the DAG represents a conditional distribution, i.e. an arrow from X to Y means that the conditional distribution of Y depends on X .

Making these substitutions into Eq. 2.9 yields Eq. 2.8. This is likely why DAGs are often called “Bayesian networks.”

2.4.1 Example: spectral sirens

For concreteness, we will now provide a generic example of using a DAG for spectral siren cosmology, inferring the mass distribution and cosmological parameters with GW data alone.

At the top level are cosmological parameters $\Lambda_c = \{H_0, \Omega_m, \Omega_\Lambda, \kappa, w\}$ and mass distribution parameters. We will choose to use a phenomenological model to describe the mass

distribution, assuming it can be described by a minimum BH mass, $m_{\min}$, maximum BH mass, $m_{\max}$, and a single power law in between, α . Therefore, $\Lambda = \{\Lambda_c, m_{\min}, m_{\max}, \alpha\}$. By putting the cosmological parameters on the same level as the mass distribution parameters, we are assuming they do not influence each other. For example, a different dark energy equation of state should not impact the maximum allowed black hole mass, which is influenced by nuclear reaction rates.

CBCs are described by two mass parameters, six spin degrees of freedom, the binary's location on the sky, its distance, and its orientation. Together, these parameters produce a true GW strain h^t which is combined with the detector noise to produce the observed strain: $h = h^t + n$. This is the data, d .

Plugging these in to Eq. 2.8 yields

$$\begin{aligned} p(\{d\}, \{\theta\}, \Lambda, \Lambda_n) = & p(\Lambda_c, m_{\min}, m_{\max}, \alpha, \Lambda_n) \\ & \times p(\{m_{1\text{ det}}, m_{2\text{ det}}, \vec{s}_1, \vec{s}_2, \text{RA}, \text{DEC}, D_L\} | \Lambda_c, m_{\min}, m_{\max}, \alpha) \\ & \times p(n | \Lambda_n) p(\{d\} | \{m_{1\text{ det}}, m_{2\text{ det}}, \vec{s}_1, \vec{s}_2, \text{RA}, \text{DEC}, D_L\}, n) \end{aligned}$$

We can now make several simplifying assumptions. First, neither cosmological parameters nor mass distribution parameters influence CBC spins, sky locations, polarization angles, coalescence phases, inclinations, or detector-frame arrival times. Second, we can choose to only infer specific cosmological parameters to which we think our data is most sensitive. Given current detector horizons, high-redshift phenomena such as curvature and the dark energy equation of state will likely be poorly constrained. We can therefore choose to fix such parameters to values inferred by other experiments. This is equivalent to placing a strict prior on these parameters, e.g. $p(\kappa) = \delta(\kappa - 0)$. Third, we assume Λ_n is known exactly

through the measured power spectral density. These assumptions yield

$$\begin{aligned}
p(\{d\}, \{\theta\}, \Lambda) = & p(H_0, \Omega_m, m_{\min}, m_{\max}, \alpha) \\
& \times p(\{m_{1\text{ det}}, m_{2\text{ det}}, D_L\} | H_0, \Omega_m, m_{\min}, m_{\max}, \alpha) \\
& \times p(n | \Lambda_n) p(\{d\} | \{m_{1\text{ det}}, m_{2\text{ det}}, D_L\}, n),
\end{aligned}$$

which is the joint distribution used in Abbott et al. [2023b].

2.4.2 *Example: constraining population synthesis parameters*

The usefulness of a DAG becomes even more apparent when we consider more complicated data-generation schemes. To demonstrate this, we use the example of using GW data to directly constrain processes that occur in binary star systems.

“Population synthesis” codes simulate these stellar processes along with the distribution of stars and stellar environments in the Universe to predict populations of objects that can be compared to observations. GW astronomers therefore frequently talk about astrophysical processes in terms of the parameters used in these population synthesis codes. This in contrast to the many phenomenological population parameters that are typically used in the GW population studies in order to infer the form of the population from the GW data.

Different studies assume different relationships between population-synthesis parameters and phenomenological parameters. Some authors directly compare the two [e.g. Baxter et al., 2021, Stevenson et al., 2019, van Son et al., 2022b], but it is increasingly common to compare the full outputs of population-synthesis simulations (i.e. distributions of CBC parameters) to the distributions recovered from GW data analysis. Some authors have even attempted to fit population-synthesis parameters directly with GW data in a form of simulation-based inference [Zevin et al., 2021, Cheng et al., 2023, Colloms et al., 2025].

We identify two possible conceptualizations of the relationship between population-synthesis

parameters and phenomenological parameters:

1. Population synthesis simulations and the parameters therein fully describe astrophysical processes. These produce a CBC population which can then be described by phenomenological parameters. This is the scenario described by the right panel of Fig. 2.1, where population synthesis parameters are at the top level and phenomenological parameters are at the second level from the top.
2. The CBC population can be described by either population synthesis parameters or phenomenological parameters, or a combination of both. This scenario is described by a DAG in which phenomenological and population-synthesis parameters are all on the top level.

We will now use DAGs to show that these conceptualizations are mathematically equivalent if one set of parameters follows deterministically from the others, but are not if they are probabilistically generated.

For simplicity, we focus on only two population-synthesis parameters: the $^{12}\text{C}(\alpha, \gamma)^{16}\text{O}$ reaction rate in massive stars, R_{CO} , and the accretion efficiency of mass transfer between a star and its compact object companion, β_{acc} . We also only use three phenomenological parameters: m_{max} , m_{min} and α , described in Section 2.4.1. We additionally assume that Λ_n is known exactly and exclude inference on n for the rest of this Section.

The joint distribution of all nodes in the DAG in the right panel of Fig. 2.1 is found by using the same rules as in Section 2.4.1: an arrow from X to Y means that the conditional distribution of Y depends only on X . This yields

$$\begin{aligned}
& p(\{h\}, \{m_1\}, \{m_2\}, m_{\text{min}}, m_{\text{max}}, \alpha, \beta_{\text{acc}}, R_{\text{CO}}) \\
&= p(\beta_{\text{acc}}, R_{\text{CO}}) p(m_{\text{min}}, m_{\text{max}}, \alpha | \beta_{\text{acc}}, R_{\text{CO}}) \\
&\quad \times p(\{m_1\}, \{m_2\} | m_{\text{min}}, m_{\text{max}}, \alpha) p(\{h\} | \{m_1\}, \{m_2\}).
\end{aligned} \tag{2.12}$$

If the phenomenological parameters $\{m_{\min}, m_{\max}, \alpha\}$ are exact deterministic functions of the population-synthesis parameters $\{R_{\text{CO}}, \beta_{\text{acc}}\}$, then $p(m_{\min}, m_{\max}, \alpha | \beta_{\text{acc}}, R_{\text{CO}}) = \delta(\{m_{\min}, m_{\max}, \alpha\} - f(R_{\text{CO}}, \beta_{\text{acc}}))$, where f is the function that parametrizes $\{m_{\min}, m_{\max}, \alpha\}$ in terms of $\{R_{\text{CO}}, \beta_{\text{acc}}\}$. An example of this is the analytic function used to calculate $m_{\min}$ from β_{acc} and four other population-synthesis parameters in van Son et al. [2022b]. We can then marginalize over either the phenomenological parameters or the population-synthesis parameters, depending on which we want to infer. For concreteness, we marginalize over the phenomenological parameters to obtain

$$\begin{aligned}
& p(\{h\}, \{m_1\}, \{m_2\}, \beta_{\text{acc}}, R_{\text{CO}}) \\
&= \int dm_{\min}, dm_{\max} d\alpha [p(\beta_{\text{acc}}, R_{\text{CO}}) \times \delta(\{m_{\min}, m_{\max}, \alpha\} - f(R_{\text{CO}}, \beta_{\text{acc}})) \\
&\quad \times p(\{m_1\}, \{m_2\} | m_{\min}, m_{\max}, \alpha) p(\{h\} | \{m_1\}, \{m_2\})] \\
&= p(\beta_{\text{acc}}, R_{\text{CO}}) p(\{m_1\}, \{m_2\} | f(\beta_{\text{acc}}, R_{\text{CO}})) p(\{h\} | \{m_1\}, \{m_2\}).
\end{aligned} \tag{2.13}$$

We now consider the second way of conceiving the data-generation problem, where population-synthesis parameters and phenomenological parameters are on the same level on the DAG. In this case, the joint distribution is

$$\begin{aligned}
& p(\{h\}, \{m_1\}, \{m_2\}, m_{\min}, m_{\max}, \alpha, \beta_{\text{acc}}, R_{\text{CO}}) \\
&= p(\beta_{\text{acc}}, R_{\text{CO}}, m_{\min}, m_{\max}, \alpha) \\
&\quad \times p(\{m_1\}, \{m_2\} | m_{\min}, m_{\max}, \alpha, \beta_{\text{acc}}, R_{\text{CO}}) p(\{h\} | \{m_1\}, \{m_2\}).
\end{aligned} \tag{2.14}$$

This is different from Eq. 2.12. Now, we again consider the scenario in which the phenomenological parameters are exact deterministic functions of the population-synthesis parameters. In this case, the two sets of parameters are redundant and

$$p(\beta_{\text{acc}}, R_{\text{CO}}, m_{\min}, m_{\max}, \alpha) = p(\beta_{\text{acc}}, R_{\text{CO}}) = p(m_{\min}, m_{\max}, \alpha). \tag{2.15}$$

Performing similar manipulations in all distributions in Eq. 2.14 yields

$$\begin{aligned}
& p(\{h\}, \{m_1\}, \{m_2\}, m_{\min}, m_{\max}, \alpha, \beta_{\text{acc}}, R_{\text{CO}}) \\
& = p(\beta_{\text{acc}}, R_{\text{CO}}) p(\{m_1\}, \{m_2\} | \beta_{\text{acc}}, R_{\text{CO}}) p(\{h\} | \{m_1\}, \{m_2\}),
\end{aligned} \tag{2.16}$$

which is the same as Eq. 2.13. The two conceptions of data generation presented above are therefore equivalent if phenomenological parameters are deterministically generated from population-synthesis parameters. This result is equivalent to the assumption that the data can be fully described either by phenomenological parameters or by population synthesis parameters.

CHAPTER 3

BRIDGING THE GAP: THE TRANSITION BETWEEN NEUTRON STAR AND BLACK HOLE MASSES

One region in which features in the CBC mass distribution may be straightforwardly found is the region between NS and BH masses. This is because nuclear physics prohibits the existence of NSs above $\sim 2.5 M_{\odot}$, and this maximum NS mass may leave an imprint on the CBC mass distribution. Additionally, electromagnetic observations of BHs in X-ray binaries suggest an absence of objects below $5 M_{\odot}$. This mismatch between the maximum observed NS mass and minimum observed BH mass points to the possibility of a “lower mass gap”: a dearth of compact objects between $\sim 2.5 M_{\odot}$ and $\sim 5 M_{\odot}$.

However, most population analyses of GW sources are not sensitive to this region because they artificially truncate their population models at $3 M_{\odot}$, including binaries that have both masses above that truncation in BBH-only population analyses, and binaries with both masses below that truncation in BNS-only analyses. This arbitrary boundary between the two populations excludes events that do not cleanly fall into these categories, and removes the possibility of probing the region near $3 M_{\odot}$, which is precisely our region of interest.

In this chapter, we model the population of all CBCs without first dividing it into sub-populations. Fitting this model to the available GW data, we find evidence for a feature near the predicted maximum NS mass, and hints at a partially-filled lower mass gap. This informs the astrophysics of compact object formation, NS equation-of-state studies, and the formation channels of GW binaries. It additionally enables a data-driven classification scheme for ambiguous systems, each binary is assigned a probability of lying above, below, or within this mass gap. We demonstrate this in Section 3.4.

I later repeated the analysis presented here on the third gravitational-wave transient catalog (GWTC-3) in Abbott et al. [2023a], finding that it is likely the lower mass gap is not entirely empty. This has important implications for specific supernova explosion

mechanisms, as they must now be able to create BHs below $\sim 5 M_{\odot}$. In Kiendrebeogo et al. [2023], Kunnumkai et al. [2024b,a], I then used the results of that analysis to predict the number of BNS, NSBH, and multimessenger events in future observing runs.

This chapter is based on *Bridging the Gap: Categorizing Gravitational-wave Events at the Transition between Neutron Stars and Black Holes*, a paper written in collaboration with Maya Fishbach, Reed Essick, Daniel E. Holz, and Shanika Galaudage, and published as Farah et al. [2022a] in the *Astrophysical Journal* [931].

3.1 Introduction

At the time this chapter was written in June of 2022, the LVK had made 46 unambiguous detections of BBH mergers [Abbott et al., 2019b, 2021b], two detections of BNS mergers [Abbott et al., 2017b, 2020a], and, most recently, two NSBH merger candidates [Abbott et al., 2021c]. However, some of the detected sources could not have obviously been ascribed to one of these source categories. For example, GW190814 [Abbott et al., 2020e] has a secondary mass of $m_2 = 2.59^{+0.08}_{-0.09} M_{\odot}$, making it either the most massive NS or the lowest mass BH detected to date. GW190917_114630, presented for the first time in GWTC-2.1 [Abbott et al., 2021f], is in a similar position. These events lie at the edges of both the NSBH and BBH populations and therefore have the potential to probe the extremes of the subpopulation to which they belong.

However, their classification remains elusive, and in the absence of detectable tides or electromagnetic counterparts that definitively identify their secondaries as NSs, other classification schemes are necessary. Some analyses have directly used NS equation of state (EOS) constraints to compute the probability that a given component mass is below the maximum allowed NS mass [Essick and Landry, 2020, Abbott et al., 2020e, 2021c]. In low latency, the LVK searches used a hard cutoff in component masses of $3 M_{\odot}$ in the third observing run¹

1. See the LVK Public Alerts User Guide: <https://emfollow.docs.ligo.org/userguide/content.h>

[Chatterjee et al., 2020] and $2.83 M_{\odot}$ in the second observing run [Abbott et al., 2019c], both motivated by EOS predictions for the maximum possible NS mass.

However, modern models for the EOS predict different values for the maximum allowed NS mass, typically ranging anywhere between $2.0\text{--}2.5 M_{\odot}$ [Margalit and Metzger, 2017, Ruiz et al., 2018, Shibata et al., 2019, Legred et al., 2021]. See, e.g., Chatziioannou [2020] for a review of recent observational constraints. Additionally, the maximum mass of the galactic population of NSs is currently estimated to be $\sim 2 - 2.6 M_{\odot}$ [Antoniadis et al., 2016, Alsing et al., 2018, Farr and Chatziioannou, 2020, Fonseca et al., 2021] by electromagnetic observations of pulsars.

Meanwhile, electromagnetic observations of BHs in X-ray binaries suggest an absence of objects below $5 M_{\odot}$ [Bailyn et al., 1998, Özel et al., 2010, Farr et al., 2011]. This mismatch between the maximum observed NS mass and minimum observed BH mass points to the possibility of a “lower mass gap”: a dearth of compact objects between ~ 2.5 and $\sim 5 M_{\odot}$.

However, this inferred lack of low mass BHs may be due to systematic errors in inclination measurements of X-ray binaries [Kreidberg et al., 2012]. Correcting for such systematic errors populates the lower mass gap, but a relative paucity of low mass BHs –as opposed to their complete absence– is still likely given the available observations. Additionally, some candidate mass gap objects have been identified in detached binaries [Giesers et al., 2018, Thompson et al., 2019, Jayasinghe et al., 2021] and by microlensing measurements [Wyrzykowski and Mandel, 2020], though Mroz and Wyrzykowski [2021] argue that the objects identified by microlensing actually have masses below the lower edge of the mass gap. It is therefore possible that there may not be a gap in the mass spectrum of compact objects despite the existence of one in the X-ray binary mass spectrum. This may shed light on the evolutionary processes of compact object binary progenitors [Mandel et al., 2021]. Regardless, finding evidence for or against a lower mass gap in the binary merger population will add useful

insight.

Interestingly, there have been very few GW detections in the lower mass gap range, as shown in Fig. 3.1. The low number of GW detections in the purported lower mass gap may be due to lower detector sensitivity to events with low masses (which are intrinsically quieter), but studies that take such selection effects into account find that the BBH distribution cannot be trivially extended to NS mass ranges. For example, Fishbach et al. [2020a] found that a single power law cannot fit the mass distribution of both BBHs and BNSs via an analysis of events in GWTC-1 [Abbott et al., 2019b]. Similarly, Abbott et al. [2021d] find that if the BBH mass spectrum extends below $\sim 4 M_{\odot}$, a deviation from a power law starting at $\sim 8 M_{\odot}$ is necessary to accommodate the BBH observations made in GWTC-2.

This chapter pursues a different approach than that of Abbott et al. [2021d]. We simultaneously model the overall distribution of all compact binaries without first dividing them into subpopulations. In so doing, we may be less susceptible to issues arising from the placement of *ad hoc* boundaries. By examining the total mass distribution, we can first ask whether there is a need to subdivide the observed events into separate categories at all, such as events that fall above, below, or within the lower mass gap, instead of asserting such subpopulations exist *a priori*.

A full description of the lower mass gap has been previously stymied by the small number of detections available. Characterizing the nature of this feature in the overall mass spectrum of compact objects can reveal the true delineation between NSs and BHs in merging binaries, providing a useful compliment to event classification based on cutoffs that are either arbitrary, motivated by external observations of arguably different populations, or reliant on theoretical predictions. With the recent influx of GW detections, enough data is available to make such classification schemes possible.

To resolve the transition between NSs and BHs, we follow the procedure of Fishbach et al. [2020a] (herein FEH20) and model the full spectrum of CBCs. We do not separate

events based on previous classifications, instead using all GW events available to determine if a global fit to component masses alone is able to find two or more distinct subpopulations. Fitting a single mass model across the entire spectrum of detected CBCs has the added benefit of allowing the inclusion of GW events that are not definitively classified into a specific source category.

We further describe our methodology and present the population models used in Sec. 3.2. We show that the transition between NSs and BHs cannot be described by a single power law and explore other morphologies in Sec. 3.3. In Sec. 3.4.2, we demonstrate that the models considered in this work are the first to be able to accommodate GW190814. We describe insights on the NSBH events in Sec. 3.4.3. We then classify all low-mass events with respect to the inferred subpopulation delineations in Sec. 3.4.1, concluding in Sec. 3.5. The appendices provide additional details on model fits.

3.2 Methods

We first describe our selection criteria for which events to include within our analysis in Sec. 3.2.1. We then describe our parametric models for the mass distribution in Sec. 3.2.2 before describing metrics that can be used to describe features in the mass spectrum, sometimes without the need for a particular parametrization, in Sec. 3.2.3. Sec. 3.2.4 describes our inference procedure.

3.2.1 *Event list*

We consider all events in GWTC-2 [Abbott et al., 2021b] that pass this threshold, as well as the two NSBH detections from the second half of the third LVK observing run (O3b) [Abbott et al., 2021c]. This results in 46 events, all of which are shown in Fig. 3.1. Parameter estimation for all events can be found at the the gravitational wave open science center

[Abbott et al., 2021i, GWOSC]².

For the purposes of estimating sensitivity, we treat the NSBH events as if they were detected in the first half of the third LVK observing run (O3a) because data from other detections during O3b were not publicly available at the time of this writing. This choice introduces a modest bias in the inferred rate of NSBH-like events, which is discussed in Section 3.6.1.

3.2.2 Population models

We consider several nested mass models. That is, we describe our conclusions in the context of different sets of assumptions about the mass distribution, each of which corresponds to a specific hyperprior. Following Fishbach and Holz [2020b], FEH20, and Doctor et al. [2020], we parametrize the joint component mass distribution with separate draws for each component mass from a one-dimensional mass distribution and a pairing function. Our one-dimensional mass model can be described by a two-piece broken power law adorned with several multiplicative features that chip off parts of the mass distribution to produce dips and additional features. That is, the one-dimensional distribution is described by

$$p(m|\lambda_{1D}) \propto h(m|m_{\min}, \eta_{\min}) n(m|\gamma_{\text{low}}, \gamma_{\text{high}}, \eta_{\text{low}}, \eta_{\text{high}}, A) \ell(m|m_{\max}, \eta_{\max}) \quad (3.1)$$

$$\times \begin{cases} (m/m_{\text{break}})^{\alpha_1} & \text{if } m < m_{\text{break}} \\ (m/m_{\text{break}})^{\alpha_2} & \text{if } m_{\text{break}} \leq m \end{cases} \quad (3.2)$$

2. Events observed during the first and second LVK observing runs can be found at <https://dcc.ligo.org/LIGO-P1800370/public>, and this work uses the “Overall” samples for BBH events observed during this time and the samples from the IMRPhenomPv2NRT_highSpin waveform for BNS event GW170817. Events observed during the first half of the third LVK observing runs can be found at <https://dcc.ligo.org/LIGO-P2000223/public>, and this work uses the samples from the PrecessingSpin-IMRHM waveform for BBH events observed during this time and the samples from the AlignedSpinInspiralTidal_HS waveform for BNS-like event GW190425. For events GW200105 and GW200115, this work uses the “Combined_PHM_high_spin” samples found at <https://dcc.ligo.org/LIGO-P2100143/public>.

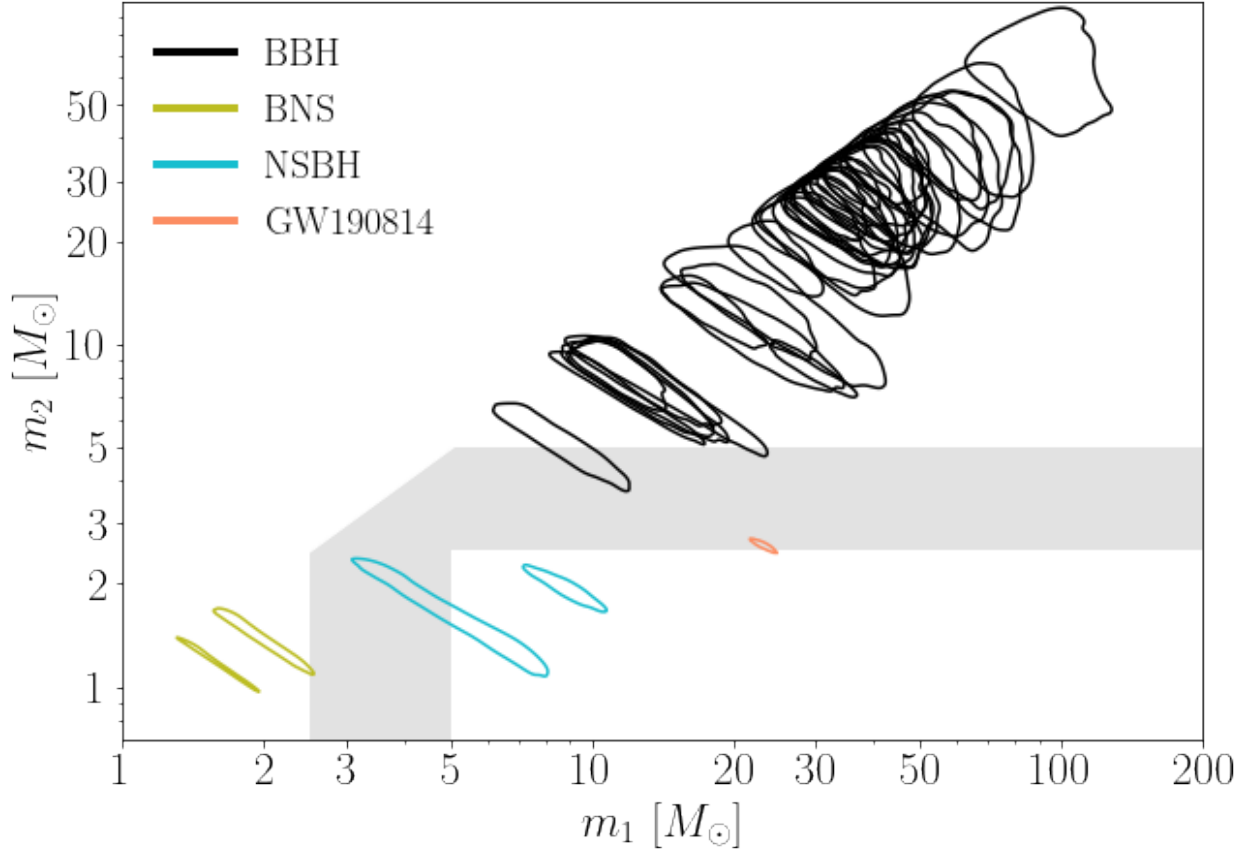


Figure 3.1: 90% posterior credible intervals for the component masses for all CBCs included in this population study assuming uniform priors in detector-frame masses. Events classified by the LVK as BBHs, BNSs and NSBHs are shown in black, green, and blue, respectively. The ambiguously classified event GW190814 is shown in orange. The grey band indicates the approximate location of the purported “lower mass gap.” GW190814 is the only event within this region at more than 90% credibility.

where

$$\ell(m|m_0, \eta) = \left(1 + \left(\frac{m}{m_0}\right)^\eta\right)^{-1} \quad (3.3)$$

is a low-pass Butterworth filter with roll-off mass m_0 and the sharpness of the roll-off set by η (larger η imply sharper roll-offs),

$$h(m|m_0, \eta) = 1 - \ell(m|m_0, \eta) = \left(1 + \left(\frac{m_0}{m}\right)^\eta\right)^{-1} \quad (3.4)$$

is a corresponding high-pass filter, and

$$n(m|\gamma_{\text{low}}, \gamma_{\text{high}}, \eta_{\text{low}}, \eta_{\text{high}}, A) = 1 - A \times h(m|\gamma_{\text{low}}, \eta_{\text{low}}) \times \ell(m|\gamma_{\text{high}}, \eta_{\text{high}}) \quad (3.5)$$

is a notch filter that subtracts a fraction of the signal (set by A) between γ_{low} and γ_{high} . Our one-dimensional mass model builds upon a basic broken power law by adding a high-pass filter at the lowest masses to allow for a fixed but smooth turn-on of the mass distribution, a low-pass filter at the highest masses to allow for a variable tapering of the mass distribution, and a notch filter in between to model a potential lower mass gap. This model has 12 parameters, which we collectively refer to as $\lambda_{1\text{D}}$. It was first presented in FEH20. However, we include an additional high-pass filter at low masses instead of a sharp cut-off at m_{min} . FEH20's model is therefore nested in ours in the limit $\eta_{\text{min}} \gg 1$.

To construct population models for both component masses, we employ a pairing function f_p that relates primary and secondary masses:

$$p(m_1, m_2|\lambda_{1\text{D}}, \beta_{\text{low}}, \beta_{\text{high}}) \propto \Theta(m_2 \leq m_1) p(m_1|\lambda_{1\text{D}}) p(m_2|\lambda_{1\text{D}}) f_p(m_1, m_2|\beta_{\text{low}}, \beta_{\text{high}}) , \quad (3.6)$$

where $\Theta(\cdot)$ is the Heaviside function that enforces the convention $m_2 \leq m_1$.

Note that the one-dimensional distribution's hyperparameters $\lambda_{1\text{D}}$ are shared between

the primary and secondary mass distributions. Note also that the one-dimensional distribution does not, in general, correspond to the marginal distribution for either the primary or secondary mass. We use the pairing function

$$f_p(m_1, m_2 | \beta_{\text{low}}, \beta_{\text{high}}) = \begin{cases} (m_2/m_1)^{\beta_{\text{low}}} & \text{if } m_2 < 5 M_{\odot} \\ (m_2/m_1)^{\beta_{\text{high}}} & \text{if } 5 M_{\odot} \leq m_2 \end{cases} \quad (3.7)$$

to allow for the possibility that most BBHs have different formation channels than NS-containing or lower-mass-gap binaries. The break point in this power law is fixed at $5 M_{\odot}$ and is not fit as a free parameter. This was done to avoid a scenario in which the large number of BBHs dominate the inference on the location of this break point. For example, contamination from relatively high total mass (but low mass ratio) BBHs could dominate the part of the pairing function parameterized by β_{low} . The choice of $5 M_{\odot}$ enforces a separation between the pairing function of BBHs and binaries that contain NS or lower-mass gap objects. We note that this is a different pairing function than that used in FEH20, which employed a single power law in mass ratio for all sources. Again, their model is nested within ours in the limit $\beta_{\text{low}} = \beta_{\text{high}}$.

For simplicity, we assume fixed, independent spin and redshift distributions. We assume sources are uniformly distributed in co-moving volume (V_c) and source frame time so that

$$p(z) \propto \frac{dV_c}{dz} \left(\frac{1}{1+z} \right) \quad (3.8)$$

We also assume spin distributions where each component is independent and follows a distribution that is uniform in magnitude $\chi_{1,2}$ and isotropic in orientation so that

$$p(|\chi|) = \text{U}(0, 1), \quad (3.9)$$

$$p(z) = \text{U}(z) \quad (3.10)$$

where $U(x, y)$ is the uniform distribution between x and y and $z_{1,2} = \cos \theta_{1,2}$ is the cosine of the tilt angle between component spin and a binary’s angular momentum.

Although some authors have shown evidence for the evolution of the mass distribution and/or the merger rate with redshift [Fishbach et al., 2021], and it is unlikely that the spin magnitude distribution is uniform or that the spin tilt distribution is isotropic [Abbott et al., 2021d], we do not expect these choices to affect the mass inference within current statistical uncertainties.

Eqs. 3.2 and 3.7 intentionally provide a lot of flexibility in the functional form of the mass distribution via 14 free parameters while encoding several features predicted by theory. Because the current data may not be able to constrain all the hyperparameters, we consider three nested models in order to demonstrate several of our conclusions more clearly. In order of most to least model freedom,

- **BROKEN POWER LAW + DIP (BPL+Dip)**: assumes the edges of low-mass features are steep, but leaves the locations as free parameters, and fixes the break in the power law to occur at the lower edge of the gap. Sets $\eta_{\min} = \eta_{\text{high}} = 50 \gg 1$ and assumes $m_{\text{break}} = \gamma_{\text{low}}$. Alternatively, at times we assume the gap is deep with a sharp lower edge but allow the upper edge to be fit as a free parameter ($A = 0.99$ and fits η_{high}).
- **BROKEN POWER LAW + DROP (BPL+Drop)**: makes similar assumptions as BPL+Dip, except for $\eta_{\text{high}} = 0$. This *de facto* models a sharp drop-off above γ_{low} and removes the feature at γ_{high} .
- **BROKEN POWER LAW (BPL)**: assumes a steep turn-on at $m_{\min}$ but removes the dip/discontinuity by setting $\eta_{\text{low}} = \eta_{\text{high}} = 0$.

An illustration of these mass models is provided in Fig. 3.2.

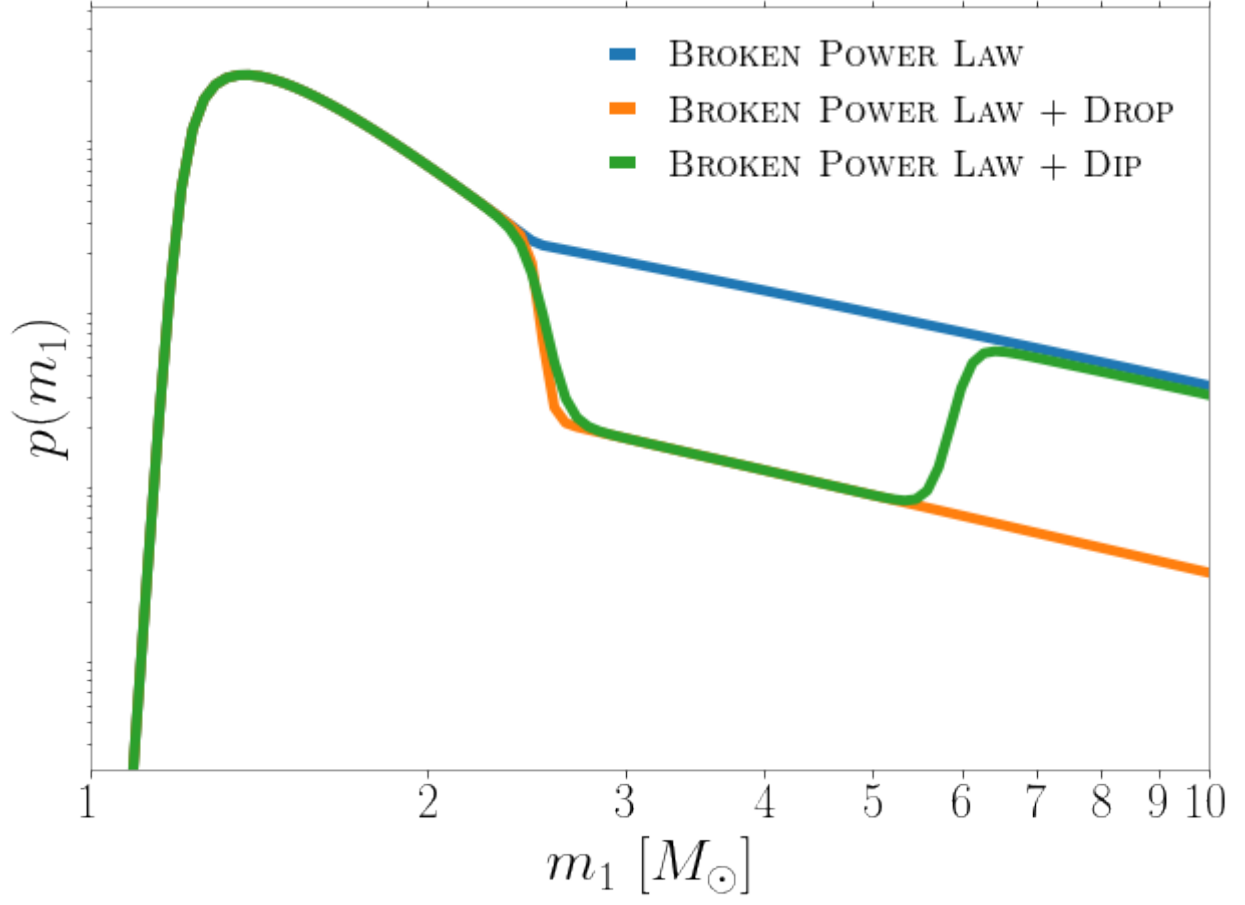


Figure 3.2: Illustration of the primary mass distribution for the three mass models used in this work. All models are depicted with the same values for m_{break} , α_1 , and α_2 .

3.2.3 Characterizing the prominence of features in the mass distribution

In the case of a mass gap with sharp edges ($\eta_{\text{low,high}} \gg 1$), the notch filter strength A is an unambiguous measure of the gap depth, and the parameters γ_{low} and γ_{high} are the precise locations of the beginning and end of the gap, respectively. With this in mind, we can quantify the evidence in favor of a gap with a Bayes factor between models where $A \neq 0$ and $A = 0$: $\mathcal{B}_{A=0}^{A \neq 0}$.

However, this assumes the shape of the gap is well known *a priori*. This is not the case, particularly for the upper edge of the gap. In the case of smooth gap edges, the depth and location of the gap are dictated by a more complex combination of hyperparameters. We therefore define a few straightforward measures of feature prominence which can be used regardless of the exact parametrization adopted.

The first is the relative probability of events between γ_{low} and γ_{high} vs. that of events between m_{min} and γ_{low} :

$$r_{\text{gap}}^{\text{NS}} = \frac{p(m \in [m_{\text{min}}, \gamma_{\text{low}}] \mid \lambda_{1\text{D}})}{p(m \in [\gamma_{\text{low}}, \gamma_{\text{high}}] \mid \lambda_{1\text{D}})}. \quad (3.11)$$

With this definition, $r_{\text{gap}}^{\text{NS}} \gg 1$ indicates a prominent drop-off in the mass distribution at γ_{low} , whereas $r_{\text{gap}}^{\text{NS}} \sim 1$ is indicative of a smoother transition above NS masses.

The second is similar: it is the relative probability of events between γ_{low} and γ_{high} vs. that in an equally wide mass interval starting at γ_{high} :

$$r_{\text{gap}}^{\text{BH}} = \frac{p(m \in [\gamma_{\text{high}}, 2\gamma_{\text{high}} - \gamma_{\text{low}}] \mid \lambda_{1\text{D}})}{p(m \in [\gamma_{\text{low}}, \gamma_{\text{high}}] \mid \lambda_{1\text{D}})}. \quad (3.12)$$

A value of $r_{\text{gap}}^{\text{BH}} \gg 1$ indicates a prominent rise in the mass distribution characteristic of a lower mass gap, whereas $r_{\text{gap}}^{\text{BH}}$ consistent with unity indicates a flat transition between NS and BH masses.

To fully characterize a mass gap, one must examine both $r_{\text{gap}}^{\text{NS}}$ and $r_{\text{gap}}^{\text{BH}}$. The former only quantifies the extent of the initial drop in the mass distribution and the latter only quantifies

the extent of a subsequent rise. For example, if the mass distribution were to monotonically decrease without a gap, $r_{\text{gap}}^{\text{NS}}$ would still be greater than unity. Just using this metric to identify a mass gap would be misleading. Similar effects would occur with $r_{\text{gap}}^{\text{BH}}$ if the mass distribution monotonically increased.

Furthermore, if the upper edge of the mass gap is sharp but the mass distribution falls off steeply after the gap, then $r_{\text{gap}}^{\text{BH}}$ may still be ~ 1 (within this scenario, one may not be able to distinguish between a “gap with sharp edges and rapid roll-off” or a “gentle roll-off with a narrow peak”). It is therefore the combination of $r_{\text{gap}}^{\text{NS}} > 1$, $r_{\text{gap}}^{\text{BH}} > 1$, and A that signals the presence of a gap.

We further discuss potential metrics of feature prominence in Section 3.6.2.

3.2.4 *Statistical framework*

We use the hierarchical Bayesian inference methodology discussed in Chapter 2. We sample from the posterior distribution in Eq. 2.1 using the approximations in Eqs. 2.4 and 2.7 to determine the shape of the mass distribution using `gwpopulation` [Talbot et al., 2019] with the nested sampling algorithm `dynesty` [Speagle, 2020].

We approximate the detectable set of signals as those that yield an optimal network signal-to-noise ratio (SNR) $\rho \geq 11$ based on networks of the LIGO Hanford and Livingston detectors during O1 and O2 as well as the LIGOs and Virgo during O3a. This SNR threshold was chosen to approximately match the SNR at which our chosen set of events is complete: the events included in this analysis follow the expected ρ^{-4} distribution above $\rho \sim 11$.

During each observing run, we use a single representative power spectral density (PSD) to model each detector’s sensitivity separately [LVK, 2015b,a, 2017b,a, 2020] and estimate the network SNR by coherently projecting simulated signals into all detectors. While this semi-analytic approach to sensitivity estimation is commonplace in the literature, we note that it makes several assumptions that may not be perfectly accurate (a single stationary

PSD for each detector, the same SNR threshold for all systems, etc.). The size of the possible systematic errors these assumptions could introduce is not known precisely, but we believe it is less than $\mathcal{O}(10\%)$ based on comparisons of our semianalytic rate estimates with fixed populations to those previously published in the literature using real searches [Abbott et al., 2019b, 2021b].

Where needed, we estimate Bayes factors via Savage-Dickey Density Ratios [Dickey and Lientz, 1970, Wagenmakers et al., 2010] using the hyperposteriors and the hyperpriors.

3.3 Population-level insights

In this section, we search for and characterize features between NS-like and BH-like masses. Following FEH20, we first determine the existence of such features by asking whether a single power law can describe the CBC mass spectrum between $1 - 10 M_\odot$ in Sec. 3.3.1. We find that it cannot. We then describe the nature of this deviation from a power law in Sec. 3.3.3 through the features provided by the parametrized models from Sec. 3.2.2. We provide the inferred posteriors for all hyperparameters in these models in Section 3.6.3.

3.3.1 *Deviation from a single power law describing BH and NS masses*

Neither FEH20 nor Abbott et al. [2021d] include GW190814, nor do they include GW190425, a BNS detected in the first half of the third LVK observing run [Abbott et al., 2020a]. NSBH-like events GW200105_162426 and GW200115_042309 (abbreviated GW200105 and GW200115), which contain secondaries near the upper limits of what can be considered a NS and primaries at lower BH masses than have been previously detected in definitive BBHs, had not been published at the time of either study. These events appear to bridge the NS and BH subpopulations, so the deviations from a single power law in the range $1 - 10 M_\odot$ may no longer be present in the most recent data. We therefore aim to determine if the inclusion of these events removes the feature found by FEH20 and Abbott et al. [2021d].

The first mass model we fit is BROKEN POWER LAW. It has one power law spectral index α_1 at NS-like masses and another α_2 at BH-like masses, with a transition between the two at m_{break} . Draws from the hyperposterior for BROKEN POWER LAW are shown in the top panel of Fig. 3.3. There is a clear transition in the mass distribution at $m_{\text{break}} = 2.4^{+0.5}_{-0.5} M_{\odot}$. This value is consistent with current constraints on the maximum mass of cold, non-spinning NS, which are typically between 2.0-2.5 $M_{\odot}$ at 90% confidence [see, e.g., Essick et al., 2020, Legred et al., 2021, Farr and Chatziioannou, 2020, Chatziioannou, 2020], although larger values are still consistent with low-density nuclear physics.

We find $\alpha_2 - \alpha_1 > 0$ at 99.3% credibility, indicating that a single power law is unable to fit the mass distribution down to low masses. The median and 90% credible limits on α_1 and α_2 are $-5.4^{+2.5}_{-2.2}$ and $-1.0^{+0.36}_{-0.4}$ respectively. This is consistent with findings in FEH20 and Abbott et al. [2021d], and therefore the low mass events GW190814, GW200105, and GW200115 do not fully bridge the BH and NS populations.

As an aside, we also note that $\beta_2 - \beta_1 > 0$ at 99.6% credibility, indicating that the pairing functions between BBHs and CBCs with low secondary mass are distinct. The median and 90% credible limits on β_1 and β_2 are $0.79^{+1.2}_{-0.66}$ and $3.5^{+2.0}_{-1.6}$ respectively. However, because of the fixed location of the break between β_1 and β_2 , we do not use these parameters to locate delineations between subpopulations.

Again, full posteriors on all hyperparameters in the BROKEN POWER LAW model can be found in Section 3.6.3.

3.3.2 *Steep decline after NS masses*

The steepness of the NS part of the power law in BROKEN POWER LAW (i.e. large magnitude of α_1) may be driven primarily by a difference in the merger rates of BNS and BBH systems, rather than by the shape of the NS distribution itself. While our particular analysis has a slight bias in the relative heights of the low mass and high mass parts of the mass distribution

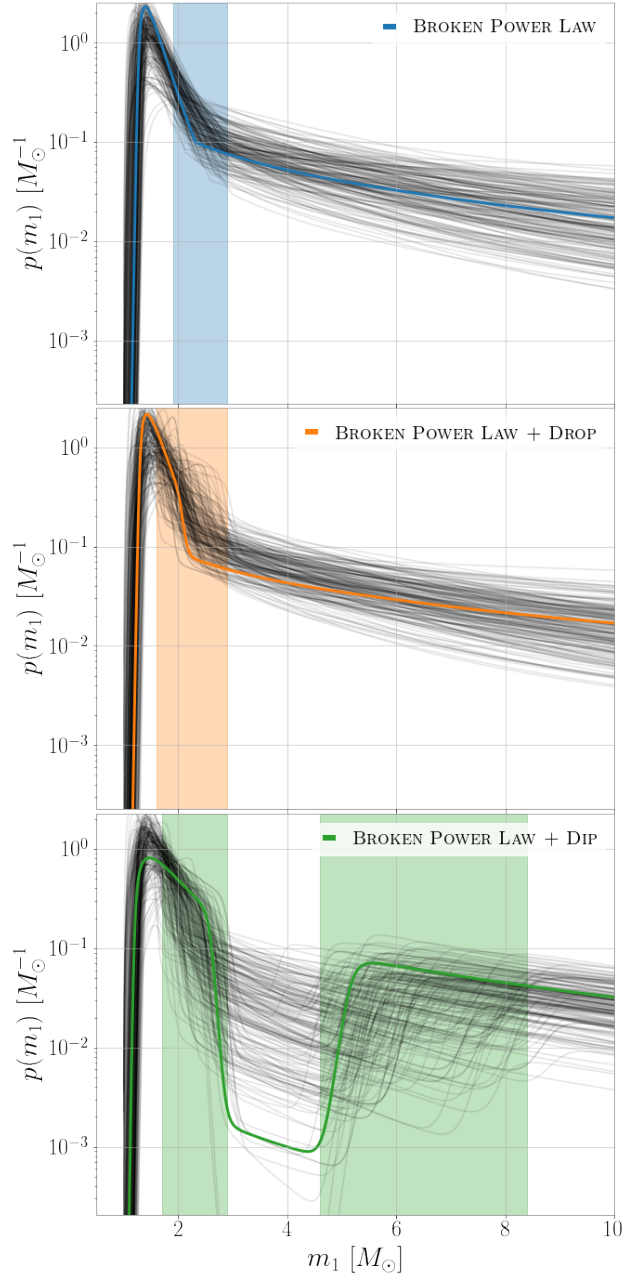


Figure 3.3: Draws from the inferred mass distributions between $1-10 M_{\odot}$ using the BROKEN POWER LAW (*top*), BROKEN POWER LAW + DROP (*middle*), and BROKEN POWER LAW + DIP (*bottom*) models. Each panel has 200 random draws from the hyperposterior (*thin black lines*). As an illustration, the draw that maximizes the overall posterior is highlighted in color. 90% credible intervals on the locations of features in the mass distribution are indicated by colored bands. All three models find a steep falloff in the mass distribution between $\sim 2.2-3 M_{\odot}$ and BROKEN POWER LAW + DIP finds a subsequent rise at $\sim 6 M_{\odot}$.

because of the choice to treat the NSBH events as detected in the first half of the third observing run, many other studies have found the BNS merger rate to be several times higher than the BBH merger rate [Abbott et al., 2019b,a, Fishbach et al., 2020a, Abbott et al., 2021b,d].

To decouple the shape of the NS distribution from the difference in merger rates between BNS and BBH systems, we turn to BROKEN POWER LAW + DROP. This model allows for a drop-off in the mass distribution at m_{break} . The extent of the discontinuity is parameterized by A , which is a free parameter. $A = 2$ corresponds to a maximal discontinuity: in this case, the rate of BH-containing events drops to zero after m_{break} . $A = 0$ corresponds to no discontinuity: in this case, BROKEN POWER LAW + DROP reduces to BROKEN POWER LAW. The resulting fit to this model is shown in the middle panel of Fig. 3.3. A clear drop-off is found after NS masses, which can be quantified by $r_{\text{gap}}^{\text{NS}}$, the ratio of probability mass below m_{break} to that in a comparable interval just above it (Eq. 3.11). We find $r_{\text{gap}}^{\text{NS}} = 4.1_{-3.0}^{+10}$, which excludes 1 to $> 90\%$ credibility.

Fig. 3.4 shows the inferred posteriors for A and α_1 under the BROKEN POWER LAW + DROP framework. The region of parameter space with both small A and low-magnitude α_1 is ruled out, confirming the need for a sharp drop-off in the mass distribution at m_{break} . This feature can be achieved either by a large negative α_1 , as in BROKEN POWER LAW, or by a large A . However, at the largest values of A , α_1 is unconstrained, indicating that the steep drop-off in the merger rate after NS masses is driving the constraints on α_1 , rather than the shape of the NS distribution itself. We therefore conclude that BROKEN POWER LAW is sufficient to describe the limited data available. This is consistent with the findings of Landry and Read [2021], who show that the mass distribution of NSs in binaries cannot be determined from GW data alone (yet).

Within BROKEN POWER LAW + DROP, the location of the break in the power law (and therefore the location of the discontinuity) is inferred to be at $m_{\text{break}} = 2.20_{-0.41}^{+0.51} M_{\odot}$,

which is consistent with but shifted lower than the corresponding value inferred by BROKEN POWER LAW ($2.4_{-0.5}^{+0.5} M_{\odot}$). This is due to the different modelling assumptions inherent in both models, and both values are equally plausible. Furthermore, given the one-dimensional posterior on A , we find that the data cannot distinguish between $A = 0$ and $A \neq 0$ and therefore have no clear preference for BROKEN POWER LAW or BROKEN POWER LAW + DROP. This conclusion is corroborated by the Bayes factors presented in Section 3.6.3, which are similar between the two models. We therefore utilize BROKEN POWER LAW for the event-level analyses in this work due to its simplicity and fewer free parameters.

3.3.3 Searching for a lower mass gap

Having confidently established a deviation from a single power law in the form of a steep decrease in the mass distribution after NS-like masses, we wish to identify other features. In particular, we search for signatures of the purported lower mass gap.

To do this, we employ the BROKEN POWER LAW + DIP model, described in Sec. 3.2.2. The fit to this model is shown in the bottom panel of Fig. 3.3 assuming $\eta_{\text{high}} = 50$. Note that the maximum-posterior draw (highlighted in green in Fig. 3.3) is not necessarily required to lie in the bulk of other draws. This is true for all three models considered but is especially apparent for BROKEN POWER LAW + DIP. Values that maximize the hyperposterior for all models can be found in Section 3.6.3.

We infer the depth of a potential mass gap through the hyperparameter A , where $A = 0$ reduces to the BROKEN POWER LAW mass model and $A = 1$ corresponds to a maximal gap. The lower and upper edges of this gap are represented by γ_{low} and γ_{high} , respectively. We infer $\gamma_{\text{low}} = 2.2_{-0.5}^{+0.7} M_{\odot}$ and $\gamma_{\text{high}} = 6.0_{-1.4}^{+2.4} M_{\odot}$. These are consistent with the values found in FEH20 ($\gamma_{\text{low}} = 2.2_{-0.5}^{+0.6} M_{\odot}$, $\gamma_{\text{high}} = 6.7_{-1.5}^{+1.0} M_{\odot}$). This agreement with a GWTC-1 analysis suggests that the addition of events in GWTC-2 and the NSBH-like events are consistent with this previously inferred population and do not fully fill the lower mass gap,

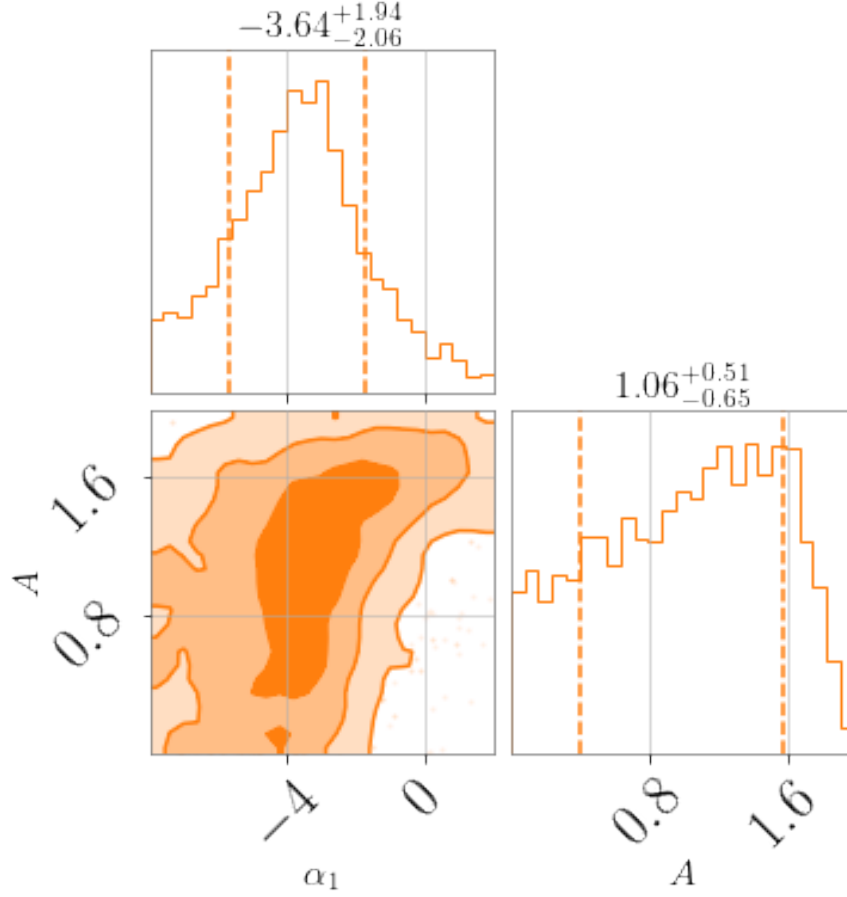


Figure 3.4: Values of the spectral index governing the low mass part of the spectrum α_1 and drop depth A inferred by the BROKEN POWER LAW + DROP model. Contours indicate 1- σ , 2- σ , and 3- σ credible regions. The two parameters are correlated and the lower right region is excluded: mild power law slopes are only allowed when A is large. Thus, the inference finds a dramatic drop-off in the merger rate, either through a steep negative initial power-law slope (α_1), or a discontinuity (large A). The two scenarios are both allowed by the data due to the inability to resolve the shape of the NS distribution.

if one exists. These events may, however, indicate that a lower mass gap might not be completely empty.

Within the framework of BROKEN POWER LAW + DIP, we find suggestions of a lower mass gap, but are unable to conclusively detect or rule out its existence. The posterior on A is shown in Fig. 3.5. It peaks at large values but does not rule out the possibility that some events occupy the gap. Nevertheless, a model with a maximal gap ($A = 1$) is favored over no gap ($A = 0$) with a Bayes factor of 55.0.

However, $r_{\text{gap}}^{\text{BH}} = 1.84^{+5.74}_{-1.14}$, which suggests the existence of a lower mass gap but does not exclude $r_{\text{gap}}^{\text{BH}} = 1$. This may be due to the fact that the power law slope governing the BH part of the spectrum is negative, which can lead to $r_{\text{gap}}^{\text{BH}} \sim 1$ even in the presence of a gap. However it is also possible that the assumption of very sharp gap edges may be driving the preference for large A within BROKEN POWER LAW + DIP.

Due to the nature of the notch filter $n(m|A, \gamma_{\text{low}}, \gamma_{\text{high}}, \eta_{\text{low}}, \eta_{\text{high}})$ used in BROKEN POWER LAW + DIP, a sharp drop-off in the mass distribution after γ_{low} will always be accompanied by a subsequent rise at γ_{high} when we set $\eta_{\text{low}} = \eta_{\text{high}} \gg 1$. Thus, if the data prefer such a sharp drop-off that is not achieved through a steep α_1 , using BROKEN POWER LAW + DIP may artificially inflate our metrics of gap significance. This systematic is due in part to the choice to set $\eta_{\text{low}} = \eta_{\text{high}} = 50$, rather than leaving the steepness of the gap edges as free parameters. We therefore seek to determine if the data actually prefer a subsequent rise in the mass distribution after the initial drop at γ_{low} by allowing η_{high} to be a free parameter. Positive values of η_{high} indicate the existence of a subsequent rise, and larger positive values imply steeper rises. To lessen the number of free parameters being fit to a region with relatively few detections, we enforce a drop in the mass distribution after NS masses by fixing $\eta_{\text{low}} = 50$ and $A = 0.99$. We are therefore asking, given the existence of a drop in the mass distribution at γ_{low} , do the data necessitate a subsequent rise?

Fig. 3.6 shows the two dimensional posteriors on γ_{high} and η_{high} , the parameters deter-

mining the location and steepness of the upper edge of the mass gap. We find $\eta_{\text{high}} > 0$ at 99.8% credibility, with lower 90% credible bound at $\eta_{\text{high}} = 2.82$, indicating that if there is a large drop in the mass distribution at γ_{low} , the data prefer a subsequent rise at γ_{high} rather than an additional drop ($\eta_{\text{high}} < 0$) or a featureless transition ($\eta_{\text{high}} = 0$).

In this model fit, $\gamma_{\text{high}} = 5.98^{+0.99}_{-0.87} M_{\odot}$, which is completely consistent with the location of the global maximum of the BBH distribution found in Abbott et al. [2021d]. They find $7.8^{+1.8}_{-2.0} M_{\odot}$ for the POWER LAW + PEAK model and $6.02^{+0.78}_{-1.96} M_{\odot}$ for the BROKEN POWER LAW model, the latter of which is more morphologically similar to the high-mass behavior of the models considered in this work.

To determine if the assumption of $A = 0.99$ is requiring a subsequent rise in the mass distribution, we repeat the analysis but set $A = 0.5$. We find nearly identical results, indicating that a subsequent rise is preferred even with a moderate drop at γ_{low} .

As a final test of our results, we perform a fit where all hyperparameters are allowed to vary in Section 3.6.2. This fit tends to prefer the existence of a lower mass gap, though the results are far from conclusive. However, due to the small number of detections at low masses, we cannot meaningfully constrain more than two hyperparameters in this region. We therefore fix $\eta_{\text{high}} = 50$ and allow A to vary for the remaining analyses, noting that a steep upper edge and variable gap depth allows for more intuitive interpretations of the event classification presented in Sec. 3.4.

3.4 Event-level insights

Building upon our population-level insights, we now place individual events in the context of the inferred features in the mass distribution. Sec. 3.4.1 discusses how we do this in general, marginalizing over both the uncertainty in individual events' parameters as well as the population hyperparameters. We then discuss the implications for specific events that previous models have not been able to accommodate: GW190814 (Sec. 3.4.2) and the two

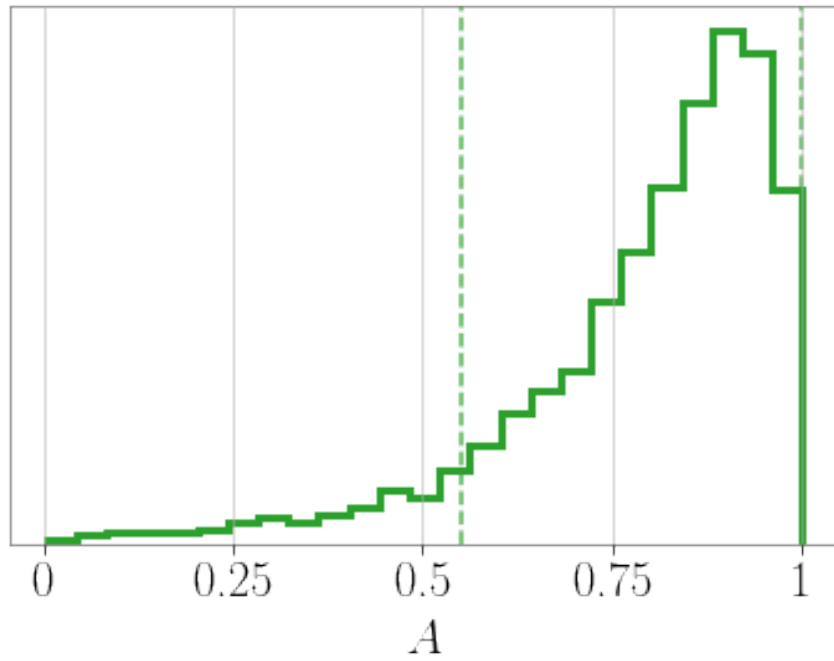


Figure 3.5: Posterior on the gap depth parameter A under the BROKEN POWER LAW + DIP model including all events. $A = 1$ indicates a maximal gap, and $A = 0$ indicates no mass gap. A peaks at 0.90, with 90% highest density interval between 0.55 and 1.0 (*vertical dashed lines*), indicating support for both a partially-filled and totally empty mass gap.

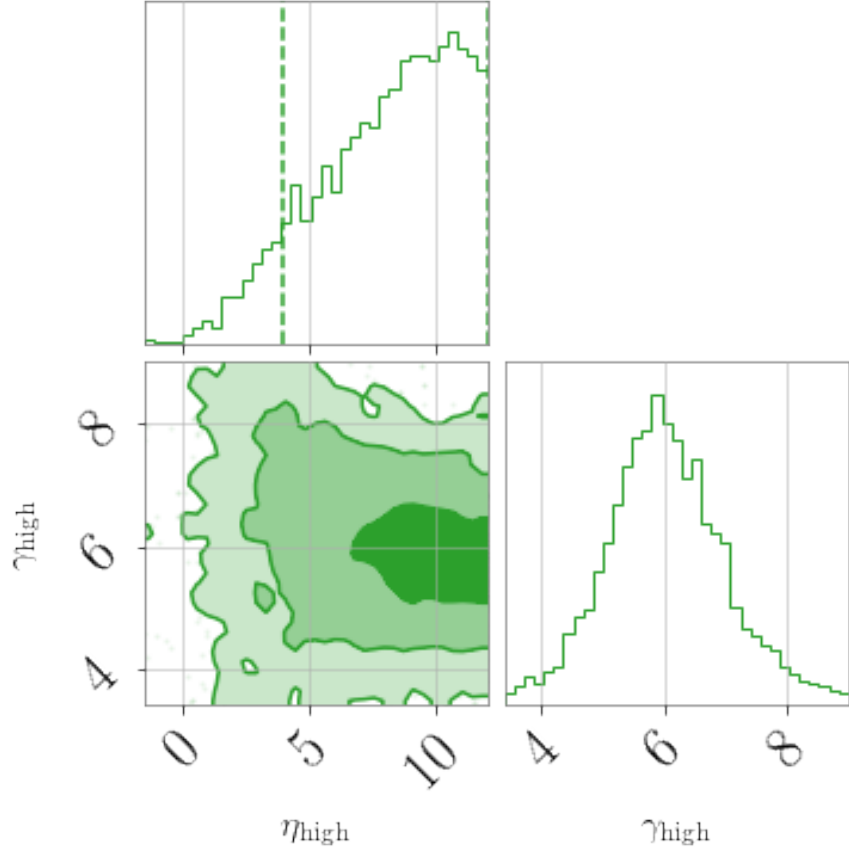


Figure 3.6: Corner plot of the parameters governing the location (γ_{high}) and steepness (η_{high}) of the upper edge of the mass gap in the case where a steep decline in the mass distribution is enforced after NS masses ($A = 0.99$). The 90% highest probability density interval for η_{high} is indicated by vertical dashed lines. η_{high} is found to be positive, indicating a rise in the mass distribution near γ_{high} after the drop enforced at γ_{low} .

NSBH-like events GW200105 and GW200115 (Sec. 3.4.3).

3.4.1 Event classification

Fits to the full spectrum of CBCs allow us to classify events with respect to features inferred from the detected population.

Tab. 3.1 uses the BROKEN POWER LAW model fit to characterize events in relation to m_{break} . It lists the probabilities that each event’s component masses lie in regions of interest, marginalized over the uncertainties in each event’s component masses and the uncertainties in the relevant model hyperparameters. For example, if one interprets m_{break} to be the maximum NS mass, the probability that an event contains a NS is $P(m_2 \leq m_{\text{break}})$ and the probability that an event is a BNS is $P(m_1 \leq m_{\text{break}})$, since $m_2 \leq m_1$ by definition. Similarly, if we consider objects with masses larger than m_{break} to be BHs, the probability that an event is a NSBH is $P(m_2 \leq m_{\text{break}} < m_1)$. Using these assumptions, we find that all probabilities presented in Tab. 3.1 are qualitatively consistent with the classifications given to each event by the LVK.

GW170817 is found here to have $P(m_1 \leq m_{\text{break}}) = 0.99$, which is consistent with the LVK’s classification of this event as a BNS. GW190425 has $P(m_1 \leq m_{\text{break}}) = 0.9$, with $< 10\%$ support for a NSBH classification. This is consistent with this event’s classification as a BNS. NSBH events GW200105 and GW200115 both have larger $P(m_2 \leq m_{\text{break}} < m_1)$ than their respective $P(m_1 < m_{\text{break}})$ or $P(m_2 > m_{\text{break}})$, so under this framework they are more likely NSBHs than BNSs or BBHs. The secondary mass of GW190814 is neither definitively a NS nor a BH under this classification scheme: the posterior on m_{break} overlaps almost completely with that of GW190814’s m_2 .

These events’ classifications in the BROKEN POWER LAW + DROP formalism are similar, though the lower inferred value of γ_{low} (and therefore m_{break}) decreases the probability that each component is below m_{break} by $\lesssim 10\%$ for the BNS and NSBH-like events, and by $\sim 20\%$

for GW190814.

Tab. 3.2 uses the BROKEN POWER LAW + DIP model fit to characterize events in relation to γ_{low} and γ_{high} . Most events that have been previously identified by the LVK as containing a NS (GW170817, GW190425, GW200115) all have $P(m_2 \leq \gamma_{\text{low}}) > 0.95$. BNS event GW170817 has $P(m_1 \leq \gamma_{\text{low}}) > 0.95$ as well, but BNS event GW190425 has $\sim 10\%$ support of its primary being in the mass gap rather than a NS, and NSBH-like event GW200105 has $\sim 10\%$ support of its secondary being in the mass gap rather than an NS. This is likely due to the relatively large uncertainties on the component masses of GW190425 and GW200105 and on γ_{low} . Events previously identified by the LVK as being consistent with NSBHs [Abbott et al., 2021c, GW200105, GW200115] receive a similar classification in this analysis, with large $P(m_2 \leq \gamma_{\text{low}})$, and near-vanishing $P(m_1 \leq \gamma_{\text{low}})$.

Additionally, the component masses of three events have considerable support in the mass gap: the secondary of GW190814, the secondary of GW190924_021846, and the primary of GW200115. While a totally empty gap is not ruled out—none of the three have more than 75% support in the gap—the combined effect of these three events is to decrease the inferred depth of the lower mass gap. The secondary mass of GW190814 remains ambiguously classified: it has non-negligible support of being either in the lower mass gap or being a NS. As the posterior on m_2 for GW190814 is well constrained, the ambiguous classification is instead due to the inference’s inability to determine whether the event should be positioned in the first power law or in the notch filter of BROKEN POWER LAW + DIP. This results in bimodality in γ_{low} ’s posterior distribution, which we discuss further in Sec. 3.4.2. As more detections of GW events constrain the global mass distribution, uncertainties in the locations of γ_{low} and γ_{high} will decrease, allowing for a more definitive classification of this event.

We further examine three events that were published as “exceptional” by the LVK.

Probability	$P(m_2 \leq m_{\text{break}})$	$P(m_1 \leq m_{\text{break}})$	$P(m_2 \leq m_{\text{break}} < m_1)$
Interpretation	$P(\text{contains a NS})$	$P(\text{is a BNS})$	$P(\text{is a NSBH})$
GW170817	> 0.99	0.99	0.01
GW190425	> 0.99	0.90	0.09
GW190814	0.35	< 0.001	0.35
GW200105	0.88	< 0.001	0.88
GW200115	0.94	< 0.001	0.94

Table 3.1: Events classified in the BROKEN POWER LAW formalism. Values are the probabilities that each event’s component masses lie in a given region of parameter space, considering uncertainty in the component masses as well as in the inferred values of m_{break} . If m_{break} is interpreted as the delimiter between NS and BH masses, the probabilities can be understood as labeled in the second row. The uncertainties in these probabilities are estimated to be less than 1%. Only events with > 5% probability of containing a NS according to this classification are included in this table.

46

Probability	$P(m_2 \leq \gamma_{\text{low}})$	$P(m_1 \leq \gamma_{\text{low}})$	$P(\gamma_{\text{low}} \leq m_2 \leq \gamma_{\text{high}})$	$P(\gamma_{\text{low}} \leq m_1 \leq \gamma_{\text{high}})$
Interpretation	$P(\text{contains a NS})$	$P(\text{is a BNS})$	$P(m_2 \text{ in mass gap})$	$P(m_1 \text{ in mass gap})$
GW170817	> 0.99	> 0.99	< 0.001	< 0.001
GW190425	> 0.99	0.87	0.01	0.13
GW190814	0.26	< 0.001	0.74	< 0.001
GW190924_021846	< 0.001	< 0.001	0.49	0.05
GW200105	0.81	< 0.001	0.19	0.02
GW200115	0.96	< 0.001	0.04	0.44

Table 3.2: Events classified in the BROKEN POWER LAW + DIP formalism. Values are the probabilities that each event’s component masses lie in a given region of parameter space, considering uncertainty in the component masses as well as in the inferred values of γ_{low} and γ_{high} . If γ_{low} and γ_{high} are interpreted as the lower and upper edges of the mass gap and γ_{low} is therefore interpreted as the maximum observed NS mass in merging binaries, the probabilities can be understood as labeled in the second row. The uncertainties in these probabilities are estimated to be less than 1%. Only events with > 5% probability of being in one of these categories are included in this table.

3.4.2 *GW190814*

GW190814 was observed in the first half of the third observing run [Abbott et al., 2020e] and has the most asymmetric mass ratio of any gravitational wave event to date. Its secondary has a mass of $m_2 = 2.59^{+0.08}_{-0.09} M_\odot$, leaving the nature of this component ambiguous: depending on the choice of maximum allowed NS mass, it can either be classified as a low-mass BH or the most massive NS observed to date. Previous “leave-one-out” studies [Abbott et al., 2021d, Essick et al., 2022] have definitively identified GW190814 as an outlier with respect to the detected BBH population. Additionally, Landry and Read [2021] find that including the secondary of GW190814 in the population of NSs in merging binaries considerably changes the inferred mass distribution when compared to a fit that excludes this event, though uncertainties in both fits are large. It is therefore desirable to find a model that can account for such an extreme event.

We perform a similar study with GW190814 by comparing the mass distributions inferred with the full event list to those inferred by excluding GW190814. We find that in the cases of BROKEN POWER LAW, BROKEN POWER LAW + DROP, and BROKEN POWER LAW + DIP, the inferred distributions without GW190814 are consistent with those with GW190814. In the frameworks of BROKEN POWER LAW and BROKEN POWER LAW + DROP, the posterior of m_{break} is largely unchanged with the exclusion of this event, but the posterior distribution of β_1 broadens and shifts to higher values when excluding GW190814. In the framework of BROKEN POWER LAW + DIP, the posterior of γ_{low} when GW190814 is included in the inference is in some disagreement with that inferred without GW190814. When GW190814 is included, the posterior on γ_{low} is bimodal, with the lower mode corresponding to putting GW190814’s secondary in the gap and the higher mode corresponding to locating the gap at higher masses than GW190814’s secondary. Excluding GW190814 keeps only the lower mode. Additionally, the posterior on A peaks at a slightly higher value when GW190814 is excluded. Nonetheless, the inferred values with and without GW190814 are still consistent

with one another, as can be seen in Fig. 3.7. Therefore, both BROKEN POWER LAW and BROKEN POWER LAW + DIP successfully account for this event.

Note that such leave-one-out analyses can introduce biases that may lead to the false classification of the left-out event as an outlier unless modifications to the likelihood are used to mitigate this effect [Essick et al., 2022]. However, since these modifications tend to bring the hyperposterior inferred without the event in question closer to that inferred with all events, we assume the conclusions derived from Fig. 3.7 will remain the same with the use of the methods presented in Essick et al. [2022].

3.4.3 *GW200105 and GW200115*

GW200105 and GW200115 are the most definitive detections of NSBH-like systems via gravitational waves to date, observed by the LVK during the second half of the third observing run [Abbott et al., 2021c]. GW200105 has component masses $m_1 = 8.9^{+1.2}_{-1.5} M_\odot$ and $m_2 = 1.9^{+0.3}_{-0.2} M_\odot$, and GW200115 has component masses $m_1 = 5.7^{+1.8}_{-2.1} M_\odot$ and $m_2 = 1.5^{+0.7}_{-0.3} M_\odot$. Both events therefore have components near the edges of the purported lower mass gap, and analyzing them in the context of the BROKEN POWER LAW + DIP model is helpful in understanding their classification. Likewise, these events play a large role in probing the values of γ_{low} and γ_{high} inferred in this analysis.

We first determine if GW200105 and GW200115 are consistent with the rest of the population of CBCs by performing a model fit to BROKEN POWER LAW + DIP without these events and comparing it to the population inferred with these events. Section 3.6.1 shows these two fits to the data. We find that the two inferred populations are consistent with one another, and that the NSBH events have the effect of constraining γ_{low} and γ_{high} towards a narrower gap. This effect, as well as the fact that both events have $P(\gamma_{\text{low}} \leq m_i \leq \gamma_{\text{high}}) < 0.5$, indicate that GW200105 and GW200115 are both consistent with being “gap-straddling binaries,” though there is still a definite possibility that either or both events have one com-

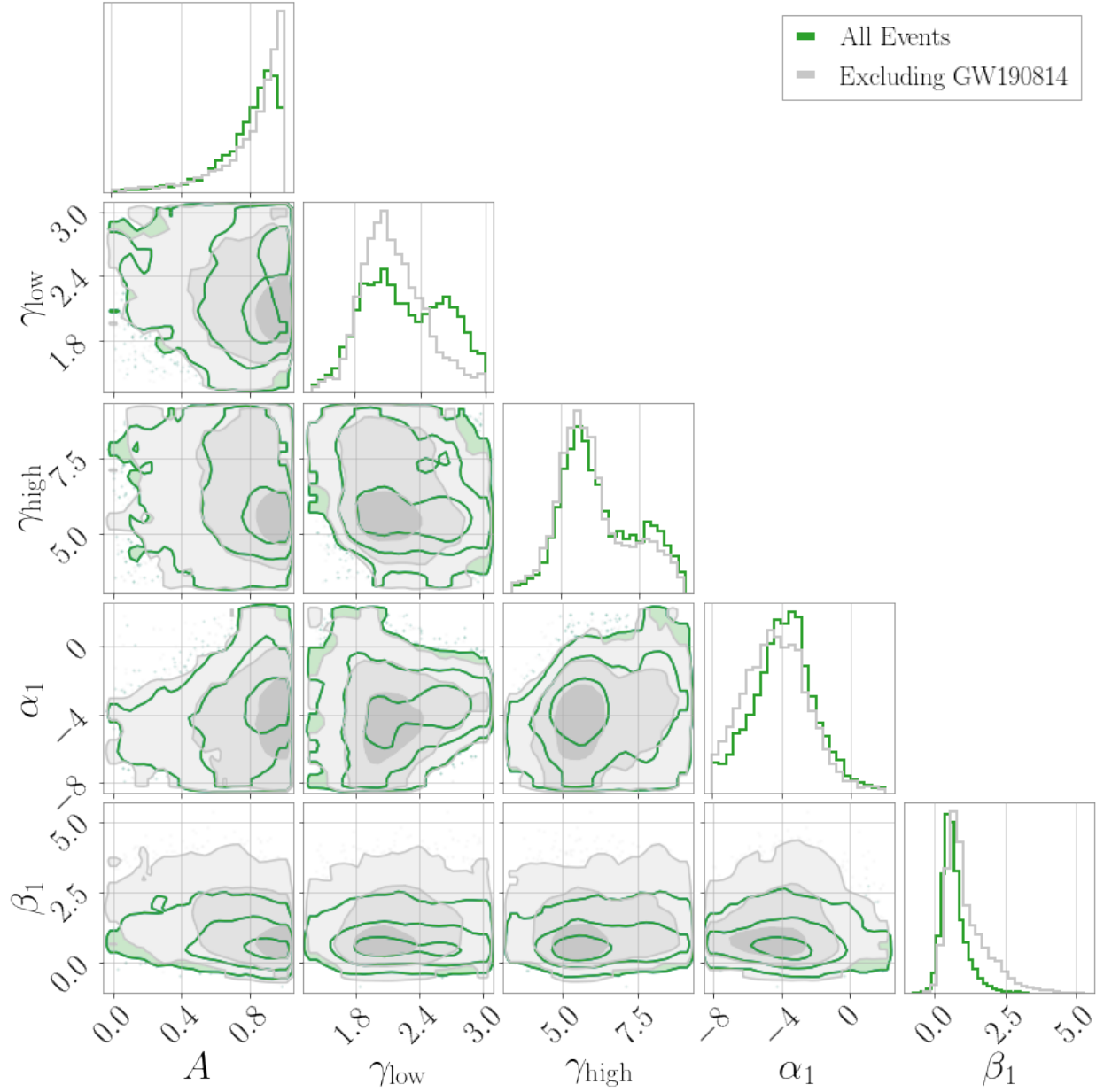


Figure 3.7: Effects of leaving out event GW190814 under the BROKEN POWER LAW + DIP model. Green contours correspond to the inference when GW190814 is included, and grey contours correspond to the inference when it is excluded. Almost all hyperparameters remain unchanged with the inclusion vs exclusion of GW190814, indicating that this event is consistent with the inferred population using this model. The only parameter that exhibits a noticeable change is γ_{low} , where the bimodality of the posterior distribution disappears with the exclusion of GW190814.

ponent in the lower mass gap.

3.5 Discussion

A lower mass gap between NSs and BHs has been suggested by observations of X-ray binaries that find a lack of BHs below $5 M_{\odot}$ [Bailyn et al., 1998, Özel et al., 2010, Farr et al., 2011] despite predictions for the maximum allowed NS mass by the NS equation of state [Margalit and Metzger, 2017, Ruiz et al., 2018, Shibata et al., 2019, Chatziioannou, 2020, Legred et al., 2021] of $\sim 2 - 2.5 M_{\odot}$. Such a dearth of systems just above the maximum NS mass is in contrast with theoretical studies of compact object formation scenarios which predict a continuous distribution of supernova remnant masses that have a smooth transition from NSs to BHs [e.g., Fryer and Kalogera, 2001, Fryer et al., 2012, Kushnir, 2015, Zevin et al., 2020, Drozda et al., 2020].

In this chapter, we explore whether there is evidence for a lower mass gap in the existing gravitational wave data. Inferences on the edges of such a gap have previously been carried out by examining the extrema of separate BH and NS distributions. However, this approach lessens the amount of information available to both analyses, and can lead to biased inferences. It is insufficient to simply look at the largest NS mass and smallest BH mass in the catalog, and make statements about a mass gap. *The best estimates for individual compact object masses can only be achieved through an analysis of the full population, including systems on both sides of the gap as well as those in the gap (if any).* Fitting a single model across the entire spectrum of compact binary coalescences allows us to find features that are difficult to infer from fits to the BBH, BNS, or NSBH mass distributions alone.

We present global fits to the mass distribution of NSs and BHs in merging binary systems in Fig. 3.3. All models considered find a sharp drop in the mass distribution at $\sim 2 - 3 M_{\odot}$,

either via a steep power law (*top panel*) or a sharper “discontinuity” (*middle and bottom panels*). Although this feature has been determined solely from the mass distribution of GW systems, it presumably relates to the maximum mass of NSs in GW binaries. It is suggestive that this break occurs precisely in the mass range predicted from equation-of-state calculations Legred et al. [2021]. More detections at NS masses will inform the shape of the mass spectrum below $3 M_{\odot}$, allowing us to better resolve the nature of this drop-off.

For models that allow for a feature in the mass distribution after the initial “neutron star” drop, we find a subsequent rise at $\sim 4.5 - 8.5 M_{\odot}$ (bottom panel of Fig. 3.3). While this hints at a potential mass gap between NSs and BHs, we cannot confidently detect nor rule out the existence of such a feature with the limited detections and low sensitivity in that region.

As we emphasize above, understanding a single GW event can only be done in the context of the full underlying population. Given our population model which bridges the gap between NSs and BHs, we are able to analyze individual events in the context of this model, and thereby classify them as BNS, NSBH, BBH, or binaries with one component in the mass gap.

Using this population-based classification scheme, we find that GW200105 likely straddles the mass gap, with its primary component confidently higher than the inferred upper edge, and the secondary likely falling below the edge, though uncertainties on the lower edge location make the exact classification uncertain. We similarly reaffirm that GW200115 is an NSBH, with its secondary component below the lower edge of the mass gap and its primary having roughly equal support below and above the upper edge. Our inference is unable to determine if GW190814’s secondary mass lies below or above the lower edge of the mass gap, but slightly prefers it to lie within the mass gap. This ambiguity is driven by uncertainty in the location of the lower edge of the gap. Thus, the nature of GW190814 may emerge as future data constrain global population models. We re-emphasize that these results are

based solely on GW data; the existence of baryons has not been assumed in this analysis. Nonetheless, the conclusions are broadly consistent with expectations from NS EOS studies as well as analyses of galactic X-ray binaries.

All of the models considered in this work can accommodate the full list of observed events without any obvious outliers. In particular, we find that GW190814 is consistent with the full set of compact binary coalescences, even if we are not yet able to identify to which subpopulation it belongs. In this work we focused on the component mass distribution to identify the component objects of binaries. However, future investigations of the mass ratio distribution may help us distinguish between NSBH and BBH population with NSBH systems in recent population synthesis work that tend to prefer mass ratios < 0.5 [Broekgaarden et al., 2021].

The delineation between NSs and BHs is a fundamental observable of GW binary populations. Observable constraints will inform the astrophysics of compact object formation, NS equation-of-state studies, and the formation channels of GW binaries [e.g. Fryer et al., 2002, Belczynski et al., 2012, Fryer et al., 2012, Müller et al., 2016, Breivik et al., 2019, Shao and Li, 2021, Liu et al., 2021]. For example, the confident identification of a feature in the overall mass distribution that is distinct from the maximum NS mass inferred from nuclear theory would have important implications for specific supernova explosion mechanisms [Fryer et al., 2012, Kiziltan et al., 2013]. From GW observations alone, we have identified a sharp drop in the mass distribution of compact objects at $\sim 2\text{--}3 M_\odot$, as might be expected from the maximum allowed mass of NSs. There is insufficient data to confidently constrain the existence of a subsequent mass gap between NSs and BHs. However, the LVK has now nearly completed the fourth LVK observing run (O4), putting the GW community in a position to definitively resolve or disprove this mass gap.

3.6 Additional analysis details

3.6.1 Hyperposteriors inferred with and without NSBH events

In Fig. 3.8 we present a corner plot of the posteriors for relevant hyperparameters inferred by including and excluding the NSBH events GW200105 and GW200115. We find these events are broadly consistent with the rest of the CBC population. The bottom row shows the effect on the overall rate of mergers. It illustrates the effects of treating these two events as if they were detected in the first half of the third observing run. This choice introduces a systematic bias on the rate inference of 23%, which is well within the statistical uncertainties of the rate. Nonetheless, we do not quote merger rates of NSBH-like events, leaving this calculation to future analyses that are able to include the full set of GW detections in the third observing run.

3.6.2 Most flexible model fit

As a final check of our results, we perform an analysis where all hyperparameters are simultaneously fit. A corner plot for this analysis is shown in Fig. 3.9. Notably, we find $\gamma_{\text{low}} = 2.23^{+0.43}_{-0.38} M_{\odot}$, $\gamma_{\text{high}} = 6.34^{+1.18}_{-1.03} M_{\odot}$, and $A = 0.89^{+0.08}_{-0.26}$, all of which are broadly consistent with the corresponding hyperposteriors inferred using BROKEN POWER LAW + DIP.

A prior range for $\eta_{\{\text{high}, \text{low}\}}$ that included both negative values and 50 was too large to allow convergence of our sampler given the relatively low number of detections, so we chose uniform priors between -4 and 12. Though neither hyperparameter is particularly well-measured, we find both η_{low} and η_{high} to be positive with $> 90\%$ credibility, indicating the existence of a dip in the mass distribution between γ_{low} and γ_{high} : $\eta_{\text{low}} = 7.94^{+2.70}_{-3.63}$ and $\eta_{\text{high}} = 8.54^{+2.36}_{-3.52}$. Similarly, we find a Bayes Factor of $\mathcal{B}_{\eta_{\text{high}}=0}^{\eta_{\text{high}}=12} = 13$ in favor of a rise in the mass distribution at γ_{high} . However, both hyperposteriors are approximately uniform above ~ 8 , indicating that the inference is unable to resolve the steepness of the gap edges

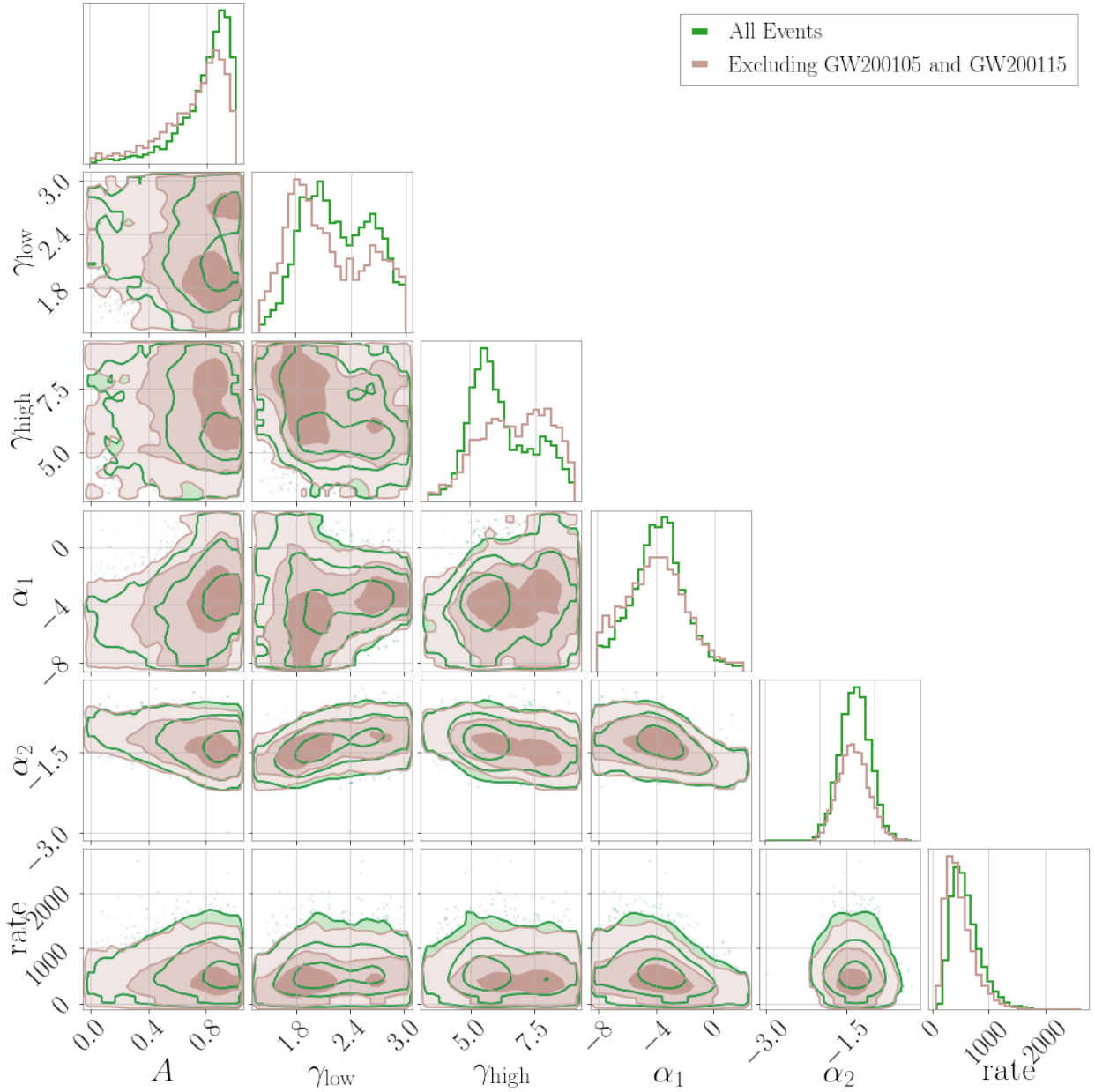


Figure 3.8: A subset of hyperposteriors for the BROKEN POWER LAW + DIP model inferred with (*green*) and without (*brown*) the events GW200105 and GW200115. The mass distribution does not change significantly with the exclusion of these events, indicating that the NSBHs are consistent with the rest of the detected population. When the NSBHs are included, the upper edge of the mass gap, γ_{high} , is slightly better constrained and support for a wider gap (low γ_{low} and high γ_{high}) diminishes. This suggests that these events probe the gap location. Hyperparameters on which the exclusion of the NSBH events has no effect are not included in this plot.

beyond a power law that is increasing (or decreasing, in the case of η_{low}) faster than $\sim m^8$. We therefore conclude that fixing the gap edges to be relatively sharp ($\eta_{\{\text{high}, \text{low}\}} = 50$) is allowed by the data. We additionally find that η_{min} peaks near the upper end of the prior bound, validating that a steep turn-on at the lowest masses is preferred by the data.

As discussed in Section 3.2.3, the depth and location of the mass gap are not well-defined in terms of one hyperparameter when the edges of the gap are allowed to be smooth. It is therefore desirable to employ metrics for the existence of a mass gap that are both easily interpreted and independent of model parameters. These could be used for any mass model, not just the ones presented here.

Potentially the simplest such metric is the existence of a local minimum in the mass distribution between $1\text{--}10 M_{\odot}$. A lack of support for such a dip in the mass distribution in this region may indicate either a lack of a lower mass gap or an inability to resolve one if it does exist. We can therefore construct Bayes factors for the existence of a local minimum that are defined as the ratio of the percentage of draws from the hyperposterior that find a local minimum to the percentage of draws from the hyperprior that find a local minimum.

Within this most flexible model fit, we find that 80% of draws from the hyperposterior have support for a local minimum between $1 M_{\odot}$ and $10 M_{\odot}$, compared to 49.1% of prior draws. This gives a Bayes factor of 1.63 in favor of a local minimum, showing that a lower mass gap is slightly but not definitively preferred by the data.

An important caveat to this statistic, however, is that the existence of a local minimum does not always map to the existence of a mass gap. For example, a shallow-but-declining power law meeting a shallow-but-increasing power law would result in a local minimum, but not a feature that one would expect if there were an astrophysically-induced dearth of events within a certain range. This phenomenon inflates the number of prior draws that produce a local minimum relative to those that produce a mass gap feature, thus decreasing the resulting Bayes factor if the posterior does in fact resemble a mass gap.

More complex metrics can be constructed to mitigate this problem, but most are subject to similar issues and are less easily interpreted. Therefore, we caution against using such statistics as the single assessment of feature significance, and instead recommend they be paired with complementary probes that are more well-suited to specific models.

3.6.3 *Model comparison*

In Figs. 3.10, 3.11, and 3.12 we show corner plots of BROKEN POWER LAW, BROKEN POWER LAW + DROP and BROKEN POWER LAW + DIP, respectively. At the top of each column in these corner plots is listed the median and 68% credible intervals for each one-dimensional marginal distribution. Note that this is not the same as the set of overall median values, as there are significant correlations between the hyperparameters. The maximum hyperposterior value for each model is provided in Table 4 of Farah et al. [2022a]. These are the values used to produce the colorful line in each panel of Fig. 3.3. We additionally provide Bayes factors (ratios of marginal likelihoods) and maximum likelihood ratios for each model relative to BROKEN POWER LAW in Table 4 of Farah et al. [2022a]. Three of the models have similar Bayes factors $\mathcal{B}$, with a spread in $\mathcal{B}$ of $\lesssim 20\%$. However, BROKEN POWER LAW + DIP is favored by a factor of ~ 10 . While not conclusive, this difference between Bayes factors suggests that a mass gap with sharp edges may be preferred. The ratio of maximum likelihoods for each model is provided to help disentangle the effect of the priors on the Bayes factors. The two models which allow for a lower mass gap seem to be preferred over those that do not, with maximum likelihoods that are larger by factors of ~ 20 –50. Therefore, the low Bayes factor of the most flexible model may partially be due to its large prior volume, i.e. an “Occam penalty.” Nonetheless, the level of discrepancy seen in Table 4 of Farah et al. [2022a] is not large enough to rule out any of the models considered based on their predictive accuracy.

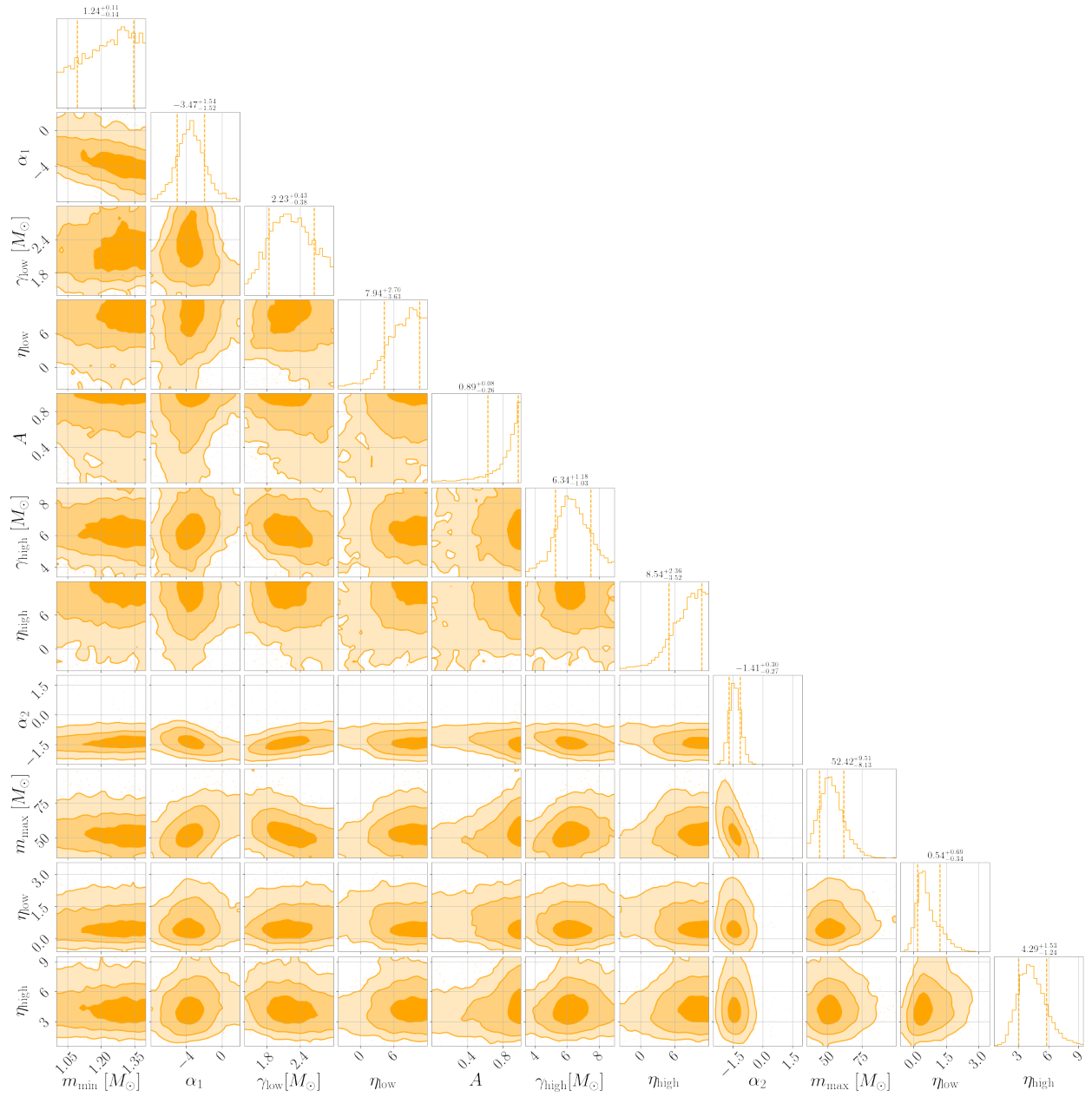


Figure 3.9: Corner plot for an analysis where all hyperparameters are simultaneously fit. On the diagonal are the one-dimensional marginal distributions for each hyperposterior. Above each column are written the median and 68% credible intervals for the corresponding hyperposterior.

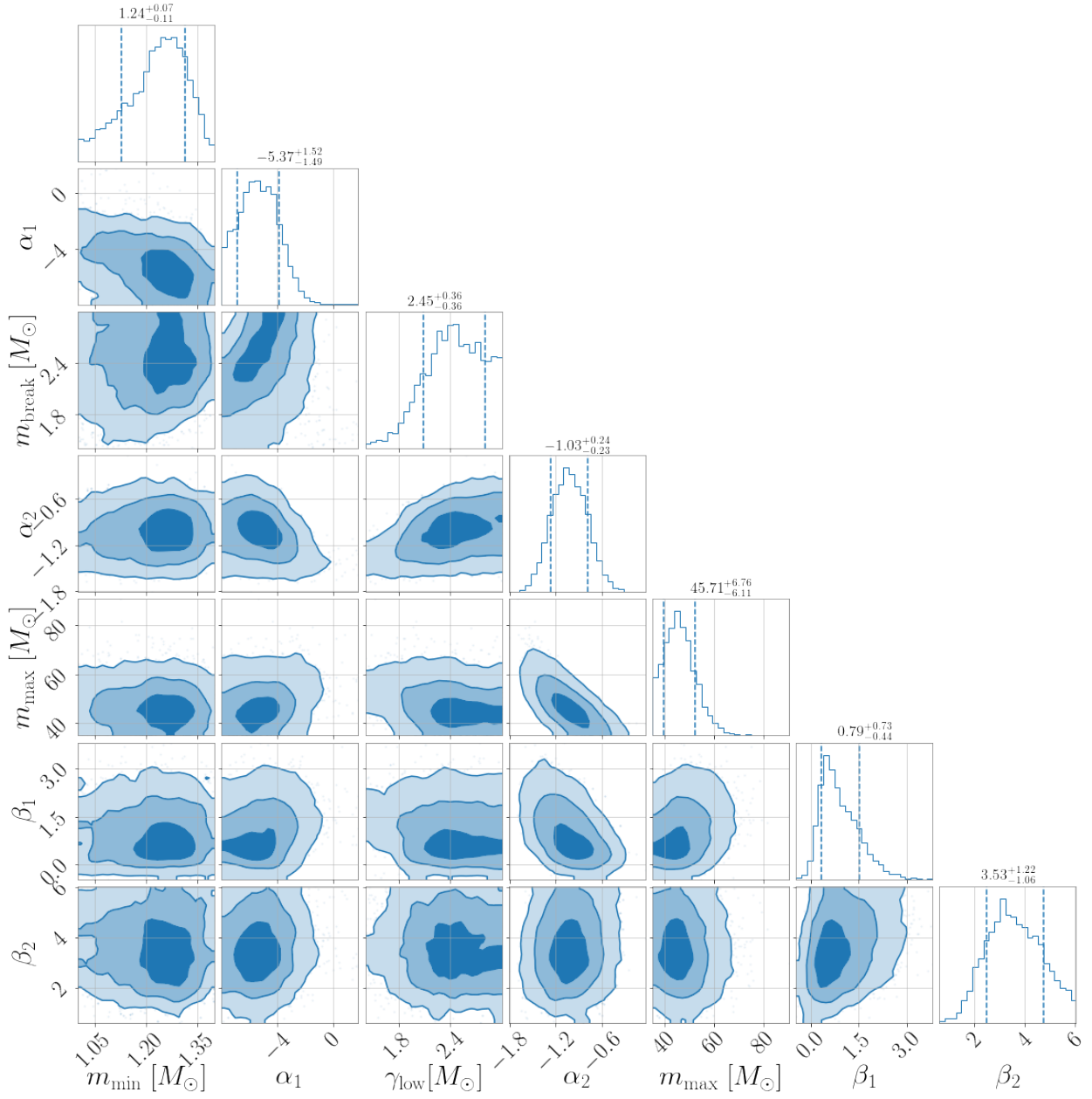


Figure 3.10: Corner plot for the BROKEN POWER LAW model. On the diagonal are the one-dimensional marginal distributions for each hyperposterior. Above each column are written the median and 68% credible intervals for the corresponding hyperposterior.

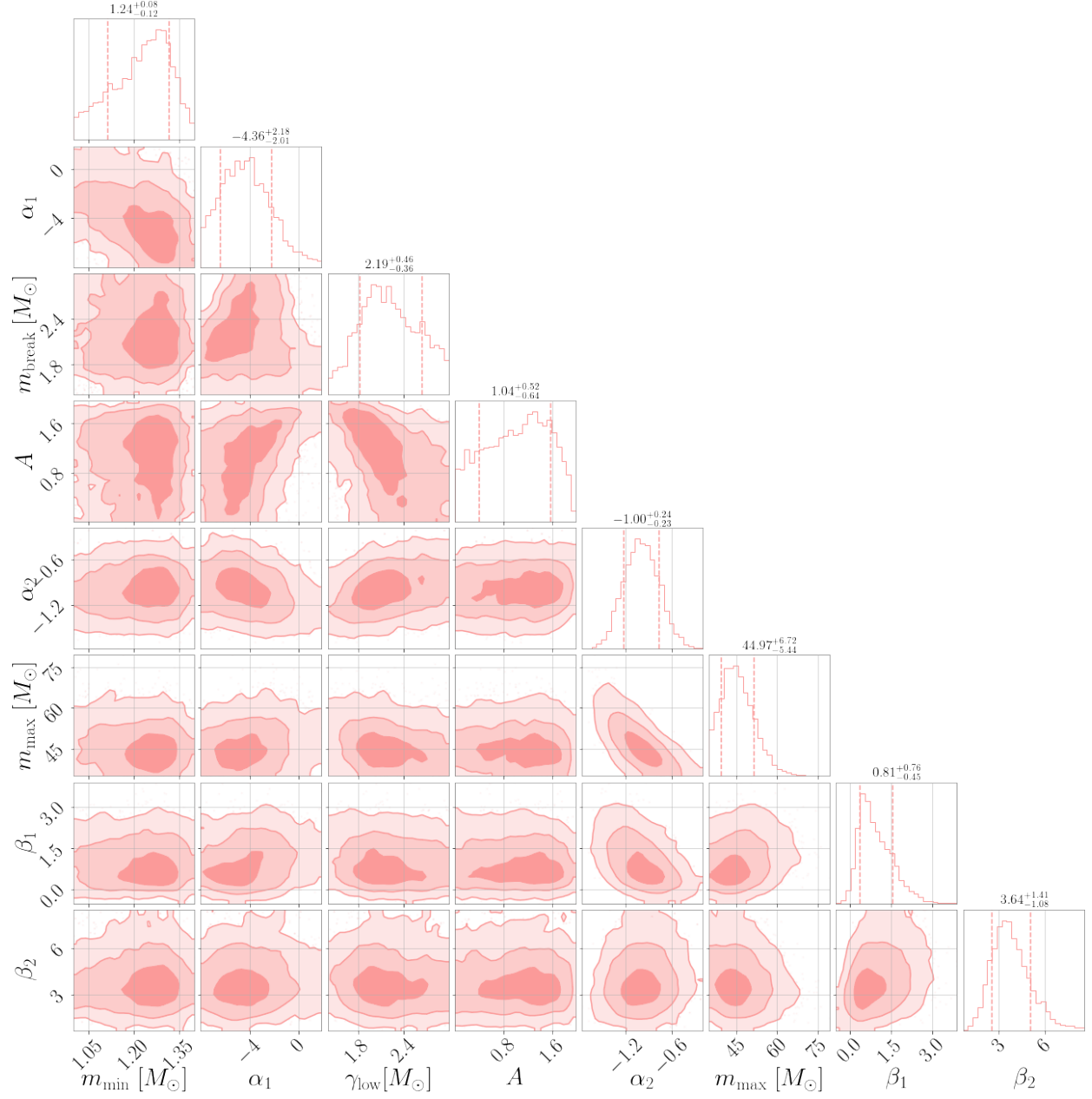


Figure 3.11: Same as Fig. 3.10 but for BROKEN POWER LAW + DROP.

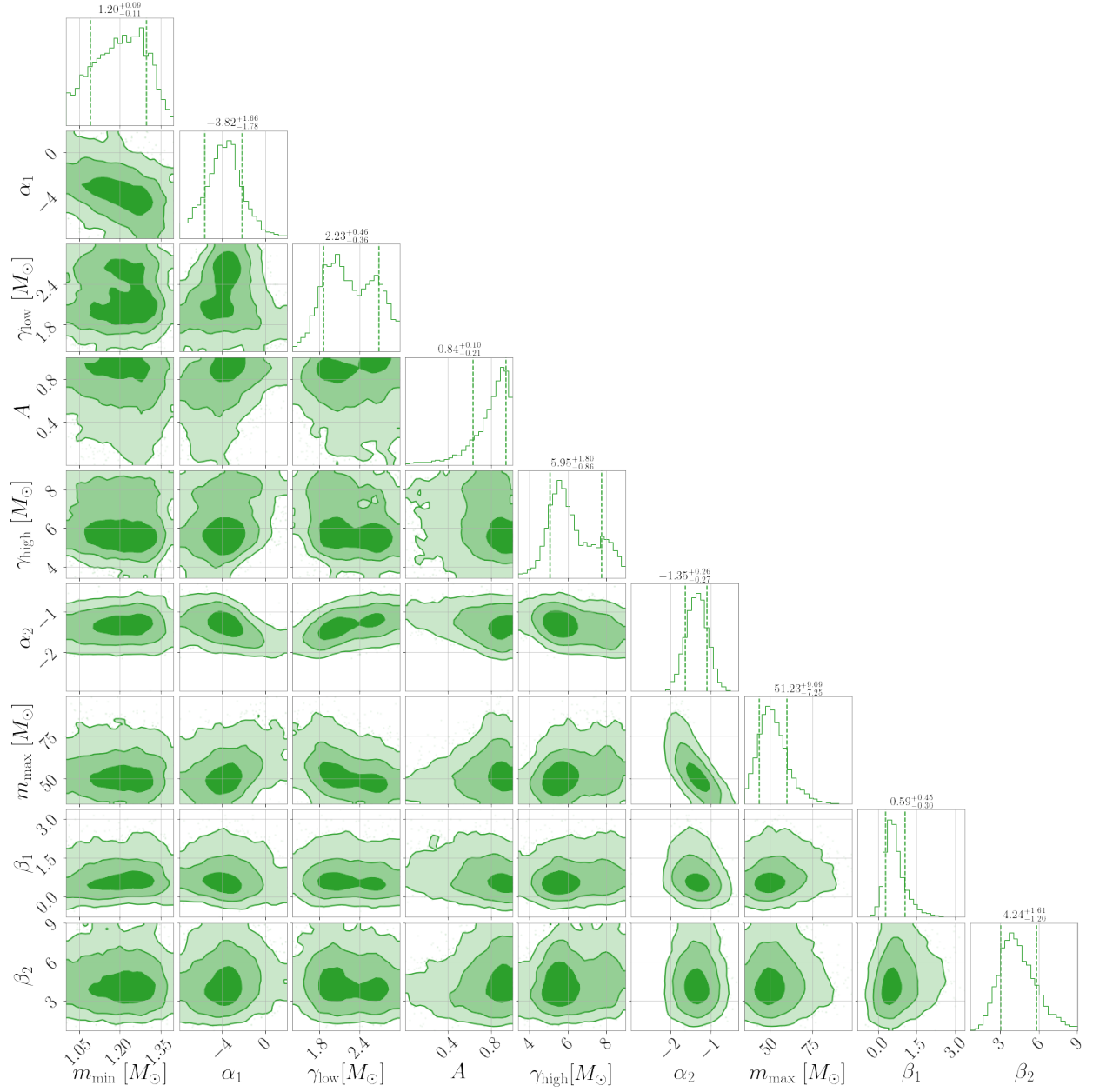


Figure 3.12: Same as Fig. 3.10 but for BROKEN POWER LAW + DIP.

CHAPTER 4

THINGS THAT MIGHT GO BUMP IN THE NIGHT: SIGNIFICANCE OF STRUCTURE IN THE BINARY BLACK HOLE MASS DISTRIBUTION

The previous chapter scrutinized the region between NS and BH masses, motivated by theoretical and observational predictions for features in that mass range. We will now turn our attention to a higher-mass region: the BBH population.

As detailed in Chapter 1, features in the BBH mass spectrum are shaped by massive stellar physics, dynamical encounters in dense environments, and binary interactions. The complicated and theoretically uncertain nature of these physical processes has motivated many GW astronomers to search for features in the BBH part of the mass spectrum in a weakly-modeled way. These “non-parametric” methods have considerably more flexibility than the parametric methods used to search for specific predicted features. While this added flexibility allows for the identification and characterization of previously-unpredicted features, it can also obfuscate model systematics and make non-parametric models susceptible to overfitting. Thus, it is necessary to understand if the features identified by non-parametric models are due to overfitting or model systematics, as opposed to being real characteristics of the GW data.

In this chapter, we make this distinction for population analyses of BBHs in GWTC-3. We do this by performing the same population analysis that was done on GWTC-3, but on hundreds of simulated GW catalogs with a known underlying distribution with no features. Since these simulated catalogs have the same number of events, selection function, and measurement uncertainty as GWTC-3, we can determine how often analyses on these catalogs produce features at the same statistical significance as those identified in GWTC-3.

We found that only two (peaks at $\sim 10 M_{\odot}$ and $\sim 35 M_{\odot}$) of the three features identified

in GWTC-3’s BBH distribution could not be reproduced by fits to featureless catalogs, implying that the third feature (a dip at $\sim 14 M_\odot$) could be due to model systematics or overfitting. The two irreproducible features are therefore likely imprints of astrophysical processes, but more data and/or different models are required to determine if the same is true for the third feature.

This chapter is based on *Things That Might Go Bump in the Night: Assessing Structure in the Binary Black Hole Mass Spectrum*, a paper written in collaboration with Bruce Edelman, Michael Zevin, Maya Fishbach, Jose María Ezquiaga, Ben Farr, and Daniel Holz, and published in the *Astrophysical Journal* [955] as Farah et al. [2023]. While we focus on non-parametric models here, the methods presented in this chapter can also be extended to parametric models, and, since the publication of Farah et al. [2023], have been used to validate parametric and non-parametric analyses alike.

4.1 Introduction

In Section 1.1 of Chapter 1, we discussed how the mass distribution of GW sources is shaped by astrophysical processes, and how population analyses of the current data set have revealed structure beyond a single power law in the distribution of BBH primary masses. The current picture of the BBH mass distribution is therefore a decreasing power law from low to high masses, with a global maximum at $m_1 \sim 10 M_\odot$, a potential underdensity at $m_1 \sim 14 M_\odot$, and an overdensity at $m_1 \sim 35 M_\odot$. This can be seen in Fig. 4.1, where we plot the results of fitting two parameteric models and one semi-parametric model to the BBHs in GWTC-3.

While the existence of this substructure appears robust, its interpretation is less clear. Plausible explanations for this substructure include (1) Poisson noise, (2) modeling systematics, or (3) astrophysical signatures from one or several formation channels. We aim to disentangle the first two possibilities from the third using the POWER LAW + SPLINE model [Edelman et al., 2022], one of the semi-parametric models used to identify the substructure

reported in Abbott et al. [2023a].

Poisson noise would be caused by the fact that the fiducial BBH analysis in Abbott et al. [2023a] includes only 69 events over a mass range that spans more than an order of magnitude, so observations may appear to be clumped at some masses even if the underlying distribution is smooth. We first determine if this explanation accounts for the data by simulating catalogs of BBHs whose underlying distribution does not have the features of interest, applying the analysis previously performed on GWTC-3, and determining how often such features are spuriously found. We develop several metrics comparing observations to simulated data in order to assess the statistical significance of the “bumps” in the primary mass distribution found by Abbott et al. [2023a], Edelman et al. [2022]. All of the metrics derived in this work answer the same general question: how often do we infer the existence of a feature when analyzing observations of a true population *without* that feature? In this sense, these metrics are analogous to frequentist p -values, as lower values correspond to more significant features in the data. Readers familiar with gravitational wave data analysis might find it useful to think of these metrics as false alarm rates because they quantify how often noise resembles the observed signal.

A similar frequentist analysis on a large number of mock catalogs was performed by Sadiq et al. [2022] on the peak at $\sim 35 M_{\odot}$ using an adaptive kernel density estimator (aKDE) to find features in samples drawn from featureless mass models, as well as from a model with a single peak. They account for selection effects, but not measurement uncertainty. They find that an aKDE is able to identify peaks in the data, and that the peak at $\sim 35 M_{\odot}$ found in GWTC-2 is statistically significant within the aKDE model.

The second effect mentioned above, model systematics, could also plausibly cause spurious inference of features beyond a power law. It is potentially concerning that the models considered in Abbott et al. [2023a] that find peaks and troughs in the mass distribution are inherently “bumpy”: both POWER LAW + PEAK [Talbot and Thrane, 2018] and MULTI

SOURCE employ a smoothed power law with a Gaussian component, FLEXIBLE MIXTURES is a linear combination of Gaussian components, and POWER LAW + SPLINE employs a smoothed power law under a cubic spline modulation. The question is then whether these “bumpy” models can recover sharp features or if they instead create peaks and troughs that are morphologically dissimilar to the true distribution. This is most easily addressed by cross-checking with independent models such as BROKEN POWER LAW [Abbott et al., 2021d, 2023a] and the auto-regressive model presented in Callister and Farr [2024].

Inaccuracies in the selection function are also known to cause systematic biases when inferring the underlying population [e.g. Malmquist, 1922, 1925]. These biases could, in principle, also cause an incorrect inference of structure in the astrophysical distribution of BBH masses. However, selection effects in GW detectors are remarkably well-characterized, so we expect this effect to be subdominant to Poisson uncertainty. As the number of events grows, so will our accuracy in the estimation of the selection function [Farr, 2019, Essick and Farr, 2022].

We provide posterior samples from our simulated catalogs in an accompanying data release [Farah et al., 2022b], and also provide a publicly-available `python` package, `GWMockCat` [Farah et al., 2022c], to create similar samples according to user-defined populations.¹

Section 4.2 provides a demonstrative example: it foregoes a full fit to the astrophysical population of sources, and compares the observed distribution of masses to possible observed distributions given an underlying power law in primary mass, (incorrectly) assuming no measurement uncertainty. This analysis suggests that the observed peak at $\sim 35 M_{\odot}$ is statistically significant, but that all other features beyond a simple power law might be explainable by Poisson noise. This motivates a thorough study using a full hierarchical Bayesian analysis on simulated event posteriors, which we carry out in Section 4.3. Section 4.4 summarizes our conclusions and discusses their implications for the astrophysical

1. The data release can be found at <https://doi.org/10.5281/zenodo.7411991>, and `GWMockCat` can be installed at <https://git.ligo.org/amanda.farah/mock-PE>.

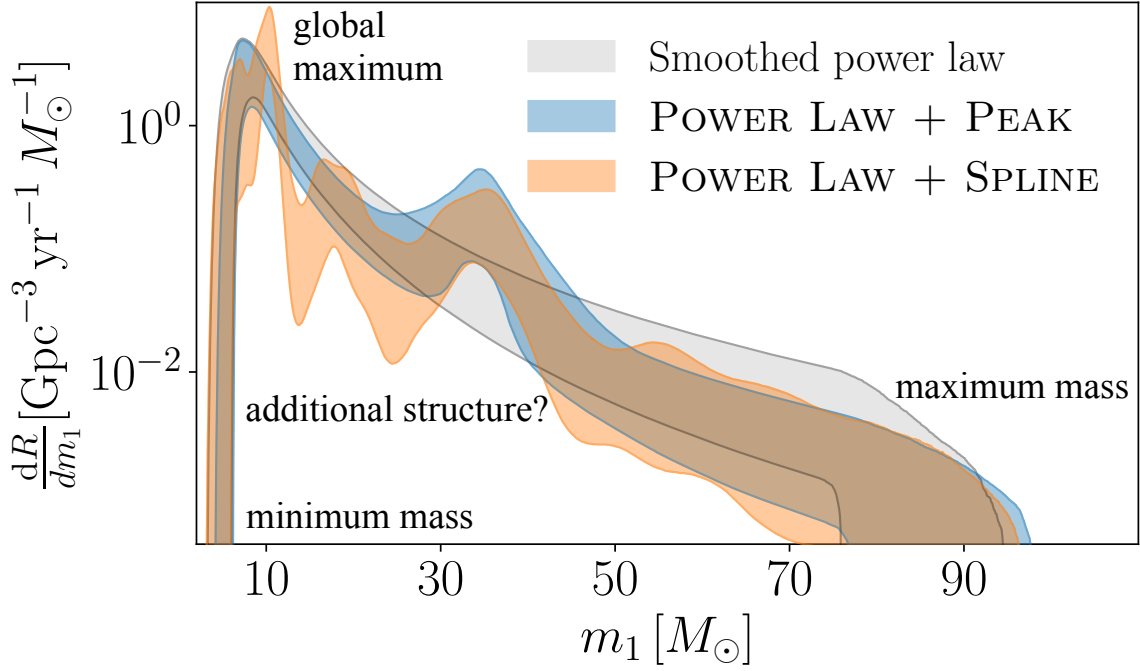


Figure 4.1: Distribution of primary BBH masses inferred using GWTC-3 and three different population models. The smoothed power law model (*grey*) consists of a single power law slope between a minimum and maximum mass, with the merger rate set to exactly zero outside of those bounds. It also includes a smoothing parameter at the low-mass end that allows for an offset between the minimum BH mass and the global maximum of the distribution. The POWER LAW + PEAK model is similar to the smoothed power law, but also includes a Gaussian component. The POWER LAW + SPLINE model adds a cubic spline modulation to a smoothed power law to allow for additional substructure. We seek to determine if the perturbations beyond a power law found by POWER LAW + SPLINE and other semi-parametric models can be explained by random associations in the data due to a finite number of observations, or if they are features of the true underlying distribution.

origin of the gravitational waves observed thus far by the LVK.

4.2 Motivation

To construct a simple test of feature significance and motivate further study, we first avoid a fit to the mass distribution and instead consider the *observed* distribution of primary masses and its resemblance to one that would result from a simple power law. The observed population differs significantly from the astrophysical one, as current gravitational wave detectors

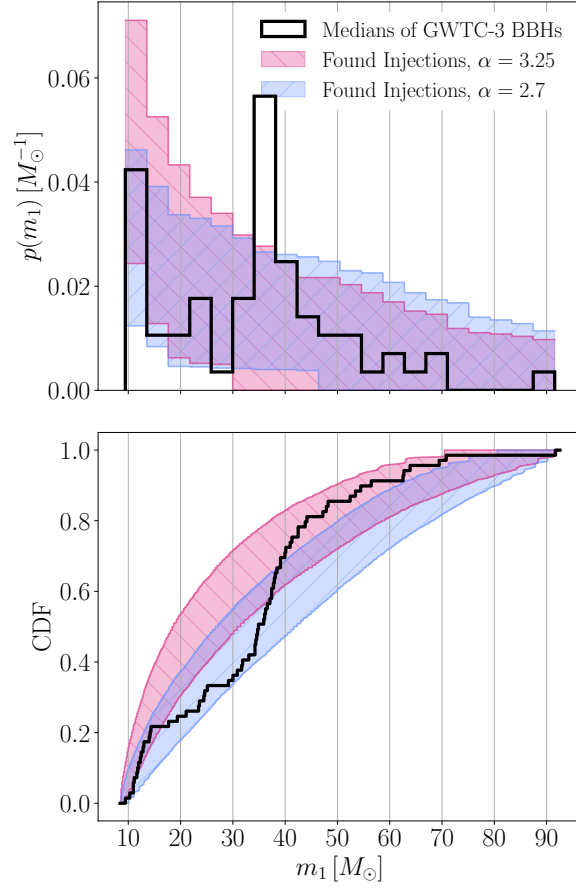


Figure 4.2: Observed source-frame primary mass distributions. Black solid lines contain the median *a posteriori* values for the binary black holes in GWTC-3. Pink and blue bands indicate the 90% credible interval on the *observed* distributions predicted from astrophysical distributions that are power laws in primary mass with spectral index $\alpha = 3.25$ and $\alpha = 2.7$, respectively. The top panel shows a histogram of observed primary masses. For GWTC-3’s distribution to be consistent with the null distributions, we expect its bin heights in the top panel to be within the 90% credible intervals in 18 out of the 20 bins. The uncertainties in these predicted distributions are due only to Poisson noise resulting from a finite number of observations, rather than modeling uncertainty or uncertainty in parameter estimation. Therefore, the cumulative distribution functions in the bottom panel are similar to a conventional posterior predictive check, but with only one source of uncertainty. The large deviations of the black curve from the shaded bands in some regions indicate the difficulty that a single power law with Poisson shot noise has in fully explaining the observations. However, many of the apparent excursions from a power law are well-contained within the predicted bands.

are subject to selection biases that favor the detection of closer and more massive systems, as well as measurement error that affects each system differently. We construct plausible observed mass distributions that could occur from detecting 69 BBHs whose astrophysical distribution is a featureless power law in primary mass. To do this, we use the samples provided by Collaboration [2021], which were created for sensitivity estimation for the LVK’s GWTC-3 analysis. Each of these samples comes with a probability of being drawn from an assumed underlying distribution and a false alarm rate (FAR) assigned by each search used by the LVK. We can then re-weight these samples to our desired population model (in this case, a power law in m_1, q , and z) using the draw probability, and apply the same FAR threshold used in Abbott et al. [2023a] to select “found injections.” Of the $\sim 6 \times 10^4$ found injections, we resample to $N = 10^4$ independent sets of 69 draws each to directly compare to observations.

We then histogram each set of these found injections, thereby obtaining a distribution of bin heights for our mock populations. Using several thousand realizations of found injection sets enables us to construct a null distribution of bin heights and characterize the effect of Poisson noise on the shape of the observed distribution. We compare these null distributions with the observed distribution of BBH masses in GWTC-3² by assuming the primary masses are measured perfectly and using the median *a posteriori* values of their primary masses as point estimates. The result is shown in the top panel of Fig. 4.2, which plots the 90% credible interval on the observed null distributions, along with the distribution of median primary masses of GWTC-3’s BBHs. For the null distributions, we consider two power law spectral indices as representative examples: $\alpha = 2.7$ and $\alpha = 3.25$. These are chosen to represent a range of plausible values for the BBHs in GWTC-3: a power law fit to GWTC-3 yields

2. For all comparisons to real observations, we use the publicly available posterior samples for GWTC-2.1 and GWTC-3. We use samples generated with the IMRPhenomXPHM waveform and a prior proportional to the square of the luminosity distances (i.e. the samples were not “cosmologically reweighted”). To make the most direct comparison with Abbott et al. [2023a], we keep events with secondary mass larger than $3 M_\odot$ and FAR less than 1 yr^{-1} , resulting in 69 events.

$\alpha = 2.98^{+0.16}_{-0.28}$, where the bounds represent $1\text{-}\sigma$ deviations.

To obtain a more quantitative measure, we compare bin heights from the found injections, h_{inj} , to the bin heights of observed events in GWTC-3, $h_{\text{GWTC-3}}$, obtaining for each bin i the fraction of simulated bin heights that are lower than those of GWTC-3 BBHs. Explicitly,

$$r_h^i = \frac{1}{N} \sum_j^N \begin{cases} 1 & \text{if } h_{i,\text{inj}}^j < h_{i,\text{GWTC-3}} \\ 0 & \text{if } h_{i,\text{inj}}^j \geq h_{i,\text{GWTC-3}} \end{cases}, \quad (4.1)$$

where the sum is over the $N = 10^4$ sets of found injections, and r_h is defined for each bin. For example, if $r_h = 0.95$ for a given bin, the observed distribution in that bin is larger than would be expected from a featureless power law 95% of the time. A value of r_h approaching unity corresponds to a “bump” in the observed mass distribution, and a value of r_h approaching zero is indicative of a “dip.”

Note that the comparison between the null distributions and GWTC-3 are occurring at each bin, rather than across all bins. We do this because the magnitude of Poisson noise depends on the value of m_1 : since the underlying distribution is not uniform, fewer events are expected at very high m_1 and therefore the relative standard deviation is larger. This is also a consequence of Eddington bias [Eddington, 1913]. Making comparisons at specific points in m_1 does not, however, properly correct for the look-elsewhere effect. We will address this effect in Section 4.3.

The three most significant values of r_h^i in the case of $\alpha = 3.25$ are $r_h^{15.6 M_\odot} = 0.033$, $r_h^{27.9 M_\odot} = 0.036$, $r_h^{36.1 M_\odot} > 0.999$, where the superscripts indicate the centers of the bins at which r was calculated. This means that less than 0.1% of mock populations had more events near $36.1 M_\odot$ than GWTC-3 does, 3.3% of mock populations had fewer events near $15.6 M_\odot$ than GWTC-3, and at $27.9 M_\odot$, 3.6% of mock populations had fewer events.

Repeating the exercise for $\alpha = 2.7$, we find the three most significant values of r_h^i to be $r_h^{40.2 M_\odot} = 0.935$, $r_h^{27.9 M_\odot} = 0.020$, $r_h^{36.1 M_\odot} > 0.999$. The locations of the significant

features differ when the assumed underlying distribution changes³. In either case, the bump at $\sim 35 M_\odot$ is unlikely to be due to Poisson noise, but other features may be.

To avoid the need to arbitrarily choose bins, we additionally construct a cumulative distribution function (CDF) of the primary masses and compare it to the CDFs of the null distributions, shown in the bottom panel of Fig. 4.2. This comparison is akin to a posterior predictive check in that it can highlight where the model fails to predict the data. Importantly, though, it differs from the conventional posterior predictive check because we have purposefully left out the effects of modeling uncertainty and measurement uncertainty in order to isolate the effects of Poisson noise. The prior distributions are therefore also not included, since each event is assumed to be measured with perfect accuracy.

If $\alpha = 3.25$, the null distributions are consistent with the data below $\sim 18 M_\odot$ and above $\sim 35 M_\odot$, but not between them, meaning that the $\sim 10 M_\odot$ and $\sim 35 M_\odot$ peaks can be explained by Poisson noise, but the underdensity between them could not be. On the other hand, if $\alpha = 2.7$, the null distributions are consistent with the data everywhere except for above $\sim 40 M_\odot$, suggesting that under this scenario, Poisson noise can explain all features except for the $\sim 35 M_\odot$ peak.

For both spectral indices considered, two of the three features found by Abbott et al. [2023a] can be explained by Poisson noise from a finite number of observations. However, this does not mean that exactly two of the features are the result of Poisson noise, just that no more than two can be caused by the phenomenon. Additionally, it is not clear *which* of the features are more likely to have physical origin, as this method offers no quantitative way to determine which power law slope is preferred.

3. It is also possible to determine the existence of local minimia or maxima in this observed distribution independently of the underlying power law. This can be done using a dip test for unbinned data [Hartigan and Hartigan, 1985] or the minimum number of components required for a Gaussian mixture model [McLachlan and Peel, 2000]. However, features in the observed distribution would be difficult to disentangle from selection effects, so we recommend only applying these to the astrophysical distribution, as in Tiwari [2021]. Since our principal aim is to quantify the significance of excursions from a power law, we leave such tests for future work.

Importantly, this methodology does not account for the effects of measurement error, which can cause significant biases near the edges of sharp distributions when not properly accounted for [Fishbach et al., 2020b]. We therefore turn to a full hierarchical Bayesian analysis of simulated catalogs, which will allow us to fit for population model parameters, take measurement uncertainty into account, and directly compare to metrics used in Abbott et al. [2023a].

4.3 Hierarchical analysis and results

We determine how often the features inferred in the mass distribution of BBHs would be spuriously found in data whose underlying distribution does not have those features. To do this, we construct a null distribution by simulating BBH observations that would occur if the underlying astrophysical distribution was a single power law with no substructure in a finite range. The procedure for creating synthetic BBH observations is described in Section 4.5.1. Mock observations are combined with corresponding sensitivity estimates in a hierarchical Bayesian analysis, described in Chapter 2. We analyze these simulations in the same way as the BBHs in GWTC-3 to determine how often the features observed in GWTC-3 would be found from an underlying distribution without those features.

4.3.1 *POWER LAW + SPLINE mass model*

We use the POWER LAW + SPLINE semi-parametric primary mass model as a flexible model that is easily capable of finding peaks and valleys in the mass distribution [Edelman et al., 2022, Abbott et al., 2023a]. This model parameterizes perturbations or deviations from a simpler underlying distribution with flexible cubic spline functions. Specifically, given an underlying hyper-prior for primary mass, $p(m_1|\Lambda)$, the POWER LAW + SPLINE model

describes the primary mass distribution as:

$$p_{\text{spline}}(m_1|\Lambda, \{m_i\}, \{f_i\}) \propto p(m_1|\Lambda) \exp(f(m_1|\{m_i\}, \{f_i\})) \quad (4.2)$$

where $f(m_1|\{m_i\}, \{f_i\})$ is the function describing the perturbations, which we model with a cubic spline function interpolated by introduced hyper-parameters, $\{m_i\}$, the locations of spline knots in mass space, and $\{f_i\}$, the height of the perturbation function at each knot. This describes a semi-parametric model as it includes a simple “parametric” component (the underlying distribution) in addition to a non-parametric component that models the perturbation around the simple description. For this study we use the simplest primary mass model for the underlying description, which is the TRUNCATED model, describing a power law with sharp cutoffs at the lower and upper mass bounds [Fishbach and Holz, 2017, Edelman et al., 2022]. While this model has been shown to insufficiently describe the primary mass distribution, it captures the majority of the broadest features [Abbott et al., 2021d, 2023a].

To assess the significance of peaks or valleys found with the POWER LAW + SPLINE model one can look at the posterior distribution of the perturbation heights as a function of mass. This tells us how far “off” the simple power law description is from accounting for the data. Specifically we can find what percentile $f(m_1) = 0$ falls in the posterior distribution as a function of mass. For data exactly distributed as a power law (the underlying population), the inferred perturbation function should be symmetric about 0 with widths determined by the prior distributions on the knot heights and the number of observed events. At masses where the percentile of zero perturbation approaches 100% (0%) we can say there is an over (under) density of events at these masses, compared to the underlying power law distribution. This is identical to the analysis done by Abbott et al. [2023a], who use the percentile at which the perturbation function excludes zero at a given location as a metric for how significant a feature is at that location.

4.3.2 Metrics of feature significance

As described in Section 4.3.1, the POWER LAW + SPLINE model makes use of a perturbation function constructed from cubic splines. The height of the perturbation function, $f(m_1)$, at a point in primary mass, m_1 , is then a direct measure of the deviation from a power law at that point. We can determine how often one would find spurious evidence for substructure by simulating catalogs from a power law, fitting them with the POWER LAW + SPLINE model, and examining the resulting perturbation function.

If the mock catalogs produce perturbation functions with similar amplitudes to those seen for GWTC-3, the structure in the GWTC-3 fit might be described by Poisson noise. On the other hand, if the perturbation functions produced by fits to the mock catalogs are always lower in amplitude to that of the GWTC-3 fit, the structure in the GWTC-3 data is likely to be present in the underlying distribution.

For a given mock catalog, we find the m_1 value where the median *a posteriori* value of the perturbation function is maximal. We obtain the posterior distribution of perturbation function amplitudes at that location, $g(f_{\max})$. We repeat this for all mock catalogs, obtaining a set of maximal perturbation function distributions, $\{g_j(f_{\max})\}$. These are plotted in light grey on the left panels of Fig. 4.4. The locations of the three maximal perturbation function amplitudes in the GWTC-3 fit are, from least to most Bayesian significance, $13.8 M_\odot$, $10.3 M_\odot$, and $35.7 M_\odot$. The posterior distributions of perturbation function heights at these locations are $g_{\text{GWTC-3}}(f(13.8 M_\odot))$, $g_{\text{GWTC-3}}(f(10.3 M_\odot))$, and $g_{\text{GWTC-3}}(f(35.7 M_\odot))$, and are plotted in orange in the left panels of Fig. 4.4. The amplitude of the perturbation function at $13.5 M_\odot$ is negative (i.e. it is a dip rather than a bump), so we flip its distribution about zero for more direct comparison. The same is done for all $g(f_{\max})$ whose medians are negative, as the perturbation function's prior is symmetric about zero.

4.3.3 *Simulation study*

To determine whether the features in the mass spectrum of GWTC-3 BBHs are the result of Poisson noise of a finite number of observations drawn from a featureless power law, we compare POWER LAW + SPLINE fits using the GWTC-3 catalog and 300 mock catalogs generated from a “featureless” power law. The mock catalogs considered in this section are all generated from the same underlying distribution: a truncated power law in primary mass, mass ratio, and redshift, with a smoothing at low component masses to ensure the peak of the mass distribution is not in the same location as the minimum mass. The explicit form of the mock catalogs’ population model, including values of all of its hyperparameters, can be found in Section 4.5.2.

Full parameter estimation was not performed on each event in each mock catalog; instead, we use prescriptions for generating event posteriors that reproduce the correlations between an event’s parameters, as well as the typical uncertainties seen in GWTC-3. We show in Section 4.5.3 that the prescriptions used are sufficient to reproduce population analyses such as the ones scrutinized in this work. In order for our comparisons to be consistent between featureless mock catalogs and GWTC-3, we recreate GWTC-3’s event posteriors with the same prescriptions as were used for the mock catalogs, perform a population analysis on those, and use the resulting perturbation function for all comparisons to mock catalogs. Our population re-analysis of the GWTC-3 events with mock posteriors appears consistent with the full analysis presented by the LVK in Abbott et al. 2023a. Lines labeled “GWTC-3” or depicted in orange in Figs. 4.3, 4.4, and 4.5 refer to the analysis done on the recreated version of GWTC-3. All analyses presented here were repeated using the original LVK-released parameter estimation samples in Section 4.5.3, and the same qualitative conclusions were reached, though with slightly more statistical significance.

Despite knowing the parameters of the underlying population for the mock catalogs, we allow all hyperparameters to vary when fitting POWER LAW + SPLINE to the mock catalogs.

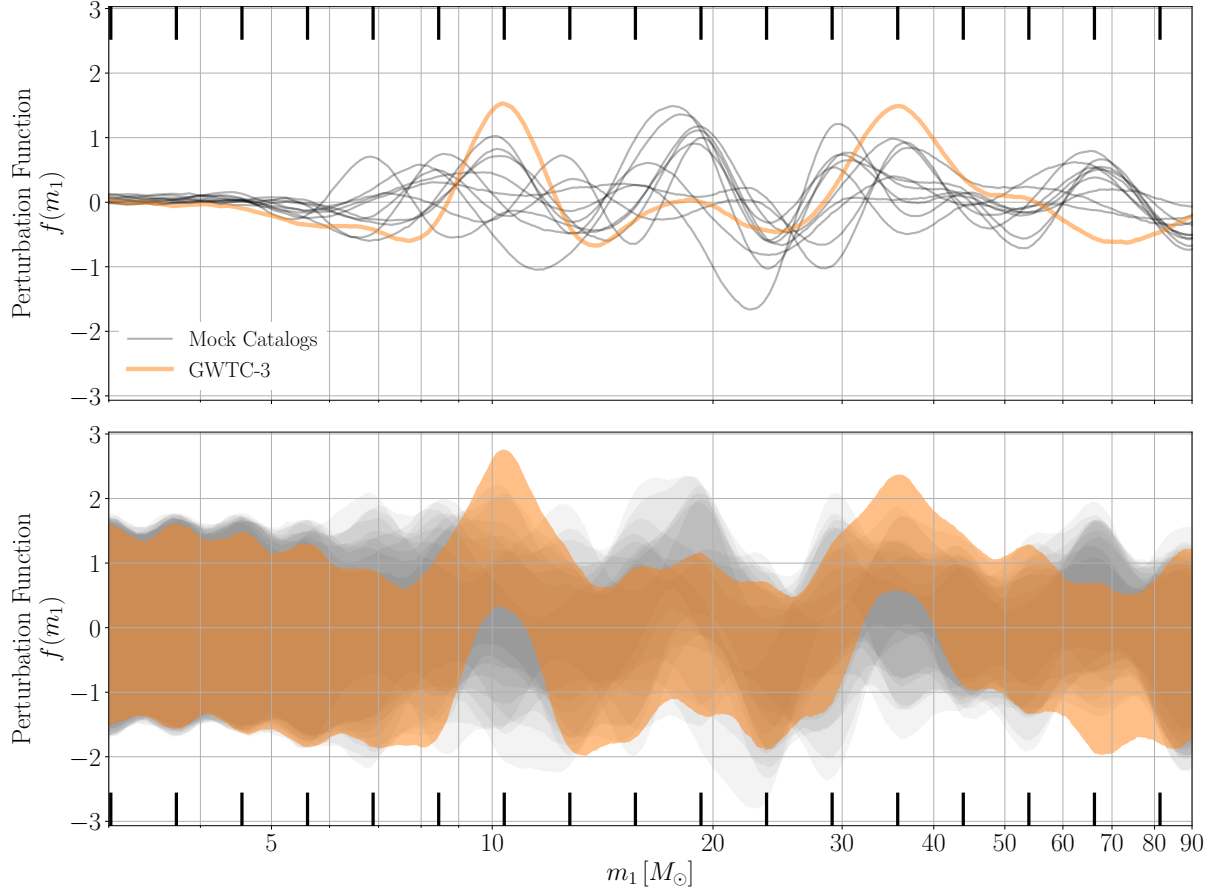


Figure 4.3: Median (*top panel*) and 90% credible interval (*bottom panel*) of the perturbation function resulting from the POWER LAW + SPLINE fit to the primary masses in GWTC-3 (*orange*) and in 10 mock catalogs (*grey*). The perturbation function multiplies a smoothed power law in primary mass to add modulations to an otherwise monotonic distribution, making it a direct measure of deviations from a power law. It is a cubic spline with knots fixed at the locations indicated by the black vertical tick marks. The prior on the perturbation heights is the unit normal distribution, as can be seen below $\sim 5 M_{\odot}$ where there are no detections to constrain the likelihood and the posterior reverts to the prior. The perturbation function corresponding to GWTC-3 events appears large in amplitude in three locations: $\sim 10 M_{\odot}$, $\sim 14 M_{\odot}$, and $\sim 35 M_{\odot}$. While the medians of the perturbation function at these distributions are comparable in amplitude, the posterior distribution at $\sim 35 M_{\odot}$ ($\sim 14 M_{\odot}$) is the most (least) tightly constrained.

The resulting perturbation functions are shown in Fig. 4.3 for 10 randomly chosen mock catalogs and GWTC-3. The perturbation functions deviate from their prior distribution in the mass range where detections exist (above $\sim 5 M_\odot$ and below $\sim 85 M_\odot$), even in the case of mock catalogs. This means that the perturbation functions are informed by the mock data despite the mock data not inherently requiring a deviation from a power law. The question still remains whether the perturbation function heights inferred from mock catalogs with no substructure are larger than those inferred from GWTC-3. While nonzero values of the perturbation function are common in the 10 mock catalog fits shown in Fig. 4.3, only a few amplitudes appear comparable in height to the three largest amplitudes of the GWTC-3 perturbation function.

To verify this, we isolate the largest amplitude perturbations for all 300 mock catalog fits and compare them to the three largest amplitude perturbations for the GWTC-3 fit. These are plotted in the leftmost panels of Fig. 4.4. The light grey curves are the posterior distributions of largest perturbation function amplitudes $\{g_j(f_{\text{max}})\}$ for each simulated catalog j . These appear to have the same general shape as one another, though with noticeable scatter. The orange curves in each panel are the posterior distributions of GWTC-3’s perturbation function $g_{\text{GWTC-3}}(f(m_1))$ at its three maximal locations: $m_1 = 13.8 M_\odot, 10.3 M_\odot$, and $35.7 M_\odot$.

The distribution for the $\sim 14 M_\odot$ dip appears qualitatively similar to that of the simulated catalogs, the $\sim 10 M_\odot$ peak appears to be slightly shifted with respect to most of the simulated catalogs but still within their range, and the $\sim 35 M_\odot$ peak is noticeably shifted towards higher values relative to the bulk of the simulated catalog distributions. This suggests that the $\sim 35 M_\odot$ peak is unlikely to be the result of Poisson noise or modeling systematics, while other features could plausibly be explained by those effects.

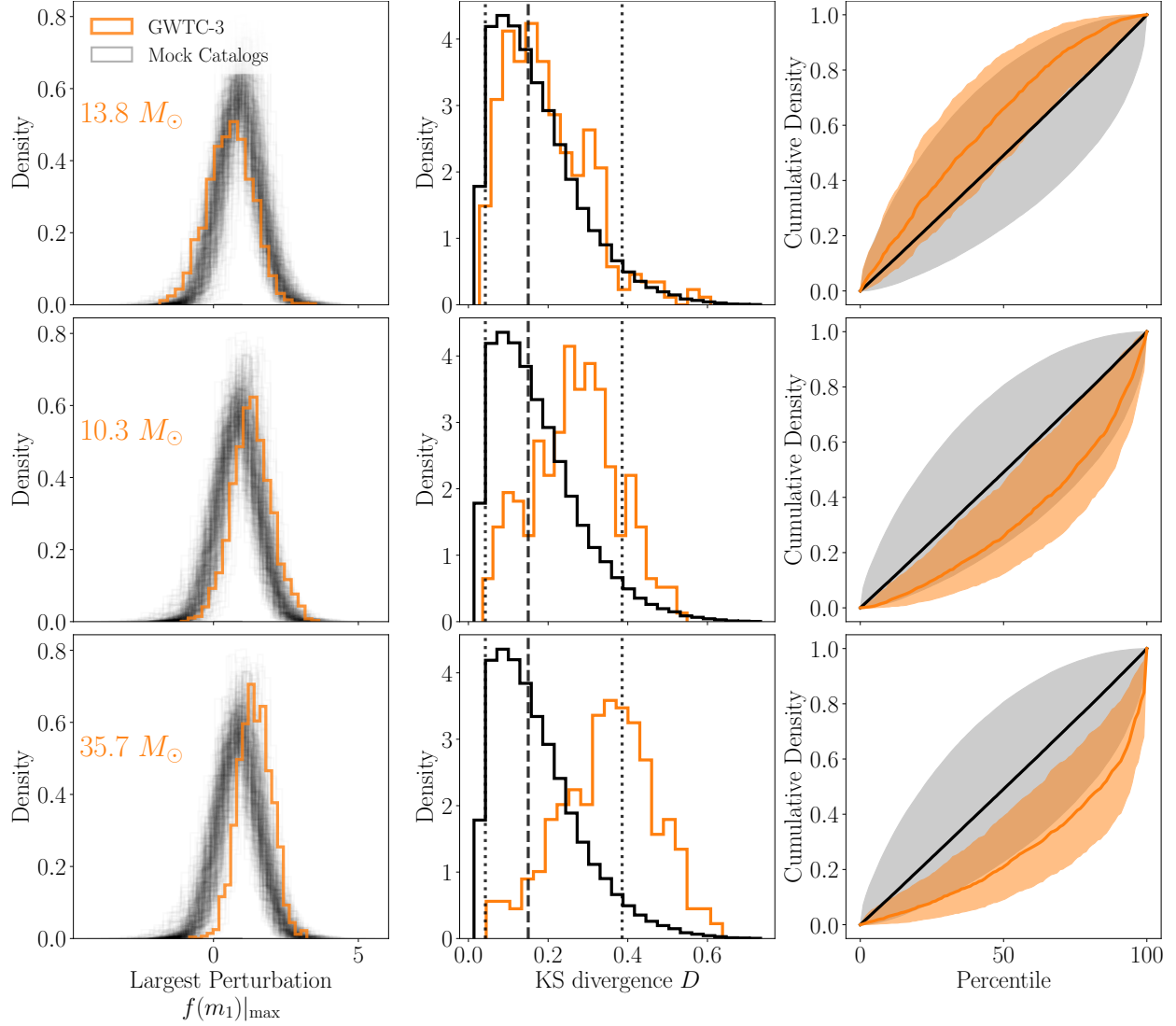


Figure 4.4: Three largest deviations from a power law observed in GWTC-3 compared to mock catalogs. *Left column:* The posterior distribution of perturbation function heights at the location where the posterior distribution is maximal for mock catalogs (*light grey*) and GWTC-3 (*solid orange*). *Middle column:* Null distribution (*black*) and GWTC-3 distribution (*orange*) of Kolmogorov–Smirnov (KS) divergences between the individual distributions in the left column. Smaller values of the KS divergence indicate more similar distributions. *Right column:* Null distribution (*black*) and GWTC-3 distribution (*orange*) of percentiles. Large deviations from the diagonal indicate a more significant rightward shift of the GWTC-3 distribution relative to the mock catalogs. Each row corresponds to a different local extremum for GWTC-3: $m_1 = 13.8 M_\odot$ (*top*), $m_1 = 10.3 M_\odot$ (*middle*), and $m_1 = 35.7 M_\odot$ (*bottom*), while the global extrema for each mock catalog are shown in all rows, along with the aggregated distribution across all mock catalogs (*solid black*). The $\sim 35 M_\odot$ peak is an outlier with respect to both the KS and percentile statistics, but the other two features are more ambiguous.

Maximum perturbation amplitude

To obtain a more quantitative measure, we derive several metrics from the distributions of maximal perturbation function amplitudes. The first uses the Kolmogorov–Smirnov (KS) test: we compute the KS divergence D between each of the $\{g_j(f_{\max})\}$ distributions to obtain a null distribution of KS divergences, shown in the solid black curve in the middle column of Fig. 4.4. We then perform a KS test between the $\{g_j(f_{\max})\}$ distributions and $g_{\text{GWTC-3}}(f(m_1))$ and obtain the orange curves in the middle column of Fig. 4.4. From this, we find that the KS divergences for mock catalogs are larger than those of GWTC-3 20%, 11%, and 7% of the time for the $14 M_\odot$, $10 M_\odot$, and $35 M_\odot$ features, respectively. This means, for example, that mock catalogs can produce perturbation function posteriors as tall as the one inferred from GWTC-3 at $\sim 35 M_\odot$ only 7% of the time. Written in terms of $g(f)$, $g_{\text{GWTC-3}}(f(14 M_\odot)) \neq g_j(f_{\max})$ to 20%, $g_{\text{GWTC-3}}(f(10 M_\odot)) \neq g_j(f_{\max})$ to 11%, and $g_{\text{GWTC-3}}(f(35 M_\odot)) \neq g_j(f_{\max})$ to 7%. Though none of these percentages are convincingly small, this indicates that the orange histograms are more statistically distinct from the black histograms in the case of the $\sim 35 M_\odot$ peak than they are in the cases of the features at $10 M_\odot$ and $14 M_\odot$.

The second metric is obtained by quantifying the shift of $g_{\text{GWTC-3}}(f(m_1))$ relative to the set of $\{g_j(f_{\max})\}$. For each point in $g_{\text{GWTC-3}}(f(m_1))$, we calculate the percentile in which it lies in each of the $\{g_j(f_{\max})\}$, obtaining the orange bands in the rightmost panels of Fig. 4.4. For comparison, we do the same for each of the $\{g_j(f_{\max})\}$ relative to each other, constructing the grey bands in the rightmost panels of Fig. 4.4. We then take the mean of the set of light orange bands and light black bands to obtain the solid orange and solid black curves, respectively. The black bands serve as null distributions, so large deviations from those indicate significant shifts. We observe a large deviation for the $\sim 35 M_\odot$ peak, a moderate deviation for the $\sim 10 M_\odot$ peak, and only a slight deviation for the $\sim 14 M_\odot$ dip. Quantitatively, $g_{\text{GWTC-3}}(f(35 M_\odot)) \geq g_j(f_{\max})$ to $83^{+17}_{-69}\%$ (90% credible interval),

meaning that the $\sim 35 M_\odot$ peak lies in the 83_{-69}^{+17} rd percentile of the mock catalogs’ largest perturbation heights. For the other features, $g_{\text{GWTC-3}}(f(10 M_\odot)) \geq g_j(f_{\text{max}})$ to $74_{-60}^{+25}\%$ and $g_{\text{GWTC-3}}(f(14 M_\odot)) \geq g_j(f_{\text{max}})$ to $34_{-32}^{+52}\%$. In comparison, the corresponding statistic for the null distributions is $g_j(f_{\text{max}}) \geq g_i(f_{\text{max}})$ to $50_{-46}^{+47}\%$.

It is not possible to draw firm conclusions from these large uncertainties. However, the central values indicate that the $\sim 35 M_\odot$ peak is noticeably shifted relative to the mock catalogs’ perturbation functions, the $\sim 10 M_\odot$ peak is moderately shifted, and the $\sim 14 M_\odot$ dip even has a slightly lower amplitude than the maximum perturbation functions typical of mock catalogs.

Inconsistency with a power law

The final metric we consider is inspired by the statistic presented in Abbott et al. [2023a], which states that “the inferred perturbation $f(m_1)$ strongly disfavors zero at both the $10 M_\odot$ and $35 M_\odot$ peak.” We therefore turn from considering the full distribution of perturbation function heights at a given location to the percentile at which it excludes zero. A perturbation function amplitude of zero is a useful reference point for several reasons. The most intuitive is that it causes the population model to behave like a featureless power law, so a posterior that excludes zero to high credibility indicates an inconsistency with a power law. Zero is also the mean of the prior predictive distribution for the perturbation function: the prior allows for equal upwards and downwards fluctuations, symmetric about zero perturbation. Similarly, a vanishing perturbation function amplitude is the state to which we expect the posterior predictive distribution to asymptote in the limit of infinite detections from an underlying power law distribution. We therefore plot the percentile at which each mock catalog excludes zero perturbation in Fig. 4.5.

We then calculate how often a simulated catalog’s perturbation function excludes zero to the same credibility as that of GWTC-3. This is the same as finding the point along the

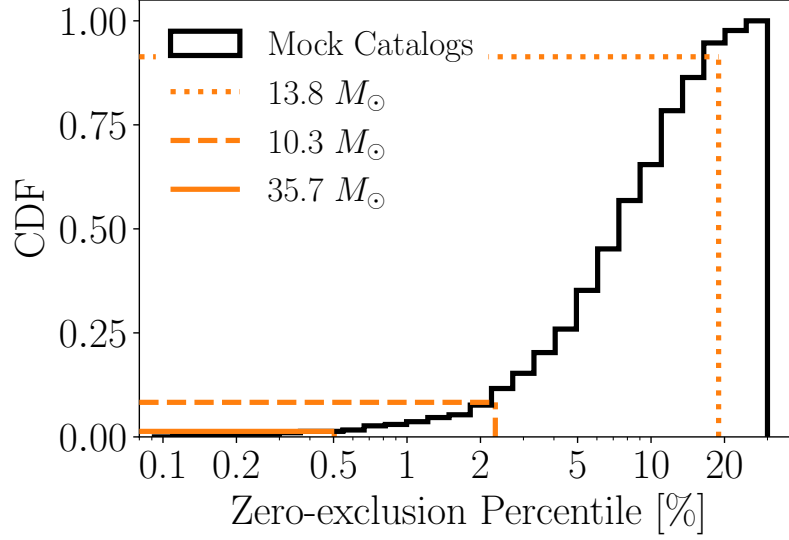


Figure 4.5: Percentile at which the posterior distribution of the perturbation function excludes zero for GWTC-3 (*orange vertical lines*) and catalogs drawn from a featureless distribution (*black histogram*). For GWTC-3, we evaluate the perturbation function’s posterior distribution at primary mass (m_1) values of $13.8 M_\odot$ (*dotted*), $10.3 M_\odot$ (*dashed*), and $35.7 M_\odot$ (*solid*). For mock catalogs, we find the primary mass value at which the perturbation function is maximal and evaluate its posterior distribution there. The values reported here are the percentage of the posterior distribution that is greater than zero at those values in m_1 . The $13.5 M_\odot$ feature excludes zero to a level comparable to some of the mock catalogs, but the other two features exclude zero to a level not reproducible by any mock catalogs.

y -axis of Fig. 4.5 at which each of the vertical orange lines hits the cumulative distribution function. 1.7%, 10.0%, and 92.7% of the $\{g_j(f_{\max})\}$ exclude zero to the same percentile as $g_{\text{GWTC-3}}(f(35 M_{\odot}))$, $g_{\text{GWTC-3}}(f(10 M_{\odot}))$, and $g_{\text{GWTC-3}}(f(14 M_{\odot}))$, respectively. The fact that, for example, $g_{\text{GWTC-3}}(f(14 M_{\odot})) < 0$ to 20.7% but 92.7% of mock catalogs have a similar or smaller statistical excursion is due in part to the difference between Bayesian credible intervals and frequentist p -values, and because our metric corrects for the look-elsewhere effect by comparing GWTC-3’s perturbation function at specific locations to all possible locations in the mock catalogs.

Combined with the metrics presented in Section 4.3.3, the results above lead us to conclude that the peak at $\sim 35 M_{\odot}$ is difficult to reproduce with featureless catalogs, but it is possible that the dip at $\sim 14 M_{\odot}$ is just a large fluctuation rather than an astrophysical feature. The peak at $\sim 10 M_{\odot}$ is difficult to reproduce with featureless catalogs, though it is easier to reproduce than the $\sim 35 M_{\odot}$ peak. We discuss the interpretation of this feature in more detail in Section 4.4.

In summary, featureless catalogs can sometimes produce features as tall as the $\sim 10 M_{\odot}$ peak, and they can sometimes produce perturbations constrained away from zero with the same credibility. The dip at $\sim 14 M_{\odot}$ could be a Poisson fluctuation because fits to featureless catalogs can easily produce perturbations as large, and as credibly constrained away from zero perturbation. The peak at $\sim 35 M_{\odot}$ is difficult to reproduce by mock catalogs in any way: its perturbation amplitude is too large and too credibly constrained away from zero.

The fact that we find one of the features explainable by Poisson noise is consistent with Section 4.2, which suggests that up to two of the excursions from a power law can be explained by Poisson fluctuations. Our conclusions are also in broad agreement with those presented in Abbott et al. [2023a], as they report confident detections for the two largest peaks in the mass distribution but only modest evidence for the dip at $\sim 14 M_{\odot}$.

4.4 Discussion

Previous analyses of the BBH mass spectrum by the LVK and others have found evidence for structure beyond a simple power law [Abbott et al., 2021d, 2023a]. There has been considerable work exploring possible astrophysical causes of these identified features. Our aim is instead to determine, from a statistical viewpoint, whether astrophysical arguments need be invoked at all.

We first demonstrate that it is only possible for up to two of the three deviations from a power law to be explained by Poisson noise about a single power law distribution. Therefore, at least one feature must be added on top of a power law to describe the data.

We then perform a more thorough analysis, simulating thousands of BBHs with measurement uncertainty, selection effects, and a known underlying distribution. We fit the POWER LAW + SPLINE model to the resulting catalogs and find that the data is inconsistent with a single power law, agreeing with the LVK result. However, we find that one of the previously identified features, an underdensity at $\sim 14 M_\odot$, may not be present in the true astrophysical distribution. Instead, it may have been the result of a Poisson fluctuation around a simple power law, or an artifact of the models used to fit the mass spectrum. The metrics constructed in this work differ from those previously used to assess the significance of features in the mass distribution because, by virtue of comparing to several simulated catalogs, they correct for the look-elsewhere effect. This is only in mild tension with the conclusions reached by Abbott et al. [2023a], as they report “modest evidence” in favor of a dip at $14 M_\odot$.

We find the other two previously identified peaks, at $\sim 10 M_\odot$ and $\sim 35 M_\odot$, unlikely to be the result of Poisson noise or modeling artifacts. Simulated catalogs coming from distributions that do not include these features can reproduce the height of the $\sim 10 M_\odot$ peak, but not its lack of support for zero perturbation. The $\sim 35 M_\odot$ peak is difficult to reproduce from featureless catalogs in any way.

Our conclusions are consistent with a recent study by Callister and Farr [2024] who fit the BBH mass distribution with an autoregressive model and find that the primary mass distribution gradually decreases as a function of mass and exhibits two local maxima with a relatively flat continuum between them. They interpret the 14 solar mass dip found by other analyses to be a flattening of the power law index at lower masses rather than a local minimum. We also find similar results to Edelman et al. [2023] who construct the mass distribution entirely from basis splines and find peaks at $\sim 10 M_\odot$ and $\sim 35 M_\odot$. The significance of the peaks near $10 M_\odot$ and $35 M_\odot$, as well as the lack of significance of the dip near $14 M_\odot$, is also in agreement with Sadiq et al. [2022] and Wong and Cranmer [2022].

The dip near $\sim 14 M_\odot$ may be a large Poisson fluctuation or an artifact of the models used to characterize it. If it is in fact a feature of the underlying distribution, it is difficult to resolve with current observations.

The peak near $\sim 10 M_\odot$ is likely an imprint of the true astrophysical distribution. Its amplitude is slightly larger than what featureless catalogs can produce with random fluctuations, and it is inconsistent with the power law that describes the rest of the BBH mass distribution at a level that only a small fraction of featureless catalogs can achieve. We therefore report moderate evidence that additional structure beyond a power law is needed to explain the peak at $\sim 10 M_\odot$.

The $\sim 10 M_\odot$ feature may either be an additional peak that is distinct from the one created by the underlying smoothed power law at $\sim 7 M_\odot$ [Abbott et al., 2023a, Edelman et al., 2022, Tiwari, 2022] or the sole peak in the region between $\sim 5 M_\odot$ and $\sim 20 M_\odot$ [Edelman et al., 2023]. These two possibilities can be seen in Fig. 4.1. The former scenario is the case where we interpret the first two peaks in the orange band as distinct from one another, therefore treating the global maximum inferred by POWER LAW + SPLINE as a different feature from the global maximum inferred by POWER LAW + PEAK. In the latter scenario, the role of the perturbation function is to shift the global maximum from the value

inferred by the power law component to a slightly higher value without removing the mass distribution’s support for $5\text{--}10\,M_\odot$ objects. A simple smoothed power law, such as that employed by the POWER LAW + PEAK model (see grey and blue bands in Fig. 4.1), may not be flexible enough to place a global maximum at $\sim 10\,M_\odot$ while also fitting the correct slope at larger masses and fitting the correct merger rate below $\sim 10\,M_\odot$, so it places its global maximum at $\sim 7\,M_\odot$. This scenario, in which there is a single local maximum below $\sim 12\,M_\odot$, is consistent with Edelman et al. [2023] and Callister and Farr [2024], both of whom find only one significant maximum between approximately $3\,M_\odot$ and $12\,M_\odot$ using fully non-parametric methods. If this interpretation is correct and the global maximum of the BBH mass distribution is indeed offset from the minimum mass by $\sim 5\,M_\odot$, the upper edge of the lower mass gap may not be as morphologically simple as previously assumed [e.g., Fishbach et al., 2020a, Farah et al., 2022a, Ezquiaga and Holz, 2022], making it potentially difficult to resolve with parametric models alone. The marginal evidence for the significance of the $\sim 10\,M_\odot$ peak is likely driven by the fact that current observations are insufficient to distinguish between these two scenarios, which is unsurprising considering the lower sensitivity of GW detectors to low-mass events relative to high-mass events.

A peak anywhere between $\sim 7\,M_\odot$ and $\sim 10\,M_\odot$ could be indicative of particular evolutionary processes that are dominant within formation environments. van Son et al. [2022b] showed that a global maximum near this value is consistent with and robustly predicted by the stable mass transfer channel in isolated binary evolution, as stability during mass transfer requires mass ratios between the donor star and accreting compact object to be relatively symmetric, and stellar companions to $\sim 10\,M_\odot$ BHs must be near this mass to form compact objects above the minimum BH mass. This may be an indication that the stable mass transfer channel operates more efficiently than the traditional common envelope channel for generating merging BBHs. If the stable mass transfer channel is indeed the cause of the global maximum in the primary mass distribution, the exact location of this global

maximum will constrain the core mass fraction, mass transfer stability, and mass transfer efficiency of this process [van Son et al., 2022b]. Though dynamical formation channels with low escape velocities, such as globular clusters, struggle to produce a global maximum at $10\ M_{\odot}$ [Antonini et al., 2022], dynamical environments with higher escape velocities may more readily produce merging BBHs with lower masses around $10\ M_{\odot}$ due to the more prevalent lower-mass BHs preferentially remaining bound to these clusters following supernova kicks.

We find that the peak centered on $35\ M_{\odot}$ is the most likely to be a feature of the true underlying distribution. This bodes well for the “spectral siren” [Farr et al., 2019, Ezquiaga and Holz, 2022] method of estimating cosmological parameters from GW observations, as this peak happens to be the most informative feature for this method since it is a well-measured, somewhat sharp feature in the mass distribution [Abbott et al., 2023b]. The astrophysical process that gives rise to this feature is still a topic of discussion. The key reason for including a flexible bump-like feature in the phenomenology of parametric models, such as the POWER LAW + PEAK model used by the LVK [Talbot and Thrane, 2018], was to accommodate a potential build-up of BHs with masses just below the pair instability mass gap, as pulsational pair instability supernovae are predicted to efficiently shed material from high-mass stars with cores in the mass range of $M_{\text{core}} \sim 45 - 65$ [Woosley, 2017, 2019, Marchant et al., 2019, Renzo et al., 2020]. It is difficult to reconcile the locations of the local maxima found in the BBH primary mass distribution with predictions of the pair instability process in the cores of massive stars. The largest uncertainty determining the location of the lower edge of the pair instability mass gap is the $^{12}\text{C}(\alpha, \gamma)^{16}\text{O}$ reaction rate, which determines the abundance of oxygen in stellar cores [e.g., Farmer et al., 2019]. Higher $^{12}\text{C}(\alpha, \gamma)^{16}\text{O}$ reaction rates lead to a higher oxygen abundance in the stellar core, which will ignite explosively during core collapse and lead to (pulsational) pair instability supernovae occurring at lower core masses. However, even at 3σ deviations above the median measured value of the $^{12}\text{C}(\alpha, \gamma)^{16}\text{O}$ reaction rate,

the lower end of the mass gap only reaches $\approx 38 M_{\odot}$ [Farmer et al., 2020]. This is above where the measured overdensity in the observed mass spectrum occurs. This may be an indication that the peak at $35 M_{\odot}$ is the result of certain isolated binary evolution scenarios [e.g. chemically homogeneous evolution, see du Buisson et al., 2020, Zevin et al., 2021, Bavera et al., 2022], another BBH formation channel entirely (e.g. globular clusters, see Antonini et al. 2022), or that stellar evolution models are missing particular ingredients that can shift the location of the pair instability gap (relaxing the assumption that the exploding stars are hydrogen-free, adjustments to convective overshooting, see e.g. Iorio et al. 2022).

Additionally, several studies have suggested that the observed peaks in the BBH mass distribution can be explained by successive generations of hierarchical mergers [Tiwari, 2021, Mahapatra et al., 2022, Tiwari, 2022], though no correlation has been detected in the spin distribution of BBHs [Biscoveanu et al., 2022a], which is also necessitated by the hierarchical merger formation scenario [Gerosa and Berti, 2017, Fishbach et al., 2017, Rodriguez et al., 2019, Kimball et al., 2020a, Doctor et al., 2020, 2021, Gerosa et al., 2021]. Additionally, for these peaks to correspond to hierarchical mergers of the same population, the dominant hierarchical pairing would have to be the first generation BH with a third generation BH [Mahapatra et al., 2022, Tiwari, 2022], whereas the dominant pairing predicted by Rodriguez et al. [2019] is a first generation BH generation with a second generation BH. While it is certainly possible that GWTC-3 contains hierarchical mergers (e.g. Abbott et al. 2020d, though also see Fishbach and Holz 2020a), the relative fraction of events formed this way is likely too small to form the structure observed in the primary mass distribution [Kimball et al., 2020b], and some fine-tuning may be needed to avoid a cluster catastrophe [Zevin and Holz, 2022]. The exact physical reason for the overdensity at $35 M_{\odot}$ therefore remains unclear. However, we confirm that it is a robust signature in the observational data; future observing runs will help to constrain its precise location, width, and possible redshift evolution.

4.5 Additional analysis details

The following software packages were used in this chapter: `gwpopulation` [Talbot et al., 2019], `bilby` [Ashton et al., 2019], `dynesty` [Speagle, 2020], `numpy` [Harris et al., 2020], `xarray` [Hoyer and Hamman, 2017], `matplotlib` [Hunter, 2007], `pandas` [McKinney, 2010]

4.5.1 Generation of mock observations in *GWMockCat*

We describe the process used to simulate gravitational wave event posteriors in mass and redshift, based on the procedure developed in Fishbach et al. [2020b].

This process neglects the generation of spin posteriors as this work only seeks to understand the significance of features in the mass distribution, and individual-event likelihoods are approximately separable in spin and primary mass for BBHs, and we do not model any spin populations in this work. However, spin and mass parameters are not totally uncorrelated for low-mass or high mass ratio events, so future work attempting to validate features seen in the mass ratio distribution, NSBH, or BNS populations should consider simulating spin parameters as well. A lightweight, publicly-available `python` package that can reproduce these mock posteriors and generate similar catalogs from arbitrary underlying populations and detector sensitivities is available for download and installation⁴. The package is called *GWMockCat*, and installation instructions, examples, and documentation are available in the git repository. Several packages exist to draw events from BBH population models [Belczynski et al., 2008, Breivik et al., 2020, Riley et al., 2022, e.g.], some of which also simulate GW detector selection effects [Karathanasis et al., 2022b]. *GWMockCat* complements these by additionally simulating event-level posteriors without the need to run full parameter estimation inference, saving significant computational time.

To create realizations of catalogs that would reasonably result from a known underlying astrophysical population, $p(m_1, m_2, z)$, we first make independent draws of the event param-

4. <https://git.ligo.org/amanda.farah/mock-PE>

eters, $\{m_1, m_2, z\}$, from that population model. Each draw corresponds to a potential event in the catalog, although we draw many more potential events than we wish to keep since not all events generated from the astrophysical distribution will ultimately be detected. We then convert each event’s redshift z and source-frame component masses to a detector-frame (redshifted) chirp mass, $\mathcal{M}_{\text{det}}$, and symmetric mass ratio, η . The symmetric mass ratio and source-frame chirp mass $\mathcal{M}$ are related to the source-frame component masses via

$$\eta = \frac{m_1 m_2}{(m_1 + m_2)^2} \quad (4.3)$$

$$\mathcal{M} = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}}. \quad (4.4)$$

All detector-frame masses are related to their source-frame values via $M_{\text{det}} = M(1 + z)$, where M can describe any parameter with units of mass (e.g. m_1, m_2 , or $\mathcal{M}$).

We then utilize the basic procedure outlined in Fishbach et al. [2018] and Fishbach et al. [2020b] to assign “observed” parameters for each event, using measurement uncertainty that is correlated across parameters and a mock parameter estimation likelihood. We first calculate an optimally oriented signal-to-noise ratio (SNR) ρ_{opt} from the true event parameters using a characteristic power spectral density (PSD) of the LIGO Livingston detector in O3 [Abbott et al., 2020b]. ρ_{opt} is the SNR an event would have if it were “optimally oriented” with respect to the detector, that is, directly overhead and with its angular momentum vector pointed along the line of sight [Chen et al., 2021]. In reality, GW sources have varying sky positions and angular momentum vectors. The effect on the SNR of a source’s deviation from the optimal orientation can be summarized by a multiplicative constant, Θ , such that

$$\rho = \rho_{\text{opt}} \Theta, \quad (4.5)$$

where Θ is between zero and unity.

GW sources are typically assumed to be distributed isotropically in sky position and

orientation. For a single detector, this yields a corresponding distribution for Θ , described in Finn and Chernoff [1993]. Therefore, for each event i , we assign a true value $\hat{\Theta}_i$ drawn from this distribution and use it to calculate the event's true single-detector SNR $\hat{\rho}$. The set of *true* parameters for each potential event in the catalog is then $\hat{\theta}_i = \{\hat{\mathcal{M}}_{\text{det}}, \hat{\eta}, \hat{\rho}, \hat{\Theta}\}_i$.

Given the true parameters, the basic procedure of generating samples from the posterior distribution of each event is to draw an observation from each event's likelihood, use that observation as the central value of the posterior distribution, and then to draw samples from that posterior, assuming a prior.

To obtain *observed* parameters, θ_i^{obs} , we need the likelihood, $\mathcal{L}_{\text{total}}(\theta_i^{\text{obs}}|\hat{\theta}_i)$. We model each event's likelihood as

$$\mathcal{L}_{\text{total}}(\theta_i^{\text{obs}}|\hat{\theta}_i) = \mathcal{L}_{\mathcal{M}}(\mathcal{M}_{\text{det},i}^{\text{obs}})\mathcal{L}_{\eta}(\eta_i^{\text{obs}})\mathcal{L}_{\Theta}(\Theta_i^{\text{obs}})\mathcal{L}_{\rho}(\rho_i^{\text{obs}}), \quad (4.6)$$

where

$$\begin{aligned} \mathcal{L}_{\mathcal{M}}\left(\ln\left(\mathcal{M}_{\text{det},i}^{\text{obs}}\right) \mid \ln\left(\hat{\mathcal{M}}_{\text{det},i}\right), \rho_i^{\text{obs}}\right) &= \mathcal{N}\left(\ln\left(\mathcal{M}_{\text{det},i}^{\text{obs}}\right) \mid \mu = \ln\left(\hat{\mathcal{M}}_{\text{det},i}\right), \sigma = \sigma_i^{\mathcal{M}}\left(\rho_i^{\text{obs}}\right)\right) \\ \mathcal{L}_{\eta}\left(\eta_i^{\text{obs}} \mid \hat{\eta}_i, \rho_i^{\text{obs}}\right) &= \mathcal{N}\left(\eta_i^{\text{obs}} \mid \mu = \hat{\eta}_i, \sigma = \sigma_i^{\eta}, \left(\rho_i^{\text{obs}}\right)\right) \\ \mathcal{L}_{\Theta}\left(\Theta_{\text{obs},i} \mid \hat{\Theta}_i, \rho_i^{\text{obs}}\right) &= \mathcal{N}\left(\Theta_i^{\text{obs}} \mid \mu = \hat{\Theta}_i, \sigma = \sigma_i^{\Theta}\left(\rho_i^{\text{obs}}\right)\right) \\ \mathcal{L}_{\rho}\left(\rho_i^{\text{obs}} \mid \hat{\rho}_i\right) &= \mathcal{N}\left(\rho_i^{\text{obs}} \mid \mu = \hat{\rho}_i, \sigma = \sigma_i^{\rho}\right). \end{aligned} \quad (4.7)$$

Here, $\mathcal{N}(\mu, \sigma)$ is the normal distribution with mean μ and standard deviation σ .

The standard deviations are determined by assuming the uncertainties on all parameters except for the SNR scale inversely with ρ_{obs} [Veitch et al., 2015]. In stationary, Gaussian noise, we expect the matched-filter SNR in a single detector to have unit variance [Allen et al., 2012], i.e. $\sigma_i^{\rho} = 1$ for all i . We therefore draw ρ_{obs} for each event from $\mathcal{L}^{\rho}(\rho_i^{\text{obs}}|\rho_i)$. This observed SNR will serve as the detection statistic that determines whether each event is

observable. We assume events that pass an SNR threshold of $\rho_{\text{obs},i} > 8$ in a single detector are detected. In this way, we allow for events near threshold to fluctuate above or below threshold, emulating the actual noise process in the detectors. Of the events that make it through detection, we randomly select 69 of them to constitute a mock catalog with the same number of BBHs as were analyzed by Abbott et al. [2023a]. The standard deviations for $\mathcal{M}_{\text{det}}$, η , and Θ of the detected events are calculated via

$$\begin{aligned}\sigma_i^{\mathcal{M}}(\rho_i^{\text{obs}}) &= u_{\mathcal{M}}/\rho_i^{\text{obs}} \\ \sigma_i^{\eta}(\rho_i^{\text{obs}}) &= u_{\eta}/\rho_i^{\text{obs}} \\ \sigma_i^{\Theta}(\rho_i^{\text{obs}}) &= u_{\Theta}/\rho_i^{\text{obs}},\end{aligned}\tag{4.8}$$

where we have chosen $u_{\mathcal{M}} = 0.08 M_{\odot}$, $u_{\eta} = 0.022$, and $u_{\Theta} = 0.21$ to match uncertainties in these parameters typical of events observed in O3.

Observed values for all parameters are drawn from Equation 4.6 with standard deviations defined in Equation 4.8. With θ_i^{obs} in hand, we are now ready to construct a posterior distribution. We apply the following priors:

$$\begin{aligned}\pi(\mathcal{M}_{\text{det}}) &= U(0 M_{\odot}, 500 M_{\odot}) \\ \pi(\eta) &= U(0, 0.25) \\ \pi(\Theta) &= U(0, 1) \\ \pi(\rho) &= U(0, 300),\end{aligned}\tag{4.9}$$

where $U(x_1, x_2)$ is the uniform distribution with lower bound x_1 and upper bound x_2 . The bounds on η and Θ are chosen because those parameters are only physically defined in the domains $(0, 0.25]$ and $[0, 1]$, respectively. Neither $\mathcal{M}$ nor ρ are defined below zero, but the upper bounds were chosen somewhat arbitrarily: they must only be large enough that the likelihood has minimal support above them. The posterior distributions for each parameter

are then Gaussians centered on the observed value, with standard deviations defined in Equation 4.8. They are therefore the same as the distributions in Equation 4.7, but with the role of the true and observed values switched. We then simulate multiple-dimensional posterior samples for each event by drawing 5000 independent samples⁵ of detector-frame chirp mass, symmetric mass ratio, and Θ from the posterior. Explicitly,

$$\begin{aligned}
\log \mathcal{M}_{\text{det},i} &\sim \mathcal{N}(\mu = \ln(\mathcal{M}_{\text{det},i}^{\text{obs}}), \sigma = \sigma_i^{\mathcal{M}}) \\
\eta_i &\sim \mathcal{N}(\mu = \eta_i^{\text{obs}}, \sigma = \sigma_i^{\eta}) \\
\Theta_i &\sim \mathcal{N}(\mu = \Theta_i^{\text{obs}}, \sigma = \sigma_i^{\Theta}) \\
\rho_i &\sim \mathcal{N}(\mu = \rho_i^{\text{obs}}, \sigma = \sigma_i^{\rho}).
\end{aligned} \tag{4.10}$$

Realistic correlations between other parameters such as component masses and redshift are obtained by transforming samples in $\{\mathcal{M}_{\text{det}}, \eta, \Theta, \rho\}$ -space to $\{m_1, m_2, z\}$ -space. When necessary, we convert between luminosity distance and redshift using the cosmological parameters presented in Ade et al. [2016] so as to maintain consistency with the conventions used in [Abbott et al., 2019b, 2021b,g].

The induced prior on m_1, m_2 , and z is therefore not uniform in those parameters. This is reasonable, so long as users appropriately transform the prior when doing population inference on source-frame component masses and redshift. We therefore provide a module in `GWMockCat` that performs these transformations. For the case of this analysis, we opt to re-weight the samples to a prior that is uniform in detector-frame component mass and proportional to the square of the luminosity distance in order to mimic the priors used in the standard LVK analysis [Abbott et al., 2019b, 2021b,g].

5. We use 5000 samples to optimize the speed of population inference while also ensuring the number of effective samples used for Monte Carlo sums in the population inference always satisfies the criterion outlined in Farr [2019]. That criterion has since been shown to be insufficient and has been superseded by Essick and Farr [2022], but we utilize the former for consistency with the analysis performed in Abbott et al. [2023a]. However, users of the `GWMockCat` package can easily modify the number of posterior samples to suit their needs.

The fact that Equation 4.6 is separable up to dependence on $\rho_{\text{obs},i}$ means that once $\rho_{\text{obs},i}$ is calculated for a given event, samples for $\mathcal{M}_{\text{det}}, \eta_{\text{obs}}, \Theta_{\text{obs}}$, and ρ_{obs} can be drawn independently from each other. This approximate independence is due, in part, to the fact that detector-frame chirp mass, symmetric mass ratio, SNR, and Θ are the best-measured parameters of any compact binary coalescence signal. This fact saves considerable computational resources, allowing for many mock event posteriors to be generated quickly on a single CPU⁶.

We generate sensitivity estimates along with our mock catalogs to ensure that the selection function is calculated consistently to the event selection criteria [Essick and Fishbach, 2024]. To do this, we draw 2×10^7 independent samples in m_1, m_2, z , and Θ from the following distribution:

$$p(m_1, m_2, z, \Theta) \propto m_1^\alpha \left(\frac{m_2}{m_1} \right)^\beta \frac{dV_c}{dz} (1+z)^{\kappa-1} p(\Theta), \quad (4.11)$$

where we have chosen $\alpha = 2.35$, $\beta = 1.70$, and $\kappa = 2.7$, and $p(\Theta)$ is the distribution described in Finn and Chernoff [1993] which corresponds to isotropically oriented sources that are also isotropically positioned on the sky. We truncate this distribution below $m_2 = 1 M_\odot$, above $m_1 = 200 M_\odot$, and above $z = 4$, and confirm that there are no mock posterior samples outside of those ranges. We will refer to these draws as “injections.” We then calculate an optimally-oriented SNR for each injection using the same PSD as was used for the mock observations, and compute a true SNR using Equation 4.5. We emulate noise fluctuations in SNR in the same way we do for mock observations, namely by using Equation 4.7, so that each injection has a corresponding observed SNR. Injections can then be subject to the same selection criteria as our mock observations when performing a population inference (in our case, $\rho_{\text{obs}} > 8$).

We validate this process by constructing a mock catalog from a known distribution with

6. For example, a catalog of 100 events can be generated in $\mathcal{O}(10)$ seconds.

fixed hyperparameters, and then fitting the same distribution to our mock catalog, but allowing the hyperparameters to vary. We then verify that the recovered hyperparameters are consistent with those used to generate the mock catalog. The result is shown in Section 4.5.2, along with additional validation studies.

4.5.2 Validation of mock catalogs

In this Section, we validate the process of creating mock event posteriors and catalogs from a known underlying population outlined in Section 4.5.1. For this process, we use the same simulated catalogs utilized in Section 4.3.3. The simulated underlying population is described by $p_{\text{mock}}(m_1, m_2, z | \Lambda_{\text{mock}})$, where Λ_{mock} is the set of hyperparameters $\{\alpha, \delta, m_{\text{min}}, m_{\text{max}}, \beta, \kappa\}$,

$$p_{\text{mock}}(m_1, m_2, z | \Lambda_{\text{mock}}) \propto p(m_1 | \alpha, \delta, m_{\text{min}}, m_{\text{max}}) p(m_2 | m_1, \beta) p(z | \kappa), \quad (4.12)$$

and the individual mass and redshift distributions are given by the following:

$$p(m_1 | \alpha, \delta, m_{\text{min}}, m_{\text{max}}) \propto \begin{cases} 0 & \text{if } m < m_{\text{min}} \\ m_1^{-\alpha} \frac{1}{1+f(m-m_{\text{min}}, \delta)} & \text{if } m_{\text{min}} \leq m < m_{\text{min}} + \delta \\ m_1^{-\alpha} & \text{if } m \geq m_{\text{min}} + \delta \\ 0 & \text{if } m > m_{\text{max}} \end{cases}, \quad (4.13)$$

$$p(m_2 | m_1, \beta) \propto \left(\frac{m_2}{m_1} \right)^\beta, \quad (4.14)$$

$$p(z | \kappa) \propto \begin{cases} 0 & \text{if } (z < 0) \cup (z > z_{\text{max}}) \\ \frac{dV_c}{dz} (1+z)^{\kappa-1} & \text{otherwise} \end{cases}. \quad (4.15)$$

This is equivalent to the POWER LAW + PEAK model in Abbott et al. [2023a] and Abbott et al. [2021d], with λ_{peak} set to 0. We will call the population model described by

Parameter	Description	Value
β	Spectral index for the power law of the mass ratio distribution.	1.70
α	Negative spectral index for the power law of the primary mass distribution.	3.14
$m_{\min}$	Minimum mass of the primary mass distribution.	$4.56 M_{\odot}$
$m_{\max}$	Maximum mass of the primary mass distribution.	$81.08 M_{\odot}$
δ	Range of mass tapering at the lower end of the mass distribution.	$5.96 M_{\odot}$
κ	Spectral index for the power law factor of the redshift distribution.	2.7

Table 4.1: Hyperparameter values for the underlying population of mock catalogs described by SMOOTHED POWER LAW (Equations 4.12–4.15).

Equations 4.12–4.15 SMOOTHED POWER LAW. We generate catalogs from the model that results from setting Λ_{mock} to the values provided in Table 4.1. These values were chosen by fitting this population model to GWTC-3 (grey band in Fig. 4.1) and obtaining the median *a posteriori* value for each hyperparameter.

We validate the mock catalogs’ generation by fitting them with SMOOTHED POWER LAW and allowing the hyperparameters to be inferred from the mock data. We then determine whether the inferred values of the hyperparameters are consistent with the values in Table 4.1. We fit 100 mock catalogs of 69 events each, 10 results of which are shown in Fig. 4.6. While there is noticeable scatter about the injected value, it is generally consistent with the recovered mass distributions: the hyperparameters of the underlying mass distribution fall within the inferred mass hyperparameters’ 90% credible intervals 89.6% of the time. We therefore conclude that any biases that the mock posterior generation process introduces in the mass distribution are subdominant to the statistical uncertainties of the fit.

To further explore systematic differences caused by mock catalog generation that may be subdominant to the considerable statistical uncertainties resulting from a fit to only 69 events, we fit SMOOTHED POWER LAW to a single catalog of that is five times larger. The result is shown in Fig. 4.7. The hyperparameters of the underlying distribution seem to be

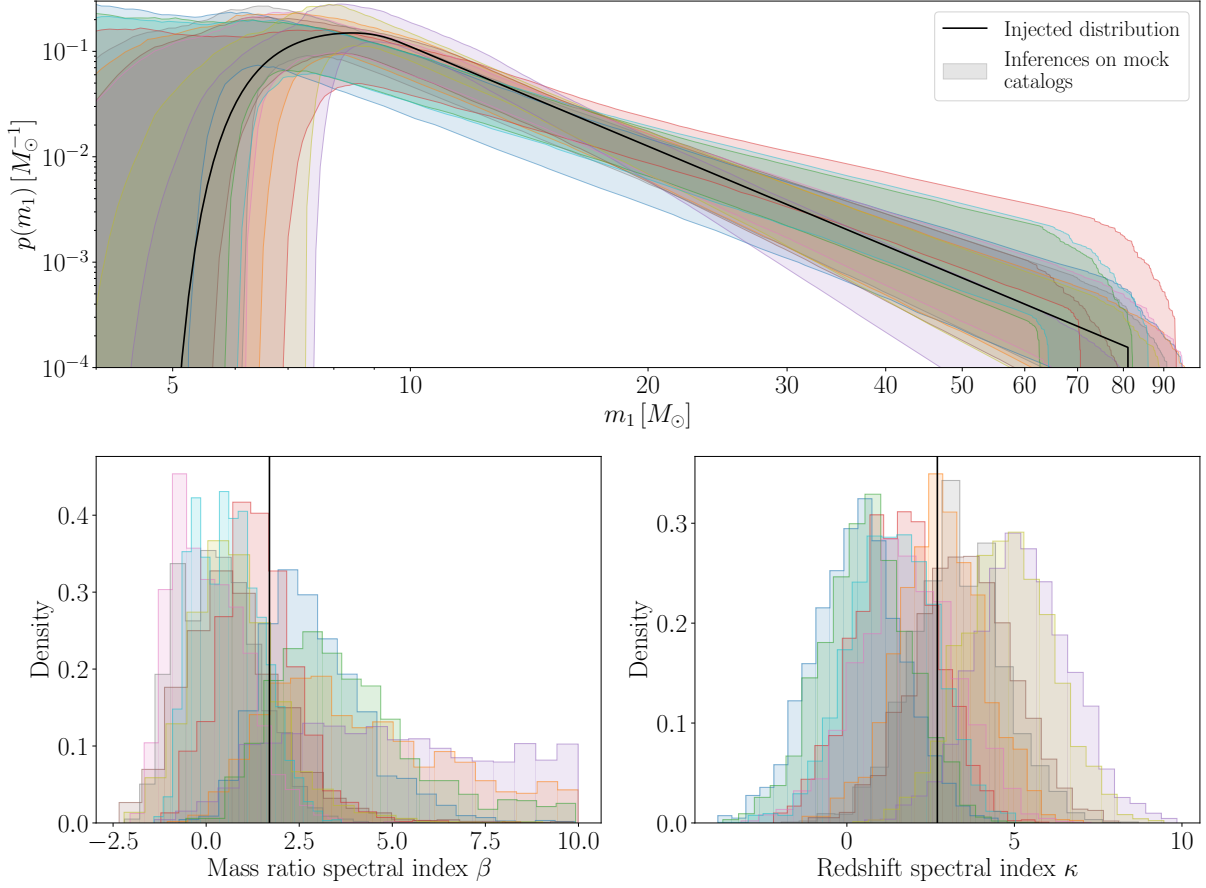


Figure 4.6: Injected (*Solid black line*) and recovered (*colored shaded bands*) distributions for 10 mock catalogs. *Top*: probability density function of primary masses. *Bottom left*: hyperposterior distribution for β , the power law spectral index of the mass ratio distribution. *Bottom Right*: hyperposterior distribution for κ , the spectral index of the power law factor in the redshift distribution.

consistent with the inferred hyperposterior, so we conclude that our mock event posterior generation process produces biases subdominant to measurement uncertainty typical of 345-event catalogs. We therefore find this method of generating mock catalogs sufficient to test the significance of features identified in the mass distribution of GWTC-3.

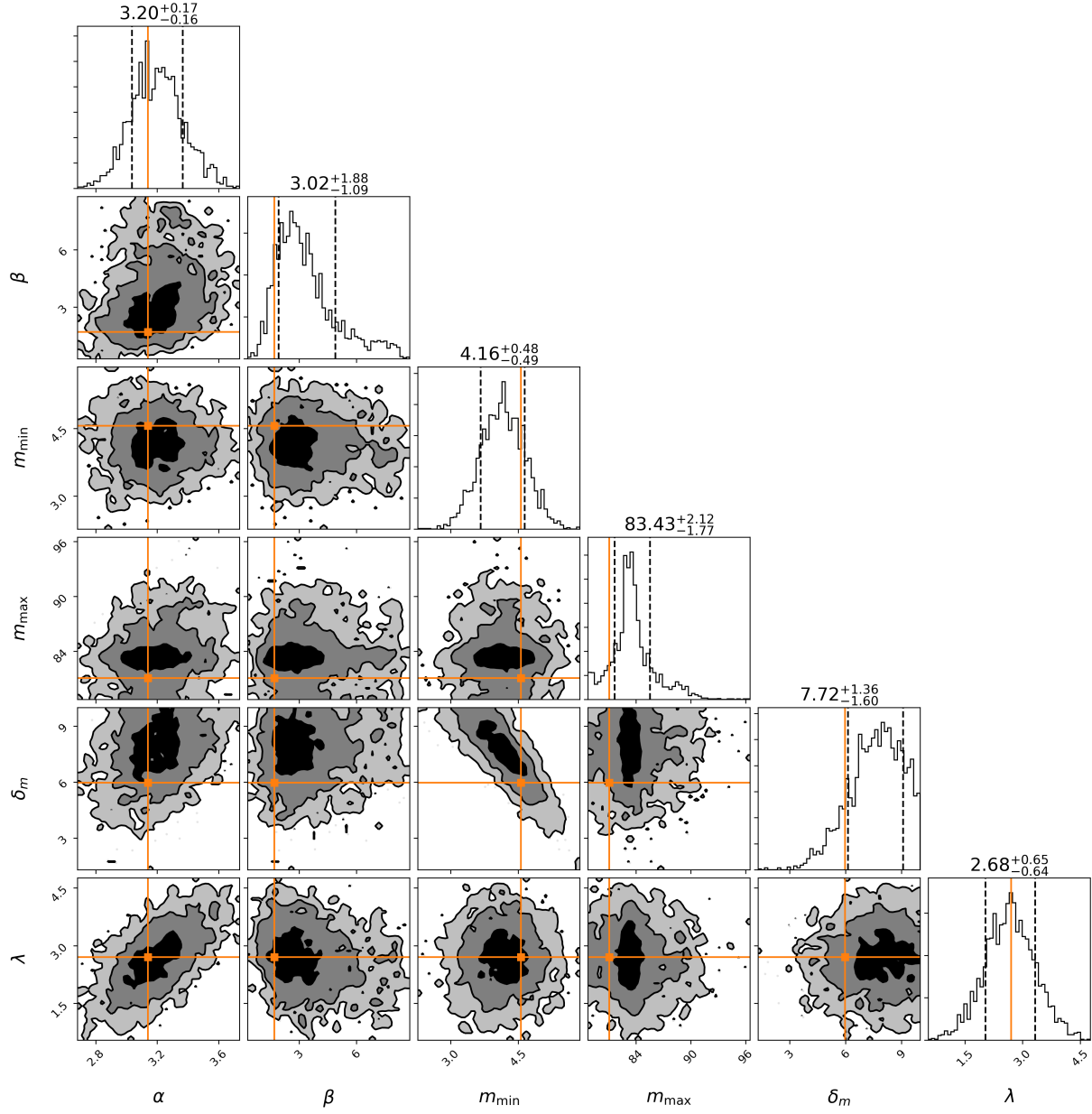


Figure 4.7: A corner plot of the inferred hyperposterior from a fit to a mock catalog with 345 events. The injected values are shown in orange. The recovered hyperposterior is consistent with the injected population.

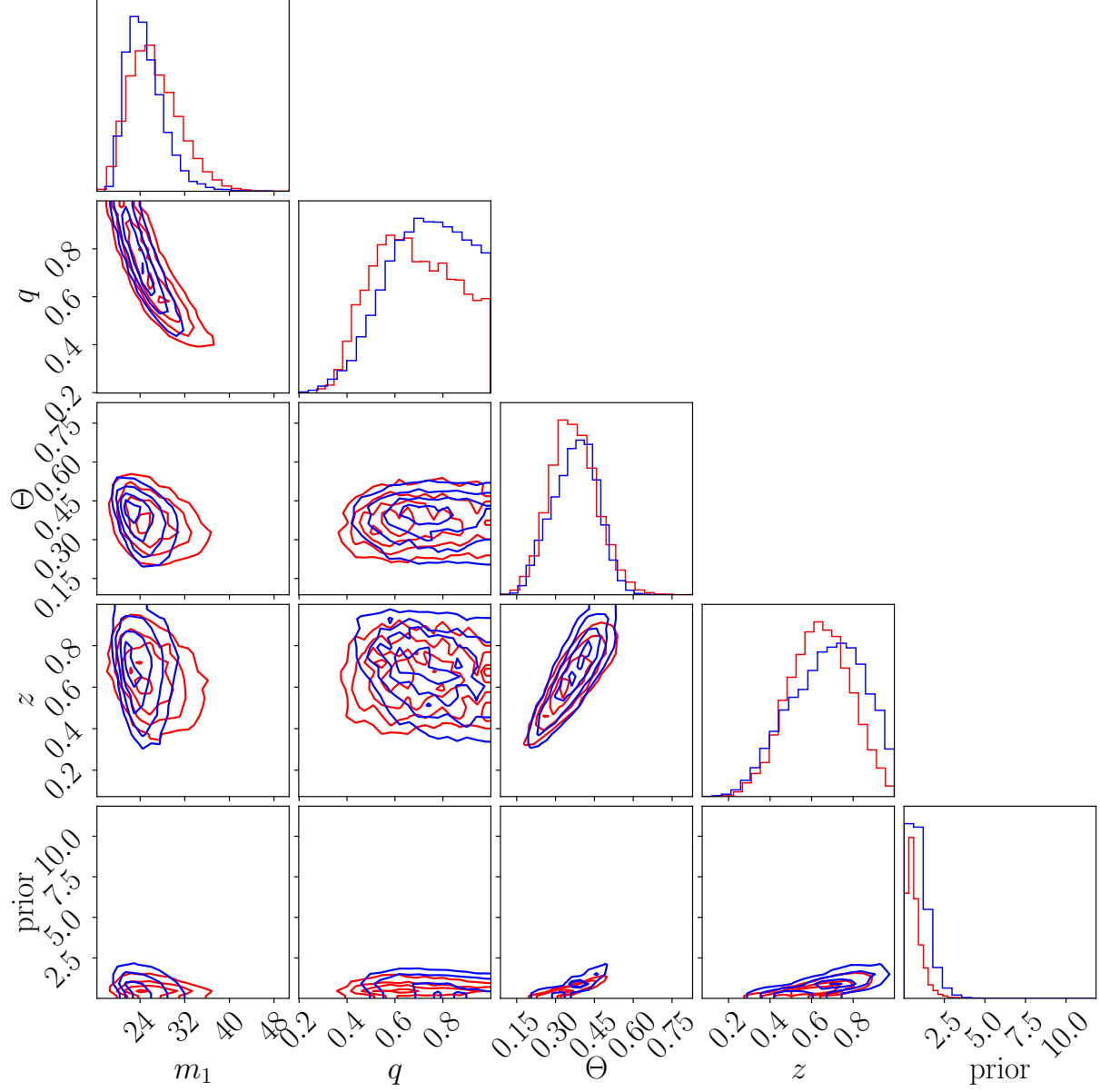


Figure 4.8: Mock posteriors simulated with **GWMockCat** (*red*) compared to posteriors made with full parameter estimation as released by the LVK (*blue*) for the event GW191215_223052. The two sets of posterior samples appear consistent to the level needed for a population analysis.

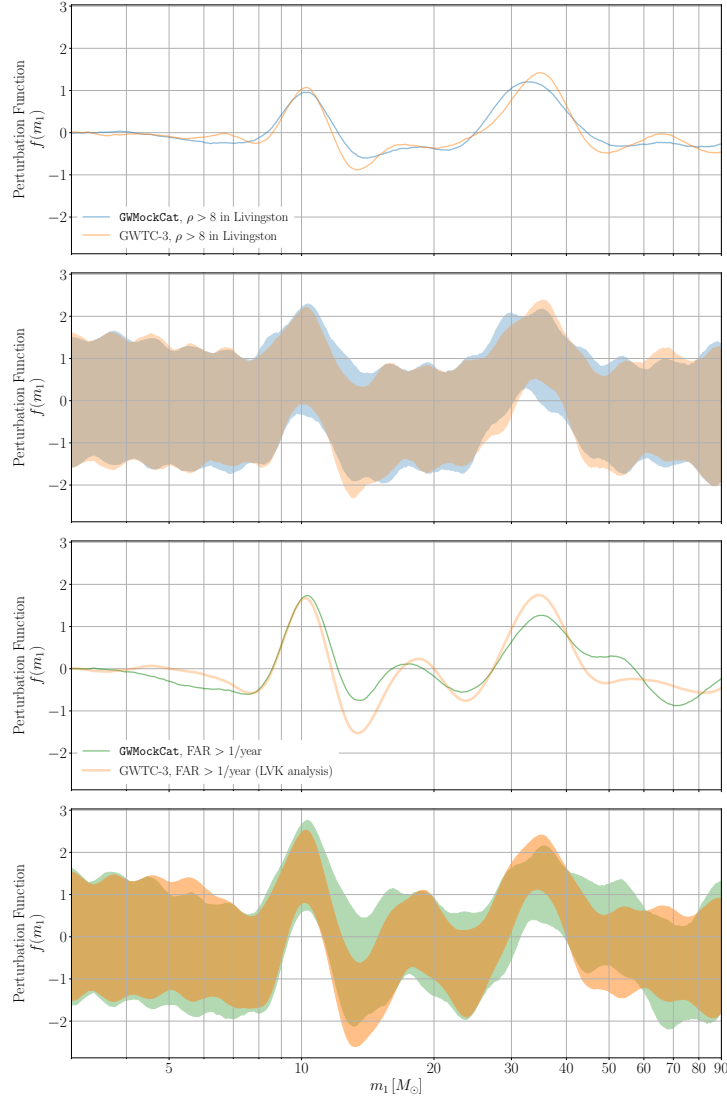


Figure 4.9: Perturbation functions of a population fit to GWTC-3 when using LVK-released posterior samples for each event (*orange*), and when using posterior samples simulated by GWMockCat (*blue and green*). The two sets of figures correspond to two selection criteria, one which is analogous to the one used for mock catalogs in this work ($\text{SNR} > 8$, *top two panels*), and one which is identical to that used for the LVK analysis ($\text{FAR} > 1$, *bottom two panels*). The main difference between the selection criteria is that they result in catalog sizes that are different by nearly a factor of two, and therefore the statistical error on the perturbation function is noticeably different between them. Using the same number of events in the simulated catalog as the LVK-released catalog results in a perturbation function that is similar in amplitude to the LVK-released population analysis. We conclude that the prescriptions used in GWMockCat are sufficient for the purposes of population analyses. We use the green curves in the bottom two panels for the analysis presented in the body of this chapter, and the orange curves in the bottom two panels for the analysis presented in this Section.

4.5.3 Accuracy of mock catalogs when used in a population analysis

As a second check of **GWMockCat**’s ability to simulate catalogs accurately enough to be used in a population analysis of the mass distribution of BBHs, we recreate GWTC-3 with **GWMockCat**. We then compare **POWER LAW + SPLINE**’s fit to this mock catalog with its fit to the posterior samples released by the LVK for GWTC-3. We find that the two resulting mass distributions are consistent, and therefore conclude that the approximate prescriptions used in **GWMockCat** are sufficient to probe the mass distribution of BBHs, at least for current GW detector sensitivities.

To simulate GWTC-3, we re-weight all GWTC-3 events to the same prior as used for sampling in **GWMockCat**, namely uniform in detector-frame chirp mass, uniform in symmetric mass ratio, and uniform in sky angle, as defined in Equation 4.9. We then take the mean of the re-weighted detector frame chirp mass, symmetric mass ratio, sky angle, and single-detector SNR posteriors as the observed parameters θ_i^{obs} for each event. Using the priors defined in Equation 4.9 and likelihoods defined in Equations 4.7 and 4.8, we construct mock posterior distributions on detector frame chirp mass, symmetric mass ratio, sky angle, and single-detector SNR. We then convert these parameters to source-frame component masses and redshift. Finally, we re-weight back to the standard prior used in LVK’s parameter estimation process. The end result of this is shown in Fig. 4.8 for GW191215_223052, an event that was chosen at random from the catalog and happened to have a primary mass and redshift near the mode of the detected events. The mock posteriors appear consistent with the true posteriors, and degeneracies between parameters seem to be suitably captured. This behavior is qualitatively similar for all simulated events, though some had mock posteriors that were slightly more consistent with the full parameter estimation posterior samples, and some had mock posteriors that were slightly less consistent.

Any population analysis must define criteria for inclusion in the population. We apply two different criteria and report the results of both in Fig. 4.9. The first, a cut at $\text{SNR} > 8$

in Livingston, is chosen to be analogous to the detection criteria of the other mock catalogs, which use a single-detector SNR cut since it is not possible to run all of the pipelines necessary to produce a FAR on mock data. The second criterion, a false alarm rate (FAR) cut at 1/yr was used to be consistent with the analysis done by the LVK on GWTC-3 [Abbott et al., 2023a]. Either choice is reasonable because the selection function is known with respect to both SNR and FAR. We can therefore use the publicly-released sensitivity estimates [Collaboration, 2021] to reconstruct the underlying, or astrophysical distribution of BBHs from either of these catalogs. All of the metrics of feature significance presented in Section 4.3 make use of this astrophysical distribution when comparing GWTC-3’s population fit to that of mock catalogs.

The first criterion (single-detector SNR cut) produced a final catalog with many fewer events than the second criterion (FAR cut), with the former resulting in 36 events and the latter resulting in 69 events. Therefore, the population analysis performed on the catalog selected by SNR has much wider hyper-posteriors than the one performed on the catalog selected by FAR. However, *these two catalogs do not appear to be systematically biased with respect to one another, nor are they systematically biased with respect to the LVK-released analysis.* This is again because the selection function is known with respect to both of these criteria, and therefore the reconstructed astrophysical distributions are consistent.

Interestingly, the mock catalog with only 36 events in it still finds the peak at $\sim 35 M_{\odot}$ to be significant, with the perturbation function excluding zero to $< 5\%$. However, other features are not well enough resolved to appear significant with only 36 events. We find very little difference between the full parameter estimation perturbation function fit (orange bands in Fig. 4.9) and the mock catalog (blue band) in the case of a single-detector SNR cut (top two panels). We take this to mean that for high-SNR events, the mock parameter estimation is sufficient for use in population analyses of the mass distribution, at least for O3-like detector sensitivities.

Using the same events in the simulated catalog as the LVK-released catalog results in a perturbation function that is similar in amplitude to the LVK-released analysis (lower two panels of Fig. 4.9). However, the width of the perturbation function’s hyperposterior is slightly inflated in the mock catalog case. This may be because the mock parameter estimation scales event posterior widths inversely with SNR, so mock posterior widths are over-estimated with respect to full parameter estimation for the events that meet the FAR threshold but have low SNR.

In order to be as consistent as possible in our comparisons of mock catalogs to GWTC-3, we use the perturbation function obtained by analyzing the `GWMockCat` version of GWTC-3 for all comparisons to mock catalogs in Section 4.3. In this Section, we repeat the analysis performed in Section 4.3, but instead used the perturbation function released by the LVK in Abbott et al. [2023a] in lieu of the perturbation function obtained by fitting `POWER LAW + SPLINE` to the `GWMockCat` version of GWTC-3. In other words, the main text uses the green curves in Fig. 4.9, and we repeat the analysis using the orange curves in the bottom two panels of Fig. 4.9 in this Section.

We find that using the LVK-released perturbation function increases the Bayesian significance of all features relative to the `GWMockCat` reproduction. This is consistent with the conclusions reached in Fig. 4.9: all hyperposteriors narrow slightly when using the LVK-released version of parameter estimation, but there is no systematic shift as a function of primary mass or any other parameter. When using the LVK-released perturbation function, None of the 300 $\{g_j(f_{\max})\}$ exclude zero to the same percentile as $g_{\text{GWTC-3}}(f(35 M_{\odot}))$ or $g_{\text{GWTC-3}}(f(10 M_{\odot}))$, and 1.3% of the $\{g_j(f_{\max})\}$ exclude zero to the same percentile as $g_{\text{GWTC-3}}(f(14 M_{\odot}))$. These values are smaller than those presented in Section 4.3.3, but lead to the same conclusions: the $10 M_{\odot}$ and $35 M_{\odot}$ peaks are difficult to reproduce with featureless catalogs, but the $14 M_{\odot}$ dip is not. Performing full parameter estimation on mock catalogs would also likely narrow the hyperposteriors for those catalogs, in turn increasing

the significance of the peaks seen in their the perturbation functions. If this were to be the case, the fraction of mock catalogs that can reproduce features in the GWTC-3 distribution would likely be similar to those found in Section 4.3.3. A reproduction of the analysis presented in Section 4.3.3 with the LVK-released perturbation function also finds similar results to that done on the `GWMockCat` perturbation function. We therefore conclude that the results presented Section 4.3 are robust to the procedure used for parameter estimation of GWTC-3 events.

CHAPTER 5

NO NEED TO KNOW: ASTROPHYSICS-AGNOSTIC COSMOLOGY WITH THE MASS SPECTRUM OF GRAVITATIONAL-WAVE SOURCES

Having established the existence of two robust features in the BBH mass distribution, we can use them to measure the expansion history of the Universe. Mass distribution features are “spectral sirens,” as they provide self-calibrating redshift measurements directly associated with luminosity distances, as discussed in Chapter 1, Section 1.2. This is particularly useful in the context of the current Hubble tension, a disagreement between multiple methods of measuring the local expansion rate of the Universe [Freedman, 2021]. Additionally, spectral sirens are sensitive to the expansion history of the Universe over a wide range of cosmological redshift, as they only require GW observations. These are observable to higher redshifts than those at which electromagnetic observations are complete, which makes spectral sirens uniquely positioned to probe the nature of dark energy.

However, spectral sirens are subject to systematic bias arising from mismodeling of the mass distribution. Such mismodeling is likely, as an accurate model of the mass distribution would require knowledge of which mechanisms form the majority of CBC, as well as each physical process within them. As theoretical uncertainties are large for all of these processes, it is desirable to avoid assuming any particular morphology for the mass distribution when doing spectral siren cosmology. Non-parametric techniques provide a clear solution to this problem.

In this chapter, we fit a flexible, non-parametric model for the mass distribution of compact binaries simultaneously with cosmological parameters. We thereby forego physically-motivated modeling of the CBC mass distribution in favor of unbiased estimation of cosmological parameters. We apply this method to a simulated catalog of 1,000 GW signals,

consistent with expectations for the next LIGO-Virgo-KAGRA collaboration observing run.

It has been commonly presumed that *a priori* theoretical predictions for features in the mass distribution are necessary for spectral siren analyses. However, we find that despite our model’s flexibility, both the source mass model and cosmological parameters are correctly reconstructed. This demonstrates that spectral sirens can accurately infer cosmological parameters without prior astrophysical knowledge of CBC formation.

This astrophysically-agnostic spectral siren technique will be essential to arrive at precise and unbiased cosmological constraints from GW source populations observed by future GW detectors. This chapter is based on *No Need to Know: Toward Astrophysics-free Gravitational-wave Cosmology*, a paper written in collaboration with Thomas Callister, Jose María Ezquiaga, Michael Zevin, and Daniel Holz, and published in the *Astrophysical Journal* [978].

5.1 Introduction

Central to the spectral siren methodology is knowledge of the compact binary mass distribution. However, first principles models for mass distributions of merging compact binaries are not available: significant theoretical uncertainties exist about the shapes, locations, and very existence of predicted features in the mass distribution [e.g. Zevin et al., 2017, Mapelli, 2020, Zevin et al., 2021, Marchant and Bodensteiner, 2024]. This includes uncertainties about potentially dominant features, such as the existence of a “pile-up” due to pulsational-PISN, a PISN-driven upper mass gap [Farmer et al., 2019], the maximum NS mass [Fryer and Kalogera, 2001, Alsing et al., 2018], and the existence of a putative lower mass gap between NSs and black holes [Özel et al., 2010, Farr et al., 2011, Farah et al., 2022a]. To this end, spectral siren cosmology relies on *simultaneously* measuring a source-frame mass spectrum alongside cosmological parameters. This is typically accomplished by adopting a phenomenological, parametric model for the mass distribution, usually composed of power

laws and Gaussians [e.g. Abbott et al., 2023b, Mastrogiovanni et al., 2023b].

Such parametric modeling of the compact binary mass distribution raises its own set of dangers. It is well known that different parametric models can generically yield very different constraints on cosmological parameters [Abbott et al., 2021d, 2023a]. This is problematic: because the measured mass distribution serves as the template by which to extract redshifts, a mismodeled mass distribution would introduce systematic errors in inferred redshifts and, in turn, systematically bias any resulting cosmological inference [Ezquiaga and Holz, 2022, Mastrogiovanni et al., 2021, Pierra et al., 2024].

The situation is demonstrated in Fig. 5.1, in which we perform spectral siren cosmology on a simulated population of binary black holes. We reconstruct the mass distribution using two parametric models, one that contains the true simulated mass distribution and one that does not. While the former yields a measurement of H_0 consistent with the true underlying value, the latter does not. Such systematic biases may already be relevant, as cosmological measurements by the LIGO-Virgo-KAGRA are known to depend on the choice of mass model used [Abbott et al., 2023b]. Furthermore, these biases may become a dominant source of uncertainty in the near future [Pierra et al., 2024]. The prospects of such a dominant systematic uncertainty is troubling. If prior knowledge of the mass distribution’s morphology (whether an exact theoretical prediction or knowledge of the correct parametric family of models) is a prerequisite for the spectral siren method, the effectiveness of such a technique would be significantly hampered.

In this work, we explicitly demonstrate that no prior knowledge of the shape of the CBC mass spectrum is necessary to use the spectral siren methodology. We do this by inferring $H(z)$ with a flexible, non-parametric model for the mass distribution of CBCs (blue shaded band in Fig. 5.1). This model makes minimal prior assumptions about the shape of the mass distribution, enabling it to accurately infer a wide range of morphologies and remain agnostic to the astrophysical processes that give rise to features in the mass distribution.

Despite its flexibility, our approach is able to consistently obtain unbiased measurements of cosmological parameters, showing that non-parametric methods are not only sufficient for a spectral siren measurement, they can also mitigate systematic effects in the measurement caused by model misspecification.

The non-parametric mass model’s ability to recover injected cosmological parameters demonstrates that the information in the spectral siren measurement does not come from the enforcement of specific features in the mass distribution. Rather, it is provided by the assumption that either all CBCs follow a common mass distribution, or that any evolution of the mass distribution with redshift does not exactly mimic cosmology [Ezquiaga and Holz, 2022].

We find that our non-parametric model allows for a 5.7% measurement of H_0 and a 6.4% measurement of $H(z = 0.9)$ during the fifth LVK observing run (O5), when the detectors will reach their design sensitivity. We highlight measurements of H_0 within a Λ CDM universe in order to benchmark the accuracy and precision of our non-parametric method, as well as explore the role of spectral sirens in elucidating the Hubble tension. However, a primary utility of spectral siren measurements will be in constraining $H(z)$ under different cosmological models, and at redshifts that are relatively inaccessible by electromagnetic observations, especially with next-generation gravitational-wave detectors [Ezquiaga and Holz, 2022, Chen et al., 2024].¹

This chapter is organized as follows: Section 5.2.1 describes the simulated dataset. Section 5.2.2 introduces the spectral siren method, demonstrating how cosmological parameters are inferred from the mass distribution of GW sources. Section 5.2.3 describes the non-parametric mass distribution we develop for use within the spectral siren method. In Section 5.3, we present the results of using parametric and non-parametric mass distributions, as well as projections for future constraints on $H(z)$. We discuss the implications of our

1. See example spectral siren cosmological inference using parametric mass spectrum models: https://github.com/ezquiaga/spectral_sirens.

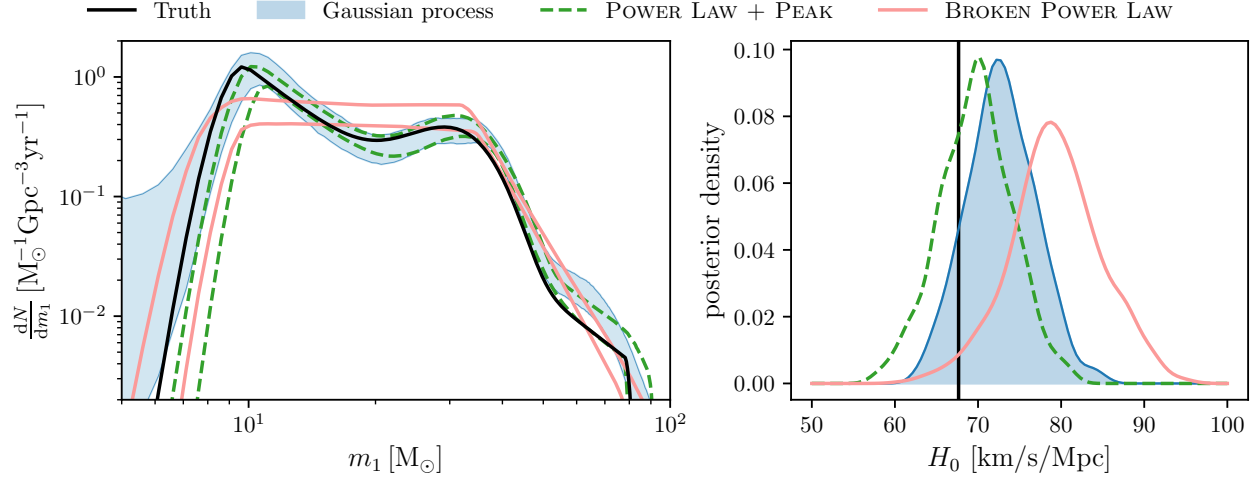


Figure 5.1: Spectral siren measurement for a simulated catalog with the correct parametric model (POWER LAW + PEAK, green dashed), a deliberately incorrect parametric model (BROKEN POWER LAW, pink solid) and the non-parametric model presented in this work (Gaussian process, blue filled). The left panel shows the recovered source frame primary mass distribution for each model, and the simultaneously-inferred posteriors on H_0 are shown in the right panel. The bands in the left panel represent the 90% credible intervals. The mass distribution and H_0 value used to generate the data are shown by a solid black line in each panel. The deliberately incorrect parametric model fails to recover the true mass distribution and therefore produces an estimate of H_0 that is offset from the true value, whereas both the correct and non-parametric models recover the mass distribution and H_0 . As the true mass distribution is unknown for real observations, using a non-parametric model mitigates systematic uncertainty that would otherwise arise from mismodeling the CBC population.

results and outline future work in Section 5.4.

This study was carried out using the reproducibility software `showyourwork` [Luger et al., 2021], which leverages continuous integration to programmatically download the data from `zenodo.org`, create the figures, and compile the manuscript. The git repository associated to this study allows anyone to re-build the entire manuscript, and is publicly available at <https://github.com/afarah18/spectral-sirens-with-GPs>.

5.2 Cosmology with an astrophysically-agnostic mass model

The spectral siren method functions by identifying the relationship between luminosity distance and redshift that causes all source frame masses to follow a distribution that smoothly

varies as a function of time. This allows for the simultaneous inference of both a mass distribution and cosmological parameters, even if the form of the source-frame mass distribution is not known in advance. As described above, though, strongly parametrized models for the mass distribution yield biased measurements of cosmological parameters if they poorly approximate the true mass distribution of compact binaries. We aim to circumvent such biases and instead model the population of GWs sources in a flexible and astrophysically-agnostic way.

There exist several non-parametric methods developed for this purpose [Tiwari, 2021, Edelman et al., 2022, Sadiq et al., 2022, Rinaldi and Del Pozzo, 2022, Edelman et al., 2023, Mandel et al., 2019, Ray et al., 2023, Callister and Farr, 2024]. While well-suited to infer the GW source population with a fixed cosmology, several of these methods employ fixed features in source frame mass, such as bin edges [Mandel et al., 2019, Ray et al., 2023] or spline nodes [Edelman et al., 2022]. Since these locations were chosen with a fixed cosmology, they risk causing the inference to prefer the cosmological parameters assumed when choosing the feature locations. Indeed, Mastrogiovanni et al. [2021] show how using fixed features can significantly bias cosmological inference within the spectral siren methodology. We therefore opt for a model of the source frame mass distribution that foregoes the need to define such features.

For this purpose, we construct a model with a Gaussian process (GP), a common tool for non-parametric inference. GPs define a random space of functions in which any subset of function values are jointly Gaussian-distributed [Rasmussen and Williams, 2006]. Their smoothness properties make them widely useful in GW data analysis for regression problems, such as modeling time-domain waveforms [Doctor et al., 2017, Huerta et al., 2018] and the NS EOS [Landry and Essick, 2019], density estimation problems, such as estimating posterior densities of single-event parameters from parameter estimation samples [D’Emilio et al., 2021], and as a prior on histogram bin heights for population inference [Mandel et al., 2017,

Li et al., 2021b, Ray et al., 2023].

Our use case is slightly different from previous analyses: we utilize a GP as a prior on the functions that describe the primary mass distribution of CBCs. This choice encodes very little prior information about the shape of the mass distribution, besides enforcing that it must be smooth.

5.2.1 *Simulated data*

To demonstrate the effectiveness of our GP-based mass distribution in agnostically inferring both cosmological parameters and the population properties of GW sources, we apply our methodology to a simulated dataset. By generating a catalog of GW sources from a known population and cosmological model, we are able to quantify the accuracy of our method. The use of simulated data also enables us to make projections for future datasets and safely ignore dimensions such as spin that do not impact cosmological measurements but would otherwise be important to simultaneously fit to avoid biases in population inference of real data, as real data exhibit correlations between distance, mass, and spin measurements [Biscoveanu et al., 2021].

We design our simulated catalog to match the characteristics of the data expected from one year of observation in O5. The BBHs in this catalog are drawn from an underlying population described by the POWER LAW + PEAK mass distribution presented in Talbot and Thrane [2018] and used in Abbott et al. [2019a, 2023a], and follow the redshift distribution presented in Callister et al. [2020], with hyperparameters consistent with those found in Abbott et al. [2023a]. We assume the mass distribution does not evolve across the redshift range to which the O5 detectors will be sensitive. This assumption is consistent with current data [Fishbach et al., 2021, van Son et al., 2022a, Abbott et al., 2023a], in which no redshift evolution of the black hole mass function is detected. At the same time, a redshift-dependent mass function is a generic astrophysical prediction, due either to

changing evolutionary environments or evolving mixture fractions between distinct compact binary formation channels [Neijssel et al., 2019, van Son et al., 2022a, Torniamenti et al., 2024, Ye and Fishbach, 2024]. We discuss this possibility further in Section 5.4, but leave it primarily for future work.

We use the `GWMockCat` [Farah et al., 2023] package to apply O5-like selection effects to the drawn BBHs, generate realistic measurement uncertainty, and produce sensitivity estimates that are consistent with the simulated GW signals. We will use the term “event” to refer to GW signals that pass the criteria for detection. This process results in a catalog of $N_{\text{ev}} = 1003$ GW signals that pass the criteria for detection, hereon called events. Additional details of the data simulation, including the form of the injected population, are described in Section 5.5.1.

5.2.2 *The spectral siren method*

To simultaneously infer cosmological parameters and the population of GW sources, we employ a hierarchical Bayesian analysis. As described in Chapter 2, this allows us to undo the selection effects of GW detectors to obtain a true, astrophysical population and constrain the cosmic expansion history.

In this work, we have restricted our analysis to the BBH primary mass distribution. However, the method can be trivially extended to the full mass distribution of CBCs [e.g. Fishbach et al., 2020a, Ezquiaga and Holz, 2022]. Additionally, Section 5.5.2 demonstrates that including the mass ratio distribution in the fit does not change the results of the analysis presented here, so long as the mass ratio distribution is accurately modeled.

We fix $\Omega_\Lambda = 1 - \Omega_M$, $\Omega_r = 0$ and $w = -1$ and use uniform priors on H_0 and Ω_M , corresponding to a flat Λ CDM cosmology.

The process by which we sample the likelihood in Equation 2.1 is outlined in Section 5.5.3.

5.2.3 Gaussian process-based mass distribution

In this Section, we give an overview of the non-parametric mass model developed for this work. Further details on this model, including an introduction to GPs and a discussion of their properties is given in Section 5.5.3.

With the GP approach, the hyper-parameters describing the mass distribution are the rate at each event-level posterior sample’s source frame mass, and the rate at each found injection’s source frame mass. The GP is $p(\Lambda)$, the prior on population parameters (except in the case of cosmological parameters, which all have uniform priors). This is demonstrated in Fig. 5.2, where the left panel shows draws from a GP, which are prior draws for the population inference. The smooth appearance of individual draws from the population prior, as well as the absence of overdensities at specific source frame mass values in the full prior distribution shown in Fig. 5.2 demonstrate that we have successfully fulfilled our goal to construct a model without predefined features in source-frame mass. Combined with the population likelihood in Equation 2.1 the prior illustrated in the left panel of Fig. 5.2 gives the population posterior in the right panel.

The lack of data just below the minimum black hole mass and just above the maximum black hole mass, combined with the fact that GW detectors are sensitive to objects at those masses, causes the GP to learn a relatively low merger rate at the edges of the mass distribution. On the other hand, there is both a lack of data and little detector sensitivity at masses above $\sim 100 M_\odot$ and below $\sim 5 M_\odot$. The mass distribution is therefore uninformed in this region and the GP reverts to its prior distribution, which resembles random scatter around the mean differential merger rate. Similar effects can be seen in other non-parametric methods [Edelman et al., 2023, Callister and Farr, 2024]. The combination of these two effects results in what appears as an uptick in the merger rate below $\sim 5 M_\odot$ and above $\sim 100 M_\odot$. However, this reversion to the prior is uninformative for the H_0 constraint and does not affect inference on cosmological parameters. Additionally, the posterior on H_0 is distinct from its

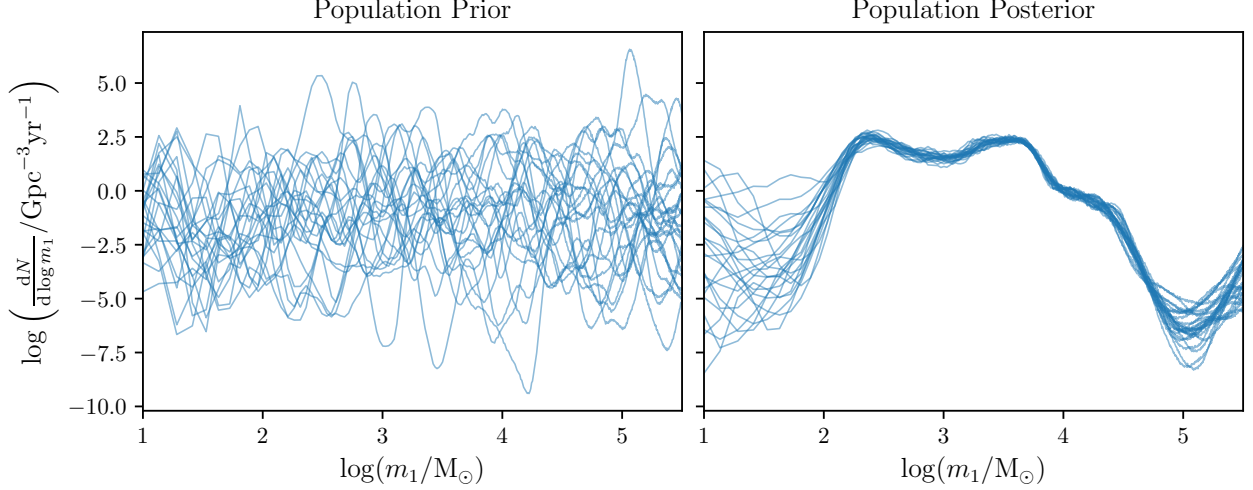


Figure 5.2: Draws from the Gaussian process (GP) used to model the mass distribution. The left panel shows prior draws from the GP and the right panel shows posterior draws once the population inference is performed on the simulated data. The posterior draws in the right panel are a subset of those used to create the 90% credible intervals in Fig. 5.1.

prior distribution (uniform in the range $[30 \text{ km}^{-1} \text{ s}^{-1} \text{ Mpc}, 120 \text{ km}^{-1} \text{ s}^{-1} \text{ Mpc}]$, indicating that the data is informative despite the flexibility of the population model.

We note that other non-parametric methods may be adapted to avoid predefined features, such as fitting for the locations of their features simultaneously with the rest of the inference [e.g. Tiwari, 2021] or, in the case of splines, by using a smoothing function that allows for features to occur at arbitrary locations appropriate smoothing [e.g. Edelman et al., 2023].

The smoothness of a given GP is determined by its kernel, which is a function that defines the covariance between input points in the GP (in our case, two source frame mass values). It defines the notion of similarity between adjacent points and thereby encodes our assumptions about the smoothness of the source frame mass distribution [Rasmussen and Williams, 2006]. Kernels themselves have parameters that determine their properties. In our use case, these are one level further removed from hyper-parameters, so we adopt the terminology used in Callister and Farr [2024] and call them “hyper-hyper-parameters.” We fit these hyper-hyper-parameters along with the hyper-parameters Λ to minimize prior assumptions about the form of the mass distribution.

5.3 Results

In this Section we show that fitting an incorrect functional form to the mass distribution of CBCs biases the inference of cosmological parameters when using the spectral siren methodology. We then demonstrate that our flexible model alleviates this bias without the need to know the morphology of the mass distribution *a priori*. We illustrate this explicitly by using three different models for the source-frame mass distribution to infer the cosmic expansion rate from the simulated catalog described in Section 5.2.1 and Section 5.5.1. The three models are as follows:

1. POWER LAW + PEAK, which includes the true mass distribution within its hyper-prior,
2. the BROKEN POWER LAW model presented in Abbott et al. [2021d], as we do not employ a high-mass truncation, which does not include the true mass distribution within its hyper-prior, and
3. the flexible, GP-based model described in Section 5.2.3, which is able to closely approximate the morphology of the true mass distribution, along with many other morphologies.

For all models considered in this work, we assume the form of the redshift distribution used to generate the data, described in Equation 5.3. We have performed the analysis both with a fixed redshift distribution and while simultaneously fitting for the redshift distribution and find no qualitative differences in our conclusions: fitting for the redshift distribution broadens the inferred posteriors on H_0 equally for all mass models, but does not affect their mean values.

The results of fitting each model to the same dataset are shown in Fig. 5.1. The left panel shows the inferred source-frame mass distribution for each of the considered models, and the

right panel shows the corresponding posteriors on H_0 . We indicate the true underlying source mass distribution and H_0 value with solid black lines in each panel.

The fits presented in Fig. 5.1 are representative results from a single simulated catalog. These provide insight into the full statistical results presented below. In particular, it can be seen that the BROKEN POWER LAW (orange curve) is inconsistent with the true value of H_0 : in the run shown in Fig. 5.1, the true value of H_0 is offset from the mean of the posterior by 2.1σ . In contrast, the POWER LAW + PEAK and GP-based models (green and blue curves) are consistent with the underlying truth. These models recover mean values of H_0 that are offset from the true value at 0.5σ and 1.2σ , respectively. Additionally, the mass distribution inferred with the GP-based model closely resembles the true, simulated distribution. This indicates that using models that cannot accurately approximate the true mass distribution will lead to a noticeable systematic bias in the estimation of cosmological parameters.

This bias is not due to the need to know the morphology of the mass distribution *a priori*, as the GP-based model recovers the correct value of H_0 despite making minimal assumptions about the mass distribution. In reality, we do not know the true functional form of the mass distribution, so it may be desirable to use a non-parametric approach to avoid potential systematic errors introduced by choosing a parametric model that likely does not contain the true mass distribution within its hyper-prior.

To obtain a quantitative measure of the systematic bias introduced by mis-modeling the mass distribution, we repeat the parametric analyses with 50 separate simulated catalogs of $\sim 1,000$ events each. We find that the BROKEN POWER LAW model produces an over- or under-estimate of H_0 at greater than 1σ 92% of the time, and the POWER LAW + PEAK model reaches the same level of bias only 54% of the time, meaning that the BROKEN POWER LAW model produces a bias more than three times as often as the POWER LAW + PEAK model. Additionally, we show in Section 5.5.4 that the BROKEN POWER LAW model

typically overestimates H_0 whereas the POWER LAW + PEAK model produces a roughly equal number of over- and underestimates of H_0 . This demonstrates that mis-modeling the mass distribution can introduce statistically significant systematic biases into measurements of cosmological parameters.

Collectively, our results indicate that *prior knowledge of the shape of the mass distribution is not required to perform an unbiased spectral siren measurement, so long as strong assumptions about the shape of the mass distribution are not made.*

5.3.1 Projections for future measurements

Fig. 5.1 demonstrates an expected 5.7% (1σ uncertainties) measurement of H_0 after one year of O5 using the GP-based spectral siren method, and a 7% measurement with parametric spectral sirens, demonstrating comparable statistical uncertainties. However, we note that the precision reached in O5 may lessen depending on the actual level of measurement uncertainty in individual events, and the existence (or lack thereof) of a maximum mass feature, which we have assumed to be present in our simulated dataset. These numbers are estimated from a fit to a single simulated catalog, but we find similar levels of statistical uncertainty from fits to different catalog realizations. By the time of O5, the GW detector network is projected to detect BBHs up to redshift ≈ 3 , with most sources lying near redshift $\simeq 1.2$ [Chen et al., 2021]. Additionally, next-generation detectors will be sensitive to sources up to redshift ~ 100 . This means that future GW observations will be more sensitive to $H(z \gtrsim 1)$ than to H_0 , and can therefore constrain several additional cosmological parameters [Chen et al., 2024]. We demonstrate this by repeating the same GP-based spectral siren analysis while also simultaneously fitting for the local matter density, Ω_M . The result is shown in Fig. 5.3. We emphasize that these precise measurements over a wide range in redshift enable precision estimation of additional cosmological parameters governing $H(z)$.

We find O5 observations to be most sensitive to $H(z = 0.9)$, which is measured at 6.4%.

The left panel of Fig. 5.3 demonstrates a strong anti-correlation between the Ω_M and H_0 posteriors, resulting in similarly informative constraints on the two parameters. This is because H_0 controls the y -intercept of the $H(z)$ curves on the right panel, while Ω_M informs the slope of those curves; the same measurement of $H(z \neq 0)$ can be obtained by increasing the slope while decreasing the y -intercept, and vice versa. Similar behavior can be observed in current measurements of the BBH redshift distribution, which exhibits a tightening of the posterior at $z \sim 0.2$ with current observations [Abbott et al., 2023a, Callister and Farr, 2024].

Next-generation detectors will be sensitive to a larger range of redshifts [ET steering committee, 2020, Evans et al., 2021], and will therefore break the degeneracy between cosmological parameters and allow for tighter constraints on both Ω_M and H_0 . However, the small cosmological volume (and thus low number of mergers) at low redshift will generally limit the constraining power of spectral sirens at $z = 0$, potentially making this method more sensitive to cosmological parameters that affect higher redshifts. Combining spectral sirens with other methods that are sensitive to the local expansion rate, such as those that employ electromagnetic counterparts, may increase the precision of GW standard sirens at all redshifts [e.g. Chen et al., 2024].

5.4 Discussion

GWs are unique cosmic messengers in that they carry both redshift and distance information, making them remarkably clean probes of the Universe’s expansion history calibrated directly by the theory of general relativity. However, the current method of determining GW redshifts via the distribution of their source masses (i.e. spectral sirens) employs an assumption on the shape of their population, typically encapsulated by simplified parametric functions. This choice, often presumed to be necessary or fundamental to the method, may introduce a systematic bias to a measurement of the cosmological parameters that is otherwise appealing

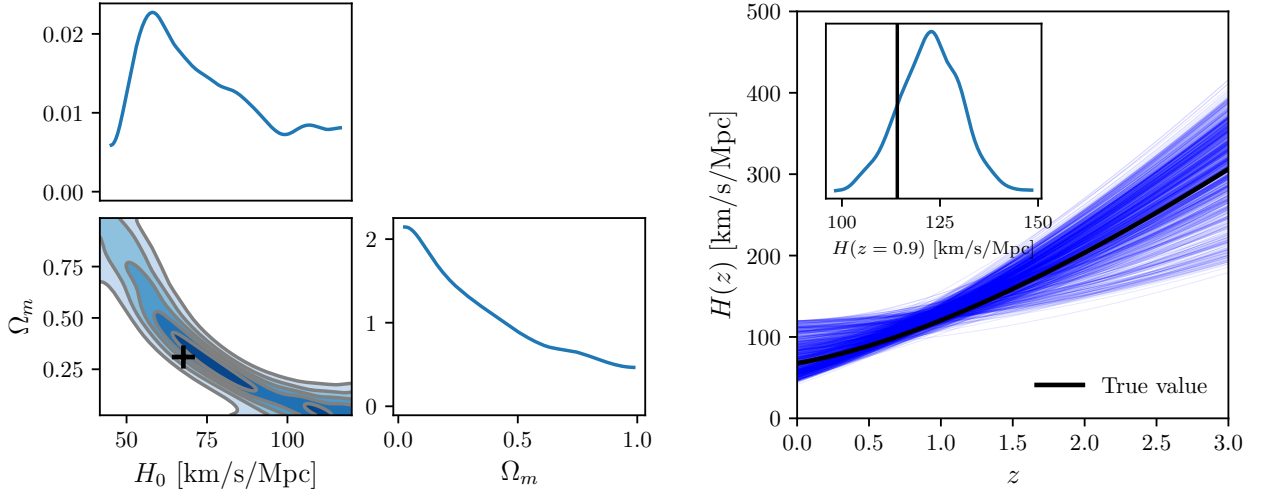


Figure 5.3: Projected constraints on multiple cosmological parameters after one year of observing at the LVK’s design sensitivity, using the Gaussian process-based spectral siren method. The right panel shows the inferred expansion history of the universe, $H(z)$. It will be measured most precisely at $z = 0.9$, as can be seen by the narrowing of the inferred $H(z)$ curves there. The inset shows the posterior on $H(z = 0.9)$. Black solid lines indicate the true value of $H(z)$ in both the inset and main panel. The left panel shows the two-dimensional posterior on H_0 and Ω_M , with the true value indicated by a black “+”. The two parameters are strongly degenerate because of the multiple ways of measuring $z = 0.9$. Spectral sirens are particularly well suited for measuring cosmological parameters that affect the Universe at $z \gtrsim 0.2$.

for its unique elegance and simplicity.

In this work we show that a specific choice of mass distribution is unnecessary to arrive at informative and accurate posteriors on H_0 and Ω_M . We do this by accurately measuring these cosmological parameters with a highly flexible model for the mass distribution. This reinforces the notion that the information in the spectral siren measurement comes from the assumption that all GW sources come from the same population, a far less stringent statement than the assumption that we understand the exact astrophysical processes which give rise to that population (i.e. the physics governing compact binary formation and evolution).

Spectral sirens are particularly useful in the context of the current Hubble tension: a disagreement between multiple methods of measuring the local expansion rate of the universe [Freedman, 2021]. As this tension can only be explained by non-standard physics or yet-unknown systematic uncertainties in either the cosmic distance ladder or cosmic microwave background measurements, direct and independent probes of the local expansion—such as those presented here—may help determine whether the current discrepancy represents a fundamental crack in our standard model of physics and/or cosmology.

Neither the mechanism from which compact binaries were formed, nor the physical processes within the mechanism, have been conclusively determined. The true functional form of the mass distribution therefore remains elusive, meaning that systematic uncertainties arising from an incorrect choice for the form of the mass distribution are inevitable. With current observations, these systematic effects are likely smaller than statistical uncertainties. However, next-generation detectors will herald sufficient GW observations to substantially decrease statistical uncertainty in these measurements; for example, Pierra et al. [2024] show that incorrect assumptions about the shape of the mass distribution can lead to $\sim 3\sigma$ systematic biases in H_0 with catalogs of 2,000 events, although this bias may be an overestimate as it does not include measurement uncertainty of the GW parameters. Thus, non-parametric approaches may be preferable to avoid the systematic errors associated with choosing a

parametric model.

In parallel with this work, Magaña Hernandez and Ray [2024] performed a spectral siren analysis on public LVK data using histogram bins defined at fixed locations in source-frame mass to flexibly measure the BBH mass distribution [as originally presented in Mandel et al., 2017, Ray et al., 2023]. Both our method and theirs employ GPs. In our approach, the black hole mass function is *itself* described by a Gaussian process, with no predefined features. In Magaña Hernandez and Ray [2024], on the other hand, the mass distribution is fundamentally a binned piecewise constant function, with a GP prior governing the relative heights of these bins. While this binned model is flexible, sharp bin edges at predefined locations nevertheless constitute unphysical features. These features may yield overly optimistic constraints on cosmological parameters, possibly explaining the increase in precision that Magaña Hernandez and Ray [2024] find relative to the parametric analysis on the exact same dataset performed in Abbott et al. [2023b]. Magaña Hernandez and Ray [2024] also include a binned reconstruction of the BBH merger rate as a function of redshift, which we fix or model parametrically in our own work. We stress, though, that this is not an intrinsic limitation of our method; one could straightforwardly adopt a second Gaussian process to govern the overall scaling of the merger rate with redshift.

More interesting is the possibility that the black hole mass distribution itself evolves with redshift. As demonstrated in Ezquiaga and Holz [2022], the degeneracy between an evolving mass distribution and the expansion of the universe can be broken provided that there are multiple features present in the mass distribution, and that we do not live in a fine-tuned universe where the evolutionary effects governing the CBC mass distribution perfectly mimic the effects of cosmological redshift. The first condition (multiple features) is known to be met in current data [Abbott et al., 2023a], with three robust features [Farah et al., 2023]: a maximum black hole mass and overdensities at $\sim 10 M_{\odot}$ and $\sim 35 M_{\odot}$. It is also predicted by many population synthesis studies, [e.g. Mapelli and Giacobbo, 2018,

Giacobbo et al., 2018, van Son et al., 2023]. The latter scenario—with features identically and monotonically shifting to higher masses with increasing redshift—would be astrophysically unlikely as the locations of features in the mass distribution are each thought to be governed by fundamentally different physical processes [Mapelli, 2020]. Extensions to the non-parametric methods presented here will allow for arbitrary redshift evolution. Additionally, since Gaussian processes naturally scale to multiple data dimensions [Rasmussen and Williams, 2006], the method presented here can easily be generalized to fit the redshift dependence of the mass distribution. Future work will extend the method developed here to mass distributions that are allowed to evolve with redshift. When applying an evolving mass distribution model to a dataset that exhibits evolution, we expect errors on cosmological parameters to broaden. Noticeable redshift evolution in the mass distribution is expected to occur above $z \simeq 1.5$, [van Son et al., 2022a], which is not yet relevant in current data or in expected O5 data, but will certainly be visible in next-generation detectors.

Because GW observations are the only data input to spectral sirens, they are sensitive to the expansion history of the universe over a wide range of cosmological redshift, rather than just the local expansion H_0 . We have shown this by simultaneously measuring Ω_M and H_0 ; the method can be trivially expanded to constrain additional cosmological parameters that govern $H(z)$, such as the dark energy EOS parameter, w . It is also possible to use a non- Λ CDM cosmological model for $H(z)$, with different parameters entirely. It may also be possible to forego the need for a parametric representation of $H(z)$ altogether, for example by using bins in luminosity distance. A similar technique using redshift bins is commonly used in cosmological measurements using galaxy surveys and type 1a supernovae [e.g. Perlmutter and Schmidt, 2003, Anderson et al., 2014, Aghamousa et al., 2016], but we have not pursued that possibility in this work. Such a method would likely increase uncertainties on $H(z)$, especially for $z \gtrsim 1.5$. We find that for the fifth LVK observing run, spectral sirens will be most constraining at $z = 0.9$. This redshift is larger than the expected redshifts of detectable

electromagnetic counterparts of BNS mergers [Kiendrebeogo et al., 2023], implying that upgrades to current GW detectors will allow the spectral siren method to probe $H(z)$ at otherwise unexplored distances by GWs.

Proposed next-generation GW detectors such as Cosmic Explorer and Einstein Telescope will be sensitive to CBCs across cosmic time (out to $z \sim 100$) [ET steering committee, 2020, Evans et al., 2021]. Future cosmological surveys, such as those enabled by the Nancy Grace Roman Space Telescope and the Vera Rubin Observatory, are expected to be able to precisely measure $H(z)$ to $z \sim 3$ [Spergel et al., 2015], making the spectral siren method uniquely positioned to measure the expansion of the universe at high redshift. When combined with bright and dark siren methods, the low-redshift expansion history will also be well-constrained [Chen et al., 2024]. Sensitivity to high redshifts is a feature of the spectral siren method in general, and non-parametric methods such as the one presented here will be imperative to avoid systematic biases in spectral siren cosmology.

5.5 Additional analysis details

5.5.1 Details of data simulation

The exact form of the injected population is

$$\frac{dN}{dm_1 dz} \propto p(m_1|\bar{\Lambda}_m)p(z|\bar{\Lambda}_z, H_0, \Omega_M), \quad (5.1)$$

where

$$p(m_1, m_2|\bar{\Lambda}_m) \propto \mathcal{S}(m_{\min}, m_{\max}) \left(f_{\text{peak}} e^{-\frac{1}{2}(\frac{m_1 - \mu}{\sigma})^2} \mathcal{N}_g + (1 - f_{\text{peak}}) m_1^\alpha \mathcal{N}_{\text{pl}} \right), \quad (5.2)$$

$$p(z|H_0, \Omega_M) \propto \frac{dV_C}{dz} \frac{1}{1+z} \frac{(1+z)_z^\alpha}{1 + \left(\frac{1+z}{1+z_p} \right)^{\alpha_z + \beta_z}}. \quad (5.3)$$

Here, $V_C(H_0, \Omega_M)$ is the comoving volume for given cosmological parameters H_0 and Ω_M , and $\bar{\Lambda}_m = \{\alpha, m_{\min}, m_{\max}, \mu, \sigma, f_{\text{peak}}\}$ are the (hyper-)parameters describing the power law in primary mass, minimum and maximum black hole mass, Gaussian peak location and width, and fraction of events in the Gaussian peak, respectively. $\mathcal{S}$ is a smoothing function at low and high masses, and $\mathcal{N}_{\text{pl}}$ and $\mathcal{N}_{\text{g}}$ are the normalizations between $m_{\min}$ and $m_{\max}$ for the power law component and truncated Gaussian component, respectively. The smoothing creates support for masses below $m_{\min}$ and above $m_{\max}$. $\bar{\Lambda}_z = \{z_p, \alpha_z, \beta_z\}$ are the parameters governing the peak of the redshift distribution, low- z power law slope, and high- z power law slope. When generating the simulated events, we have fixed $\alpha = -2.7$, $m_{\max} = 78 M_{\odot}$, $m_{\min} = 10 M_{\odot}$, $\mu = 30 M_{\odot}$, $\sigma = 7.0 M_{\odot}$, $f_{\text{peak}} = 0.05$, $z_p = 2.4$, $\alpha_z = 1.0$, and $\beta_z = 3.4$. These choices correspond to the maximum *a posteriori* values obtained by an analysis of GWTC-3 data using the POWER LAW + PEAK model [Abbott et al., 2023a]. For simplicity, we assume a uniform mass ratio distribution. Thus, $\theta = \{m_1, z\}$. We consistently apply these assumptions to the data generation process and the population inference. We do not fit for or simulate spins, as they do not redshift and hence do not carry additional cosmological information, and we fix the distribution of mass ratios.

We use cosmological parameters $H_0 = 67.66 \text{ km s}^{-1} \text{ Mpc}$, $\Omega_M = 0.3$, and $\Omega_{\Lambda} = 1 - \Omega_M$, consistent with those found by Ade et al. [2016]. We emphasize, however, that the choice of cosmological parameters for data generation is arbitrary and does not impact the results, since we are concerned only with the ability of our method to recover the injected values. Throughout the data generation and inference, we use the approximations presented in Adachi and Kasai [2012] to efficiently convert between D_L and z for a given set of cosmological parameters Ω_M and H_0 .

We do not include NSs in our simulation set, as their contribution to the spectral siren measurement is expected to be subdominant in O5. However, if a lower mass gap between the heaviest NSs and lightest black holes exists, it will provide an additional feature with

which to inform the measurement, and will be the most informative feature for spectral siren measurements with next-generation detectors [Ezquiaga and Holz, 2022].

After passing the simulated events through projected detector selection effects, the resulting catalog has 1003 events, consistent with the numbers projected for O5 by Kiendrebeogo et al. [2023]. We use the software package `GWMockCat` [Farah et al., 2023] to simulate posterior samples for these events with measurement uncertainties typical of those expected from O5 detectors. `GWMockCat` also simulates a set of software injections, which we use to estimate selection effects in the inference. To determine the detectability of both injections and simulated events in O5, we use the projected O5 LIGO power spectral density [Abbott et al., 2020b] for a single detector to calculate observed signal-to-noise ratios (SNRs) ρ_{obs} , and we consider events and injections with a single-detector SNR $\rho_{\text{obs}} > 8$ to be detectable. The full procedure for this mock data generation process is described in Fishbach and Holz [2017], Farah et al. [2023], Essick and Fishbach [2024].

5.5.2 *Effects of fitting for secondary mass*

We examine the effect of only fitting the distribution of primary masses on our results. To do so, we perform two spectral siren analyses: one that includes a fit to mass ratio, and one that does not. Both analyses use the parametric, POWER LAW + PEAK model for the distribution of primary masses. We model the distribution of mass ratios with a power law, similarly to the majority of analyses presented in Abbott et al. [2021d, 2023a].

We use the same set of simulated events for both analyses. These are generated in the same way as described in Section 5.2.1. The inference results are shown in Fig. 5.4. We find the posteriors on H_0 to be similar between the two cases. Additionally, the recovered mass spectra are nearly identical. We therefore conclude that fitting for mass ratio does not significantly impact our main conclusions.

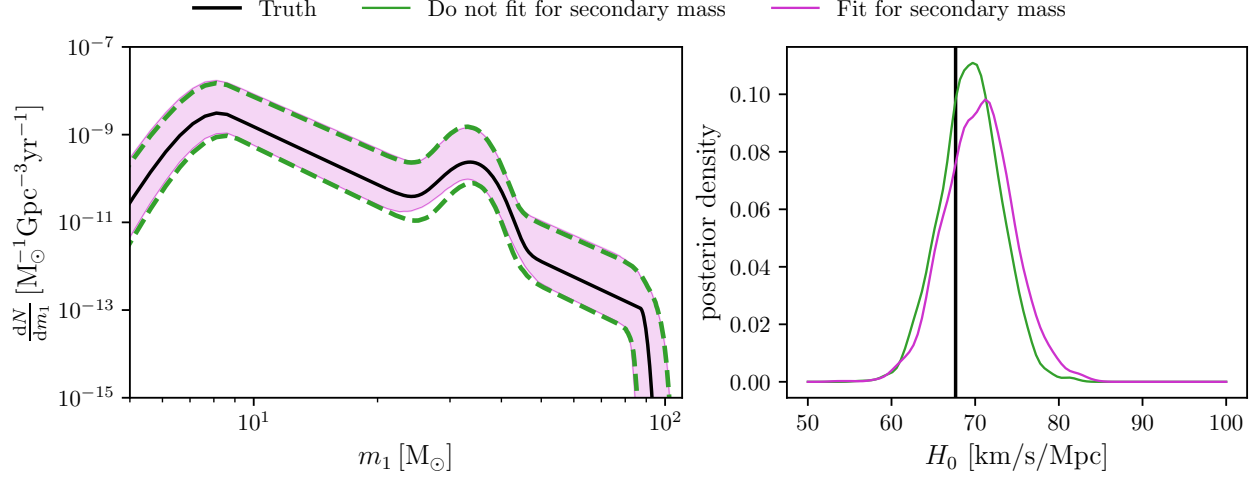


Figure 5.4: Comparison of a parametric spectral siren analysis performed while fitting for the distributions of primary and secondary masses (magenta solid) and primary masses only (green dashed). The posterior on H_0 is relatively unchanged between the two, and only minor differences are observed in the mass spectra.

5.5.3 Gaussian process-based mass distribution

In this Section, we discuss the properties of the GP-based mass distribution and describe our modeling choices in more detail.

Practically, the difference in the inference of the population when using a GP versus other modeling choices is that the population model ($dN/dt_{\text{det}} d\theta_i(\theta_i; \Lambda)$ in Equation 2.1) is determined directly by a realization of the GP, rather than by a handful of hyper-parameters Λ and evaluated on an analytical function. In other words, when using parametric models, $dN/dt_{\text{det}} d\theta_i(\theta_i; \Lambda)$ is calculated by evaluating a specific functional form described by a small set of hyper-parameters. With the GP approach, the hyper-parameters describing the mass distribution are the rate at each event-level posterior sample’s source frame mass, and the rate at each found injection’s source frame mass. The GP is $p(\Lambda)$, the prior on population parameters (except in the case of cosmological parameters, which all have uniform priors).

Because the GP is defined only at specific data points, we have $N_{\text{ev}}M + N_{\text{inj}}$ mass hyper-parameters, where M is the number of posterior samples per event and N_{inj} is the number of injections used to calculate the selection function [see e.g. Vitale et al., 2020, Essick and

Fishbach, 2021]. In this way, our GP-based mass distribution is similar to the autoregressive population models used in Callister and Farr [2024]. Indeed, an autoregressive process is a GP with a specific choice of kernel.

The kernel is a function that defines the covariance between input points in the GP (in our case, two source frame mass values). It defines the notion of similarity between adjacent points and thereby encodes our assumptions about the smoothness of the source frame mass distribution [Rasmussen and Williams, 2006]. We use a Matérn kernel [Handcock and Stein, 1993, Stein, 1999] with $\nu = 5/2$, but have repeated the analysis with $\nu = 3/2$ and ∞ , finding little impact on the results, except that the $\nu = \infty$ case (also called the squared exponential kernel) produces a slightly more jagged mass distribution. In addition to the mean, Matérn kernels have two parameters that determine their properties: a length scale l and a variance s . In our use case, these are one level further removed from hyper-parameters, so we adopt the terminology used in Callister and Farr [2024] and call them “hyper-hyper-parameters.” We fit these hyper-hyper-parameters along with the hyper-parameters Λ to minimize prior assumptions about the form of the mass distribution. We use penalized-complexity priors on the hyper-hyper-parameters to enforce that the model does not create small-scale structure uninformed by data, thereby avoiding over-fitting [Simpson et al., 2017, Simpson, 2022]. Explicitly, the priors on l and s are Fréchet and Gamma distributions, respectively, and are defined to have less than 5% support for correlation lengths smaller than the average spacing between event-level posterior means.

The time to evaluate a GP is notorious for scaling as the cube of the number of data points, making GPs unwieldy with large data sets, such as the $\mathcal{O}(10^9)$ posterior samples and software injections expected for O5. We therefore make two approximations to a full GP to increase computational efficiency. First, for each likelihood evaluation, we evaluate a full GP on a regular grid between $0.1 M_\odot$ and $250 M_\odot$ and then interpolate it at each data point. Second, we use the quasi-separability of Matérn kernels to analytically perform the

transformation between covariance matrix and GP draw. This second step is done using the `QuasisepSolver` module [Foreman-Mackey et al., 2017] in the `tinygp` code base [Foreman-Mackey, 2021], and requires data to be sortable (i.e. one-dimensional).

Algorithmically, each posterior evaluation contains the following steps:

1. Draw cosmological parameters H_0 and Ω_M from uniform prior distributions.
2. Convert the luminosity distances and detector-frame masses of each event posterior sample to redshifts and source-frame masses according to the cosmology specified by step 1.
3. Draw hyper-hyper-parameters l and s from the penalized-complexity priors described above.
4. Draw a single GP realization with a kernel defined by l and s . This is defined on a regular grid of source-frame masses and evaluated using the `QuasisepSolver` in `tinygp`.
5. Interpolate the GP at each event posterior sample and injection source-frame mass (from step 3).
6. Calculate the population likelihood according to Equation 2.1.

We perform these steps within `numpyro` [Bingham et al., 2019, Phan et al., 2019], sampling the posterior using the no-u-turn sampler for Hamiltonian Monte Carlo [Hoffman and Gelman, 2011]. This can be seen explicitly in the source code accompanying this paper, in the `scripts/priors.py` script. At each likelihood evaluation, we calculate the effective number of parameter estimation samples and found injections. After sampling, we determine if the effective number of samples exceed the efficiency criteria detailed in Essick and Farr [2022]. All results presented in this work exceed these efficiency criteria without the need to excise regions of parameter space with low numbers of effective samples.

5.5.4 *Biases induced by the BROKEN POWER LAW model*

Fig. 5.5 shows recovered posteriors on H_0 as inferred from 50 mock catalogs using both the BROKEN POWER LAW and POWER LAW + PEAK parametric models. We find that the POWER LAW + PEAK model recovers H_0 posteriors that are largely symmetric about the true value, whereas the BROKEN POWER LAW model typically finds more support for larger H_0 values. We note that [Pierra et al., 2024] report an underestimate of H_0 when fitting their Broken Power Law model to a single dataset generated with POWER LAW + PEAK, but the prior bounds on the maximum mass parameter of their model do not extend below the injected value of that parameter, which in turn does not allow for the possibility of an overestimated H_0 . It is therefore not possible to directly compare the direction of our observed H_0 offset with theirs, despite the similarities between our chosen population models.

Regardless, this systematic offset demonstrates that incorrect parametric models unfortunately induce a systematic bias in cosmological inference, not an increase in statistical uncertainty. Other parametric models may show different trends depending on which parts of the mass distribution they incorrectly model and how. Therefore, while we are able to identify the shortcomings of BROKEN POWER LAW when POWER LAW + PEAK describes the true mass distribution, it is not possible in practice to know which direction an incorrect parametric model will bias H_0 . Mitigating the bias by using more flexible models may thus be the simplest solution to this problem.

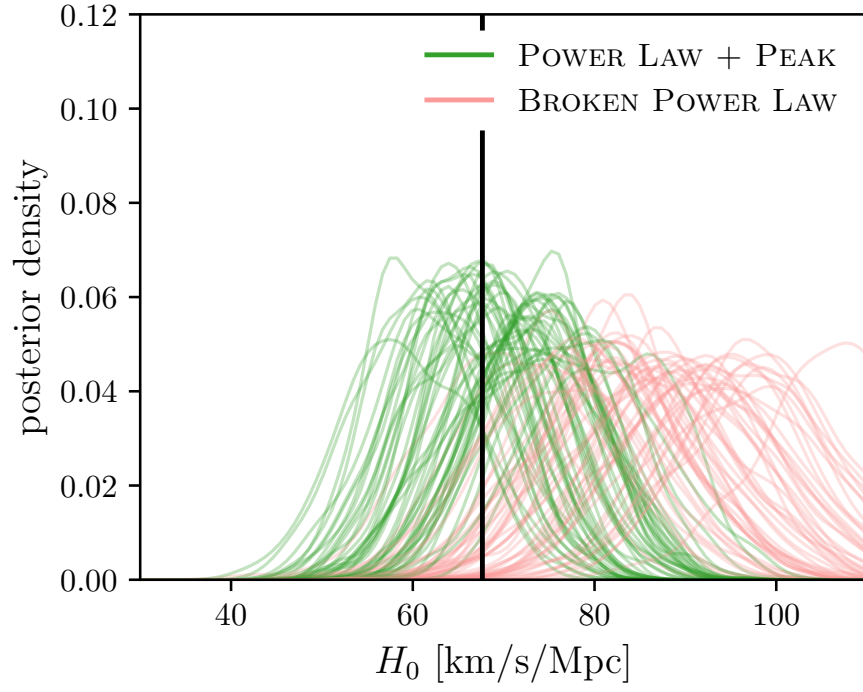


Figure 5.5: Recovered posteriors on H_0 from spectral siren analyses on 50 mock catalogs with the POWER LAW + PEAK (green) and BROKEN POWER LAW (pink) parametric models. The true, injected value is shown by a vertical black line.

CHAPTER 6

TWO OF A KIND: THE TWO-DIMENSIONAL BINARY BLACK HOLE MASS DISTRIBUTION

Analyses modeling the population of merging BBHs have generally focused on characterizing the distribution of primary masses in the binaries. Even for non-parametric methods such as those presented in Chapters 4 and 5, simplistic prescriptions are typically used for the distribution of secondary masses. However, this chapter argues that neglecting to model the secondary mass distribution with the same level of complexity as the primary mass distribution forgoes valuable astrophysical information. Namely, the secondary mass distribution and its relationship to the primary mass distribution can provide a fundamental observational constraint on the formation history of BBHs.

This is due to the fact that if both black holes experience similar stellar evolutionary processes prior to collapse, as might be expected in dynamical formation channels, the primary and secondary mass distributions would show similar features. If they follow distinct evolutionary pathways (for example, due to binary interactions that break symmetry between the initially more massive and less massive star), their mass distributions may differ.

This chapter presents the first analysis of the BBH population that explicitly fits for the secondary mass distribution. We find that the data is consistent with a $\sim 30 M_{\odot}$ peak existing only in the distribution of *secondary* rather than primary masses. This would have major implications for our understanding of the formation of these binaries. Alternatively, the data is consistent with the peak existing in both component mass distributions, a possibility not included in most other previous studies. In either case, the peak is observed at $31.4^{+2.3}_{-2.6} M_{\odot}$, which is shifted lower than the value obtained in previous analyses of the marginal primary mass distribution, placing this feature in further tension with expectations from a pulsational PISN pileup.

This chapter is based on *Two of a Kind: Comparing Big and Small Black Holes in*

Binaries with Gravitational Waves, a paper written in collaboration with Maya Fishbach and Daniel Holz, and published in the *Astrophysical Journal* [962].

6.1 Introduction

The BBH mass distribution is typically parameterized by the primary mass m_1 , the larger of the two component masses in the binary, and either the secondary mass m_2 , or the mass ratio $q = m_2/m_1$. Within these parametrizations, it is typically assumed either that the primary and secondary masses follow the same underlying distribution [Fishbach and Holz, 2020b, Doctor et al., 2020, Farah et al., 2022a, Edelman et al., 2023, Abbott et al., 2023a, Sadiq et al., 2023], or that the primary mass distribution has distinct features from the secondary mass distribution. The latter assumption implies that primary and secondary mass are physically meaningful labels, and is typically achieved by modeling the secondary mass distribution as a single power law between some minimum mass and m_1 [Fishbach and Holz, 2017, Kovetz et al., 2017, Talbot and Thrane, 2018, Baxter et al., 2021, Abbott et al., 2021d, Tiwari, 2021, Abbott et al., 2023a, Edelman et al., 2022, Callister and Farr, 2024, Godfrey et al., 2023]. The former assumption—that m_1 and m_2 follow the same underlying distribution—implies that the two-dimensional mass distribution is symmetric in primary and secondary mass, meaning that m_1 and m_2 can be interchanged without changing the mass distribution. Neither of these assumptions have been explicitly verified with the available data. In this work, we aim to determine if both components in merging BBHs follow the same underlying distribution or if there is a physical distinction between primary and secondary masses.

Understanding whether the primary and secondary masses in BBH mergers follow the same distribution will provide insight into their formation histories. For BBH mergers that are dynamically assembled in dense environments, we expect that both component BHs are drawn from the same population of stellar remnants (i.e. m_1 and m_2 are not physically meaningful labels), and that the two-dimensional BBH mass distribution will therefore be

symmetric in m_1 and m_2 . For BBH mergers that originate from binary stars that formed and evolved in relative isolation (“field binaries”), there may be a physical distinction between primary and secondary BH masses. To start off, the progenitor stars in the binary are drawn from the binary IMF, which may not be symmetric between the two components [e.g. Grudić et al., 2023]. Furthermore, the two stars exchange mass during binary stellar evolution. In each phase of mass transfer, one component acts as the donor and the other as the accretor, depending on their initial masses. Mass transfer affects the donor and the accretor stars in different ways, which can impact the masses of the resulting BHs following stellar collapse [Laplace et al., 2021]. On the other hand, both BH progenitors are expected to have undergone binary stripping and therefore experience similar supernova physics, potentially washing out any significant differences between the mass distributions of first- and second-born BHs [van Son et al., 2022a, Schneider et al., 2023]. However, the degree to which these distributions are similar or different depends on uncertain mass loss and accretion physics [van Son et al., 2022b].

Supernova kicks may also be different for first versus second-born components in *merging* binaries, because kicks determine whether or not the binary can merge within the age of the Universe [Kalogera, 1996, Gallegos-Garcia et al., 2022]. Because natal kicks are related to the remnant mass via the supernova prescription [Fryer et al., 2012, Mandel et al., 2021], different preferences for the natal kick magnitudes between first-born and second-born BHs may cause the primary and secondary mass distributions to differ in merging binaries formed through isolated binary evolution [e.g. Oh et al., 2023].

In short, binary stellar evolution consists of several processes that can break the symmetry between the population of initially more massive stars, which generally correspond to the first born and more massive (primary) BHs, and the population of initially less massive stars, which generally correspond to the second born and less massive (secondary) BHs. However, if mass inversion occurs in some systems, some initially less massive stars will end up as the

more massive BH by the time of merger, and the distribution of primary BH masses will have contributions from both the second-born and first-born BHs [Olejak and Belczynski, 2021, Hu et al., 2022, Broekgaarden et al., 2022, Zevin and Bavera, 2022]. If mass inversion happens in exactly half of merging BBH systems, the primary and secondary component mass distributions may be indistinguishable even if the first and second born distributions differ.

From a data analysis perspective, knowing that the primary and secondary mass distributions are the same allows us to measure a single set of model parameters: those of the shared distribution. This may allow features of the distribution to be measured with higher precision since both components in the binary will contribute to the measurement of each feature, rather than just the primary mass. Furthermore, disentangling the role of primary and secondary masses aids the physical interpretations of such features.

For example, Abbott et al. [2023a] found that the mass distribution exhibits a peak at primary masses $m_1 \sim 35 M_\odot$. The astrophysical origin of this overdensity in the mass distribution is uncertain, although it may be related to (pulsational) PISN [e.g. Heger and Woosley, 2002, Fishbach and Holz, 2017, Talbot and Thrane, 2018, Farmer et al., 2019]. A necessary ingredient towards understanding the origin of the $m_1 \sim 35 M_\odot$ peak is to first understand whether the peak appears exclusively among primary BBH masses, or whether secondary masses also display a peak, indicating that secondary BHs also experience the astrophysical process that leads to a mass pileup.

This chapter is organized as follows. In Section 6.2 we describe the different population models considered in this work. In Section 6.3 we present the results of fitting each of the models to the GWTC-3 [Abbott et al., 2021g], finding that while the primary and secondary masses of merging BBHs are consistent with following the same underlying distribution, it is also possible that the secondary mass distribution has a more prominent peak than does the primary mass distribution. In Section 6.4 we discuss possibilities for future observations. In

Section 6.5 we summarize our conclusions and present possible astrophysical interpretations of our results. The appendices discuss the fundamental differences between commonly-used parametrizations for the two-dimensional mass distribution and include details about the population models and statistical framework.

6.2 Population models

Because our aim is to learn whether the primary and secondary masses are consistent with being drawn from the same distribution, we describe the primary and secondary mass distributions separately, but according to the same functional form. We model each of the one-dimensional mass distributions as a mixture model between a smoothed power law component and a Gaussian component in order to make direct comparisons to the POWER LAW + PEAK model used by the LIGO-Virgo-KAGRA collaboration (LVK) to describe the distribution of primary masses [Talbot and Thrane, 2018, Abbott et al., 2021d, 2023a].

There are several ways to construct a two-dimensional mass model for the binary from this one-dimensional mass model for the component masses. However, as we discuss in Section 6.6.1, in order to explicitly compare primary and secondary mass distributions, it is necessary to use the pairing function framework first introduced for GW population modeling in Fishbach and Holz [2020b]. Explicitly,

$$p(m_1, m_2 | \Lambda) = p_1(m_1 | \Lambda_1) p_2(m_2 | \Lambda_2) f(q; \beta_q), \quad (6.1)$$

where $p_1(m_1 | \Lambda_1)$ is the underlying distribution of primary masses, $p_2(m_2 | \Lambda_2)$ is the underlying distribution of secondary masses, $f(q)$ is a pairing function that depends on the mass ratio of the system¹, Λ_1 and Λ_2 are the hyper-parameters² describing the underlying

1. In principle, the pairing function can be parameterized in terms of any observable parameter (e.g., total mass).

2. We use the term “hyper-parameter” for a model parameter that describes the population of merging

ing primary and secondary mass distributions, respectively, and $\Lambda = \{\Lambda_1, \Lambda_2, \beta_q\}$ is the set of all mass model hyper-parameters. In this work, we use a pairing function of the form $f(q; \beta_q) = q^{\beta_q} \Theta(q \leq 1)$, though other forms may provide a better fit to the data [e.g. Farah et al., 2022a].

As illustrated in Fig. 6.1, different choices for the relationship between Λ_1 and Λ_2 result in distinct morphologies for the two-dimensional mass distribution. Below, we list each of the variations we consider in this work, along with the panels in Fig. 6.1 to which they correspond. A table describing these variations in terms of choices for the hyper-prior is given in Section 6.6.2.

- **PAIRING:SYMMETRIC:** This model sets $\Lambda_1 = \Lambda_2$, making the distribution symmetric under the transformation $m_1 \leftrightarrow m_2$. It corresponds to the first row of Fig. 6.1: any feature in one of the distributions has to be present in both, so it always makes two bands that connect on the diagonal.
- **PAIRING:GENERIC:** This model allows the hyper-parameters describing the $\sim 35 M_\odot$ peak to differ between $p_1(m_1|\Lambda_1)$ and $p_2(m_2|\Lambda_2)$. It can produce any of the panels in Fig. 6.1, and is the only model that can produce a scenario such as that illustrated in the middle panel, where the feature in m_1 has a different amplitude and location than the feature in m_2 .
- **PAIRING:NO PEAK IN $p_2(m_2)$:** This model sets all $\Lambda_2 = \Lambda_1$, except for the parameter governing the height of the $\sim 35 M_\odot$ peak, which we set to vanish for the secondary mass distribution but fit as a free parameter for the primary mass distribution. This model corresponds to the bottom row of Fig. 6.1: it is only capable of having a peak in the primary mass distribution, so can only produce vertical bands in a two-dimensional mass distribution. In Section 6.6.1, we show that **PAIRING:NO PEAK IN $p_2(m_2)$** ap-

compact binaries, in contrast to the parameters describing each individual GW detection, such as one system's masses and spins.

proximately mimics the behavior of the commonly-used POWER LAW + PEAK model from, e.g., Abbott et al. [2023a], which we refer to as LVK 2023.

The different columns in Fig. 6.1 correspond to different power law spectral indices, β_q , for the mass ratio-dependent pairing function. The leftmost panels show models where components in the binary are allowed to pair up independently of mass ratio, the middle column shows a model where components have a slight preference to pair up with partners that are equal mass, and the rightmost panel shows the case where components are very “picky”: they almost always pair up with equal mass partners [Fishbach and Holz, 2020b]. When components pair nearly independently of mass ratio, the asymmetric models produce noticeably different distributions than the symmetric models. However, in the case of very picky binaries, the two scenarios become difficult to tell apart. There is therefore a degeneracy between the steepness of the pairing function and the existence of distinct features in the two mass distributions [see Tiwari, 2023, for a discussion of this phenomenon in terms of Jacobian transformations].

In all models considered in this work, the redshift distribution is modeled as a power law with spectral index κ [Fishbach et al., 2018]. We use the DEFAULT spin model from Abbott et al. [2021d, 2023a] to describe the spin magnitudes and tilts of each component. These are the same redshift and spin distributions used with LVK 2023 in Abbott et al. [2023a]. The explicit form of the full population model is given in Section 6.6.2.

Using these parameterized models, we construct a hierarchical Bayesian inference (described in Chapter 2) to determine the appropriate population-level parameters for the mass distribution, Λ , given the observed set of data $\{D_j\}$ for N observed events [Loredo, 2004, Mandel et al., 2019]. We model the data as an inhomogeneous Poisson process with the rate density (number of events per unit time per single-event-parameter hypervolume) given by

$$\frac{d\mathcal{N}}{dm_1 dm_2, ds_1 ds_2 dz} = \mathcal{R} p(z) p(s_1, s_2) p(m_1, m_2 | \Lambda), \quad (6.2)$$

where $\mathcal{R}$ acts as a normalizing constant that sets the overall magnitude of the rate.

6.3 Results

We fit each model described in Section 6.2 to the BBHs in GWTC-3. The resulting two-dimensional mass distributions for the PAIRING:SYMMETRIC and PAIRING:GENERIC models are shown in Fig. 6.2. These plots represent an average over the hyper-posterior for each model; this average is sometimes referred to as a posterior predictive distribution (PPD). The contours differ in morphology to those illustrated in Fig. 6.1 because the actual distribution of BBH component masses exhibits two peaks: one at $\sim 10 M_\odot$ and another at $\sim 35 M_\odot$ [Abbott et al., 2021d, Tiwari, 2021, Edelman et al., 2022, Sadiq et al., 2022, Abbott et al., 2023a, Edelman et al., 2023, Farah et al., 2023, Ray et al., 2023], whereas we only place one peak in the models shown in Fig. 6.1. The peak at $\sim 10 M_\odot$ creates bands in all panels that have very little vertical extent because the peak is proximal to the minimum BH mass. Nonetheless, fits using both models exhibit vertical and horizontal bands, indicating peaks in both $p_1(m_1)$ and $p_2(m_2)$. While the PAIRING:SYMMETRIC model requires this behavior, the PAIRING:GENERIC model does not, meaning that the secondary mass distribution appears to exhibit its own feature at $\sim 35 M_\odot$. Furthermore, the fact that PAIRING:GENERIC produces a PPD similar to that of PAIRING:SYMMETRIC indicates that the data support consistent primary and secondary mass distributions.

The bands in all panels do not go the full extent of parameter space but rather taper off away from the diagonal, indicating a preference for equal-mass binaries either through a pairing function or a mass ratio distribution that favors $m_1 \approx m_2$.

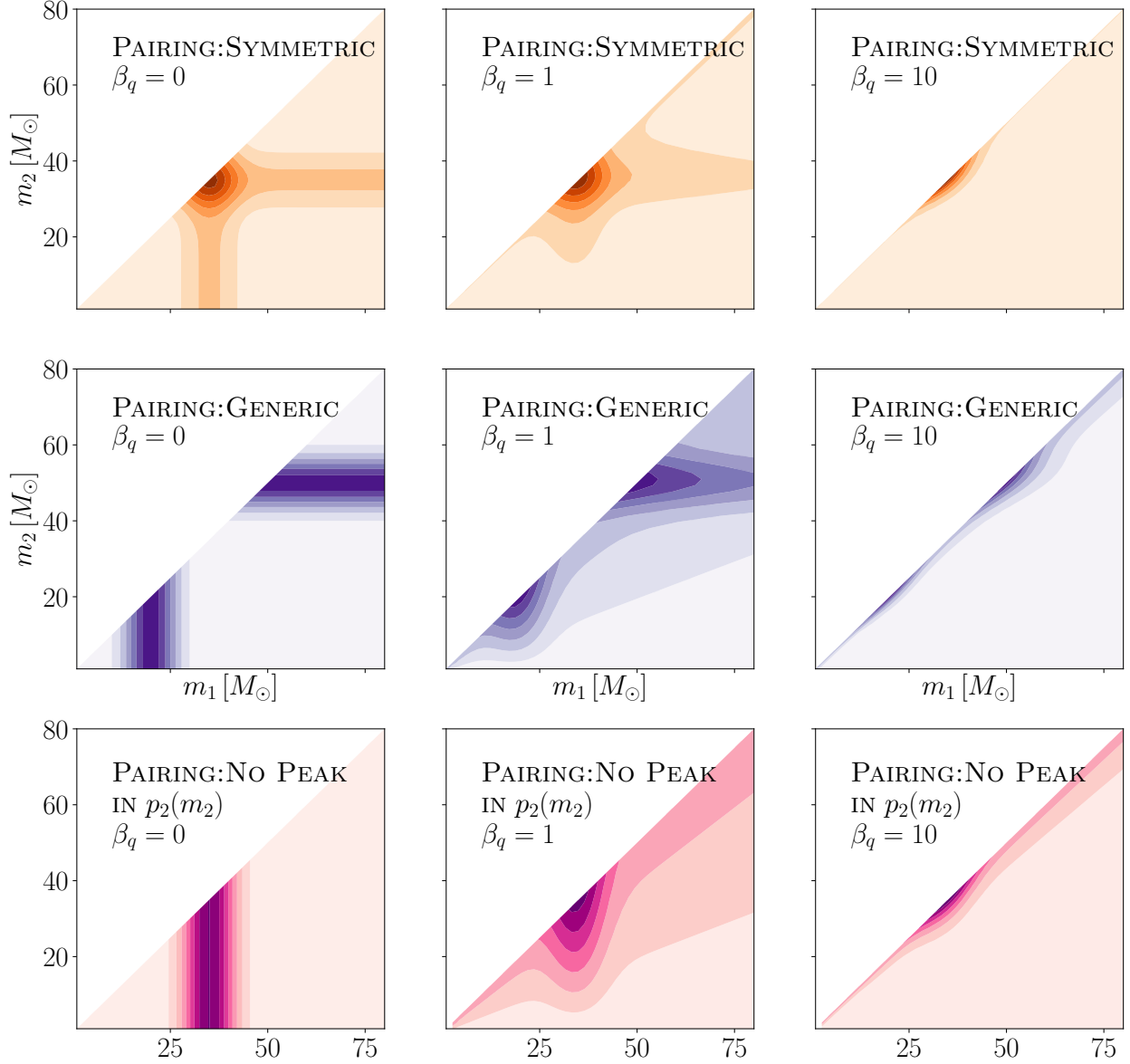


Figure 6.1: Illustration of some possible two-dimensional mass distributions under the models considered in this work. These distributions are not indicative of specific astrophysical predictions, but are instead meant to be illustrative of the morphologies accessible by current models. Overdensities/peaks in the mass distribution appear as darker filled contours in these figures. The different columns correspond to different power law spectral indices for the mass ratio-dependent pairing function, β_q . In the case where components strongly prefer to pair with nearly equal-mass partners, it becomes difficult to determine whether a feature appears only in one component mass distribution (as in the PAIRING:NO PEAK IN $p_2(m_2)$ model, *bottom row*) or in both (PAIRING:SYMMETRIC, *top row*). The diagonal contours in the middle and right columns are caused by a preference for equal-mass binaries and follow lines of constant mass ratio. The goal of this chapter is to distinguish between the different scenarios illustrated in this figure.

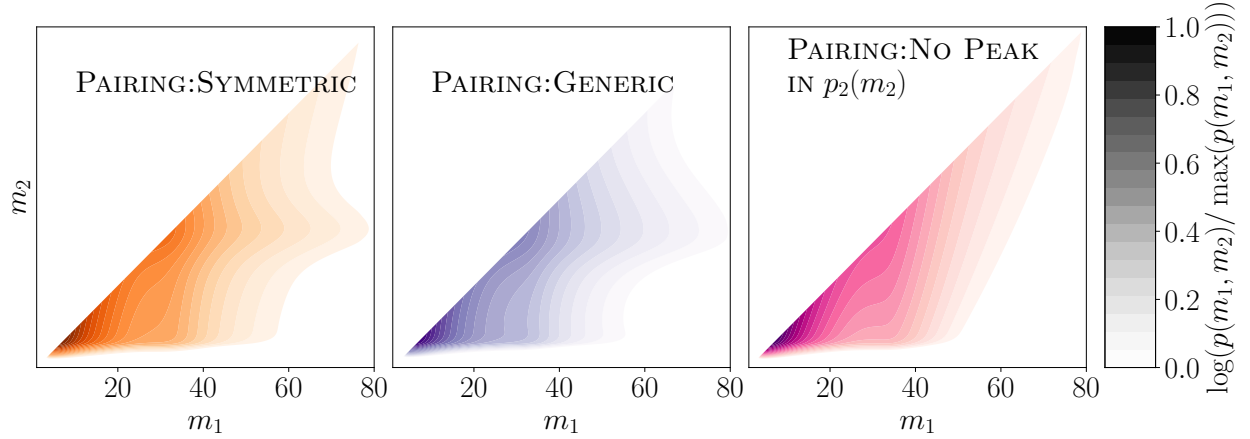


Figure 6.2: Two-dimensional PPDs for PAIRING:SYMMETRIC (*left, orange*), PAIRING:GENERIC (*middle, purple*), and PAIRING:NO PEAK IN $p_2(m_2)$ (*right, pink*), all fit to the BBHs in GWTC-3. Darker colors indicate a higher rate of sources, and each panel is individually normalized to its maximum value. All three models find a higher rate of events near the $m_1 = m_2$ diagonal, as well as peaks at $m_1 \sim 10 M_\odot$ and $m_1 \sim 35 M_\odot$. PAIRING:SYMMETRIC and PAIRING:GENERIC both find peaks in m_2 as well, as indicated by horizontal bands in the two leftmost panels, whereas PAIRING:NO PEAK IN $p_2(m_2)$ is unable to model features in the m_2 direction. Additionally, PAIRING:SYMMETRIC and PAIRING:GENERIC display peaks in similar locations despite the fact that PAIRING:SYMMETRIC requires that both m_1 and m_2 follow the same underlying distribution but PAIRING:GENERIC is able to model each component separately. This suggests that both components may be drawn from the same underlying distribution, up to a pairing function.

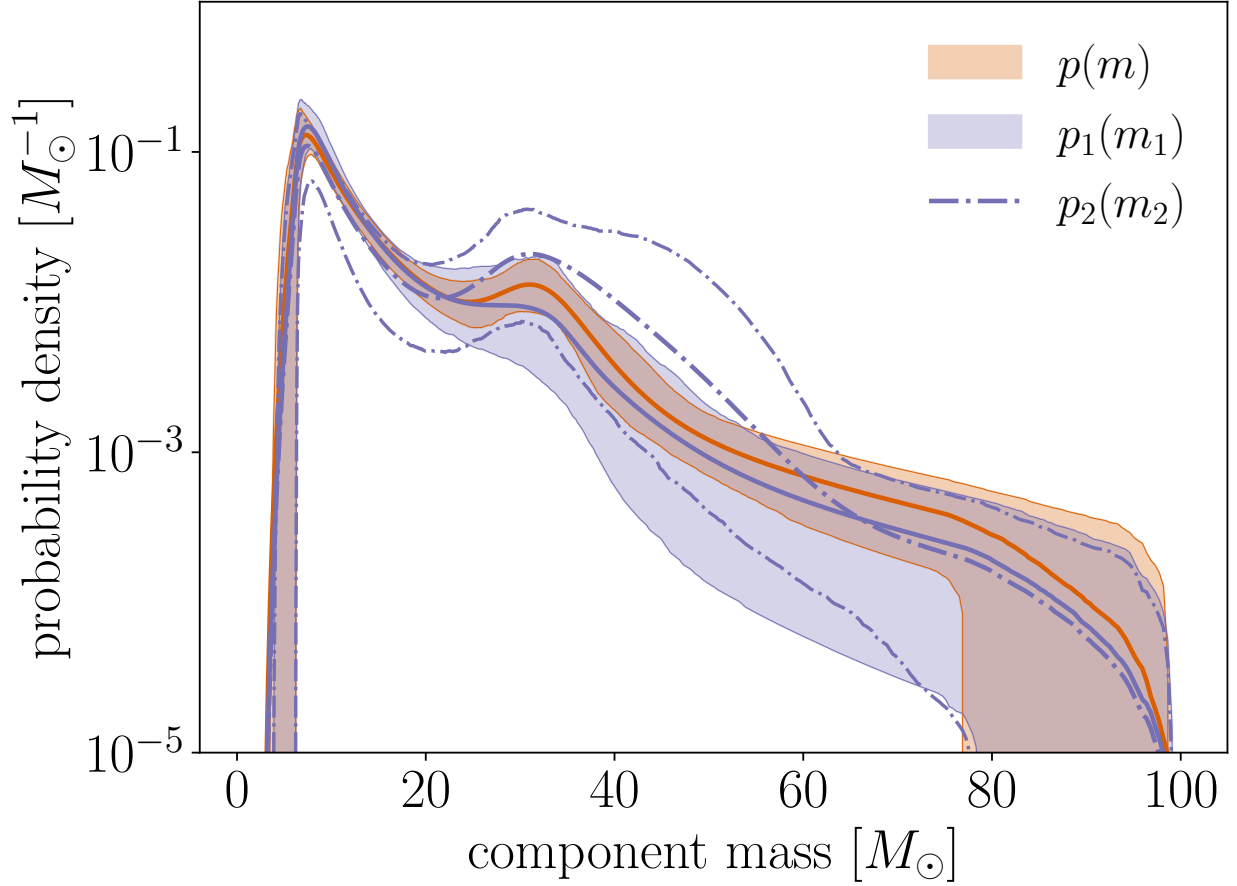


Figure 6.3: Underlying distribution of primary and secondary masses for the PAIRING:SYMMETRIC (*orange*) and PAIRING:GENERIC (*purple*) models. Under the PAIRING:SYMMETRIC model, $p_1(m_1) = p_2(m_2) \equiv p(m)$, so only $p(m)$ is shown. $p(m)$ under PAIRING:SYMMETRIC is more tightly constrained than $p_1(m_1)$ (*purple filled band*) or $p_2(m_2)$ (*dot-dashed lines*) under PAIRING:GENERIC as it has twice as many observations per free parameter. These distributions are constructed to describe the population of BHs *before* the function that pairs components into merging binaries is applied. All three underlying distributions are consistent with one another, though $p_2(m_2)$ appears to have a large peak at $\sim 35 M_\odot$ while $p_1(m_1)$ has some support for no peak in that region. $p(m)$ does not have support for no peak, but its peak is constrained to be small, while the peak in $p_2(m_2)$ is less well-constrained and may be larger in amplitude. This hints at the possibility that the peak identified in the primary mass distribution by the LVK 2023 formalism may have been driven by a peak in $p_2(m_2)$ rather than $p_1(m_1)$, though hyper-parameter uncertainties within PAIRING:GENERIC are too large to definitively determine this, and the data is consistent with $p_1(m_1) = p_2(m_2)$.

6.3.1 *Primary and secondary masses are consistent with having the same underlying distribution*

The underlying distributions (i.e. before a pairing function is applied) of the primary and secondary masses are shown in Fig. 6.3 for the two pairing function models. There is a region of overlap between the primary and secondary mass distributions under the PAIRING:GENERIC model, indicating that the primary and secondary mass distributions are consistent with one another. As expected, this region of overlap also coincides with $p(m)$, the distribution describing both component masses in the PAIRING:SYMMETRIC model.

Models not explicitly parameterized in terms of a pairing function are unable to produce underlying distributions such as those shown in Fig. 6.3, so for the sake of comparison to previous work, we turn to conditional³ m_2 distributions, $p(m_2|m_1 = C)$, where C can be any number in the domain of the m_1 distribution. Fig. 6.4 shows these conditional distributions for the PAIRING:SYMMETRIC, PAIRING:GENERIC and LVK 2023 models, the latter of which does not use a pairing function. The curves in Fig. 6.4 are averaged over the hyper-posterior for each model.

For all values of C , the inferred distribution under the PAIRING:GENERIC model behaves similarly to that of the PAIRING:SYMMETRIC model, indicating consistency between the primary and secondary mass distributions. In particular, under the PAIRING:SYMMETRIC and PAIRING:GENERIC models, we see a peak in the conditional secondary mass distribution $p(m_2|m_1 = C)$ when $C \geq 35 M_\odot$ (solid and dotted lines). While m_2 generally prefers to be near m_1 , as indicated by an upward trend in all models, the two pairing function models have more support for m_2 being within the peak than being nearly equal to m_1 when m_1 is large (dotted lines). This behavior is in contrast to the LVK 2023 model, which forces the conditional secondary mass distribution $p(m_2|m_1 = C)$ to monotonically

3. While marginal distributions are typically used for the purpose of comparing PPDs from several models on a single plot, they are not sensitive to differences between two-dimensional mass distributions when equal masses are preferred (see Section 6.6.2), so we use conditional distributions instead.

increase for all values of C because it does not allow for a peak in m_2 . While the PAIRING:GENERIC model can replicate this behavior, it instead recovers similar distributions to the PAIRING:SYMMETRIC model. The Bayes factor of PAIRING:GENERIC relative to PAIRING:SYMMETRIC is $\mathcal{B}(\text{PAIRING:GENERIC})/\mathcal{B}(\text{PAIRING:SYMMETRIC}) = 0.38$, indicating an inability to distinguish between the two models, with perhaps a modest preference for the PAIRING:SYMMETRIC model.

6.3.2 Improved constraints on peak location

If the primary and secondary mass distributions are identical, we can better measure the properties of features in that common distribution by using measurements of both m_1 and m_2 . Fig. 6.5 shows the one-dimensional hyper-posterior for the location, μ , of the Gaussian peak in the mass distribution under the PAIRING:SYMMETRIC and LVK 2023 models. We measure μ with a standard deviation of $1.50 M_\odot$ under the PAIRING:SYMMETRIC model compared to $2.08 M_\odot$ under the LVK 2023 model, representing a 28% improvement. This increase in precision is similar to that expected from using twice the number of independent events ($1 - 1/\sqrt{2} = 0.29$), because now both m_1 and m_2 contribute to the inference, rather than just m_1 .

Furthermore, the PAIRING:SYMMETRIC model recovers a lower value of $\mu = 31.4^{+2.3}_{-2.6} M_\odot$ (median and 90% credible interval) compared to the LVK 2023 result of $\mu = 33.6^{+2.6}_{-4.0}$, because it refers to a feature in the underlying $p(m)$ distribution rather than the marginal m_1 distribution. Features in $p(m)$ appear at larger masses in the marginal m_1 distribution and lower masses in the marginal m_2 distribution due to the constraint that $m_1 > m_2$.

If the $\sim 35 M_\odot$ peak is believed to be a feature of the supernova remnant mass distribution, it is best to use its location in the underlying $p(m)$ distribution rather than its location in the marginal distribution. For example, analyses wishing to compare this feature with expectations from a pair-instability supernova pileup [e.g. Stevenson et al., 2019, Farmer

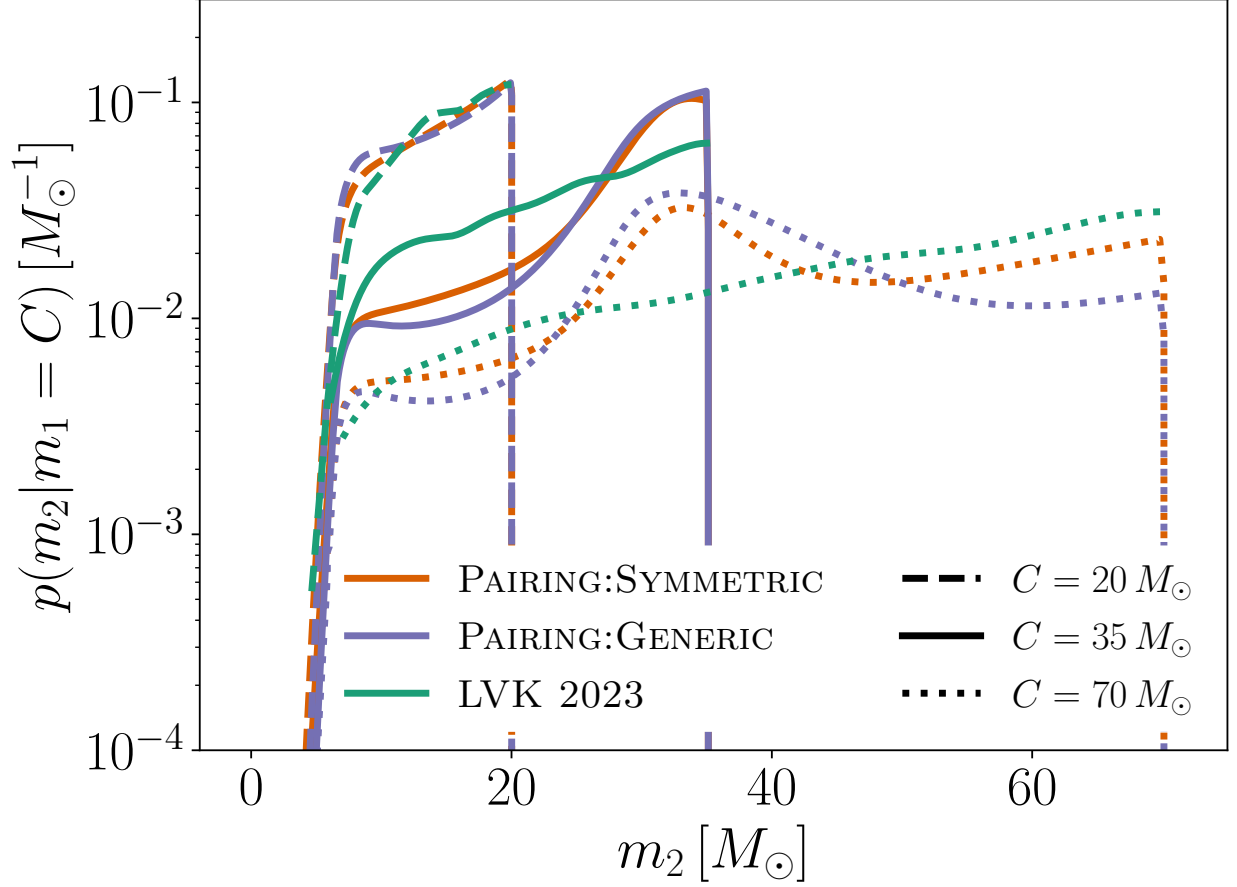


Figure 6.4: Conditional PPDs for the models considered in this work. We show the secondary mass distributions conditioned on $m_1 = 20 M_\odot$ (*dashed*), $m_1 = 35 M_\odot$ (*solid*) and $m_1 = 70 M_\odot$ (*dotted*) to exemplify when the primary mass is below, inside, and above the Gaussian peak, respectively. Orange lines correspond to PAIRING:SYMMETRIC, purple lines to PAIRING:GENERIC, and green lines to LVK 2023. Lines denote the mean posterior probability, marginalized over hyper-parameter uncertainty, and credible intervals are omitted for clarity. When m_1 is below the peak, all models agree. When m_1 is above or within the peak, PAIRING:SYMMETRIC and PAIRING:GENERIC exhibit a larger preference for m_2 to be in the peak than the LVK 2023 model does because the latter is constructed to behave like a power law in the range $[m_{\min}, m_1]$. The fact that PAIRING:GENERIC and PAIRING:SYMMETRIC approximately agree on the peak location and height indicates that the primary and secondary masses may follow the same underlying distribution up to a pairing function.

et al., 2020] should use the value presented here ($\mu = 31.4_{-2.6}^{+2.3} M_{\odot}$), whereas studies aiming to connect the peak with binary physics may wish to use the value presented by Abbott et al. [2023a] ($\mu = 33.6_{-4.0}^{+2.6}$). Interestingly, this lower peak location is in further tension with the latest theoretical predictions for a pair-instability pileup [Farag et al., 2022].

When possible, though, it is best to compare theoretical predictions directly to the two-dimensional mass distribution (such as those in Fig. 6.2) rather than to the values of specific hyper-parameters, since hyper-parameters have different meanings in different models. Correspondingly, it is important that models used to fit the data are intentionally constructed to allow for features in both the primary and secondary mass distribution. Many of the current parametric [Fishbach and Holz, 2017, Talbot and Thrane, 2018, Baxter et al., 2021, Abbott et al., 2021d, 2023a] and non-parametric [Tiwari, 2021, Edelman et al., 2022, Callister and Farr, 2024, Godfrey et al., 2023] BBH mass distribution models enforce asymmetry between m_1 and m_2 , excluding the possibility that the primary and secondary mass distributions share the same features⁴ (with a few exceptions, e.g. Edelman et al. 2023, Ray et al. 2023, Sadiq et al. 2023). It is possible to construct a mass distribution that allows for features in both the primary and secondary mass distribution without using a pairing function [e.g. Ray et al., 2023], but directly comparing the primary and secondary mass distributions using these parametrizations is less straightforward.

Another application of our improved measurement of the peak location is “spectral siren” cosmology, which uses such features in the mass distribution to infer the expansion history of the universe [Chernoff and Finn, 1993, Messenger and Read, 2012, Taylor and Gair, 2012, Farr et al., 2019, Ezquiaga and Holz, 2021, 2022, Abbott et al., 2023b]. Previous analyses found the location of this peak to be the parameter most correlated with the Hubble constant

4. Callister and Farr [2024] and Godfrey et al. [2023] use flexible models to fit $p(q)$ rather than using power laws, but they do not use a pairing function and instead fit the m_1 and q distributions independently, therefore assuming the product of the marginal distributions equals the two-dimensional distribution. This is the same fundamental choice as is described in Equation 6.3, and therefore assumes that m_1 and m_2 follow different distributions.

[see Figures 5 and 13 of Abbott et al., 2023b], so an improved constraint on μ should therefore have a relatively large effect on the constraints of cosmological parameters if all other mass distribution parameters remain equally-well constrained.

6.3.3 Evidence for a peak in the secondary mass distribution

Consistency between the primary and secondary mass distributions implies a peak in the secondary mass distribution at $m_2 \sim 30 M_\odot$, as this feature has already been robustly identified in the primary mass distribution [Abbott et al., 2021d, Tiwari, 2021, Abbott et al., 2023a, Edelman et al., 2023, Callister and Farr, 2024, Farah et al., 2023]. However, the data is also consistent with differing primary and secondary mass distributions, as shown by the regions in which the filled and dashed bands do not overlap in Fig. 6.3. In this case, it is worthwhile to explicitly determine whether there exists a peak in the secondary mass distribution.

The left panel of Fig. 6.6 shows the posterior distribution for the parameters governing the height of the peak in m_1 and m_2 under the PAIRING:GENERIC model, λ_1 and λ_2 , respectively. Setting $\lambda_{\{1,2\}} = 0$ means that no binaries in the astrophysical population are in the Gaussian peak⁵, while $\lambda_{\{1,2\}} = 1$ means all binaries are in the Gaussian peak. The lower left and upper right regions of the two-dimensional posterior are excluded, meaning that λ_1 and λ_2 cannot both be zero, nor can they both be large. This indicates that either one of the two distributions has a large peak while the other has none, or that both distributions have moderate peaks.

In fact, the posterior on λ_2 peaks at a higher value than λ_1 . Under the LVK 2023 model, the 1D posterior on λ peaks between the marginal λ_1 and λ_2 posteriors of the PAIRING:GENERIC model. Additionally, in the PAIRING:GENERIC model, λ_1 has more support

5. The integrated fraction of events in the region near the peak is higher than the fraction indicated by λ because both the Gaussian peak and the underlying power law contribute to the rate in that region. Therefore, $\lambda_{\{1,2\}} = 0$ does not mean that there are no events with masses $\sim 35 M_\odot$.

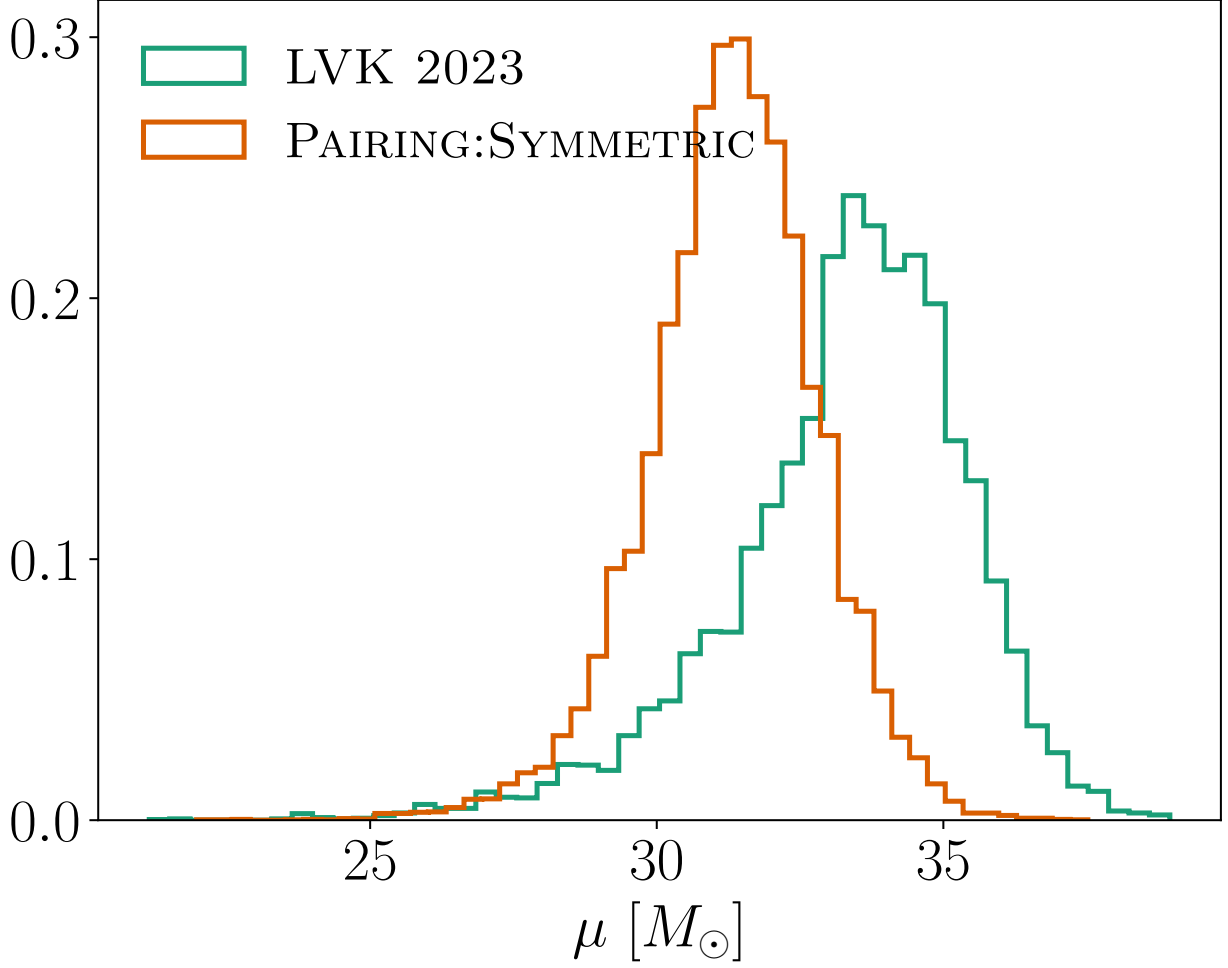


Figure 6.5: Marginal posteriors on μ , the location of the Gaussian peak, for the LVK 2023 and PAIRING:SYMMETRIC models. This parameter is more precisely measured under the PAIRING:SYMMETRIC formalism by $\sim 1/\sqrt{2}$, indicating that m_1 and m_2 independently contribute to this measurement. μ is the most correlated parameter with the Hubble constant when using “spectral sirens” [Ezquiaga and Holz, 2022, Abbott et al., 2023b], so using the PAIRING:SYMMETRIC model will presumably improve cosmological measurements. The central values of the two distributions differ because the hyper-parameter has slightly different effects on the resulting 2D mass distribution each model.

at zero relative to λ , while λ_2 has reduced support at zero. This hints at the intriguing possibility that the secondary masses may be driving the nonzero value of λ found by Abbott et al. [2021d] and Abbott et al. [2023a]. In other words, *it is possible that the $\sim 30 M_\odot$ peak exists the secondary mass distribution rather than the primary mass distribution.* However, the data are still consistent with $\lambda_1 = \lambda_2$: the dashed grey line in Fig. 6.6 intersects the $1.5\text{-}\sigma$ contour of the hyper-posterior.

As shown in Section 6.6.1, the LVK 2023 formalism behaves similarly to PAIRING:NO PEAK IN $p_2(m_2)$, which is nested within PAIRING:GENERIC when $\lambda_2 = 0$. Therefore, ruling out $\lambda_2 = 0$ would indicate that the pairing function formalism is strongly preferred. We find that $\lambda_2 = 0$ is disfavored but not ruled out: there is 4.2 times more posterior density at $\lambda_2 = 0.17$ (its median *a posteriori* value) than at $\lambda_2 = 0$. It is therefore difficult to tell with the current number of detections whether the data prefer for the peak to exist in $p_1(m_1)$, $p_2(m_2)$, or both.

These three possibilities are somewhat degenerate because of the preference for equal mass BBHs. To illustrate this, the right panel of Fig. 6.6 compares the power law spectral index, β_q , of the pairing function under PAIRING:GENERIC and PAIRING:NO PEAK IN $p_2(m_2)$. When there is no peak in $p_2(m_2)$, the β_q posterior shifts to higher values, which corresponds to a larger preference for $q \sim 1$. This is because the PAIRING:NO PEAK IN $p_2(m_2)$ model has a peak in $p_1(m_1)$, so it allows for a high fraction of secondary masses to lie within the peak by making more binaries equal-mass. For comparison, we also plot the posterior on β_q under PAIRING:SYMMETRIC to show how a different set of assumptions about λ_2 changes β_q . The shift in β_q when $\lambda_2 = 0$ is larger than when $\lambda_2 = \lambda_1$, suggesting that it is driven by an excess of events with $m_2 \sim 35 M_\odot$ rather than by other model assumptions or a different realization of the inference.

In summary, we find modest evidence for a peak in $p_2(m_2)$, suggesting that pairing function models may be preferred over the LVK 2023 formalism, though current observations

do not allow us to definitively rule out that the peak exists only in $p_1(m_1)$. The data is consistent with $\Lambda_1 = \Lambda_2$, so there is not strong evidence against the possibility that m_1 and m_2 are drawn from the same underlying distribution, up to a pairing function.

6.3.4 *BBHs are picky*

Pairing functions provide an intuitive way to quantify whether the properties of one BH in a binary influences those of its companion. Binaries that pair up independently of each component’s masses are described by a pairing function with $\beta_q = 0$, whereas a preference for equal-mass binaries is described by $\beta_q > 0$. We find $\beta_q > 0$ to $> 99.99\%$ for PAIRING:SYMMETRIC and to 99.95% for PAIRING:GENERIC. This is consistent with an earlier study by Fishbach and Holz [2020b] who use the pairing function formalism on GWTC-1 [Abbott et al., 2019b]. However, BBHs are not maximally picky: the posterior on β_q is not railing against the high end of the prior bounds.

Our inferred two-dimensional mass distribution is generally consistent with that shown in Sadiq et al. [2023], who use a non-parametric approach and assume that the population is symmetric under $m_1 \leftrightarrow m_2$. They show evidence for peaks in the mass ratio distribution at $q \sim 0.5$ when $m_1 \sim 15 M_\odot$ and $m_1 \sim 70 M_\odot$, which they interpret as a lack of preference for similar-mass pairings. However, we note that these peaks in the mass ratio distribution translate to peaks in the secondary mass distribution at $\sim 7 M_\odot$ and $\sim 35 M_\odot$, the same locations at which the primary mass distribution exhibits peaks. Parametrizations that assume symmetry under $m_1 \leftrightarrow m_2$ but do not use a pairing function are unable to disentangle the effects of preference for similar-mass pairings from features in one or both component mass distributions. We therefore conclude that both our results and those presented in Sadiq et al. [2023] are consistent with a preference for similar-mass pairings as well as structure in the secondary mass distribution.

In this work, we have only examined power law forms for the pairing function, but

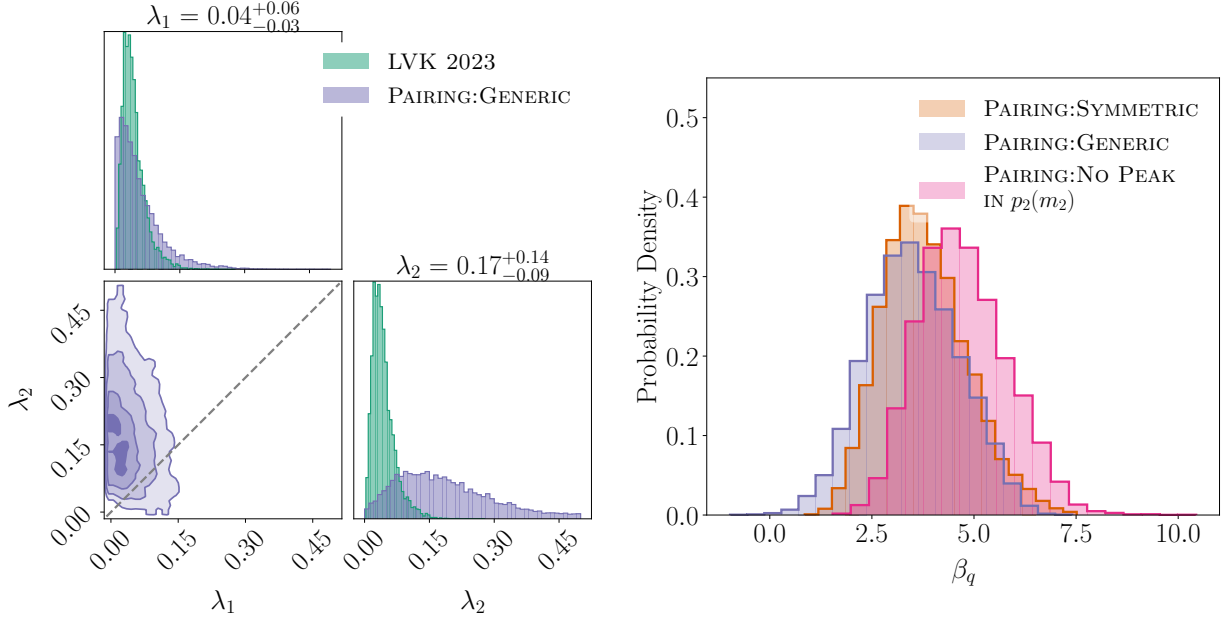


Figure 6.6: Hyperposteriors under pairing function models. *Left:* Corner plot of hyperparameters governing the height of the Gaussian peak for the primary (λ_1) and secondary (λ_2) mass distributions under the PAIRING:GENERIC model. The medians and 90% credible intervals are shown above their respective marginal distributions. The medians indicate that roughly 17% of BBHs have secondary masses in the Gaussian peak while 4% have primary masses in the peak, though this preference for a larger peak in $p_2(m_2)$ may be due to the relatively poor constraint on λ_2 rather than a true preference in the data. The lower left portion of the two-dimensional posterior is excluded, indicating that there must be a peak in either $p_1(m_1)$ or $p_2(m_2)$, and a slight preference exists for the peak to be in $p_2(m_2)$ (upper left portion of plot) or both distributions (central portion of plot), though the peak being in $p_1(m_1)$ only is not completely ruled out. For reference, the 1D posterior on λ under the LVK 2023 model is shown in green, with $\lambda = 0.04^{+0.03}_{-0.02}$ [Abbott et al., 2023a]. The dashed grey line indicates where $\lambda_1 = \lambda_2$. *Right:* Marginal posteriors on β_q , the hyperparameter controlling the steepness of the pairing function under the PAIRING:SYMMETRIC, PAIRING:GENERIC, and PAIRING:NO PEAK IN $p_2(m_2)$ models. When we set $\lambda_2 = 0$, β_q increases, indicating a preference for equal-mass components. This behavior is likely caused to accommodate an excess of events with $m_2 \sim 35 M_\odot$, which can either be caused by a peak in $p_2(m_2)$ or with a peak in $p_1(m_1)$ and a strong preference for equal-mass binaries (see discussion in Section 6.6.1).

more complex models are under investigation and will be presented in future work. If the pairing function is a more complex function of mass ratio, or if it depends on multiple parameters, such as primary mass or spin [e.g. Farah et al., 2022a], it may become easier to distinguish between pairing function models and models parameterized similarly to LVK 2023, as complex pairing functions may find more support away from the $m_1 = m_2$ diagonal. The same is true if the marginal mass ratio distribution has multiple modes, or if mass ratio is allowed to correlate with other parameters [e.g. effective spin, Callister et al., 2021].

6.4 Looking to the future

The LVK 2023 and pairing function models differ most in their predictions for the number of events in the region $m_1 \in [35, 40] \cap m_2 \in [m_{\min}, 30]$, because they disagree on whether to place m_2 in the Gaussian peak or somewhat evenly throughout the available parameter space. We can therefore determine which model will be preferred in the future by counting the number of detected events in that region.

It is expected that 260^{+330}_{-150} BBHs will be detected by the end of O4, and 870^{+1100}_{-480} BBHs will be detected by the end of O5 [Kiendrebeogo et al., 2023]. With 260 (870) total BBH detections, we expect 23.6 ± 4.6 (79.0 ± 8.5) BBHs to fall in the region $m_1 \in [35, 40] \cap m_2 \in [m_{\min}, 30]$ under the LVK 2023 model and 12.6 ± 3.5 (42.3 ± 6.3) under the PAIRING:SYMMETRIC model. This means that by the end of O4, we will be able to distinguish the LVK 2023 and PAIRING:SYMMETRIC models to $> 2\sigma$, and we will be able to distinguish them to $> 4\sigma$ by the end of O5.

Of course, it will be necessary to perform a full hierarchical analysis, as measurement uncertainties of detected systems will cause events to scatter into and out of this region. Additionally, in a hierarchical Bayesian context, posterior predictive checks performed on the event-level parameters (such as the true masses of individual events) are less sensitive than those performed on the observed data, such as the strain or the maximum likelihood

event parameters [Sinharay and Stern, 2003, Gelman, 2006, Bayarri and Castellanos, 2007, Lored, 2013]. This is why, for example, Fishbach et al. [2020b] performs posterior predictive checks on the maximum-likelihood values of the observed and predicted events rather than the posterior distributions of those events. Therefore, more sensitive posterior predictive checks may be able to distinguish between the two frameworks with fewer events than we project here.

6.5 Summary and implications for astrophysics of merging BBHs

We present the first analysis of the secondary mass distribution of merging BBHs. This allows us to explore whether the primary and secondary component masses in merging BBHs are drawn from the same distribution, or whether the BBH mass distribution is asymmetric in $m_1 \leftrightarrow m_2$ as is commonly assumed in BBH population studies. We find the data to be consistent with two possibilities: either the primary and secondary mass distributions are similar, or a larger peak exists in the secondary mass distribution than the primary mass distribution. In either scenario, a peak likely exists in the secondary mass distribution. This possibility is not considered in many previous analyses of the BBH population, which fix the secondary mass distribution to be a power law [e.g. Abbott et al., 2019a, 2021d, 2023a] or assume that m_1 and m_2 are interchangeable [e.g. Fishbach and Holz, 2020b, Sadiq et al., 2023]. The existence of this secondary-mass peak has implications for the formation channels of merging BBHs and the origin of the $\sim 35 M_\odot$ peak.

If the mass distribution is indeed symmetric under $m_1 \leftrightarrow m_2$ as in our preferred model PAIRING:SYMMETRIC, this may imply that a large fraction of merging BBHs are formed through dynamical assembly, in which the two component BHs are born through a similar process and then find each other in a dense stellar environment. In this case, the pairing function and its dependence on mass ratio and/or total mass may encode valuable information about dynamical processes like mass segregation and binary exchanges.

The appearance of a peak in both primary and secondary mass distributions is consistent with an origin in the BH remnant mass function. However, the peak location at $31.4^{+2.3}_{-2.6} M_{\odot}$ is in tension with predictions for pair-instability supernovae, so this feature is likely caused by another astrophysical process. As another application of our work, we suggest that future spectral siren measurements consider the PAIRING:SYMMETRIC model when inferring the Hubble constant, because it provides an improved measurement of the peak location relative to the LVK 2023 model and the peak location is strongly correlated with the inferred value of H_0 .

On the other hand, the data remain consistent with an asymmetric BBH mass distribution $p(m_1, m_2)$, in which $p_1(m_1) \neq p_2(m_2)$, as in the PAIRING:GENERIC model. This would imply that a significant fraction of merging BBHs originate from field binaries, in which “primary” and “secondary” are physically meaningful labels if they tend to correspond to the first-born or second-born BH in a binary. The component mass distributions $p_1(m_1)$ and $p_2(m_2)$ would then encode the binary IMF and the highly uncertain physics of binary stellar evolution. Specifically, the location and prominence of features in each component mass distribution may provide insight into how mass loss versus mass accretion affect stellar evolution and collapse, including their effect on supernovae and the BH remnant mass function. For example, when mass transfer is unstable, a peak is expected at $\sim 15 M_{\odot}$ in the primary mass distribution, but near the minimum BH mass in the secondary mass distribution, though the relative locations of these peaks depends on common-envelope and supernova kick physics [van Son et al., 2022a].

Notably, if the BBH mass distribution is asymmetric, we find that *it is possible that no peak exists in the primary mass distribution, and the previously-identified peak in primary mass is actually driven by an overdensity in the secondary mass distribution*. A peak that is more prominent in $p_2(m_2)$ than $p_1(m_1)$ may imply that PISN are more frequent or are shifted to lower masses among second-born BHs.

The degree of asymmetry in the BBH mass distribution provides insight into the frequency of mass inversion in binary stars, providing a complementary probe to BH spins [Mould et al., 2022]. If mass inversion never occurs, m_1 will typically be the remnant of the donor star. However, if mass inversion is the norm, m_1 will be the remnant of the accretor star. A perfectly symmetric BBH distribution under isolated binary evolution would imply that mass inversion happens exactly 50% of the time, causing the primary and secondary mass distributions to be indistinguishable even though binary physics imparts different distributions on the first-born versus second-born BH, but this is statistically unlikely. Determining whether primary and secondary BH masses follow the same underlying distributions is therefore a novel and promising probe of formation channels. However, this test should be interpreted within the full BBH population inference, as theoretical models should make consistent predictions for the mass, spin and redshift distributions.

We have only explored parametric models in this work, and can therefore only comment on the features in the mass distribution explicitly parameterized by our model. Future investigations may find it beneficial to use a non-parametric approach that separately fits the primary and secondary mass distributions by employing a pairing function.

We additionally find that BHs in merging binaries have a strong preference to pair with similar-mass BHs for all forms of the secondary mass distribution considered in this work. This is consistent with the results presented in Fishbach and Holz [2020b], though we more confidently exclude the scenario in which BBHs pair independently of mass ratio since we now have many more detected BBHs.

By the end of O5, we expect to confidently distinguish between the different scenarios presented here for the BBH mass distribution. In this work, we found that secondary masses in merging BBH systems likely display a peak at $\sim 35 M_\odot$, whereas previous results identified this peak exclusively among primary masses. With a few hundred additional BBH observations, we expect to determine whether both component mass distributions have a

peak at the same location and if the peak is more prominent among secondary masses.

Directly incorporating the distribution of secondary masses can serve as an important tool to constrain formation mechanisms of BBHs.

6.6 Additional Analysis Details

The following software packages were used in this chapter: `gwpopulation` [Talbot et al., 2019], `bilby` [Ashton et al., 2019], `scipy` [Virtanen et al., 2020], `corner.py` [Foreman-Mackey, 2016].

6.6.1 Comparison of Common Mass Distribution Parametrizations

Differences between mass ratio distributions and pairing functions

Given a form for the primary mass distribution, there are several ways to construct a two-dimensional mass distribution. We discuss two possibilities here that are common in the literature, showing the different effects they have on the resulting two-dimensional mass distribution.

The main qualitative differences between the two parametrizations are illustrated in Fig. 6.7. The top row has three examples of mass distributions that can be described by a model of the form

$$p(m_1, m_2 | \Lambda_m, \beta) = p(m_1 | \Lambda_m) p(q | m_1, m_{\min}, \beta) \quad (6.3)$$

for different values of β . Here, m_1 is the mass of the heavier component in the binary, m_2 is the mass of the lighter component, and $q = m_2/m_1 \leq 1$ is the mass ratio. This can be equivalently written as

$$p(m_1, m_2 | \Lambda_m, \beta) = p(m_1 | \Lambda_m) p(m_2 | m_1, \beta), \quad (6.4)$$

since $m_2 = qm_1$. The parametrization described in Equation 6.3 is used by all parametric models presented in Abbott et al. [2021d] and Abbott et al. [2023a] that were used to model the primary mass distribution of BBHs, such as BROKEN POWER LAW and POWER LAW + PEAK, and is also commonly used in other analyses [e.g., Fishbach and Holz, 2017, Kovetz et al., 2017, Talbot and Thrane, 2018, Tiwari, 2021, Edelman et al., 2022, Callister and Farr, 2024, Godfrey et al., 2023]. For the remainder of this Section, we will refer to models parameterized by Equation 6.3 as “CONDITIONED-Q,” since they require the mass ratio distribution to be explicitly conditioned on primary mass.

The bottom row of Fig. 6.7 has three examples of mass distributions that can be described using a “pairing function”, f [Fishbach and Holz, 2020b]. Models with pairing functions have the form

$$p(m_1, m_2 | \Lambda) = p_1(m_1 | \Lambda_1) p_2(m_2 | \Lambda_2) f(q; \beta_q) \quad (6.5)$$

where $f(q)$ is a pairing function that depends on the mass ratio of the system⁶, and $\Lambda = \{\Lambda_1, \Lambda_2, \beta_q\}$ is the set of all model hyper-parameters. In this work, we use a pairing function of the form $f(q; \beta_q) = q^{\beta_q} \Theta(q \leq 1)$, though other forms may provide a better fit to the data [e.g. Farah et al., 2022a]. In the examples illustrated in Fig. 6.7, the primary and secondary mass distributions are equivalent, so $p_1(m) = p_2(m) \equiv p(m)$. Alternatively, $\Lambda_1 = \Lambda_2$. This describes a situation in which there is a single underlying mass distribution from which both components are drawn. The pairing function then describes how likely the two components are to be combined in a merging binary based on their mass ratio. A pairing function that prefers equal mass binaries will cause a marginal mass ratio distribution that has more support near $q = 1$, but the inverse is not necessarily true.

The parametrization in Equation 6.5 factorizes the possibility that components in binaries prefer to be near-equal mass and the possibility that the primary and secondary masses have

6. In principle, the pairing function can be parameterized in terms of any observable parameter (e.g., total mass).

distinct probability distributions. In other words, pairing functions allow us to model the secondary mass separately from the primary mass, while also allowing for the possibility that component BHs prefer to pair with similar-mass BHs.

In Fig. 6.7, a peak at $35 M_{\odot}$ is placed in both models to show the effects of such features in both cases. One consequence of CONDITIONED-Q models is that features such as Gaussian peaks can only appear in the primary mass distribution. This is shown by the vertical bands in the top row of Fig. 6.7 and lack of horizontal bands, since a band in the vertical (horizontal) direction are caused by a peak or dip in the primary (secondary) mass distribution for a range of secondary (primary) masses. For pairing function models, features can appear in $p_1(m_1)$, $p_2(m_2)$, or both. We have illustrated the case in which the same feature appears in both component mass distributions, and this appears as both the vertical and horizontal bands in the bottom row of Fig. 6.7. If Λ_1 and Λ_2 are allowed to differ, features could appear in only one of these distributions. This would cause there to only be horizontal bands if features only existed in $p_2(m_2)$, and only vertical bands if features only existed in $p_1(m_1)$. Features are also able to appear in different locations in $p_1(m_1)$ vs $p_2(m_2)$ under the pairing function formalism. However, the CONDITIONED-Q formalism only allows for bands in the vertical direction, meaning that it is not flexible enough to capture true underlying distributions with features in $p_2(m_2)$. The behavior of the CONDITIONED-Q formalism can in general be approximated by the pairing function formalism, while the opposite is not true.

The different columns in Fig. 6.7 correspond to different power law spectral indices for the mass ratio distribution (*top row*, β) and mass ratio-dependent pairing function (*bottom row*, β_q). The top row's leftmost panel has a uniform mass ratio distribution, the middle panel has a mass ratio distribution that mildly favors equal-mass binaries, and the right panel's mass ratio distribution strongly favors equal mass binaries. Analogously, the bottom row's leftmost panel shows a model where components in the binary are allowed to pair up independently of mass ratio, the middle panel shows a model where components have

a slight preference to pair up with partners that are equal mass, and the rightmost panel shows the case where components are “picky”: they almost always pair up with equal mass partners [Fishbach and Holz, 2020b]. When the mass ratio distribution is broad, or when components pair near-independently of mass ratio, the CONDITIONED-Q models produce noticeably different distributions than the pairing function models. However, in the case of very picky binaries or a steeply rising mass ratio distribution, the CONDITIONED-Q and pairing function models become difficult to tell apart, and likely explain the data equally well. There is therefore a degeneracy between the steepness of the pairing function and the existence of distinct features in the two mass distributions [see Tiwari, 2023, for a discussion of this phenomenon in terms of Jacobian transformations].

Fortunately, as we will show in Section 6.3.4, we measure $\beta \sim 3.5$ and $\beta_q \sim 1$, so the data lie somewhere between the middle and rightmost columns.⁷ This means that differentiating between the two scenarios will be difficult, but possible given enough data.

Mass distributions of compact objects are often visualised through a plot of the marginal component mass distributions. The marginal m_2 distribution is defined as

$$p(m_2|\Lambda) = \int dm_1 p(m_1, m_2|\Lambda), \quad (6.6)$$

where $p(m_1, m_2|\Lambda)$ can be parameterized in the way described in Equation 6.3 or Equation 6.5. However, in all but the top left panel in Fig. 6.7, the *marginal* secondary mass distribution exhibits a peak between $30 M_\odot$ – $40 M_\odot$, even though the secondary mass distribution is not able to have any features on its own. This is because features in the primary mass distribution induce features in the marginal secondary mass distribution if equal-mass binaries are preferred: when a binary’s primary mass is within the peak, its secondary mass is

7. The value for β under CONDITIONED-Q is different than the value found for β_q under the pairing function models because the two parameters cause different behaviors in the 2D mass distribution within their respective models. A low value for β does not imply that BBHs are not picky, just that the marginal mass ratio distribution rises slowly. We discuss this in more detail in Section 6.6.1.

also likely to be in that peak simply because $m_2 \simeq m_1$ is preferred. It is therefore difficult to distinguish between these different scenarios by looking at the marginal distributions alone. Instead, we analyze two-dimensional distributions such as the ones illustrated in Fig. 6.7, as well as the secondary mass distributions conditioned at various values of m_1 . The latter can be thought of as one-dimensional slices of the former.

It is our goal to determine whether the data prefer models described by pairing functions or by CONDITIONED-Q functions. We also aim to determine if the primary and secondary masses are drawn from the same distribution, and whether we can draw physical insights from features that appear in the primary mass distribution, secondary mass distribution, or both.

“Pickiness”

Under the CONDITIONED-Q formalism, we cannot determine how BHs in binaries choose their companions, though we can gain some insight from their marginal mass ratio distribution. CONDITIONED-Q parametrizes the mass ratio distribution as a power law with spectral index β , where $\beta > 0$ corresponds to mass ratio distributions with more support for similar-mass binaries. Abbott et al. [2023a] find $\beta = 1.1^{+1.7}_{-1.3}$, which similarly indicates a preference for equal-mass binaries. Note that the value for β under CONDITIONED-Q is noticeably different than the value found for β_q under the pairing function models because the two parameters cause different behaviors in the 2D mass distribution within their respective models. A low value for β does not imply that BBHs are not picky, just that the marginal mass ratio distribution rises slowly. As shown in Section 6.6.2, the pairing function models and CONDITIONED-Q model all produce near-identical marginal mass ratio distributions, despite different values for β_q and β . Low values for β are partially due to the fact that the low- q end of the marginal mass ratio distribution will always be suppressed because of the existence of a minimum BH mass. This minimum mass makes it impossible to get

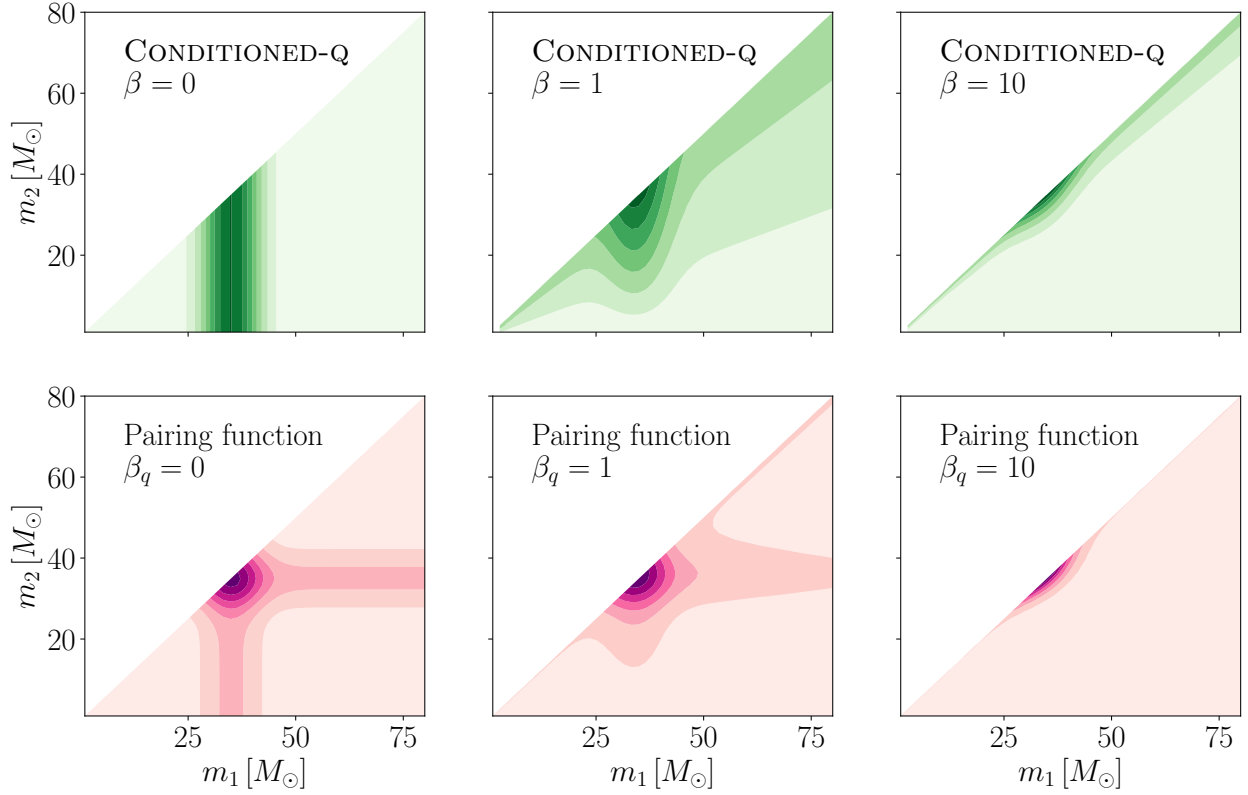


Figure 6.7: Illustration of some possible two-dimensional mass distributions under the commonly used “CONDITIONED-Q” formalism described by Equation 6.3 (*top row*) and the pairing function formalism described by Equation 6.5 (*bottom row*). Overdensities/peaks in the mass distribution appear as darker filled contours in these figures. The CONDITIONED-Q formalism is only able to produce models with features in the m_1 distribution, as shown by the vertical bands in the top row, whereas the pairing function formalism can model features in either m_1 or m_2 , or both. The different columns correspond to different power law spectral indices for the mass ratio distribution (*top row*, β) and mass ratio-dependent pairing function (*bottom row*, β_q). In the case of a steeply rising mass ratio distribution, or if components strongly prefer to pair with nearly equal-mass partners, the CONDITIONED-Q model and the pairing function model become difficult to tell apart and likely explain the data equally well, as shown by the two panels in the leftmost column. The diagonal contours in the middle and right columns are caused by a preference for equal-mass binaries and follow lines of constant mass ratio.

extreme mass ratios unless m_1 is very large, and the m_1 distribution has very little support for $m_1 \gtrsim 60 M_\odot$. Therefore, β does not need to be large in order for the marginal mass ratio distribution to strongly disfavor unequal-mass binaries. In fact, the existence of a minimum mass plus a tapering at high m_1 means that even when $\beta < 0$, the marginal mass ratio distribution rises towards $q = 1$.

Mimicking CONDITIONED-Q with a pairing function model

We show that the PAIRING:NO PEAK IN $p_2(m_2)$ model approximates the morphology of the LVK 2023 model, which uses the CONDITIONED-Q formalism.

The right two panels of Fig. 6.11 show the two-dimensional PPDs for the LVK 2023 and PAIRING:NO PEAK IN $p_2(m_2)$ models. Both exhibit vertical bands and no horizontal bands, and have a similar-magnitude drop in merger rate moving away from the diagonal.

The leftmost panel of Fig. 6.11 shows the conditional m_2 distribution for both models. While the slopes of the two models differ slightly, a peak in m_2 appears in neither. We therefore find PAIRING:NO PEAK IN $p_2(m_2)$ to be an appropriate proxy for the behavior of the LVK 2023 model for the purposes of this work. A future measurement ruling out $\lambda_2 = 0$ with high confidence would therefore serve as evidence for the pairing function models over LVK 2023-like models.

6.6.2 General Population Model

Base model

We model the two-dimensional mass distribution by constructing separate one-dimensional distributions for the primary and secondary masses and combining draws from these two distributions according to a pairing function, as in Equation 6.5. Fishbach and Holz [2020b] find the pairing function is most informative when parameterized by the mass ratio of the

binary, so we adopt a pairing function described by a power law in mass ratio.

Because our aim is to learn whether the primary and secondary masses are consistent with being drawn from the same distribution (up to a pairing function and subject to the constraint that $m_1 \geq m_2$), we describe the primary and secondary mass distributions separately, but according to the same functional form. We model each of the 1-dimensional mass distributions as a mixture model between a smoothed power law component and a Gaussian component G in order to make direct comparisons to the POWER LAW + PEAK model used by the LIGO-Virgo-KAGRA's (LVK) population analysis to describe the distribution of primary masses Talbot and Thrane [2018], Abbott et al. [2023a]. Explicitly,

$$\begin{aligned}
p(m_1|\Lambda_1) &\propto \left[(1 - \lambda_1)\Theta(m_1 > m_{\min,1})\Theta(m_1 < m_{\max,1}) \left(\frac{\alpha_1 + 1}{m_{\max,1}^{\alpha_1+1} - m_{\min,1}^{\alpha_1+1}} \right) m_1^{-\alpha_1} \right. \\
&\quad \left. + \lambda_1 G(m_1|\mu_1, \sigma_1) \right] \times S(m_1|m_{\min,1}, \delta_1) \\
p(m_2|\Lambda_2) &\propto \left[(1 - \lambda_2)\Theta(m_2 > m_{\min,2})\Theta(m_2 < m_{\max,2}) \left(\frac{\alpha_2 + 1}{m_{\max,2}^{\alpha_2+1} - m_{\min,2}^{\alpha_2+1}} \right) m_2^{-\alpha_2} \right. \\
&\quad \left. + \lambda_2 G(m_2|\mu_2, \sigma_2) \right] \times S(m_2|m_{\min,2}, \delta_2).
\end{aligned} \tag{6.7}$$

Here, $G(m_{\{1,2\}}|\mu_{\{1,2\}}, \sigma_{\{1,2\}})$ is a normalized Gaussian distribution with mean $\mu_{\{1,2\}}$ and width $\sigma_{\{1,2\}}$. The parameter $\lambda_{\{1,2\}}$ is a mixing fraction determining the relative prevalence of mergers the power law and Gaussian components, and $S(m_{\{1,2\}}, m_{\min,\{1,2\}}, \delta_{\{1,2\}})$ is a smoothing function which rises from 0 to 1 over the interval $(m_{\min,\{1,2\}}, m_{\min,\{1,2\}} + \delta_{\{1,2\}})$. Λ_1 is then the set of hyper-parameters $\{m_{\min,1}, m_{\max,1}, \alpha_1, \lambda_1, \mu_1, \sigma_1, \delta_1\}$ and $\Lambda_2 = \{m_{\min,2}, m_{\max,2}, \alpha_2, \lambda_2, \mu_2, \sigma_2, \delta_2\}$.

In all models considered in this work, the redshift distribution is modeled as a power law with spectral index κ [Fishbach et al., 2018] such that

$$p(z) \propto \frac{dV_c}{dz} \left(\frac{1}{1+z} \right) (1+z)^\kappa. \tag{6.8}$$

We use the DEFAULT spin model from Abbott et al. [2021d, 2023a] to describe the spin magnitudes and tilts of each component. These are the same redshift and spin distributions used with the POWER LAW + PEAK mass model in the analysis presented in Abbott et al. [2023a].

A fit to the model described above is provided in Section 6.6.2, though we focus on specific variations nested within this more general model for the remainder of this work.

Fit to general base model

We present results of a fit to the most general form of the pairing function model described in Section 6.6.2 (Equation 6.7). Here, we do not fix any parameters to be equal between Λ_1 and Λ_2 , and instead infer them separately.

Fig. 6.8 shows the posterior of all mass-related hyper-parameters within this model in the form of a corner plot. Because of the large number of free parameters, several are not well-constrained. Nonetheless, all parameters describing $p_1(m_1)$ are consistent with those describing $p_2(m_2)$. Additionally, strong correlations exist between α_1, α_2 , and β_q , making it difficult to meaningfully constrain all three parameters. We therefore set $\alpha_1 = \alpha_2$ in the PAIRING:GENERIC model.

Fig. 6.9 shows the underlying distributions, $p_1(m_1)$ and $p_2(m_2)$ inferred by the most general version of the base model. As expected, the constraints on these distributions weaken relative to that of the PAIRING:GENERIC model. However it is still clear that $p_1(m_1)$ and $p_2(m_2)$ are consistent with one another and that it is possible that $p_2(m_2)$ has a larger peak than $p_1(m_1)$ does. Notably, there now seems to be very little evidence for a peak in $p_1(m_1)$, whereas a clear peak still exists in $p_2(m_2)$.

These results are consistent with what has been presented using the PAIRING:GENERIC model in the main body of this work.

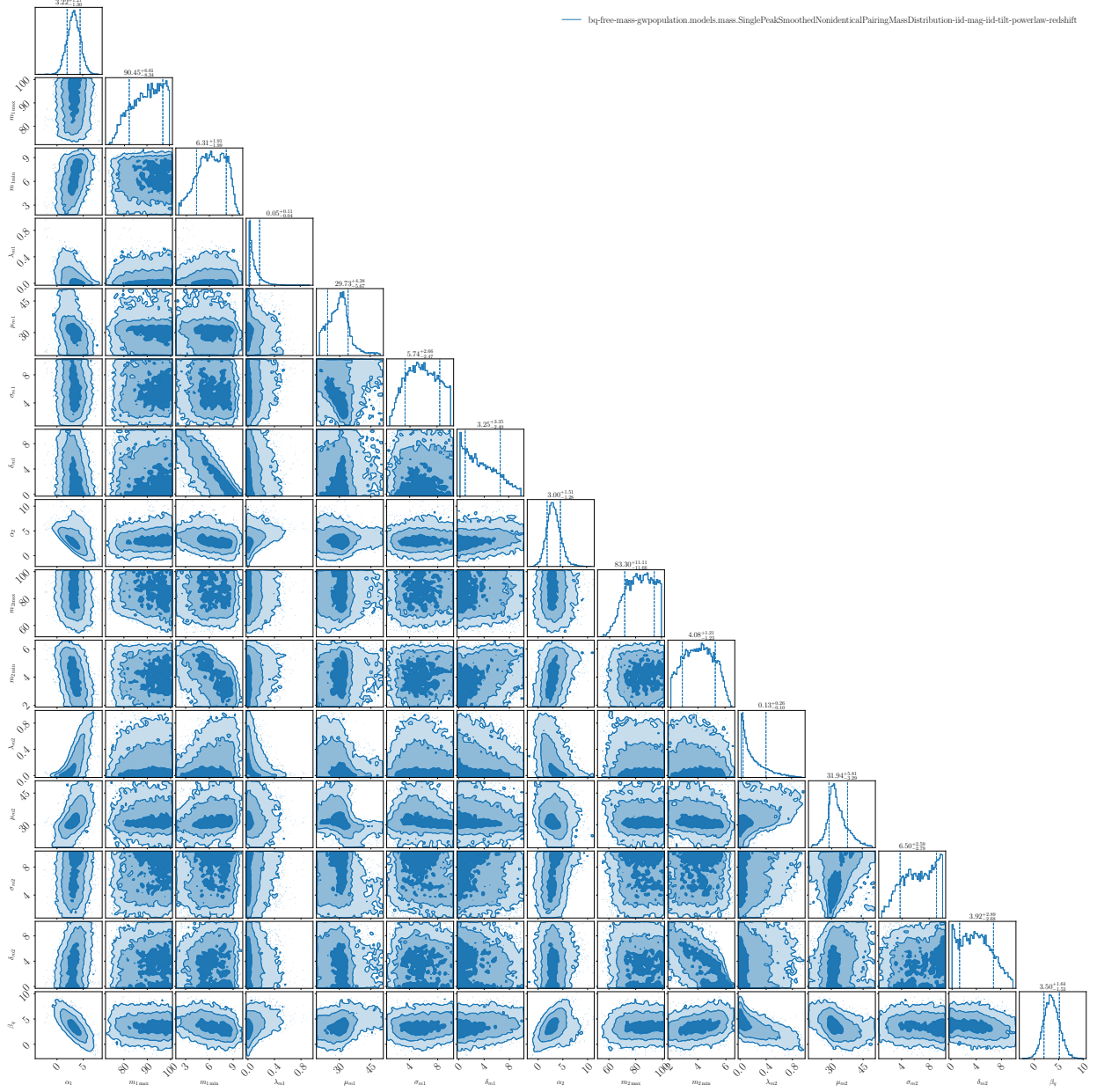


Figure 6.8: Corner plot of all mass-related hyper-posteriors in the most general form of the base model. Power law spectral indices governing the slope of the primary and secondary mass distributions are degenerate with one another and with the pairing function power law, so only two of the three parameters can be meaningfully constrained at a time. All other parameters that perform the same function in $p_1(m_1)$ and $p_2(m_2)$ are consistent between the two underlying distributions.

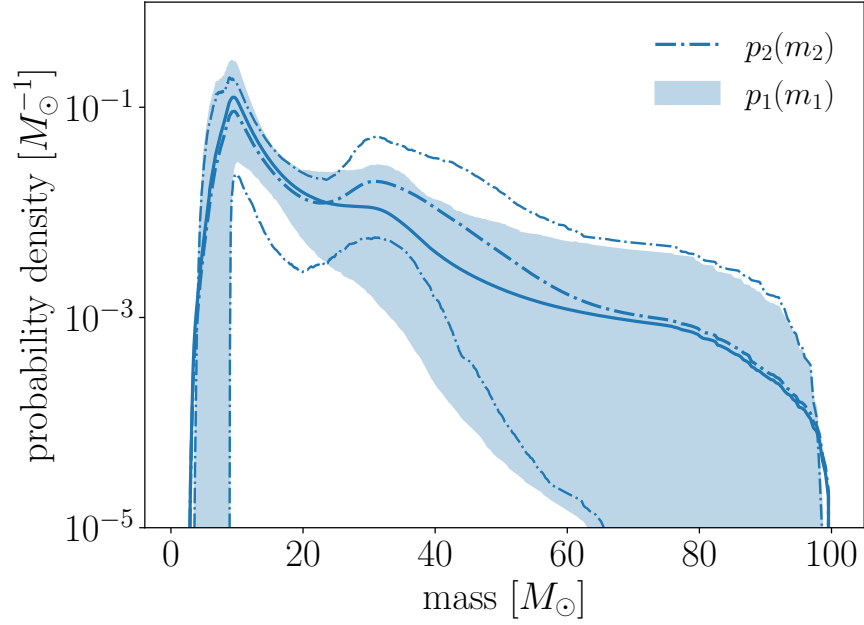


Figure 6.9: Underlying distributions (i.e. before a pairing function is applied) of the primary (*shaded band*) and secondary (*dot-dashed lines*) masses under the most general form of the base model. Despite fitting all parameters separately between the two distributions, they appear consistent with one another. However, both are relatively poorly constrained. Interestingly, the support for a peak in $p_1(m_1)$ lessens in this more generalized model, while the support for a peak in $p_2(m_2)$ remains the same as in PAIRING:GENERIC.

Model comparison

In this Section, we compare the various models considered in this work, including the CONDITIONED-Q model from LVK 2023. Table 1 of Farah et al. [2024] lists the hyper-prior choices made to construct each model, along with descriptions of each hyper-parameter.

Fig. 6.10 shows the m_1 , m_2 , and q PPDs, marginalized over all other parameters. These show a high level agreement between the different models, despite their two-dimensional PPDs differing in Fig. 6.2. Notably, the marginal secondary mass distribution exhibits a peak even for the CONDITIONED-Q model. As discussed in Section 6.6.1, this is caused by the preference for equal masses: if m_1 is in the $\sim 35 M_\odot$ peak, a preference for $m_2 \approx m_1$ will cause m_2 to also preferentially be in that region, causing more probability density for $m_2 \sim 35 M_\odot$ than elsewhere. Marginal PPDs are therefore not very sensitive to whether m_2 prefers its own features independently of the existence of a preference for equal mass binaries. Instead, we examine the underlying component mass distributions explicitly modeled by PAIRING:SYMMETRIC and PAIRING:GENERIC, as well as the values of informative hyper-parameters in these models to determine whether $p_2(m_2)$ differs from $p_1(m_1)$. For comparisons to CONDITIONED-Q, we plot secondary mass distributions conditioned on various values of m_1 .

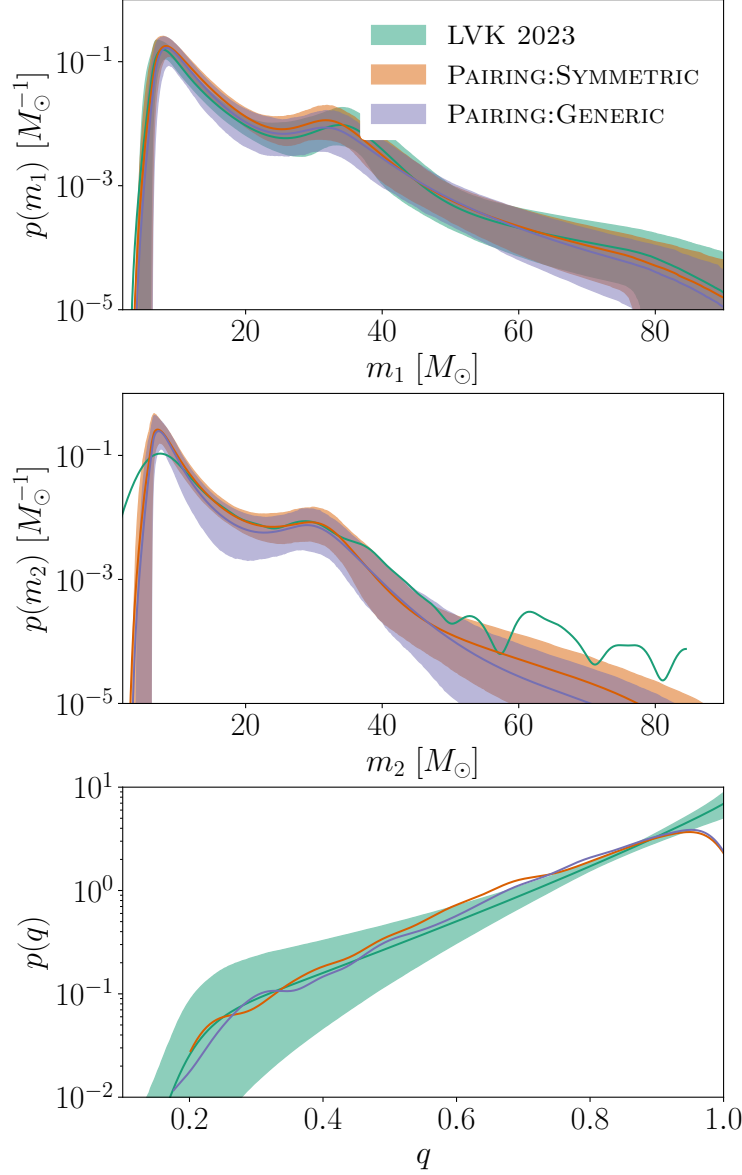


Figure 6.10: Marginal PPDs of primary mass (*top*), secondary mass (*middle*) and mass ratio (*bottom*) for CONDITIONED-Q (*green*), PAIRING:SYMMETRIC (*orange*) and PAIRING:GENERIC (*purple*). Solid lines are the mean value and shaded regions represent the 90% credible interval. These are obtained by integrating the 2D PPDs shown in Fig. 6.2 along each dimension in turn. Since the CONDITIONED-Q model is parameterized in terms of m_1 and q , we reconstruct its marginal m_2 distribution by sampling from the joint distribution and creating a kernel density estimate, causing some artificial “wiggles” in these plots. The same is true for the marginal q distributions of the pairing function models. Despite the 2D PPDs appearing different between CONDITIONED-Q and the other two models, the marginal PPDs all appear similar. This is because features in the primary mass distribution induce features in the marginal secondary mass distribution if equal-mass binaries are preferred. Therefore, marginal PPDs are not sensitive to the features in the 2D mass distribution in which we are interested.

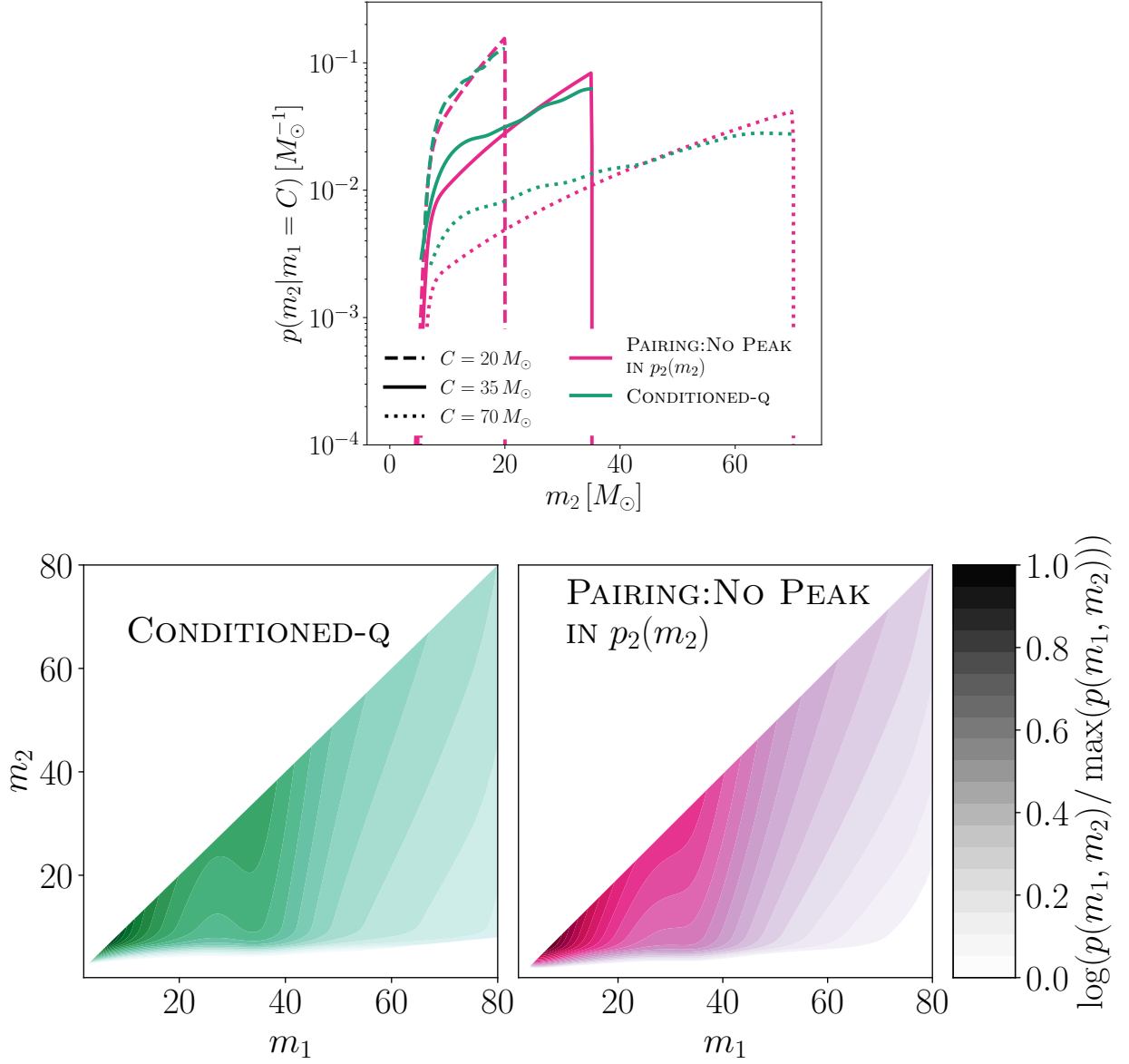


Figure 6.11: Two-dimensional (*bottom*) and conditional m_2 (*top*) PPDs for the CONDITIONED-Q and PAIRING:NO PEAK IN $p_2(m_2)$ models.

CHAPTER 7

DON'T JUST LEAVE ONE OUT: PROBING EXTREMAL GRAVITATIONAL-WAVE EVENTS WITH COARSE-GRAINED LIKELIHOODS

The majority of this dissertation has focused on gaining insight from features in the inferred GW source mass distribution itself. However, this population also serves as a helpful reference point for understanding individual events. For example, events that are inconsistent with the population are classified as outliers. Outliers are understood as originating from a different population than the majority of other observations, or, equivalently, as a symptom of population models that are inadequate at describing all of the observed data. In either understanding of outliers, they are useful hints of yet-undescribed subpopulations.

The two most common methods of identifying outliers with respect to the inferred population of GW sources are the “leave-one-out” approach, wherein the population fit with all events in a catalog is compared to a fit with all events except for the suspected outlier, and the PPD approach, wherein the inferred population is evaluated on its ability to predict the suspected outlier. In this chapter, we show that the “leave-one-out” approach is biased because population fits will always change under the exclusion of the most extreme data point, even if that data point is not an outlier. We also point out the ambiguity inherent in the PPD approach: the population that should predict the suspected outlier could either be the one inferred with or without that event. We then propose a solution to both of these problems in the form of coarse-grained population likelihoods. When applying this coarse-grained likelihood to GWTC-2, we found that GW190814 is an outlier with respect to the BBH population, but other events are not, consistent with previous studies done by the LVK. This means that full-spectrum mass models such as the one presented in 3 remain the only analyses that can accommodate GW190814.

This chapter is based on *Probing Extremal Gravitational-wave Events with Coarse-grained Likelihoods*, a paper written in collaboration with Reed Essick, Shanika Calaudage, Colm Talbot, Maya Fishbach, Eric Thrane, and Daniel Holz, and published in the *Astrophysical Journal* [926]. I was the second author of this publication, and my contribution was mainly to Section 7.3, where I used the method developed by Reed Essick to analyze events in GWTC-2. Sections of the original paper to which I did not contribute have been removed from this chapter.

7.1 Introduction

At the time of this chapter’s writing, GWTC-2 had just been released. This new catalog brought to light several interesting features within the distributions of masses and spins of compact objects in coalescing binary systems [Abbott et al., 2021d]. In particular, GWTC-2 set new records for the largest BH mass [GW190521, Abbott et al., 2020d], smallest BH mass [GW190814, Abbott et al., 2020e]¹, and most asymmetric mass ratios [GW190814 and GW190412, Abbott et al., 2020c,e]. Of immediate interest is whether these objects are merely the most extreme events observed from a single population: the most extreme examples in a catalog become more extreme as the size of the catalog grows [Fishbach et al., 2020b]. Alternatively, these events may be inconsistent with the population inferred from the rest of the detected events, and thus are *true outliers*. The observation of such outliers could suggest the first example from an as-of-yet unmodeled or entirely new (sub)population. However, it may simply indicate that the current phenomenological population models are simply a poor description of nature.

The interpretation of the most extreme GW events has significant astrophysical implications. The most massive BBH events probe the PISN mass gap [Fishbach and Holz,

1. It is possible that the secondary object in GW190814 is actually an unusually massive NS [Essick and Landry, 2020].

2017, Talbot and Thrane, 2018], a theoretically-proposed dearth of BHs between $\sim 50\text{--}120 M_{\odot}$ [Heger and Woosley, 2002]. Observing BBH systems near the pair-instability gap can inform our knowledge of nuclear reaction rates [Farmer et al., 2020], beyond standard model physics [Croon et al., 2020, Baxter et al., 2021], or the boundaries of mass gaps in general [Fishbach and Holz, 2020a, Edelman et al., 2021, Nitz and Capano, 2021, Ezquiaga and Holz, 2021] and applications thereof [e.g., Farr et al., 2019]. At the other extreme, BBH with small component masses can inform our knowledge of supernova physics [e.g., Fryer and Kalogera, 2001, Belczynski et al., 2012, Zevin et al., 2020].

Several authors have posited that the most massive BHs of GWTC-2, which appear to sit in the PISN mass gap, form a separate population from the “main population” observed to date, invoking formation scenarios such as hierarchical mergers [e.g., Abbott et al., 2020f, Kimball et al., 2020a, Tagawa et al., 2021] or primordial BHs [e.g., Franciolini et al., 2021, De Luca et al., 2021]. Indeed, in order to explain the wide range of observed BBH properties, many authors have argued that multiple formation channels are active and that the BBH population consists of multiple subpopulations [Abbott et al., 2021d, Ng et al., 2021, Zevin et al., 2021]. More than anything else, the variety of interpretations of the GWTC-2 events highlight the excitement within the GW community as new discoveries are routinely made whenever more data is recorded. However, they also emphasize the need for thoughtful consideration of the methods employed to assess whether individual events are consistent with existing population models.

Motivated by the “leave-one-out” consistency checks in Abbott et al. [2019a], Fishbach et al. [2020b], Abbott et al. [2021d], and elsewhere, we introduce a general procedure to investigate the effect of individual events on inferred populations. Specifically, we derive a “coarse-grained” analysis that retains, in a controlled way, only a subset of the total information available about some detected events. We are then able to determine whether the population inferred from this “coarse-grained” data is consistent with the population inferred

from the original data.

In particular, previous approaches compared populations inferred with all N events in a catalog to populations inferred with only $N - 1$ events, arguing that if these were similar then the event that was “left out” is consistent with the rest of the population, as its inclusion does not significantly change the inferred population. If the contrary were true, the event would be considered a true outlier. However, such tests do not account for the way in which the excluded event was selected. It was left out specifically because it was extremal in some dimension, and this selection may artificially inflate the apparent significance of differences in inferred populations, particularly in the presence of sharp boundaries within the population model. Our approach improves upon this by self-consistently accounting for how the selected events are chosen in a controlled way while clearly defining the subset of the available information that is retained. We show that our approach naturally reproduces previous techniques when a random event is excluded from the analysis. It is precisely because the excluded events are typically not chosen at random that previous techniques can introduce biases.

There is already a healthy literature proposing leave-one-out analyses as tests for model misspecification within Bayesian inference. For example, Vehtari et al. [2017] constructs a cross-validation likelihood from analyses that leave out each event in a catalog one at a time. This approach attempts to limit the possible biases associated with leaving out only extremal events, similar to the motivation for our coarse-grained approach. Other authors have considered tempering the likelihood in order to combat model misspecification, which is at times referred to as coarsening [Miller and Dunson, 2015]. In this approach, the likelihood is raised to the power of an inverse temperature $\beta = 1/T \in (0, 1]$, thereby artificially inflating statistical uncertainties. Several procedures exist to select β [e.g., Miller and Dunson, 2015, Thomas and Corander, 2019, and references therein], all of which attempt to optimize the balance between systematic error from model misspecification and additional statistical error

from tempered likelihoods. Our coarse-grained likelihoods are similar in spirit, but differ in that we eschew the use of *ad hoc* tempering in favor of marginalizing over the data and parameters from an individual event subject to a precisely specified (but looser) constraint on the event’s parameters. That is, we specify the size and placement of the “coarse grain” used to approximate the event’s likelihood.

Using this coarse-grained likelihood, we revisit several astrophysically interesting events from GWTC-2 that were discussed in detail in Abbott et al. [2021d]. To wit, we demonstrate that

- GW190814, with a secondary mass of $m_2 \sim 2.6 M_\odot$ and mass ratio $q = m_2/m_1 \sim 0.1$, is an outlier in m_2 (too small to be consistent with the main BBH population) but is not an outlier in mass ratio.
- GW190412 is not an outlier, and its mass ratio [$q \sim 0.28$; Abbott et al., 2020c] is in the tail of the main BBH population.
- GW190521 is not an outlier with respect to the main BBH population under the preferred phenomenological mass models considered in Abbott et al. [2021d]. It is only marginally inconsistent under the simplest mass model considered, which is disfavored for other reasons as well.

While we believe the quantitative details of our analysis improve upon previous methods, all of our resulting astrophysical conclusions are in agreement with Abbott et al. [2021d].

We derive our coarse-grained leave-one-out formalism from first principles in Section 7.2, and we apply this method to public LVK data in Section 7.3. We conclude in Section 7.4.

7.2 Coarse-grained leave-one-out likelihoods

We derive a procedure for coarse-graining our inference of the population in Sec. 7.2.1, while in Sec. 7.2.2 we discuss the implications of different possible procedures to choose the amount

of information retained in the coarse-grained inference. We describe how to quantify these consistency tests with a single statistic in Sec. 7.2.3.

7.2.1 Derivation of coarse-grained inference

To begin, let us assume we have N events that are each described by a set of m parameters $\theta \in \mathcal{R}^m$. We further assume these events belong to the same astrophysical population described by the hyperparameters Λ_0 :

$$\theta_i \sim p(\theta|\Lambda_0) \quad \forall \quad i. \quad (7.1)$$

If we write our model for the differential Poisson number density of signals in the universe as $d\mathcal{N}/d\theta = Rp(\theta|\Lambda)$, our joint distribution over the data $\{D_i\}$, single-event parameters $\{\theta_i\}$, hyperparameters, and the rate R is then

$$p(\{D_i\}; \{\theta_i\}; \Lambda, R) = p(\Lambda)p(R)R^N e^{-R\mathcal{E}(\Lambda)} \prod_{i=1}^N p(D_i|\theta_i)p(\theta_i|\Lambda)\Theta(\rho(D_i) \geq \rho_{\text{thr}}), \quad (7.2)$$

where Θ is the Heaviside function, $\rho(D)$ is the detection statistic used to select events for the catalog with threshold ρ_{thr} , and

$$\begin{aligned} \mathcal{E}(\Lambda) &= \int d\theta p(\theta|\Lambda) \int dD p(D|\theta)\Theta(\rho(D) \geq \rho_{\text{thr}}) \\ &= \int d\theta p(\theta|\Lambda) P(\text{det}|\theta) \\ &= P(\text{det}|\Lambda) \end{aligned} \quad (7.3)$$

is the expected fraction of events that are detectable within a population. Note that Eq. 7.2 is neither a likelihood function nor a posterior distribution, but instead is a joint distribution for both the data and the parameters. Furthermore, if we assume a uniform-in-log prior for

the rate ($p(R) \sim 1/R$) and marginalize, we obtain

$$p(\{D_i\}; \{\theta_i\}; \Lambda) = p(\Lambda) \prod_{i=1}^N \left[\frac{p(D_i|\theta_i)p(\theta_i|\Lambda)\Theta(\rho(D_i) \geq \rho_{\text{thr}})}{\mathcal{E}(\Lambda)} \right] \quad (7.4)$$

We note that the Θ in each factor within the product does not impact the inference as it is guaranteed to be one; the data is axiomatically detectable for detected events. However, these factors are important in what follows.

If we wish to construct a posterior for only Λ , we marginalize over $\{\theta_i\}$ and condition on the observed data to obtain

$$p(\Lambda|\{D_i\}) \propto p(\Lambda) \prod_{i=1}^N \left[\frac{\int d\theta_i p(D_i|\theta_i)p(\theta_i|\Lambda)}{\mathcal{E}(\Lambda)} \right]. \quad (7.5)$$

All these expressions have become commonplace within the GW community, and we refer the reader to the many reviews in the literature for more details [e.g., Mandel et al., 2019, Thrane and Talbot, 2019, Vitale et al., 2020].

We are particularly interested in the effect that subsets of events have on the inferred population, and whether those effects can be used to determine if particular events are outliers. However, if we simply remove potential outliers, we may bias the inferred population by preferentially removing the most extreme events. This, in turn, may artificially inflate the significance of any changes in the inferred population. As such, we introduce a coarse-graining procedure to replace traditional “leave-one-out” analyses. This procedure retains only a limited amount of information about a particular event. In this way, the analysis can naturally account for how the event-of-interest was selected (e.g., as the most extreme event in some dimension) while still discarding as much information about that event as possible.

To wit, let us assume that the j^{th} event is almost certainly within some region of param-

eter space $\mathcal{S}$ so that

$$\int d\theta_j p(\theta_j|D_j, \Lambda) \Theta(\theta_j \in \mathcal{S}) \geq 1 - \epsilon \quad \forall \quad \{\Lambda : p(\Lambda) \neq 0\}. \quad (7.6)$$

In other words, for any population described by Λ with nonvanishing support in the hyperprior $p(\Lambda)$, the inferred posterior on the event parameters θ_j is contained within the region $\mathcal{S}$ with high $(1 - \epsilon)$ credibility. We are typically interested in the case $\epsilon \ll 1$, which is to say we are confident that event j is contained within $\mathcal{S}$. There is a uniquely defined smallest $\mathcal{S}$ that satisfies this requirement, which in general depends on the choice of ϵ (smaller ϵ require larger $\mathcal{S}$). Additionally, we only assert that $\mathcal{S}$ is at least as large as the minimal region; it may be larger. Choosing a larger region corresponds to retaining less information about the j^{th} event.

At times, it is possible to identify individual events that are clearly separated from the rest of the population, in which case it may be straightforward to identify an appropriate choice for $\mathcal{S}$. An example could be defining $\mathcal{S}$ as the region with masses smaller than the minimum mass observed in a set. However, more complicated boundaries are possible, as hypersurfaces in multi-dimensional spaces could divide events along a nontrivial slice that depends on multiple parameters simultaneously. For example, we may define $\mathcal{S}$ to be a region with a secondary mass or mass ratio smaller than the corresponding minima observed in a set (see Sec. 7.3.1). We discuss several procedures to choose $\mathcal{S}$ in Sec. 7.2.2.

Of course, an extremal event is not necessarily inconsistent with the population determined by the other $N - 1$ events; some event is always the most extreme of any set. Indeed, the j^{th} event may still be drawn from the same $p(\theta|\Lambda_0)$ and simply be the most extreme example in the catalog. We therefore test the null-hypothesis that all N events are drawn from the same distribution by comparing the inferred population when we include all N events to the population inferred from $N - 1$ events while accounting for the knowledge that $\theta_j \in \mathcal{S}$. If the population inferred from the coarse-grained analysis is inconsistent with the full dataset,

we reject the null hypothesis that the j^{th} event is drawn from the same distribution as the other $N - 1$ events. With this method, we can be assured that we do not overestimate the significance of differences in the inferred populations because we include knowledge that the j^{th} event was selected in a particular way.

We now consider how to self-consistently limit the information about the j^{th} event within our inference. Because $\theta_j \in \mathcal{S}$ almost surely ($\epsilon \ll 1$), we can add a term to Eq. 7.4 without affecting the overall inference for Λ :

$$p(\{D_i\}; \{\theta_i\}; \Lambda) \simeq p(\Lambda) \left(\frac{p(D_j|\theta_j)p(\theta_j|\Lambda)\Theta(\rho(D_j) \geq \rho_{\text{thr}})\Theta(\theta_j \in \mathcal{S})}{\mathcal{E}(\Lambda)} \right) \quad (7.7)$$

$$\times \prod_{i \neq j}^{N-1} \left[\frac{p(D_i|\theta_i)p(\theta_i|\Lambda)\Theta(\rho(D_i) \geq \rho_{\text{thr}})}{\mathcal{E}(\Lambda)} \right] \quad (7.8)$$

where we have included an additional factor of $\Theta(\theta_j \in \mathcal{S})$ since this is one almost everywhere there is posterior support for θ_j (to within the precision specified in Eq. 7.6). This is completely analogous to the way in which terms representing the detectability of data, i.e., $\Theta(\rho(D) \geq \rho_{\text{thr}})$, are always one for the detected events. We have also replaced “=” with “ $\simeq$ ” to denote that strict equality only holds in the limit $\epsilon \rightarrow 0$. However, in what follows, we assume that ϵ is small enough to be negligible and retain the “=” in what follows.

Furthermore, the only information we wish to retain about the j^{th} event is that it almost certainly is within $\mathcal{S}$, and therefore we marginalize over both D_j and θ_j to effectively forget everything else about the event. This yields

$$\int dD_j d\theta_j p(D_j|\theta_j)p(\theta_j|\Lambda)\Theta(\rho(D_j) \geq \rho_{\text{thr}})\Theta(\theta_j \in \mathcal{S}) \quad (7.9)$$

$$\begin{aligned} &= \int d\theta_j \Theta(\theta_j \in \mathcal{S})p(\theta_j|\Lambda) \int dD_j p(D_j|\theta_j)\Theta(\rho(D_j) \geq \rho_{\text{thr}}) \\ &= \int d\theta_j \Theta(\theta_j \in \mathcal{S})p(\theta_j|\Lambda)P(\text{det}|\theta_j) \\ &= P(\text{det}, \theta_j \in \mathcal{S}|\Lambda), \end{aligned} \quad (7.10)$$

which is just the probability that the event was detected and had parameters within $\mathcal{S}$ given the underlying population model. We additionally note that this term only appears in Eq. 7.7 within a ratio, and the divisor of that ratio is $\mathcal{E}(\Lambda) = P(\text{det}|\Lambda)$. Therefore the coarse-graining procedure yields an overall factor of

$$\frac{P(\text{det}, \theta_j \in \mathcal{S}|\Lambda)}{P(\text{det}|\Lambda)} = P(\theta_j \in \mathcal{S}|\text{det}, \Lambda). \quad (7.11)$$

This has the appealing interpretation that the only information retained about the j^{th} event is the probability that it belongs to a particular part of parameter space (our “coarse grain”) given that it was detected and came from a particular population.² We discuss implications and implicit assumptions made by the choice of $\mathcal{S}$ in more detail in Sec. 7.2.2. Briefly, we note that most “agnostic” procedures for defining $\mathcal{S}$ will not depend on D_j or θ_j , and so we explicitly write $\mathcal{S} = \mathcal{S}(\{\theta_{i \neq j}\})$ below.

Putting everything together, we marginalize the coarse-grained likelihood over $\{\theta_{i \neq j}\}$ and condition on $\{D_{i \neq j}\}$ to obtain a coarse-grained posterior for Λ .

$$p(\Lambda|\{D_{i \neq j}\}, \rho(D_j) \geq \rho_{\text{thr}}; \theta_j \in \mathcal{S}(\{\theta_{i \neq j}\})) \propto \quad (7.12)$$

$$p(\Lambda) \int \left(\left[\prod_{i \neq j}^{N-1} d\theta_i \right] \left[\prod_{k \neq j}^{N-1} \frac{p(D_k|\theta_k)p(\theta_k|\Lambda)}{\mathcal{E}(\Lambda)} \right] P(\theta_j \in \mathcal{S}(\{\theta_{i \neq j}\})|\text{det}, \Lambda) \right) \quad (7.13)$$

We note that the marginalization over $\{\theta_{i \neq j}\}$ does not factor as nicely as it does in Eq. 7.5 because $\mathcal{S}(\{\theta_{i \neq j}\})$ may depend on all the events (except the j^{th} event) and is included within the integral. Methods to calculate Eq. 7.12 efficiently when one already has access to samples from $p(\Lambda|\{D_{i \neq j}\})$ or $p(\Lambda|\{D_i\})$ are discussed in Section 7.5.

2. Note that Eq. 7.11 refers to the probability that a detected event from a population came from $\mathcal{S}$, whereas Eq. 7.6 refers to the probability that a specific event (the j^{th} observed event, which produced data D_j) came from $\mathcal{S}$.

7.2.2 Implications from the choice of $\mathcal{S}$

Sec. 7.2.1 presents the coarse-grained inference for Λ based on knowledge that $\theta_j \in \mathcal{S}$. This holds for arbitrary $\mathcal{S}$ as long as Eq. 7.6 is satisfied; the choice of $\mathcal{S}$ is up to the analyst. We now consider the implications of different choices for $\mathcal{S}$ and what assumptions they represent.

To begin, if θ_j is well determined by D_j and the minimum allowable $\mathcal{S}$ is small (particularly in comparison to the extent of the prior $p(\theta|\Lambda)$), the coarse-graining procedure can still retain most, if not all, the relevant information about the j^{th} event. In this way, defining $\mathcal{S}$ in relation to the minimum allowable $\mathcal{S}$ implicitly assumes some knowledge of D_j , which is undesirable for a leave-one-out analysis. To avoid this, we instead recommend choosing $\mathcal{S}$ based on only *a priori* theoretical predictions of astrophysical interest or the data from the $N - 1$ other events. In general, this implies that one should write $\mathcal{S} = \mathcal{S}(\{\theta_{i \neq j}\})$ to explicitly show that it does not depend on either the data or parameters from the j^{th} event, as we do within Eq. 7.12.

One possible data-driven procedure for choosing $\mathcal{S}$ is particularly appealing in the special case where the parameters of the j^{th} event are cleanly separated from the rest of the events. If the j^{th} event is the most extreme event in some dimension x , we can simply define $\mathcal{S}(\{x_{i \neq j}\}) : x \geq \max_{i \neq j} \{x_i\}$ as the region where x is larger than the second-largest detected event.³ This choice is natural in the sense that it only depends on $\{D_{i \neq j}\}$ and is as generous as possible subject to the knowledge that the j^{th} event is extremal; it does not retain any information about how much larger x_j is than $\max_{i \neq j} \{x_i\}$.

Such a choice allows an analyst to pose questions such as, “how big should the largest individual event in a catalog of N events be given the observation of $N - 1$ events and the knowledge that one was larger?” Comparing the predicted largest event to what was actually observed naturally defines a metric that can be used as a quantitative consistency check (see Sec. 7.2.3). Indeed, the choice of $\mathcal{S}$ precisely specifies the information used when computing

3. One could equivalently consider the smallest event.

p -values for the null hypothesis that all N events in a catalog were drawn from the same population. Nonetheless, it is important to remember that this procedure is not unique, just as the definition of a null hypothesis is not unique, and other choices of $\mathcal{S}$ may be able to more naturally answer other questions.

In this vein, one might also consider choosing $\mathcal{S}$ based on prior theoretical motivations, so long as the selected region is compatible with Eq. 7.6. This can provide a perfectly natural way to define alternative tests of the null hypothesis and is particularly relevant when the events are not cleanly separated. For example, even before observing any data, we may identify the region $m_1 > 50 M_\odot$ as interesting from the standpoint of PISN theory, and scrutinize any events that fall in this region as potential outliers.

We note that the inability to define $\mathcal{S}$ based on the next-most extreme event only arises when the j^{th} event is not cleanly separated from the other events, and therefore it is not unambiguously the most extreme event. One may take that ambiguity itself as evidence that the event cannot be an outlier, and therefore argue that our machinery may not be needed in such situations. Nonetheless, we can still construct a perfectly self-consistent coarse-grained inference even when the minimal $\mathcal{S}$ surrounding the j^{th} event’s parameters overlaps significantly with the inferred parameters of the other $N - 1$ events. The coarse-grained event need not actually be extremal.

As a limiting case, we note that choosing $\mathcal{S}$ that spans the entire parameter space ($\mathcal{S} \rightarrow \mathcal{R}^m$) implies that $P(\theta_j \in \mathcal{S} | \text{det}, \Lambda) \rightarrow 1 \forall \Lambda$. That is, if we know nothing about the parameters of the j^{th} event, then the coarse-grained inference is equivalent to neglecting that event altogether and performing the “standard” inference for Λ with the other $N - 1$ events.⁴ This is intuitive in that, if we picked an event to coarse-grain at random without first considering its parameters, then the smallest allowable $\mathcal{S}$ that was certain to contain the event would be the entire (detectable) parameter space. At the same time, excluding an

4. The corresponding inference of the rate would still correctly account for the fact that we detected N events in total, which is not necessarily true of other leave-one-out analyses.

event at random should not introduce any biases within the inference for Λ , and one expects to perform the standard inference with only $N - 1$ events. We see, then, that performing a standard inference with $N - 1$ events after excluding an extremal event, *without* incorporating the knowledge that the excluded event was extremal, is not a self-consistent procedure because it does not correctly reflect the data selection procedure. This inconsistency leads to biases.

7.2.3 *Quantifying inconsistencies with inferred populations*

In order to test the null hypothesis that the event in question is consistent with the $N - 1$ other events, we follow a similar procedure to Fishbach et al. [2020b] and construct a p -value statistic. We marginalize over both the uncertainty in the population hyperparameters, inferred from the coarse-grained inference, and the measurement uncertainty in the parameters of the N observations. For each hyperposterior sample Λ from the coarse-grained inference, we

- i) draw N synthetic detected events from the corresponding population described by those hyperparameters,
- ii) reweigh the N observed events' posteriors for $\{\theta_i\}$ to what would be obtained under the population prior described by that Λ . For the $N - 1$ events that were not excluded, this recovers their marginal posterior for $\{\theta_{i \neq j}\}$ under the joint distribution of Eq. 7.7, while for the excluded event we obtain the population-informed posterior for θ_j ,
- iii) draw one θ_{obs} sample for each observed event, and
- iv) compare the most extreme θ_{syn} out of the N synthetic events to the most extreme θ_{obs} .

Repeating this procedure for many hyperposterior samples Λ , the p -value is obtained as the fraction of synthetic catalogs that produce an event that is at least as extreme as the most

extreme observed event. Although such statistics are not necessarily uniformly distributed even under the null hypothesis (see, e.g., Seth et al. [2019]), if the resulting marginalized p -value is small, we reject the null hypothesis and conclude that the excluded event is inconsistent with the main population.

Fishbach et al. [2020b] estimated p -values by scattering the maximum likelihood estimates by synthetic noise realizations and then comparing the maximum likelihood from the observed event to the synthetic distribution under the null hypothesis. Our approach differs in that we do not scatter the maximum likelihood estimate of our synthetic detections, but instead use the true values of the detected events. However, we still marginalize over the actual uncertainty in the observed events’ true parameters stemming from detector noise. Both procedures, then, account for measurement uncertainty but answer slightly different questions.

7.3 Astrophysical results

Using the coarse-grained inference described in Sec. 7.2, we revisit several astrophysical events from GWTC-2. We are specifically interested in evidence that individual events are incompatible with the main BBH population (the phenomenological distribution that describes the majority of detected BBH systems) inferred in Abbott et al. [2021d]. We consider GW190814 (Sec. 7.3.1), GW190412 (Sec. 7.3.2), and GW190521 (Sec. 7.3.3) in turn. Our analysis reweighs publicly available population hyperposteriors samples [The LIGO Scientific Collaboration and The Virgo Collaboration, 2020a].

In what follows, we focus on results with the POWERLAW+PEAK mass model from Talbot and Thrane [2018], Abbott et al. [2021d]. Unless otherwise noted (i.e., GW190521 in Sec. 7.3.3), astrophysical conclusions are unchanged when we assume different mass models. Furthermore, we also consider fixed $\mathcal{S}$ in each case. Specific choices for $\mathcal{S}$ are listed in each section and are either motivated by the other $N - 1$ events or theoretical expectations for

astrophysical systems.

7.3.1 *GW190814 is an outlier in secondary mass*

We begin by considering GW190814 [Abbott et al., 2020e, The LIGO Scientific Collaboration and The Virgo Collaboration, 2020c], the BBH system with the smallest secondary mass and the most extreme mass ratio observed to date. Indeed, GW190814’s secondary is so small ($m_2 \sim 2.6 M_\odot$) that there has been significant discussion about whether it could have been a NS [see, e.g., Essick and Landry, 2020], with the common consensus that the system is likely incompatible with a slowly-rotating NS. For simplicity, we eschew the question of whether both components of GW190814 were actually BHs, and instead focus on whether their masses are compatible with the distribution inferred from the rest of the BBH events in GWTC-2.

We adopt the mass models explored in [Abbott et al., 2021d], all of which include a cut-off at low masses with variable degrees of sharpness.

We define $\mathcal{S}$ for our analysis of GW190814 as follows

$$\mathcal{S}_{\text{GW190814}} : (m_2 \leq 5.0 M_\odot) \text{ OR } (q \leq 0.28). \quad (7.14)$$

The $m_2 \leq 5.0 M_\odot$ boundary is chosen to match the median posterior estimate of GW190924, the event with the second smallest secondary mass after GW190814. The $q \leq 0.28$ boundary is chosen to match the median posterior estimate of GW190412, the event with the second smallest mass ratio after GW190814. Both GW190924 and GW190814 are highlighted in blue in Fig. 7.1. This defines an “L-shaped” region in the (m_2, q) plane spanning the lowest values for both dimensions (see Fig. 7.1). We again note that this choice of $\mathcal{S}$ is not unique, and one could instead choose to define $\mathcal{S}$ with bounds on only m_2 or only q . Defining $\mathcal{S}_{\text{GW190814}}$ in terms of only m_2 or only q does not affect our conclusions, and defining $\mathcal{S}_{\text{GW190814}}$ in this

way allows us to be as agnostic as possible about GW190814 in some sense. Importantly, we note that, with this definition of $\mathcal{S}_{\text{GW190814}}$, population models that do not have support for masses below $m_2 \leq 5.0 M_\odot$ are still allowed as long as they support $q \leq 0.28$, and *vice versa*.

Fig. 7.1 summarizes our conclusions. Specifically, we see that the $(N - 1)$ -event coarse-grained hyperposterior resembles the full N -event hyperposterior for β_q , the hyperparameter that controls the extent of the population model for q , while it more closely resembles the $(N - 1)$ -event hyperposterior that neglects the coarse-grained correction for both $m_{\min}$ and δ_m , which control the minimum mass allowed in the population. We note that $m_{\min}$ sets an absolute lower bound for the allowed masses within a population, and therefore we would expect a reasonable amount of probability that $m_{\min} \leq 3 M_\odot$ in the $(N - 1)$ -event analyses if GW190814 was consistent with the population inferred from the other events. While $m_{\min} \leq 3 M_\odot$ is not excluded by the $(N - 1)$ -event analysis, it is not particularly favored, either. This suggests that *GW190814's small q is consistent with the rest of the events*, as population models already contain plenty of support for small q , but *GW190814 is inconsistent with the rest of the events because of its small m_2* .

We further quantify this by estimating a p -value that the smallest event out of an N -event catalog would have m_2 less than or equal to any event in GWTC-2, given the $(N - 1)$ -event coarse-grained hyperposterior. We find $P \leq 0.056\%$ at 90% confidence.⁵ As such, we reject the null hypothesis that GW190814 was drawn from the same population as the other $N - 1$ events. Abbott et al. [2021d] reach the same conclusion, and, following their example, we exclude GW190814 from the catalog as we explore whether other individual events are inconsistent with the remaining detections.

5. We can only bound the p -value from above as we did not find a single instance where synthetic catalogs generated $\min\{m_2\}$ smaller than GW190814's secondary after ~ 5000 trials. Similarly, we find $P \leq 0.024\%$ with the $(N - 1)$ -event analysis that neglects coarse graining.

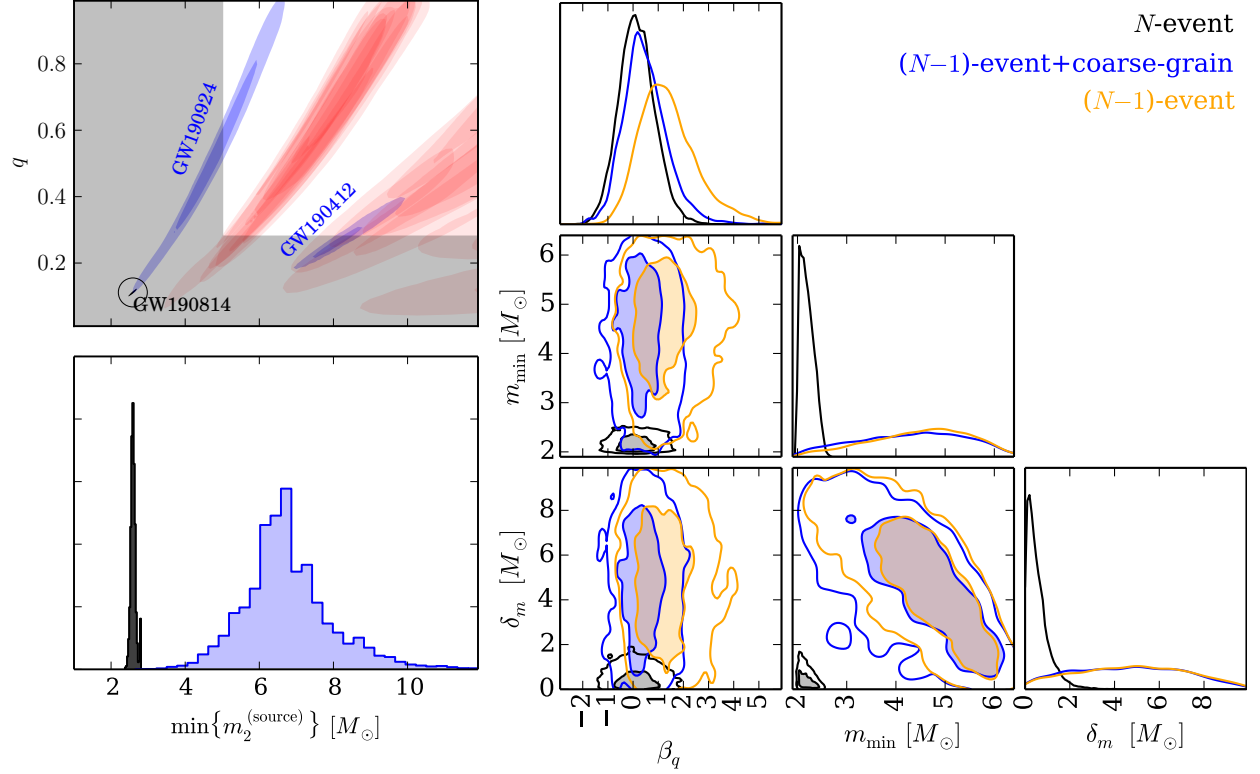


Figure 7.1: (*top left*) Depiction of $\mathcal{S}_{\text{GW190814}}$ (*shaded region*, Eq. 7.14), chosen to be the region ($m_2 \leq 5.0 M_\odot$) or ($q \leq 0.28$) based on the approximate median m_2 and q of GW190924 and GW190412, respectively. Colored shaded regions show 50% and 90% highest-probability-density credible regions from each event assuming flat priors in component masses and uniform priors in comoving volume for (*black ellipse*, circled in lower left corner) GW190814, (*blue*) GW190924 and GW190412, as well as (*red*) all other BBH systems considered in Abbott et al. [2021d]. (*bottom left*) Distribution of (*black*) GW190814's m_2 and (*blue*) the smallest expected m_2 in catalogs of 45 events based on the $(N - 1)$ -event coarse-grained hyperposterior for the POWERLAW+PEAK mass model. (*right*) Hyperposteriors inferred with different amounts of information about GW190814. Contours in the joint distributions denote 50% and 90% credible regions.

7.3.2 *GW190412 is not an outlier*

We next investigate GW190412 [Abbott et al., 2020c], which, with $q \sim 0.28$, is the only event in GWTC-2 besides GW190814 that is inconsistent with $q = 1$. As a reminder, we have already removed GW190814 from the set of events against which we compare GW190412. In this case, we define

$$\mathcal{S}_{\text{GW190412}} : q \leq 0.8. \quad (7.15)$$

as a conservative boundary for systems that have asymmetric masses. Fishbach and Holz [2020b] estimate that 90% (99%) of detected BBHs will have $q \geq 0.73$ (0.51) based on population models of the 10 BBH events in GWTC-1. Our choice for $\mathcal{S}_{\text{GW190412}}$ is even more conservative: we give the coarse-grained inference very little information about GW190412 itself, and will therefore most easily identify it as an outlier. Fig. 7.2 demonstrates the results.

Again, we are primarily interested in the low-mass (and low mass-ratio) behavior of the model and focus on the inferred values of β_q , $m_{\min}$, and δ_m . In this case, we find that all inferences agree remarkably well for $m_{\min}$ and δ_m , which is unsurprising since GW190412’s masses are not particularly extreme when considered individually. However, we observe better agreement between the N -event and $(N - 1)$ -event coarse-grained hyperposteriors for β_q than between either and the $(N - 1)$ -event inference that neglects coarse-graining. Excluding the event with smallest q shifts the inferred hyperposterior towards values that favor equal-mass systems (larger β_q).

The coarse-grained inference produces a p -value of $P = 22\%$ ⁶ for the smallest observed q to be as small or smaller than that of any BBH event in GWTC-2 (excluding GW190814). We therefore conclude that *GW190412 is consistent with the population inferred from rest of the BBH events within GWTC-2* (except GW190814). It is simply the event with the most

6. The $(N - 1)$ -event analysis without coarse graining yields $P = 19\%$.

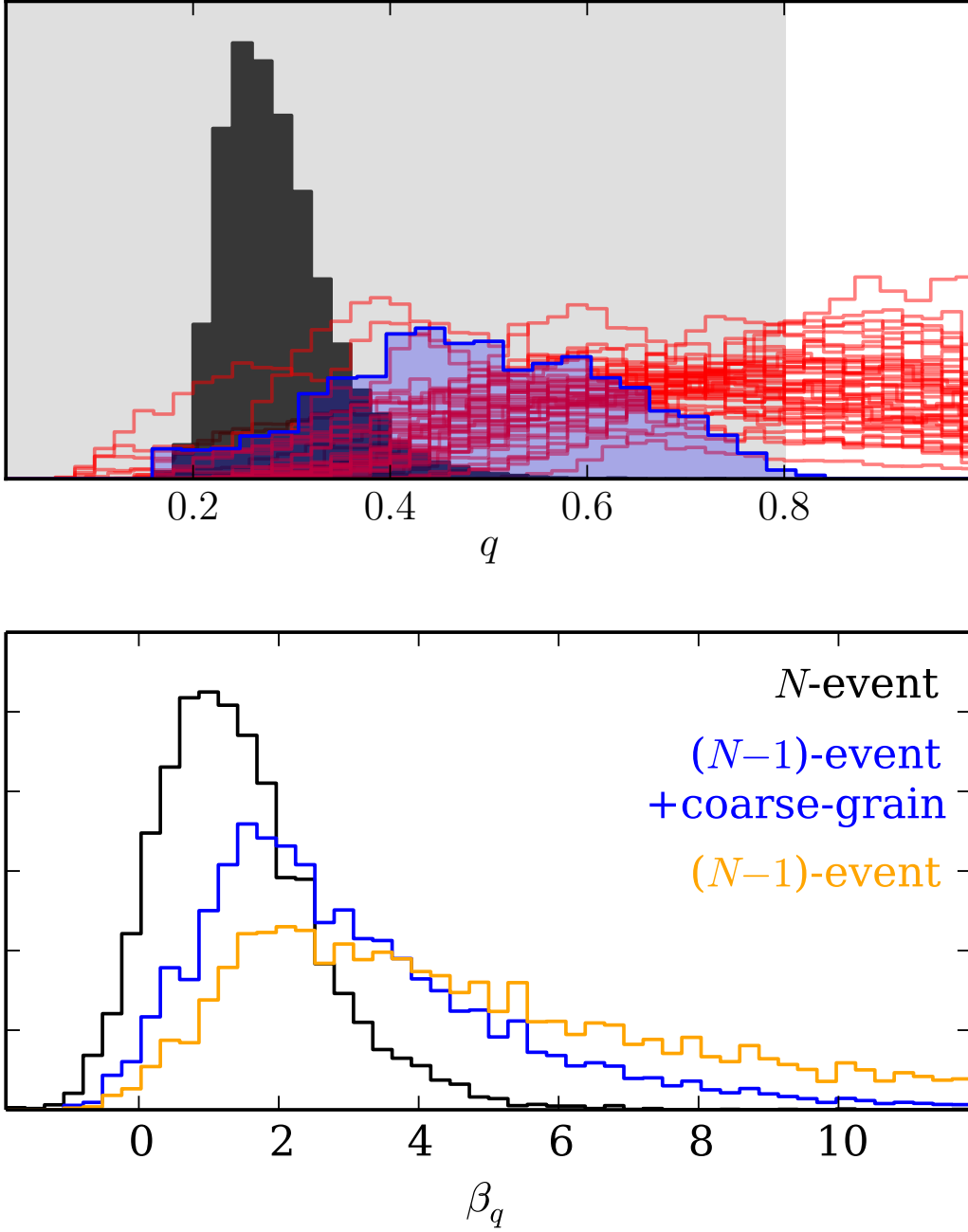


Figure 7.2: (*top*) Depiction of (*shaded grey*) $\mathcal{S}_{\text{GW190412}}$ (Eq. 7.15) with (*shaded black*) GW190412 and (*red*) all other events considered in Abbott et al. [2021d] (except GW190814) with flat priors on component masses as well as (*blue*) the prediction for the smallest q out of 44 events based on the $(N-1)$ -event coarse-grained hyperposterior with the POW-ERLAW+PEAK mass model. (*bottom*) Hyperposteriors inferred with different amounts of information about GW190412.

extreme q from that population, in agreement with both Abbott et al. [2020c] and Abbott et al. [2021d].

7.3.3 *GW190521 is not an outlier*

Finally, we consider GW190521 [Abbott et al., 2020d,f, The LIGO Scientific Collaboration and The Virgo Collaboration, 2020b]. This event is remarkable for its large component masses, although we note that it does not unambiguously have the largest component masses of any system in GWTC-2; see Fig. 7.3. In this case, we cannot easily define $\mathcal{S}_{\text{GW190521}}$ in terms of the other events in the catalog while guaranteeing that Eq. 7.6 is satisfied. However, we note that GW190521 is of particular interest because its component masses nominally fall within the PISN mass gap.⁷ In this case, it is natural to define $\mathcal{S}_{\text{GW190521}}$ in terms of an approximate boundary defining the PISN mass gap. We take

$$\mathcal{S}_{\text{GW190521}} : m_1 \geq 50 M_{\odot} \quad (7.16)$$

as a reasonable approximation, although others have considered values as large as $65 M_{\odot}$ [Abbott et al., 2020d,f, O’Brien et al., 2021]. Again, there is nontrivial overlap between this choice for $\mathcal{S}_{\text{GW190521}}$ and the parameters inferred for several other events in GWTC-2, but this does not affect our inference.

Contrary to GW190814 and GW190412, in which higher-order modes were observed and their asymmetric mass ratios were relatively well measured, GW190521’s component mass posterior is quite broad. As such, simultaneous population inference can have a significant impact on the system’s inferred properties. For this reason, Fig. 7.3 shows the posteriors of primary masses under the POWERLAW+PEAK mass model as inferred with our coarse-grained analysis. Nonetheless, several analyses have shown that it is likely that at least one

7. See, e.g., Fishbach and Holz [2020a] for alternative interpretations with component masses that straddle the mass gap.

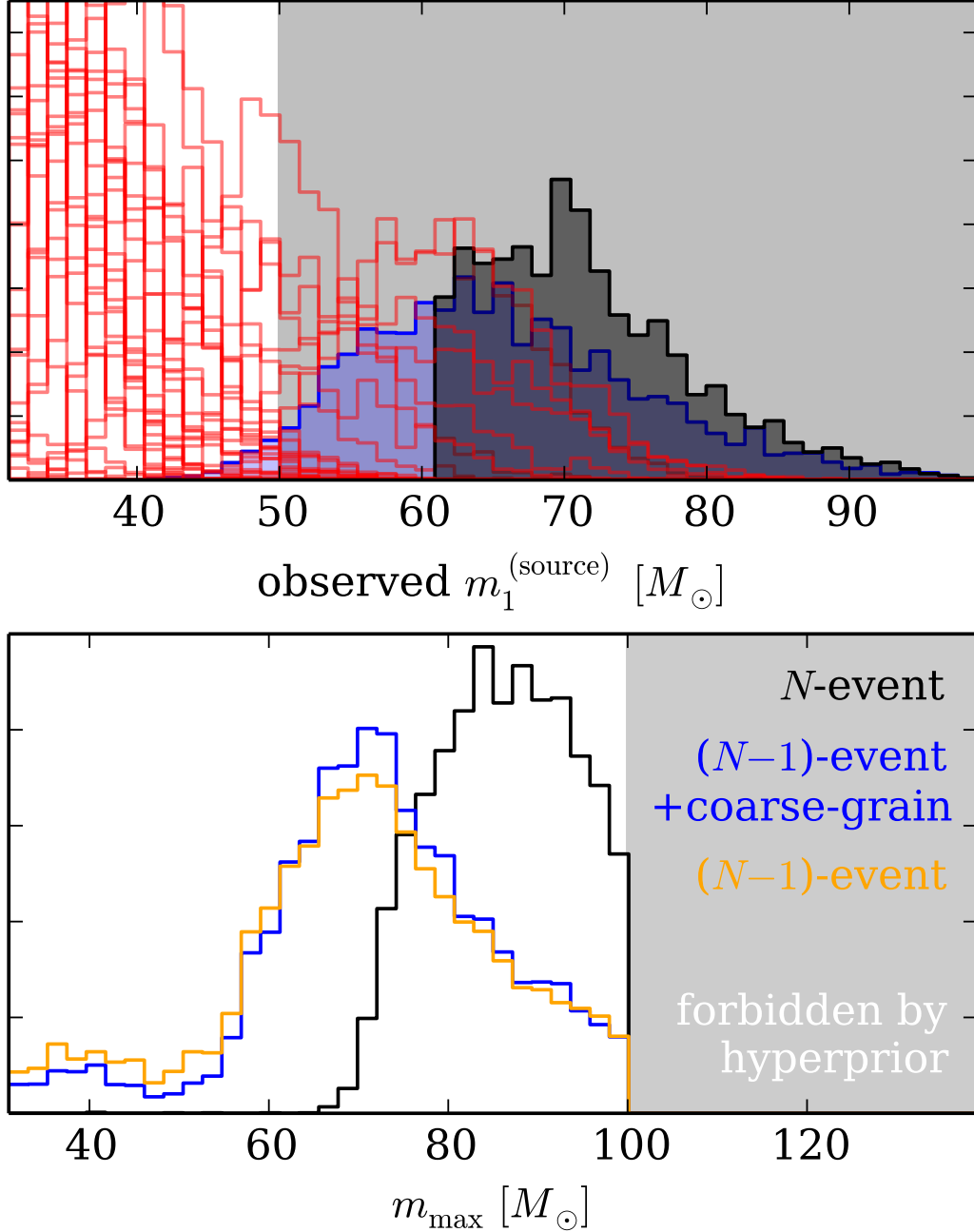


Figure 7.3: (*top*) Depiction of (*shaded grey*) $\mathcal{S}_{\text{GW190521}}$ (Eq. 7.16) with (*black shaded*) GW190521 and (*red*) all other events considered in [Abbott et al., 2021d] informed by the POWERLAW+PEAK mass model as inferred under the $(N-1)$ -event coarse-grained hyperposterior as well as the distribution of the largest predicted m_1 out of 44 events based on the $(N-1)$ -event coarse-grained hyperposterior. The hard boundary in GW190521’s distribution at $m_1 \sim 60 M_\odot$ is due to our labeling convention ($m_1 \geq m_2$) and approximately corresponds to equal-mass systems. (*bottom*) Hyperposteriors inferred with different amounts of information about GW190521. Note that m_{max} extends below $50 M_\odot$ because it only limits the power law part of POWERLAW+PEAK; the peak still has support at higher masses.

component of GW190521 had a mass $\geq 65 M_{\odot}$ [Abbott et al., 2020d,f, O’Brien et al., 2021].

Some authors have taken this to mean that GW190521 is incompatible with the population of stellar-mass BHs, but to be clear they often mean the inferred phenomenological fit from other events with an additional cut-off (not included in the fit). We are concerned with the simpler question of whether GW190521 is inconsistent with the phenomenological population model inferred from the other BBH events without imposing additional sharp cut-offs. We are not concerned with whether the current phenomenological models are compatible with PISN predictions, but instead ask whether the models are sufficient to describe the overall mass distribution and whether GW190521 is consistent with models inferred from the other events.

While our coarse-grained hyperposterior at times more closely resembles the $(N-1)$ -event hyperposterior than the N -event hyperposterior (bottom panel of Fig. 7.3), we nonetheless obtain a p -value of $P = 20\%$ that synthetic catalogs would contain a more massive event than has been observed so far. In fact, if we neglect the coarse-graining correction, we still obtain $P = 15\%$ under the POWER LAW + PEAK model. As such, and in agreement with Abbott et al. [2021d], we conclude that GW190521 is consistent with the overall mass distribution inferred from the rest of the BBH population.

One might expect other mass models, like Abbott et al. [2021d]’s TRUNCATED model with a sharp cut-off at high masses, to be less consistent with GW190521. To investigate this, we repeat our coarse-grained analysis and find $P = 11\%$ under the TRUNCATED model. Indeed, even the $(N-1)$ -event TRUNCATED hyperposterior that neglects coarse-graining only yields $P = 6\%$. As such, we may conclude that, while GW190521 is certainly not expected to be common, it also is not unambiguously inconsistent with any of the phenomenological models considered in Abbott et al. [2021d], even the simplest TRUNCATED distribution.⁸

8. While GW190521 may not be inconsistent with the mass models considered in Abbott et al. [2021d], they point out that the simple TRUNCATED model is more broadly a bad fit to the data. It overpredicts the number of massive systems that should have been observed (see also Fishbach et al. [2021]). However, even minor modifications like the POWERLAW+PEAK and BROKEN POWERLAW distributions appear to remove

That being said, the number of BBH systems detected in GWTC-2 more than quadrupled compared to GWTC-1, and both Abbott et al. [2021d] and Fishbach and Holz [2020a] point out that GW190521 does seem to be inconsistent with the truncated mass distributions inferred based on only the 10 events in GWTC-1 [Abbott et al., 2019a]. While additional GW detections have continued to surprise, we are reminded that it is important to consider new events in the context of the full catalog before drawing conclusions based on individual (apparently) exceptional events and a subset of previously observed systems.

7.4 Discussion

Within any set of observed events drawn from an unknown population, one is often interested in determining whether new events are consistent with the population inferred from the existing set. This often involves careful examination of particular events because they are extremal in some way. It is remarkable that GW astronomy has already advanced to the point where such matters are of practical importance in only a half-dozen years since the first detection [Abbott et al., 2016]. Nonetheless, we show that current approaches to answer exactly this question, which have become commonplace within the GW community, can introduce biases as they do not account for the manner in which extremal events were identified for further study.

Our method allows analysts to explicitly account for how they selected extremal events within leave-one-out analyses, representing excluded events with coarse-grained likelihoods, and clearly identifying the need to select the size and placement of the coarse grains. While we note that the exact choice of how big to make those grains is to some degree arbitrary, just as the definition of a null hypothesis is to some degree arbitrary, we propose algorithmic ways to choose the most generous grains possible.

Finally, we note that the biases introduced by excluding extremal events without ac-

such tensions.

counting for how they were selected can be particularly severe when there are sharp features in the underlying population model (e.g., mass gaps). Therefore, one must take care when analyzing population models with sharp cut-offs and attempting to assess the significance of outliers after excluding extremal events. However, even in these severe cases, we find that p -values estimated from $(N - 1)$ -event hyperposteriors are typically biased by at most a factor of a few.

Our conclusions based on our coarse-grained analysis agree with those presented in Abbott et al. [2021d], even though they did not account for the coarse-grained correction and their leave-one-out analyses may have been biased. We find that *GW190814 is an outlier* because its secondary mass is too small to be consistent with the other events. *GW190412 is not an outlier*, as its small mass ratio is simply the most extreme example from the tail of the main population. We find that *GW190521 is not an outlier* under the preferred mass models explored in Abbott et al. [2021d], and is in only moderate tension with even the simplest truncated mass models.

We again note that any population analysis will eventually face the challenge of determining whether particular events are consistent with the population inferred from the rest of the events. This problem is not unique to GW astronomy. However, as catalogs continue to rapidly grow in size, this question has become increasingly relevant. We note that another large set of events is expected with the release of the second half of the LVK collaborations' third observing run (O3b). Indeed, given the breadth of physical phenomena that GW observations can probe, it is of the utmost importance to fully characterize outlier tests. We emphasize that many of the most interesting questions in GW astronomy are specifically focused on outliers, including extreme mass, mass ratio, and spin events, and the presence of events within the putative NS-BH and PISN mass gaps. Our analysis provides a controlled way to account for event selection when examining outliers with nearly trivial additional computational cost. This will enable the robust identification of novel subpopulations with-

out fear of biasing analyses towards artificially inflated significance estimates for potential outliers.

7.5 Additional details on reweighing existing hyperposteriors

We note that Eq. 7.12 could be computationally expensive if we allow $\mathcal{S}$ to depend on the parameters of the other events. However, if $\mathcal{S}$ does not depend on the parameters of the other events, the marginalization becomes trivial. There are also significant redundancies between Eq. 7.5 and Eq. 7.12, which reduce the computational cost of reweighing existing samples and make implementing the coarse-grained likelihood a simple extension of existing likelihoods.

There are two corrections to the likelihood that may be of interest. When we have already analyzed the full set of N events and want to conduct a coarse-grained analysis *post hoc*, we note that

$$\frac{p(\Lambda|\{D_{i \neq j}\}, \rho(D_j) \geq \rho_{\text{thr}}; \theta_j \in \mathcal{S})}{p(\Lambda|\{D_i\})} \propto \frac{\int d\theta p(\theta|\Lambda) P(\text{det}|\theta) \Theta(\theta \in \mathcal{S}(\{\theta_{i \neq j}\}))}{\int d\theta_j p(D_j|\theta_j) p(\theta_j|\Lambda)} = \frac{P(\theta \in \mathcal{S}, \text{det}|\Lambda)}{p(D_j|\Lambda)}. \quad (7.17)$$

When we instead have hyperposterior samples from an $(N - 1)$ -event analysis that omitted the potential outlier, we can write

$$\frac{p(\Lambda|\{D_{i \neq j}\}, \rho(D_j) \geq \rho_{\text{thr}}; \theta_j \in \mathcal{S})}{p(\Lambda|\{D_{i \neq j}\})} \propto P(\theta \in \mathcal{S}|\text{det}; \Lambda). \quad (7.18)$$

Weighing existing hyperposterior samples by either Eq. 7.17 or 7.18, as appropriate, allows us to estimate the coarse-grained hyperposterior and quickly perform consistency tests of our null hypothesis. Such reweighing procedures are equivalent to directly sampling from Eq. 7.12 as long as enough effective samples remain to provide reliable estimates of the hyperposterior. As a demonstration, we repeat the analysis of Sec. 7.3.1 by directly sampling from Eq. 7.12,

obtaining equivalent results to the reweighed hyperposterior samples (see Fig. 7.4).

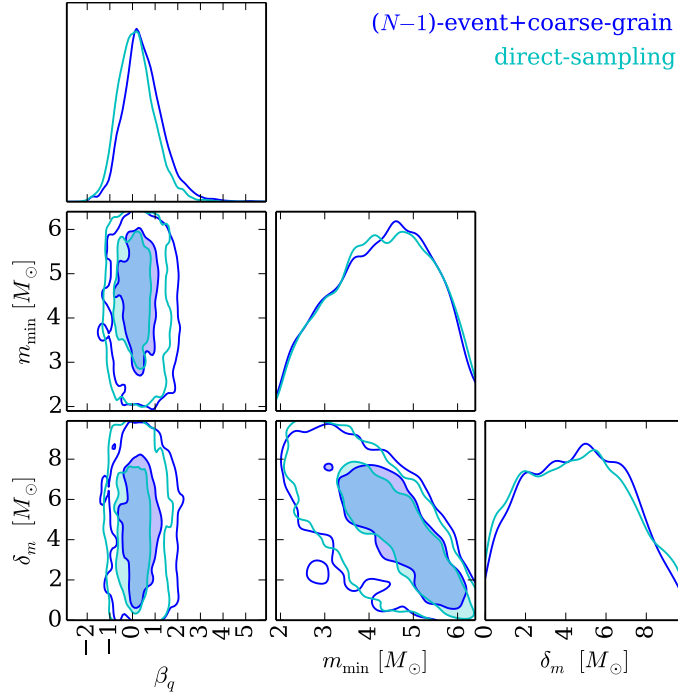


Figure 7.4: Comparison between reweighed samples and direct sampling for GW190814’s $(N-1)$ +coarse-grained hyperposterior (Sec. 7.3.1). Contours in the joint distributions denote 50% and 90% credible regions. Reweighed samples were obtained by applying Eq. 7.18 to public hyperposterior samples from an $(N-1)$ -event analysis that excluded GW190814

REFERENCES

- J. Aasi, B. P. Abbott, R. Abbott, T. Abbott, et al. Advanced LIGO. *Classical and Quantum Gravity*, 32(7):074001, March 2015. doi:10.1088/0264-9381/32/7/074001. URL <https://doi.org/10.1088/0264-9381/32/7/074001>. Publisher: IOP Publishing.
- B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, et al. Observation of Gravitational Waves from a Binary Black Hole Merger. *Phys. Rev. Lett.*, 116(6):061102, February 2016. doi:10.1103/PhysRevLett.116.061102. URL <https://link.aps.org/doi/10.1103/PhysRevLett.116.061102>. Publisher: American Physical Society.
- B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, et al. A gravitational-wave standard siren measurement of the Hubble constant. *Nature*, 551:85–88, November 2017a. ISSN 0028-0836. doi:10.1038/nature24471. URL <https://ui.adsabs.harvard.edu/abs/2017Natur.551...85A>. ADS Bibcode: 2017Natur.551...85A.
- B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, et al. GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral. *Physical Review Letters*, 119(16):161101, October 2017b. doi:10.1103/PhysRevLett.119.161101. URL <https://link.aps.org/doi/10.1103/PhysRevLett.119.161101>. Publisher: American Physical Society.
- B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, et al. Multi-messenger Observations of a Binary Neutron Star Merger. *The Astrophysical Journal*, 848:L12, October 2017c. ISSN 0004-637X. doi:10.3847/2041-8213/aa91c9. URL <https://ui.adsabs.harvard.edu/abs/2017ApJ...848L..12A>. ADS Bibcode: 2017ApJ...848L..12A.
- B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, et al. Binary Black Hole Population Properties Inferred from the First and Second Observing Runs of Advanced LIGO and Advanced Virgo. *The Astrophysical Journal Letters*, 882(2):L24, September 2019a. ISSN 2041-8205. doi:10.3847/2041-8213/ab3800. URL <https://dx.doi.org/10.3847/2041-8213/ab3800>.
- B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, et al. GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs. *Physical Review X*, 9:031040, July 2019b. doi:10.1103/PhysRevX.9.031040. URL <https://ui.adsabs.harvard.edu/abs/2019PhRvX...9c1040A>. ADS Bibcode: 2019PhRvX...9c1040A.
- B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, et al. Low-latency Gravitational-wave Alerts for Multimessenger Astronomy during the Second Advanced LIGO and Virgo Observing Run. 875(2):161, April 2019c. ISSN 0004-637X. doi:10.3847/1538-4357/ab0e8f. URL <https://doi.org/10.3847/1538-4357/ab0e8f>. Publisher: American Astronomical Society.
- B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, et al. GW190425: Observation of a Compact Binary Coalescence with Total Mass $\sim 3.4 M_{\odot}$. *The Astrophysical*

- Journal*, 892(1):L3, March 2020a. ISSN 2041-8213. doi:10.3847/2041-8213/ab75f5. URL <http://arxiv.org/abs/2001.01761>. arXiv: 2001.01761.
- B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, et al. Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA. *Living Reviews in Relativity*, 23(1):3, September 2020b. ISSN 1433-8351. doi:10.1007/s41114-020-00026-9. URL <https://doi.org/10.1007/s41114-020-00026-9>.
- B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, et al. A Gravitational-wave Measurement of the Hubble Constant Following the Second Observing Run of Advanced LIGO and Virgo. *The Astrophysical Journal*, 909(2):218, March 2021a. ISSN 0004-637X. doi:10.3847/1538-4357/abdcbb7. URL <https://dx.doi.org/10.3847/1538-4357/abdcbb7>. Publisher: The American Astronomical Society.
- R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, et al. GW190412: Observation of a binary-black-hole coalescence with asymmetric masses. *Physical Review D*, 102:043015, August 2020c. ISSN 1550-79980556-2821. doi:10.1103/PhysRevD.102.043015. URL <https://ui.adsabs.harvard.edu/abs/2020PhRvD.102d3015A>. ADS Bibcode: 2020PhRvD.102d3015A.
- R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, et al. GW190521: A Binary Black Hole Merger with a Total Mass of 150 $M_{\odot}$? *Physical Review Letters*, 125:101102, September 2020d. ISSN 0031-9007. doi:10.1103/PhysRevLett.125.101102. URL <https://ui.adsabs.harvard.edu/abs/2020PhRvL.125j1102A>. ADS Bibcode: 2020PhRvL.125j1102A.
- R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, et al. GW190814: Gravitational Waves from the Coalescence of a 23 Solar Mass Black Hole with a 2.6 Solar Mass Compact Object. *The Astrophysical Journal*, 896(2):L44, June 2020e. doi:10.3847/2041-8213/ab960f. URL <https://doi.org/10.3847/2041-8213/ab960f>. Publisher: American Astronomical Society.
- R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, et al. Properties and Astrophysical Implications of the 150 $M_{\odot}$ Binary Black Hole Merger GW190521. *The Astrophysical Journal*, 900(1):L13, September 2020f. doi:10.3847/2041-8213/aba493. URL <https://doi.org/10.3847/2041-8213/aba493>. Publisher: American Astronomical Society.
- R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, et al. GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run. *Physical Review X*, 11(2):021053, June 2021b. ISSN 2160-3308. doi:10.1103/PhysRevX.11.021053. URL <http://arxiv.org/abs/2010.14527>. arXiv: 2010.14527.
- R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, et al. Observation of Gravitational Waves from Two Neutron Star-Black Hole Coalescences. *The Astrophysical Journal*, 915:L5, July 2021c. ISSN 0004-637X. doi:10.3847/2041-8213/ac082e. URL <https://ui.adsabs.harvard.edu/abs/2021ApJ...915L...5A>. ADS Bibcode: 2021ApJ...915L...5A.

- R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, et al. Population Properties of Compact Objects from the Second LIGO-Virgo Gravitational-Wave Transient Catalog. *The Astrophysical Journal*, 913:L7, May 2021d. ISSN 0004-637X. doi:10.3847/2041-8213/abe949. URL <https://ui.adsabs.harvard.edu/abs/2021ApJ...913L...7A>. ADS Bibcode: 2021ApJ...913L...7A.
- R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, et al. Search for Lensing Signatures in the Gravitational-Wave Observations from the First Half of LIGO-Virgo's Third Observing Run. *The Astrophysical Journal*, 923:14, December 2021e. ISSN 0004-637X. doi:10.3847/1538-4357/ac23db. URL <https://ui.adsabs.harvard.edu/abs/2021ApJ...923...14A>. ADS Bibcode: 2021ApJ...923...14A.
- R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, et al. GWTC-2.1: Deep Extended Catalog of Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run. *arXiv:2108.01045 [gr-qc]*, August 2021f. URL <http://arxiv.org/abs/2108.01045>. arXiv: 2108.01045.
- R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, et al. GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run. November 2021g. URL <https://ui.adsabs.harvard.edu/abs/2021arXiv211103606T>. ADS Bibcode: 2021arXiv211103606T Type: article.
- R. Abbott, H. Abe, F. Acernese, K. Ackley, et al. Tests of General Relativity with GWTC-3, December 2021h. URL <http://arxiv.org/abs/2112.06861>. arXiv:2112.06861 [astro-ph, physics:gr-qc, physics:hep-th].
- R. Abbott, Thomas D. Abbott, Sheelu Abraham, Fausto Acernese, et al. Open data from the first and second observing runs of Advanced LIGO and Advanced Virgo. *SoftwareX*, 13:100658, January 2021i. ISSN 2352-7110. doi:10.1016/j.softx.2021.100658. URL <https://www.sciencedirect.com/science/article/pii/S2352711021000030>.
- R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, et al. Population of Merging Compact Binaries Inferred Using Gravitational Waves through GWTC-3. *Physical Review X*, 13:011048, January 2023a. doi:10.1103/PhysRevX.13.011048. URL <https://ui.adsabs.harvard.edu/abs/2023PhRvX..13a1048A>. ADS Bibcode: 2023PhRvX..13a1048A.
- R. Abbott, H. Abe, F. Acernese, K. Ackley, et al. Constraints on the Cosmic Expansion History from GWTC-3. *The Astrophysical Journal*, 949:76, June 2023b. ISSN 0004-637X. doi:10.3847/1538-4357/ac74bb. URL <https://ui.adsabs.harvard.edu/abs/2023ApJ...949...76A>. Publisher: IOP ADS Bibcode: 2023ApJ...949...76A.
- F. Acernese, M. Agathos, K. Agatsuma, D. Aisa, et al. Advanced Virgo: a second-generation interferometric gravitational wave detector. *Classical and Quantum Gravity*, 32(2):024001, December 2014. ISSN 0264-9381. doi:10.1088/0264-9381/32/2/024001. URL <https://dx.doi.org/10.1088/0264-9381/32/2/024001>. Publisher: IOP Publishing.

- Masaru Adachi and Masumi Kasai. An analytical approximation of the luminosity distance in flat cosmologies with a cosmological constant. *Progress of Theoretical Physics*, 127(1):145–152, January 2012. ISSN 1347-4081, 0033-068X. doi:10.1143/PTP.127.145. URL <http://arxiv.org/abs/1111.6396>. arXiv:1111.6396 [astro-ph].
- P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, et al. Planck 2015 results. XIII. Cosmological parameters. *Astronomy and Astrophysics*, 594:A13, September 2016. ISSN 0004-6361. doi:10.1051/0004-6361/201525830. URL <https://ui.adsabs.harvard.edu/abs/2016A&A...594A..13P>. ADS Bibcode: 2016A&A...594A..13P.
- Amir Aghamousa, Jessica Aguilar, Steve Ahlen, et al. The DESI Experiment Part I: Science, Targeting, and Survey Design, December 2016. URL <http://arxiv.org/abs/1611.00036>. arXiv:1611.00036 [astro-ph].
- T. Akutsu, M. Ando, K. Arai, Y. Arai, and S. others Araki. Overview of KAGRA: Detector design and construction history. *Progress of Theoretical and Experimental Physics*, 2021: 05A101, May 2021. doi:10.1093/ptep/ptaa125. URL <https://ui.adsabs.harvard.edu/abs/2021PTEP.2021eA101A>. ADS Bibcode: 2021PTEP.2021eA101A.
- Bruce Allen, Warren G. Anderson, Patrick R. Brady, Duncan A. Brown, and Jolien D. E. Creighton. FINDCHIRP: An algorithm for detection of gravitational waves from inspiraling compact binaries. *Physical Review D*, 85:122006, June 2012. ISSN 1550-7998 0556-2821. doi:10.1103/PhysRevD.85.122006. URL <https://ui.adsabs.harvard.edu/abs/2012PhRvD..85l2006A>. Publisher: APS ADS Bibcode: 2012PhRvD..85l2006A.
- Justin Alsing, Hector O Silva, and Emanuele Berti. Evidence for a maximum mass cut-off in the neutron star mass distribution and constraints on the equation of state. *Monthly Notices of the Royal Astronomical Society*, 478(1):1377–1391, July 2018. ISSN 0035-8711. doi:10.1093/mnras/sty1065. URL <https://doi.org/10.1093/mnras/sty1065>.
- Lauren Anderson, Eric Aubourg, Stephen Bailey, et al. The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Baryon Acoustic Oscillations in the Data Release 10 and 11 galaxy samples. *Monthly Notices of the Royal Astronomical Society*, 441(1):24–62, June 2014. ISSN 1365-2966, 0035-8711. doi:10.1093/mnras/stu523. URL <http://arxiv.org/abs/1312.4877>. arXiv:1312.4877 [astro-ph].
- John Antoniadis, Thomas M. Tauris, Feryal Özel, Ewan Barr, David J. Champion, and Paulo C. C. Freire. The millisecond pulsar mass distribution: Evidence for bimodality and constraints on the maximum neutron star mass. *arXiv:1605.01665 [astro-ph, physics:nuclear]*, May 2016. URL <http://arxiv.org/abs/1605.01665>. arXiv: 1605.01665.
- Fabio Antonini, Mark Gieles, Fani Dosopoulou, and Debatri Chattopadhyay. Coalescing black hole binaries from globular clusters: mass distributions and comparison to gravitational wave data from GWTC-3, August 2022. URL <http://arxiv.org/abs/2208.01081>. arXiv:2208.01081 [astro-ph, physics:gr-qc].

- Gregory Ashton, Moritz Hübner, Paul D. Lasky, Colm Talbot, Kendall Ackley, Sylvia Biscoveanu, Qi Chu, Atul Divakarla, Paul J. Easter, Boris Goncharov, Francisco Hernandez Vivanco, Jan Harms, Marcus E. Lower, Grant D. Meadors, Denyz Melchor, Ethan Payne, Matthew D. Pitkin, Jade Powell, Nikhil Sarin, Rory J. E. Smith, and Eric Thrane. Bilby: A User-friendly Bayesian Inference Library for Gravitational-wave Astronomy. 241(2):27, April 2019. ISSN 0067-0049. doi:10.3847/1538-4365/ab06fc. URL <https://doi.org/10.3847/1538-4365/ab06fc>. Publisher: American Astronomical Society.
- Charles D. Bailyn, Raj K. Jain, Paolo Coppi, and Jerome A. Orosz. The Mass Distribution of Stellar Black Holes. *The Astrophysical Journal*, 499:367–374, May 1998. ISSN 0004-637X. doi:10.1086/305614. URL <https://ui.adsabs.harvard.edu/abs/1998ApJ...499..367B>. ADS Bibcode: 1998ApJ...499..367B.
- Z. Barkat, G. Rakavy, and N. Sack. Dynamics of Supernova Explosion Resulting from Pair Formation. *Physical Review Letters*, 18:379–381, March 1967. ISSN 0031-9007. doi:10.1103/PhysRevLett.18.379. URL <https://ui.adsabs.harvard.edu/abs/1967PhRvL..18..379B>. ADS Bibcode: 1967PhRvL..18..379B.
- Simone S. Bavera, Tassos Fragos, Emmanouil Zapartas, Enrico Ramirez-Ruiz, Pablo Marchant, Luke Z. Kelley, Michael Zevin, Jeff J. Andrews, Scott Coughlin, Aaron Dotter, Konstantinos Kovalakas, Devina Misra, Juan G. Serra-Perez, Ying Qin, Kyle A. Rocha, Jaime Román-Garza, Nam H. Tran, and Zepei Xing. Probing the progenitors of spinning binary black-hole mergers with long gamma-ray bursts. *Astronomy and Astrophysics*, 657:L8, January 2022. ISSN 0004-6361. doi:10.1051/0004-6361/202141979. URL <https://ui.adsabs.harvard.edu/abs/2022A&A...657L...8B>. ADS Bibcode: 2022A&A...657L...8B.
- Eric J. Baxter, Djuna Croon, Samuel D. McDermott, and Jeremy Sakstein. Find the Gap: Black Hole Population Analysis with an Astrophysically Motivated Mass Function. *The Astrophysical Journal*, 916:L16, August 2021. ISSN 0004-637X. doi:10.3847/2041-8213/ac11fc. URL <https://ui.adsabs.harvard.edu/abs/2021ApJ...916L..16B>. ADS Bibcode: 2021ApJ...916L..16B.
- M. J. Bayarri and M. E. Castellanos. Bayesian Checking of the Second Levels of Hierarchical Models. *Statistical Science*, 22(3):322–343, August 2007. ISSN 0883-4237, 2168-8745. doi:10.1214/07-STS235. URL <https://projecteuclid.org/journals/statistical-science/volume-22/issue-3/Bayesian-Checking-of-the-Second-Levels-of-Hierarchical-Models/10.1214/07-STS235.full>. Publisher: Institute of Mathematical Statistics.
- Krzysztof Belczynski, Vassiliki Kalogera, Frederic A. Rasio, Ronald E. Taam, Andreas Zezas, Tomasz Bulik, Thomas J. Maccarone, and Natalia Ivanova. Compact Object Modeling with the StarTrack Population Synthesis Code. *The Astrophysical Journal Supplement Series*, 174:223–260, January 2008. ISSN 0067-0049. doi:10.1086/521026. URL

<https://ui.adsabs.harvard.edu/abs/2008ApJS..174..223B>. ADS Bibcode: 2008ApJS..174..223B.

Krzysztof Belczynski, Grzegorz Wiktorowicz, Chris L. Fryer, Daniel E. Holz, and Vassiliki Kalogera. MISSING BLACK HOLES UNVEIL THE SUPERNOVA EXPLOSION MECHANISM. 757(1):91, September 2012. ISSN 0004-637X. doi:10.1088/0004-637X/757/1/91. URL <https://doi.org/10.1088/0004-637x/757/1/91>. Publisher: American Astronomical Society.

Krzysztof Belczynski, Daniel E. Holz, Tomasz Bulik, and Richard O’Shaughnessy. The first gravitational-wave source from the isolated evolution of two stars in the 40–100 solar mass range. *Nature*, 534(7608):512–515, June 2016. ISSN 1476-4687. doi:10.1038/nature18322. URL <https://www.nature.com/articles/nature18322>. Number: 7608 Publisher: Nature Publishing Group.

Eli Bingham, Jonathan P. Chen, Martin Jankowiak, Fritz Obermeyer, Neeraj Pradhan, Theofanis Karaletsos, Rohit Singh, Paul Szerlip, Paul Horsfall, and Noah D. Goodman. Pyro: Deep Universal Probabilistic Programming. *Journal of Machine Learning Research*, 20(28):1–6, 2019. URL <http://jmlr.org/papers/v20/18-403.html>.

Sylvia Biscoveanu, Colm Talbot, and Salvatore Vitale. Sources of systematic error in gravitational-wave measurements of the binary neutron star mass distribution. *arXiv:2111.13619 [astro-ph, physics:gr-qc]*, November 2021. URL <http://arxiv.org/abs/2111.13619>. arXiv: 2111.13619.

Sylvia Biscoveanu, Thomas A. Callister, Carl-Johan Haster, Ken K. Y. Ng, Salvatore Vitale, and Will M. Farr. The Binary Black Hole Spin Distribution Likely Broadens with Redshift. *The Astrophysical Journal*, 932:L19, June 2022a. ISSN 0004-637X. doi:10.3847/2041-8213/ac71a8. URL <https://ui.adsabs.harvard.edu/abs/2022ApJ...932L..19B>. ADS Bibcode: 2022ApJ...932L..19B.

Sylvia Biscoveanu, Philippe Landry, and Salvatore Vitale. Population properties and multimessenger prospects of neutron star-black hole mergers following GWTC-3. *Monthly Notices of the Royal Astronomical Society*, October 2022b. ISSN 0035-8711. doi:10.1093/mnras/stac3052. URL <https://ui.adsabs.harvard.edu/abs/2022MNRAS.tmp.2841B>. ADS Bibcode: 2022MNRAS.tmp.2841B.

Katelyn Breivik, Sourav Chatterjee, and Jeff J. Andrews. Constraining Compact Object Formation with 2M0521. 878(1):L4, June 2019. ISSN 2041-8205. doi:10.3847/2041-8213/ab21d3. URL <https://doi.org/10.3847/2041-8213/ab21d3>. Publisher: American Astronomical Society.

Katelyn Breivik, Scott Coughlin, Michael Zevin, Carl L. Rodriguez, Kyle Kremer, Claire S. Ye, Jeff J. Andrews, Michael Kurkowski, Matthew C. Digman, Shane L. Larson, and Frederic A. Rasio. COSMIC Variance in Binary Population Synthesis. *The Astrophysical Journal*, 898(1):71, July 2020. ISSN 0004-637X. doi:10.3847/1538-4357/ab9d85. URL [http](http://)

s://dx.doi.org/10.3847/1538-4357/ab9d85. Publisher: The American Astronomical Society.

Floor S. Broekgaarden, Edo Berger, Coenraad J. Neijssel, Alejandro Vigna-Gómez, Debatri Chattopadhyay, Simon Stevenson, Martyna Chruslinska, Stephen Justham, Selma E. de Mink, and Ilya Mandel. Impact of Massive Binary Star and Cosmic Evolution on Gravitational Wave Observations I: Black Hole-Neutron Star Mergers. *Monthly Notices of the Royal Astronomical Society*, 508(4):5028–5063, October 2021. ISSN 0035-8711, 1365-2966. doi:10.1093/mnras/stab2716. URL <http://arxiv.org/abs/2103.02608>. arXiv: 2103.02608.

Floor S. Broekgaarden, Simon Stevenson, and Eric Thrane. Signatures of mass ratio reversal in gravitational waves from merging binary black holes. *The Astrophysical Journal*, 938(1):45, October 2022. ISSN 0004-637X, 1538-4357. doi:10.3847/1538-4357/ac8879. URL <http://arxiv.org/abs/2205.01693>. arXiv:2205.01693 [astro-ph, physics:gr-qc].

Thomas A. Callister and Will M. Farr. Parameter-Free Tour of the Binary Black Hole Population. *Physical Review X*, 14:021005, April 2024. doi:10.1103/PhysRevX.14.021005. URL <https://ui.adsabs.harvard.edu/abs/2024PhRvX..14b1005C>. Publisher: APS ADS Bibcode: 2024PhRvX..14b1005C.

Thomas A. Callister, Carl-Johan Haster, Ken K. Y. Ng, Salvatore Vitale, and Will M. Farr. Who Ordered That? Unequal-Mass Binary Black Hole Mergers Have Larger Effective Spins. *The Astrophysical Journal Letters*, 922(1):L5, November 2021. ISSN 2041-8205, 2041-8213. doi:10.3847/2041-8213/ac2ccc. URL <http://arxiv.org/abs/2106.00521>. arXiv:2106.00521 [astro-ph, physics:gr-qc].

Tom Callister, Maya Fishbach, Daniel E. Holz, and Will M. Farr. Shouts and Murmurs: Combining Individual Gravitational-wave Sources with the Stochastic Background to Measure the History of Binary Black Hole Mergers. *The Astrophysical Journal*, 896:L32, June 2020. ISSN 0004-637X. doi:10.3847/2041-8213/ab9743. URL <https://ui.adsabs.harvard.edu/abs/2020ApJ...896L..32C>. ADS Bibcode: 2020ApJ...896L..32C.

Deep Chatterjee, Shaon Ghosh, Patrick R. Brady, Shasvath J. Kapadia, Andrew L. Miller, Samaya Nissanke, and Francesco Pannarale. A Machine Learning Based Source Property Inference for Compact Binary Mergers. *The Astrophysical Journal*, 896(1):54, June 2020. ISSN 1538-4357. doi:10.3847/1538-4357/ab8dbe. URL <http://arxiv.org/abs/1911.00116>. arXiv: 1911.00116.

Deep Chatterjee, Abhishek Hegade K. R., Gilbert Holder, Daniel E. Holz, Scott Perkins, Kent Yagi, and Nicolás Yunes. Cosmology with Love: Measuring the Hubble constant using neutron star universal relations. *Physical Review D*, 104:083528, October 2021. ISSN 1550-7998/0556-2821. doi:10.1103/PhysRevD.104.083528. URL <https://ui.adsabs.harvard.edu/abs/2021PhRvD.104h3528C>. Publisher: APS ADS Bibcode: 2021PhRvD.104h3528C.

- Katerina Chatziioannou. Neutron-star tidal deformability and equation-of-state constraints. *General Relativity and Gravitation*, 52(11):109, November 2020. ISSN 1572-9532. doi:10.1007/s10714-020-02754-3. URL <https://doi.org/10.1007/s10714-020-02754-3>.
- Hsin-Yu Chen, Maya Fishbach, and Daniel E. Holz. A two per cent Hubble constant measurement from standard sirens within five years. *Nature*, 562:545–547, October 2018. ISSN 0028-0836. doi:10.1038/s41586-018-0606-0. URL <https://ui.adsabs.harvard.edu/abs/2018Natur.562..545C>. ADS Bibcode: 2018Natur.562..545C.
- Hsin-Yu Chen, Daniel E. Holz, John Miller, Matthew Evans, Salvatore Vitale, and Jolien Creighton. Distance measures in gravitational-wave astrophysics and cosmology. *Classical and Quantum Gravity*, 38(5):055010, January 2021. ISSN 0264-9381. doi:10.1088/1361-6382/abd594. URL <https://dx.doi.org/10.1088/1361-6382/abd594>. Publisher: IOP Publishing.
- Hsin-Yu Chen, Jose María Ezquiaga, and Ish Gupta. Cosmography with next-generation gravitational wave detectors, February 2024. URL <https://ui.adsabs.harvard.edu/abs/2024arXiv240203120C>. Publication Title: arXiv e-prints ADS Bibcode: 2024arXiv240203120C.
- Xian Chen, Shuo Li, and Zhoujian Cao. Mass-redshift degeneracy for the gravitational-wave sources in the vicinity of supermassive black holes. *Monthly Notices of the Royal Astronomical Society*, 485:L141–L145, May 2019. ISSN 0035-8711. doi:10.1093/mnrasl/slz046. URL <https://ui.adsabs.harvard.edu/abs/2019MNRAS.485L.141C>. ADS Bibcode: 2019MNRAS.485L.141C.
- April Qiu Cheng, Michael Zevin, and Salvatore Vitale. What You Don’t Know Can Hurt You: Use and Abuse of Astrophysical Models in Gravitational-wave Population Analyses. , 955(2):127, October 2023. doi:10.3847/1538-4357/aced98.
- David F. Chernoff and Lee S. Finn. Gravitational Radiation, Inspirling Binaries, and Cosmology. *The Astrophysical Journal*, 411:L5, July 1993. ISSN 0004-637X. doi:10.1086/186898. URL <https://ui.adsabs.harvard.edu/abs/1993ApJ...411L...5C>. ADS Bibcode: 1993ApJ...411L...5C.
- LIGO Scientific Collaboration and Virgo Collaboration and KAGRA Collaboration. GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run — O3 search sensitivity estimates, November 2021. URL <https://zenodo.org/record/5546676>.
- Storm Colloms, Christopher P L Berry, John Veitch, and Michael Zevin. Exploring the evolution of gravitational-wave emitters with efficient emulation: Constraining the origins of binary black holes using normalising flows. *arXiv e-prints*, art. arXiv:2503.03819, March 2025. doi:10.48550/arXiv.2503.03819.

- D. A. Coulter, R. J. Foley, C. D. Kilpatrick, M. R. Drout, A. L. Piro, B. J. Shappee, M. R. Siebert, J. D. Simon, N. Ulloa, D. Kasen, B. F. Madore, A. Murguia-Berthier, Y. C. Pan, J. X. Prochaska, E. Ramirez-Ruiz, A. Rest, and C. Rojas-Bravo. Swope Supernova Survey 2017a (SSS17a), the optical counterpart to a gravitational wave source. *Science*, 358:1556–1558, December 2017. ISSN 0036-8075. doi:10.1126/science.aap9811. URL <https://ui.adsabs.harvard.edu/abs/2017Sci...358.1556C>. ADS Bibcode: 2017Sci...358.1556C.
- Djuna Croon, Samuel D. McDermott, and Jeremy Sakstein. New physics and the black hole mass gap. *Phys. Rev. D*, 102(11):115024, December 2020. doi:10.1103/PhysRevD.102.115024. URL <https://link.aps.org/doi/10.1103/PhysRevD.102.115024>. Publisher: American Physical Society.
- V. De Luca, V. Desjacques, G. Franciolini, P. Pani, and A. Riotto. GW190521 Mass Gap Event and the Primordial Black Hole Scenario. *prl*, 126(5):051101, February 2021. doi:10.1103/PhysRevLett.126.051101. _eprint: 2009.01728.
- Walter Del Pozzo. Inference of cosmological parameters from gravitational waves: Applications to second generation interferometers. *Physical Review D*, 86:043011, August 2012. ISSN 1550-7998/0556-2821. doi:10.1103/PhysRevD.86.043011. URL <https://ui.adsabs.harvard.edu/abs/2012PhRvD..86d3011D>. ADS Bibcode: 2012PhRvD..86d3011D.
- Virginia D’Emilio, Rhys Green, and Vivien Raymond. Density estimation with Gaussian processes for gravitational-wave posteriors. *Monthly Notices of the Royal Astronomical Society*, 508(2):2090–2097, October 2021. ISSN 0035-8711, 1365-2966. doi:10.1093/mnras/stab2623. URL <http://arxiv.org/abs/2104.05357>. arXiv:2104.05357 [astro-ph, physics:gr-qc].
- James M. Dickey and B. P. Lientz. The Weighted Likelihood Ratio, Sharp Hypotheses about Chances, the Order of a Markov Chain. *The Annals of Mathematical Statistics*, 41(1):214–226, 1970. ISSN 0003-4851. URL <https://www.mendeley.com/catalogue/ff8e9252-ad6c-37f5-9e8f-8a48de0e2a06/>. ISBN: 9781177697200 Number: 1.
- Z. Doctor, D. Wysocki, R. O’Shaughnessy, D. E. Holz, and B. Farr. Black Hole Coagulation: Modeling Hierarchical Mergers in Black Hole Populations. *The Astrophysical Journal*, 893: 35, April 2020. ISSN 0004-637X. doi:10.3847/1538-4357/ab7fac. URL <https://ui.adsabs.harvard.edu/abs/2020ApJ...893...35D>. ADS Bibcode: 2020ApJ...893...35D.
- Zoheyr Doctor, Ben Farr, Daniel E. Holz, and Michael Pürrer. Statistical gravitational waveform models: What to simulate next? *Physical Review D*, 96:123011, December 2017. ISSN 1550-7998/0556-2821. doi:10.1103/PhysRevD.96.123011. URL <https://ui.adsabs.harvard.edu/abs/2017PhRvD..96l3011D>. ADS Bibcode: 2017PhRvD..96l3011D.
- Zoheyr Doctor, Ben Farr, and Daniel E. Holz. Black Hole Leftovers: The Remnant Population from Binary Black Hole Mergers. *The Astrophysical Journal Letters*, 914(1):L18, June 2021. ISSN 2041-8205. doi:10.3847/2041-8213/ac0334. URL <https://ui.adsabs.harvard.edu/abs/2021ApJL...914L..18D>.

[//dx.doi.org/10.3847/2041-8213/ac0334](https://doi.org/10.3847/2041-8213/ac0334). Publisher: The American Astronomical Society.

- P. Drozda, K. Belczynski, R. O’Shaughnessy, T. Bulik, and C. L. Fryer. Black hole - neutron star mergers: the first mass gap and kilonovae. *arXiv:2009.06655 [astro-ph]*, September 2020. URL <http://arxiv.org/abs/2009.06655>. arXiv: 2009.06655.
- L. du Buisson, P. Marchant, Ph Podsiadlowski, C. Kobayashi, F. B. Abdalla, P. Taylor, I. Mandel, S. E. de Mink, T. J. Moriya, and N. Langer. Cosmic rates of black hole mergers and pair-instability supernovae from chemically homogeneous binary evolution. *Monthly Notices of the Royal Astronomical Society*, 499:5941–5959, December 2020. ISSN 0035-8711. doi:10.1093/mnras/staa3225. URL <https://ui.adsabs.harvard.edu/abs/2020MNRAS.499.5941D>. ADS Bibcode: 2020MNRAS.499.5941D.
- A. S. Eddington. On a formula for correcting statistics for the effects of a known error of observation. *Monthly Notices of the Royal Astronomical Society*, 73:359–360, March 1913. ISSN 0035-8711. doi:10.1093/mnras/73.5.359. URL <https://ui.adsabs.harvard.edu/abs/1913MNRAS..73..359E>. ADS Bibcode: 1913MNRAS..73..359E.
- Bruce Edelman, Zoheyr Doctor, and Ben Farr. Poking Holes: Looking for Gaps in LIGO/Virgo’s Black Hole Population. *arXiv e-prints*, page arXiv:2104.07783, April 2021. _eprint: 2104.07783.
- Bruce Edelman, Zoheyr Doctor, Jaxen Godfrey, and Ben Farr. Ain’t No Mountain High Enough: Semiparametric Modeling of LIGO-Virgo’s Binary Black Hole Mass Distribution. *The Astrophysical Journal*, 924:101, January 2022. ISSN 0004-637X. doi:10.3847/1538-4357/ac3667. URL <https://ui.adsabs.harvard.edu/abs/2022ApJ...924..101E>. ADS Bibcode: 2022ApJ...924..101E.
- Bruce Edelman, Ben Farr, and Zoheyr Doctor. Cover Your Basis: Comprehensive Data-driven Characterization of the Binary Black Hole Population. *The Astrophysical Journal*, 946:16, March 2023. ISSN 0004-637X. doi:10.3847/1538-4357/acb5ed. URL <https://ui.adsabs.harvard.edu/abs/2023ApJ...946...16E>. ADS Bibcode: 2023ApJ...946...16E.
- Reed Essick and Will Farr. Precision Requirements for Monte Carlo Sums within Hierarchical Bayesian Inference. Technical report, April 2022. URL <https://ui.adsabs.harvard.edu/abs/2022arXiv220400461E>. Publication Title: arXiv e-prints ADS Bibcode: 2022arXiv220400461E Type: article.
- Reed Essick and Maya Fishbach. On Estimating Rates from Monte-Carlo Integrals over Injection Sets, February 2021. URL <https://dcc.ligo.org/T2000100>.
- Reed Essick and Maya Fishbach. Ensuring Consistency between Noise and Detection in Hierarchical Bayesian Inference. *The Astrophysical Journal*, 962:169, February 2024. ISSN 0004-637X. doi:10.3847/1538-4357/ad1604. URL <https://ui.adsabs.harvard.edu/abs/2024ApJ...962..169E>. ADS Bibcode: 2024ApJ...962..169E.

- Reed Essick and Philippe Landry. Discriminating between Neutron Stars and Black Holes with Imperfect Knowledge of the Maximum Neutron Star Mass. *The Astrophysical Journal*, 904(1):80, November 2020. doi:10.3847/1538-4357/abbd3b. URL <https://doi.org/10.3847/1538-4357/abbd3b>. Publisher: American Astronomical Society.
- Reed Essick, Ingo Tews, Philippe Landry, Sanjay Reddy, and Daniel E. Holz. Direct astrophysical tests of chiral effective field theory at supranuclear densities. *Physical Review C*, 102(5):055803, November 2020. doi:10.1103/PhysRevC.102.055803. URL <https://link.aps.org/doi/10.1103/PhysRevC.102.055803>. Publisher: American Physical Society.
- Reed Essick, Amanda Farah, Shanika Galaudage, Colm Talbot, Maya Fishbach, Eric Thrane, and Daniel E. Holz. Probing Extremal Gravitational-wave Events with Coarse-grained Likelihoods. *The Astrophysical Journal*, 926:34, February 2022. ISSN 0004-637X. doi:10.3847/1538-4357/ac3978. URL <https://ui.adsabs.harvard.edu/abs/2022ApJ...926...34E>. ADS Bibcode: 2022ApJ...926...34E.
- et al. ET steering committee. Einstein Telescope: Science Case, Design Study and Feasibility Report. Official document ET-0028A-20, August 2020. URL <https://apps.et-gw.eu/t ds/?content=3&r=17196>.
- Matthew Evans, Rana X Adhikari, Chaitanya Afle, Stefan W. Ballmer, Sylvia Biscoveanu, Ssohrab Borhanian, Duncan A. Brown, Yanbei Chen, Robert Eisenstein, Alexandra Gruson, Anuradha Gupta, Evan D. Hall, Rachael Huxford, Brittany Kamai, Rahul Kashyap, Jeff S. Kissel, Kevin Kuns, Philippe Landry, Amber Lenon, Geoffrey Lovelace, Lee McCuller, Ken K. Y. Ng, Alexander H. Nitz, Jocelyn Read, B. S. Sathyaprakash, David H. Shoemaker, Bram J. J. Slagmolen, Joshua R. Smith, Varun Srivastava, Ling Sun, Salvatore Vitale, and Rainer Weiss. A Horizon Study for Cosmic Explorer: Science, Observatories, and Community, September 2021. URL <https://ui.adsabs.harvard.edu/abs/2021arXiv210909882E>. Publication Title: arXiv e-prints ADS Bibcode: 2021arXiv210909882E.
- Jose María Ezquiaga and Daniel E. Holz. Jumping the Gap: Searching for LIGO’s Biggest Black Holes. *The Astrophysical Journal*, 909:L23, March 2021. ISSN 0004-637X. doi:10.3847/2041-8213/abe638. URL <https://ui.adsabs.harvard.edu/abs/2021ApJ...909L..23E>. ADS Bibcode: 2021ApJ...909L..23E.
- Jose María Ezquiaga and Daniel E. Holz. Spectral Sirens: Cosmology from the Full Mass Distribution of Compact Binaries. *Physical Review Letters*, 129:061102, August 2022. ISSN 0031-9007. doi:10.1103/PhysRevLett.129.061102. URL <https://ui.adsabs.harvard.edu/abs/2022PhRvL.129f1102E>. Publisher: APS ADS Bibcode: 2022PhRvL.129f1102E.
- Ebraheem Farag, Mathieu Renzo, Robert Farmer, Morgan T. Chidester, and F. X. Timmes. Resolving the Peak of the Black Hole Mass Spectrum. *The Astrophysical Journal*, 937:112, October 2022. ISSN 0004-637X. doi:10.3847/1538-4357/ac8b83. URL <https://ui.adsabs.harvard.edu/abs/2022ApJ...937..112F>. ADS Bibcode: 2022ApJ...937..112F.

- Amanda Farah, Maya Fishbach, Reed Essick, Daniel E. Holz, and Shanika Galaudage. Bridging the Gap: Categorizing Gravitational-wave Events at the Transition between Neutron Stars and Black Holes. *The Astrophysical Journal*, 931:108, June 2022a. ISSN 0004-637X. doi:10.3847/1538-4357/ac5f03. URL <https://ui.adsabs.harvard.edu/abs/2022ApJ...931..108F>. ADS Bibcode: 2022ApJ...931..108F.
- Amanda M. Farah, Bruce Edelman, Michael Zevin, Maya Fishbach, Jose María Ezquiaga, Ben Farr, and Daniel E. Holz. Things That Might Go Bump in the Night: Assessing Structure in the Binary Black Hole Mass Spectrum. *The Astrophysical Journal*, 955:107, October 2023. ISSN 0004-637X. doi:10.3847/1538-4357/aced02. URL <https://ui.adsabs.harvard.edu/abs/2023ApJ...955..107F>. ADS Bibcode: 2023ApJ...955..107F.
- Amanda M. Farah, Maya Fishbach, and Daniel E. Holz. Two of a Kind: Comparing Big and Small Black Holes in Binaries with Gravitational Waves. *The Astrophysical Journal*, 962:69, February 2024. ISSN 0004-637X. doi:10.3847/1538-4357/ad0558. URL <https://ui.adsabs.harvard.edu/abs/2024ApJ...962...69F>. Publisher: IOP ADS Bibcode: 2024ApJ...962...69F.
- Amanda Mirna Farah, Bruce Edelman, Michael Zevin, Maya Fishbach, Jose Maria Ezquiaga, Ben Farr, and Daniel Holz. Data release for "Things that might Go bump in the night: Assessing structure in the binary black hole mass spectrum", December 2022b. URL <https://zenodo.org/record/7411991>.
- Amanda Mirna Farah, Maya Fishbach, Bruce Edelman, Michael Zevin, and Jose Maria Ezquiaga. GWMockCat, December 2022c. URL <https://zenodo.org/record/7570191>. Language: eng.
- R. Farmer, M. Renzo, S. E. de Mink, P. Marchant, and S. Justham. Mind the Gap: The Location of the Lower Edge of the Pair-instability Supernova Black Hole Mass Gap. *The Astrophysical Journal*, 887:53, December 2019. ISSN 0004-637X. doi:10.3847/1538-4357/ab518b. URL <https://ui.adsabs.harvard.edu/abs/2019ApJ...887...53F>. ADS Bibcode: 2019ApJ...887...53F.
- Robert Farmer, Mathieu Renzo, Selma de Mink, Maya Fishbach, and Stephen Justham. Constraints from gravitational wave detections of binary black hole mergers on the $\dot{\Omega}_{\rm{C}}^{\rm{12}}(\alpha, \gamma)$. *The Astrophysical Journal*, 902(2):L36, October 2020. ISSN 2041-8213. doi:10.3847/2041-8213/abbadd. URL <http://arxiv.org/abs/2006.06678>. arXiv:2006.06678 [astro-ph].
- Will M. Farr. Accuracy Requirements for Empirically Measured Selection Functions. *Research Notes of the AAS*, 3(5):66, May 2019. ISSN 2515-5172. doi:10.3847/2515-5172/ab1d5f. URL <https://dx.doi.org/10.3847/2515-5172/ab1d5f>. Publisher: The American Astronomical Society.
- Will M. Farr and Katerina Chatziioannou. A Population-Informed Mass Estimate for Pulsar J0740+6620. 4(5):65, May 2020. ISSN 2515-5172. doi:10.3847/2515-

5172/ab9088. URL <https://doi.org/10.3847/2515-5172/ab9088>. Publisher: American Astronomical Society.

Will M. Farr, Niharika Sravan, Andrew Cantrell, Laura Kreidberg, Charles D. Bailyn, Ilya Mandel, and Vicky Kalogera. THE MASS DISTRIBUTION OF StELAR-MASS BLACK HOLES. 741(2):103, October 2011. ISSN 0004-637X. doi:10.1088/0004-637X/741/2/103. URL <https://doi.org/10.1088/0004-637x/741/2/103>. Publisher: American Astronomical Society.

Will M. Farr, Maya Fishbach, Jiani Ye, and Daniel E. Holz. A Future Percent-level Measurement of the Hubble Expansion at Redshift 0.8 with Advanced LIGO. *The Astrophysical Journal*, 883(2):L42, October 2019. ISSN 2041-8205. doi:10.3847/2041-8213/ab4284. URL <https://doi.org/10.3847/2041-8213/ab4284>. Publisher: American Astronomical Society.

Lee Samuel Finn and David F. Chernoff. Observing binary inspiral in gravitational radiation: One interferometer. *Physical Review D*, 47(6):2198–2219, March 1993. doi:10.1103/PhysRevD.47.2198. URL <https://link.aps.org/doi/10.1103/PhysRevD.47.2198>. Publisher: American Physical Society.

M. Fishbach, R. Gray, I. Magaña Hernandez, H. Qi, et al. A Standard Siren Measurement of the Hubble Constant from GW170817 without the Electromagnetic Counterpart. *The Astrophysical Journal*, 871:L13, January 2019. ISSN 0004-637X. doi:10.3847/2041-8213/aaf96e. URL <https://ui.adsabs.harvard.edu/abs/2019ApJ...871L..13F>. ADS Bibcode: 2019ApJ...871L..13F.

Maya Fishbach and Daniel E. Holz. Where Are LIGO’s Big Black Holes? *The Astrophysical Journal*, 851(2):L25, December 2017. ISSN 2041-8213. doi:10.3847/2041-8213/aa9bf6. URL <http://dx.doi.org/10.3847/2041-8213/aa9bf6>.

Maya Fishbach and Daniel E. Holz. Minding the Gap: GW190521 as a Straddling Binary. *The Astrophysical Journal*, 904:L26, December 2020a. ISSN 0004-637X. doi:10.3847/2041-8213/abc827. URL <https://ui.adsabs.harvard.edu/abs/2020ApJ...904L..26F>. ADS Bibcode: 2020ApJ...904L..26F.

Maya Fishbach and Daniel E. Holz. Picky Partners: The Pairing of Component Masses in Binary Black Hole Mergers. *The Astrophysical Journal*, 891:L27, March 2020b. ISSN 0004-637X. doi:10.3847/2041-8213/ab7247. URL <https://ui.adsabs.harvard.edu/abs/2020ApJ...891L..27F>. ADS Bibcode: 2020ApJ...891L..27F.

Maya Fishbach, Daniel E. Holz, and Ben Farr. Are LIGO’s Black Holes Made from Smaller Black Holes? *The Astrophysical Journal*, 840:L24, May 2017. ISSN 0004-637X. doi:10.3847/2041-8213/aa7045. URL <https://ui.adsabs.harvard.edu/abs/2017ApJ...840L..24F>. ADS Bibcode: 2017ApJ...840L..24F.

Maya Fishbach, Daniel E. Holz, and Will M. Farr. Does the Black Hole Merger Rate Evolve with Redshift? *The Astrophysical Journal*, 863(2):L41, August 2018. ISSN 2041-8213.

doi:10.3847/2041-8213/aad800. URL <https://iopscience.iop.org/article/10.3847/2041-8213/aad800>.

Maya Fishbach, Reed Essick, and Daniel E. Holz. Does Matter Matter? Using the Mass Distribution to Distinguish Neutron Stars and Black Holes. *The Astrophysical Journal*, 899(1):L8, August 2020a. ISSN 2041-8205. doi:10.3847/2041-8213/aba7b6. URL <https://doi.org/10.3847/2041-8213/aba7b6>.

Maya Fishbach, Will M. Farr, and Daniel E. Holz. The Most Massive Binary Black Hole Detections and the Identification of Population Outliers. *The Astrophysical Journal*, 891:L31, March 2020b. ISSN 0004-637X. doi:10.3847/2041-8213/ab77c9. URL <https://ui.adsabs.harvard.edu/abs/2020ApJ...891L..31F>. ADS Bibcode: 2020ApJ...891L..31F.

Maya Fishbach, Zoheyr Doctor, Thomas Callister, Bruce Edelman, Jiani Ye, Reed Essick, Will M. Farr, Ben Farr, and Daniel E. Holz. When are LIGO/Virgo’s Big Black-Hole Mergers? *The Astrophysical Journal*, 912(2):98, May 2021. ISSN 0004-637X, 1538-4357. doi:10.3847/1538-4357/abee11. URL <http://arxiv.org/abs/2101.07699>. arXiv:2101.07699 [astro-ph, physics:gr-qc].

Emmanuel Fonseca, H. Thankful Cromartie, Timothy T. Pennucci, Paul S. Ray, Aida Yu Kirichenko, Scott M. Ransom, Paul B. Demorest, Ingrid H. Stairs, Zaven Arzoumanian, Lucas Guillemot, Aditya Parthasarathy, Matthew Kerr, Ismael Cognard, Paul T. Baker, Harsha Blumer, Paul R. Brook, Megan DeCesar, Timothy Dolch, F. Adam Dong, Elizabeth C. Ferrara, William Fiore, Nathaniel Garver-Daniels, Deborah C. Good, Ross Jennings, Megan L. Jones, Victoria M. Kaspi, Michael T. Lam, Duncan R. Lorimer, Jing Luo, Alexander McEwen, James W. McKee, Maura A. McLaughlin, Natasha McMann, Bradley W. Meyers, Arun Naidu, Cherry Ng, David J. Nice, Nihan Pol, Henri A. Radovan, Brent Shapiro-Albert, Chia Min Tan, Shriharsh P. Tendulkar, Joseph K. Swiggum, Haley M. Wahl, and Weiwei Zhu. Refined Mass and Geometric Measurements of the High-Mass PSR J0740+6620. *The Astrophysical Journal Letters*, 915(1):L12, July 2021. ISSN 2041-8205, 2041-8213. doi:10.3847/2041-8213/ac03b8. URL <http://arxiv.org/abs/2104.00880>. arXiv: 2104.00880.

Daniel Foreman-Mackey. corner.py: Scatterplot matrices in Python. *Journal of Open Source Software*, 1(2):24, June 2016. ISSN 2475-9066. doi:10.21105/joss.00024. URL <https://joss.theoj.org/papers/10.21105/joss.00024>.

Daniel Foreman-Mackey. tinygp — tinygp, 2021. URL <https://tinygp.readthedocs.io/en/stable/>.

Daniel Foreman-Mackey, Eric Agol, Sivaram Ambikasaran, and Ruth Angus. Fast and scalable Gaussian process modeling with applications to astronomical time series. *The Astronomical Journal*, 154(6):220, November 2017. ISSN 1538-3881. doi:10.3847/1538-3881/aa9332. URL <http://arxiv.org/abs/1703.09710>. arXiv:1703.09710 [astro-ph, physics:physics, stat].

- William A. Fowler and F. Hoyle. Neutrino Processes and Pair Formation in Massive Stars and Supernovae. *The Astrophysical Journal Supplement Series*, 9:201, December 1964. ISSN 0067-0049. doi:10.1086/190103. URL <https://ui.adsabs.harvard.edu/abs/1964ApJS....9..201F>. ADS Bibcode: 1964ApJS....9..201F.
- Gabriele Franciolini, Vishal Baibhav, Valerio De Luca, Ken K. Y. Ng, Kaze W. K. Wong, Emanuele Berti, Paolo Pani, Antonio Riotto, and Salvatore Vitale. Quantifying the evidence for primordial black holes in LIGO/Virgo gravitational-wave data. *arXiv e-prints*, page arXiv:2105.03349, May 2021. _eprint: 2105.03349.
- Wendy L. Freedman. Measurements of the Hubble Constant: Tensions in Perspective. *The Astrophysical Journal*, 919:16, September 2021. ISSN 0004-637X. doi:10.3847/1538-4357/ac0e95. URL <https://ui.adsabs.harvard.edu/abs/2021ApJ...919...16F>. ADS Bibcode: 2021ApJ...919...16F.
- C. L. Fryer, A. Heger, N. Langer, and S. Wellstein. The Limiting Stellar Initial Mass for Black Hole Formation in Close Binary Systems. *The Astrophysical Journal*, 578(1):335, October 2002. ISSN 0004-637X. doi:10.1086/342307. URL <https://iopscience.iop.org/article/10.1086/342307/meta>. Publisher: IOP Publishing.
- C. L. Fryer, K. Belczynski, G. Wiktorowicz, M. Dominik, V. Kalogera, and D. Holz. Compact Remnant Mass Function: Dependence on the Explosion Mechanism and Metallicity. *The Astrophysical Journal*, 749(1):91, April 2012. ISSN 0004-637X, 1538-4357. doi:10.1088/0004-637X/749/1/91. URL <http://arxiv.org/abs/1110.1726>. arXiv:1110.1726 [astro-ph].
- Chris L. Fryer and Vassiliki Kalogera. Theoretical Black Hole Mass Distributions. *The Astrophysical Journal*, 554(1):548–560, June 2001. doi:10.1086/321359. URL <https://doi.org/10.1086/321359>. Publisher: American Astronomical Society.
- Jonathan R. Gair, Archisman Ghosh, Rachel Gray, Daniel E. Holz, Simone Mastrogiovanni, Suvodip Mukherjee, Antonella Palmese, Nicola Tamanini, Tessa Baker, Freija Beirnaert, Maciej Bilicki, Hsin-Yu Chen, Gergely Dálya, Jose Maria Ezquiaga, Will M. Farr, Maya Fishbach, Juan Garcia-Bellido, Tathagata Ghosh, Hsiang-Yu Huang, Christos Karathanasis, Konstantin Leyde, Ignacio Magaña Hernandez, Johannes Noller, Gregoire Pierra, Peter Raffai, Antonio Enea Romano, Monica Seglar-Arroyo, Danièle A. Steer, Cezary Turski, Maria Paola Vaccaro, and Sergio Andrés Vallejo-Peña. The Hitchhiker’s Guide to the Galaxy Catalog Approach for Dark Siren Gravitational-wave Cosmology. *The Astronomical Journal*, 166:22, July 2023. ISSN 0004-6256. doi:10.3847/1538-3881/acca78. URL <https://ui.adsabs.harvard.edu/abs/2023AJ....166...22G>. ADS Bibcode: 2023AJ....166...22G.
- Monica Gallegos-Garcia, Christopher P L Berry, and Vicky Kalogera. Evolutionary Origins of Binary Neutron Star Mergers: Effects of Common Envelope Efficiency and Metallicity, November 2022. URL <https://ui.adsabs.harvard.edu/abs/2022arXiv221115693G>. ADS Bibcode: 2022arXiv221115693G.

- Andrew Gelman. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, 1(3):515–534, September 2006. ISSN 1936-0975, 1931-6690. doi:10.1214/06-BA117A. URL <https://projecteuclid.org/journals/bayesian-analysis/volume-1/issue-3/Prior-distributions-for-variance-parameters-in-hierarchical-models-comment-on/10.1214/06-BA117A.full>. Publisher: International Society for Bayesian Analysis.
- Davide Gerosa and Emanuele Berti. Are merging black holes born from stellar collapse or previous mergers? *Physical Review D*, 95:124046, June 2017. ISSN 1550-79980556-2821. doi:10.1103/PhysRevD.95.124046. URL <https://ui.adsabs.harvard.edu/abs/2017PhRvD..9514046G>. ADS Bibcode: 2017PhRvD..9514046G.
- Davide Gerosa, Nicola Giacobbo, and Alberto Vecchio. High mass but low spin: an exclusion region to rule out hierarchical black-hole mergers as a mechanism to populate the pair-instability mass gap. *The Astrophysical Journal*, 915(1):56, July 2021. ISSN 0004-637X, 1538-4357. doi:10.3847/1538-4357/ac00bb. URL <http://arxiv.org/abs/2104.11247>. arXiv:2104.11247 [astro-ph, physics:gr-qc].
- Nicola Giacobbo, Michela Mapelli, and Mario Spera. Merging black hole binaries: the effects of progenitor’s metallicity, mass-loss rate and Eddington factor. *Monthly Notices of the Royal Astronomical Society*, 474:2959–2974, March 2018. ISSN 0035-8711. doi:10.1093/mnras/stx2933. URL <https://ui.adsabs.harvard.edu/abs/2018MNRAS.474.2959G>. Publisher: OUP ADS Bibcode: 2018MNRAS.474.2959G.
- Benjamin Giesers, Stefan Dreizler, Tim-Oliver Husser, Sebastian Kamann, Guillem Anglada Escudé, Jarle Brinchmann, C. Marcella Carollo, Martin M. Roth, Peter M. Weibacher, and Lutz Wisotzki. A detached stellar-mass black hole candidate in the globular cluster NGC 3201. *Monthly Notices of the Royal Astronomical Society*, 475: L15–L19, March 2018. ISSN 0035-8711. doi:10.1093/mnrasl/slx203. URL <https://ui.adsabs.harvard.edu/abs/2018MNRAS.475L..15G>. ADS Bibcode: 2018MNRAS.475L..15G.
- Jaxen Godfrey, Bruce Edelman, and Ben Farr. Cosmic Cousins: Identification of a Subpopulation of Binary Black Holes Consistent with Isolated Binary Evolution, April 2023. URL <https://ui.adsabs.harvard.edu/abs/2023arXiv230401288G>. Publication Title: arXiv e-prints ADS Bibcode: 2023arXiv230401288G.
- R. Gray, C. Messenger, and J. Veitch. A pixelated approach to galaxy catalogue incompleteness: improving the dark siren measurement of the Hubble constant. *Monthly Notices of the Royal Astronomical Society*, 512:1127–1140, May 2022. ISSN 0035-8711. doi:10.1093/mnras/stac366. URL <https://ui.adsabs.harvard.edu/abs/2022MNRAS.512.1127G>. ADS Bibcode: 2022MNRAS.512.1127G.
- Rachel Gray, Ignacio Magaña Hernandez, Hong Qi, Ankan Sur, Patrick R. Brady, Hsin-Yu Chen, Will M. Farr, Maya Fishbach, Jonathan R. Gair, Archisman Ghosh, Daniel E. Holz, Simone Mastroianni, Christopher Messenger, Danièle A. Steer, and John

- Veitch. Cosmological inference using gravitational wave standard sirens: A mock data analysis. *Physical Review D*, 101:122001, June 2020. ISSN 1550-79980556-2821. doi:10.1103/PhysRevD.101.122001. URL <https://ui.adsabs.harvard.edu/abs/2020PhRvD.10112001G>. ADS Bibcode: 2020PhRvD.10112001G.
- Rachel Gray, Freija Beirnaert, Christos Karathanasis, Benoît Revenu, Cezary Turski, Anson Chen, Tessa Baker, Sergio Vallejo, Antonio Enea Romano, Tathagata Ghosh, Archisman Ghosh, Konstantin Leyde, Simone Mastrogiovanni, and Surhud More. Joint cosmological and gravitational-wave population inference using dark sirens and galaxy catalogues. *Journal of Cosmology and Astroparticle Physics*, 2023(12):023, December 2023. ISSN 1475-7516. doi:10.1088/1475-7516/2023/12/023. URL <https://dx.doi.org/10.1088/1475-7516/2023/12/023>. Publisher: IOP Publishing.
- Michael Y. Grudić, Stella S. R. Offner, Dávid Guszejnov, Claude-André Faucher-Giguère, and Philip F. Hopkins. Does God play dice with star clusters?, June 2023. URL <http://arxiv.org/abs/2307.00052>. arXiv:2307.00052 [astro-ph].
- Mark S. Handcock and Michael L. Stein. A Bayesian Analysis of Kriging. *Technometrics*, 35(4):403–410, 1993. ISSN 0040-1706. doi:10.2307/1270273. URL <https://www.jstor.org/stable/1270273>. Publisher: [Taylor & Francis, Ltd., American Statistical Association, American Society for Quality].
- Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020. ISSN 1476-4687. doi:10.1038/s41586-020-2649-2. URL <https://www.nature.com/articles/s41586-020-2649-2>.
- J. A. Hartigan and P. M. Hartigan. The Dip Test of Unimodality. *The Annals of Statistics*, 13(1):70–84, 1985. ISSN 0090-5364. URL <https://www.jstor.org/stable/2241144>. Publisher: Institute of Mathematical Statistics.
- A. Heger and S. E. Woosley. The Nucleosynthetic Signature of Population III. *The Astrophysical Journal*, 567:532–543, March 2002. ISSN 0004-637X. doi:10.1086/338487. URL <https://ui.adsabs.harvard.edu/abs/2002ApJ...567..532H>. ADS Bibcode: 2002ApJ...567..532H.
- A. Heger, C. L. Fryer, S. E. Woosley, N. Langer, and D. H. Hartmann. How Massive Single Stars End Their Life. *The Astrophysical Journal*, 591:288–300, July 2003. ISSN 0004-637X. doi:10.1086/375341. URL <https://ui.adsabs.harvard.edu/abs/2003ApJ...591..288H>. ADS Bibcode: 2003ApJ...591..288H.

- Matthew D. Hoffman and Andrew Gelman. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo, November 2011. URL <http://arxiv.org/abs/1111.4246>. arXiv:1111.4246 [cs, stat].
- Daniel E. Holz and Scott A. Hughes. Using Gravitational-Wave Standard Sirens. *The Astrophysical Journal*, 629(1):15, August 2005. ISSN 0004-637X. doi:10.1086/431341. URL <https://iopscience.iop.org/article/10.1086/431341/meta>. Publisher: IOP Publishing.
- Stephan Hoyer and Joe Hamman. xarray: N-D labeled Arrays and Datasets in Python. *Journal of Open Research Software*, 5(1):10, April 2017. ISSN 2049-9647. doi:10.5334/jors.148. URL <http://openresearchsoftware.metajnl.com/articles/10.5334/jors.148/>. Number: 1 Publisher: Ubiquity Press.
- Rui-Chong Hu, Jin-Ping Zhu, Ying Qin, Bing Zhang, En-Wei Liang, and Yong Shao. A Channel to Form Fast-spinning Black Hole–Neutron Star Binary Mergers as Multimessenger Sources. *The Astrophysical Journal*, 928(2):163, April 2022. ISSN 0004-637X. doi:10.3847/1538-4357/ac573f. URL <https://dx.doi.org/10.3847/1538-4357/ac573f>. Publisher: The American Astronomical Society.
- E. A. Huerta, C. J. Moore, Prayush Kumar, Daniel George, Alvin J. K. Chua, Roland Haas, Erik Wessel, Daniel Johnson, Derek Glennon, Adam Rebei, A. Miguel Holgado, Jonathan R. Gair, and Harald P. Pfeiffer. Eccentric, nonspinning, inspiral, Gaussian-process merger approximant for the detection and characterization of eccentric binary black hole mergers. *Physical Review D*, 97:024031, January 2018. ISSN 1550-7998 0556-2821. doi:10.1103/PhysRevD.97.024031. URL <https://ui.adsabs.harvard.edu/abs/2018PhRvD..97b4031H>. ADS Bibcode: 2018PhRvD..97b4031H.
- John D. Hunter. Matplotlib: A 2D Graphics Environment. *Computing in Science Engineering*, 9(3):90–95, May 2007. ISSN 1558-366X. doi:10.1109/MCSE.2007.55. Conference Name: Computing in Science Engineering.
- Giuliano Iorio, Guglielmo Costa, Michela Mapelli, Mario Spera, Gastón J. Escobar, Cecilia Sgalletta, Alessandro A. Trani, Erika Korb, Filippo Santoliquido, Marco Dall’Amico, Nicola Gaspari, and Alessandro Bressan. Compact object mergers: exploring uncertainties from stellar and binary evolution with SEVN, November 2022. URL <http://arxiv.org/abs/2211.11774>. arXiv:2211.11774 [astro-ph].
- T. Jayasinghe, K. Z. Stanek, Todd A. Thompson, C. S. Kochanek, D. M. Rowan, P. J. Vallely, K. G. Strassmeier, M. Weber, J. T. Hinkle, F.-J. Hambsch, D. Martin, J. L. Prieto, T. Pessi, D. Huber, K. Auchettl, L. A. Lopez, I. Ilyin, C. Badenes, A. W. Howard, H. Isaacson, and S. J. Murphy. A Unicorn in Monoceros: the \$3M_\odot\$ dark companion to the bright, nearby red giant V723 Mon is a non-interacting, mass-gap black hole candidate. *Monthly Notices of the Royal Astronomical Society*, 504(2): 2577–2602, May 2021. ISSN 0035-8711, 1365-2966. doi:10.1093/mnras/stab907. URL <http://arxiv.org/abs/2101.02212>. arXiv: 2101.02212.

- Vassiliki Kalogera. Orbital Characteristics of Binary Systems after Asymmetric Supernova Explosions. *The Astrophysical Journal*, 471:352, November 1996. ISSN 0004-637X. doi:10.1086/177974. URL <https://ui.adsabs.harvard.edu/abs/1996ApJ...471..352K>. ADS Bibcode: 1996ApJ...471..352K.
- Christos Karathanasis, Suvodip Mukherjee, and Simone Mastrogiovanni. Binary black holes population and cosmology in new lights: Signature of PISN mass and formation channel in GWTC-3, November 2022a. URL <http://arxiv.org/abs/2204.13495>. arXiv:2204.13495 [astro-ph, physics:gr-qc].
- Christos Karathanasis, Benoit Revenu, Suvodip Mukherjee, and Federico Stachurski. GWSim: A python package to create GW mock samples for different astrophysical populations and cosmological models of binary black holes, October 2022b. URL <http://arxiv.org/abs/2210.05724>. arXiv:2210.05724 [astro-ph, physics:gr-qc].
- R. Weizmann Kiendrebeogo, Amanda M. Farah, Emily M. Foley, Abigail Gray, Nina Kunert, Anna Puecher, Andrew Toivonen, R. Oliver VandenBerg, Shreya Anand, Tomás Ahumada, Viraj Karambelkar, Michael W. Coughlin, Tim Dietrich, S. Zacharie Kam, Peter T. H. Pang, Leo P. Singer, and Niharika Sravan. Updated observing scenarios and multi-messenger implications for the International Gravitational-wave Network’s O4 and O5, June 2023. URL <http://arxiv.org/abs/2306.09234>. arXiv:2306.09234 [astro-ph].
- Chase Kimball, Colm Talbot, Christopher P. L. Berry, Matthew Carney, Michael Zevin, Eric Thrane, and Vicky Kalogera. Black Hole Genealogy: Identifying Hierarchical Mergers with Gravitational Waves. *The Astrophysical Journal*, 900:177, September 2020a. ISSN 0004-637X. doi:10.3847/1538-4357/aba518. URL <https://ui.adsabs.harvard.edu/abs/2020ApJ...900..177K>. ADS Bibcode: 2020ApJ...900..177K.
- Chase Kimball, Colm Talbot, Christopher P. L. Berry, Michael Zevin, Eric Thrane, Vicky Kalogera, Riccardo Buscicchio, Matthew Carney, Thomas Dent, Hannah Middleton, Ethan Payne, John Veitch, and Daniel Williams. Evidence for hierarchical black hole mergers in the second LIGO–Virgo gravitational-wave catalog. *apj*, 915:L35, November 2020b. _eprint: 2011.05332.
- Bülent Kiziltan, Athanasios Kottas, Maria De Yoreo, and Stephen E. Thorsett. THE NEUTRON STAR MASS DISTRIBUTION. *The Astrophysical Journal*, 778(1):66, November 2013. ISSN 0004-637X. doi:10.1088/0004-637X/778/1/66. URL <https://doi.org/10.1088/0004-637x/778/1/66>. Publisher: American Astronomical Society.
- Ely D. Kovetz, Ilias Cholis, Patrick C. Breysse, and Marc Kamionkowski. Black hole mass function from gravitational wave measurements. *Physical Review D*, 95:103010, May 2017. ISSN 1550-79980556-2821. doi:10.1103/PhysRevD.95.103010. URL <https://ui.adsabs.harvard.edu/abs/2017PhRvD..95j3010K>. ADS Bibcode: 2017PhRvD..95j3010K.
- Laura Kreidberg, Charles D. Bailyn, Will M. Farr, and Vicky Kalogera. MASS MEASUREMENTS OF BLACK HOLES IN X-RAY TRANSIENTS: IS THERE A MASS GAP? *The*

Astrophysical Journal, 757(1):36, September 2012. ISSN 0004-637X. doi:10.1088/0004-637X/757/1/36. URL <https://doi.org/10.1088/0004-637x/757/1/36>. Publisher: American Astronomical Society.

Keerthi Kunnumkai, Antonella Palmese, Mattia Bulla, Tim Dietrich, Amanda M. Farah, and Peter T. H. Pang. Kilonova emission from GW230529 and mass gap neutron star-black hole mergers, September 2024a. URL <https://ui.adsabs.harvard.edu/abs/2024arXiv240910651K>. Publication Title: arXiv e-prints ADS Bibcode: 2024arXiv240910651K.

Keerthi Kunnumkai, Antonella Palmese, Amanda M Farah, Mattia Bulla, Tim Dietrich, Peter T H Pang, Shreya Anand, Igor Andreoni, Tomas Cabrera, and Brendan O Connor. Detecting electromagnetic counterparts to LIGO/Virgo/KAGRA gravitational wave events with DECAM: Neutron Star Mergers, November 2024b. URL <https://ui.adsabs.harvard.edu/abs/2024arXiv241113673K>. Publication Title: arXiv e-prints ADS Bibcode: 2024arXiv241113673K.

Doron Kushnir. Thermonuclear explosion of rotating massive stars could explain core-collapse supernovae. *arXiv:1502.03111 [astro-ph]*, February 2015. URL <http://arxiv.org/abs/1502.03111>. arXiv: 1502.03111.

Philippe Landry and Reed Essick. Nonparametric inference of the neutron star equation of state from gravitational wave observations. *Physical Review D*, 99:084049, April 2019. ISSN 1550-7998/0556-2821. doi:10.1103/PhysRevD.99.084049. URL <https://ui.adsabs.harvard.edu/abs/2019PhRvD..99h4049L>. ADS Bibcode: 2019PhRvD..99h4049L.

Philippe Landry and Jocelyn S. Read. The Mass Distribution of Neutron Stars in Gravitational-wave Binaries. *The Astrophysical Journal*, 921:L25, November 2021. ISSN 0004-637X. doi:10.3847/2041-8213/ac2f3e. URL <https://ui.adsabs.harvard.edu/abs/2021ApJ...921L..25L>. ADS Bibcode: 2021ApJ...921L..25L.

E. Laplace, S. Justham, M. Renzo, Y. Götzberg, R. Farmer, D. Vartanyan, and S. E. de Mink. Different to the core: The pre-supernova structures of massive single and binary-stripped stars. *Astronomy and Astrophysics*, 656:A58, December 2021. ISSN 0004-6361. doi:10.1051/0004-6361/202140506. URL <https://ui.adsabs.harvard.edu/abs/2021A&A...656A..58L>. ADS Bibcode: 2021A&A...656A..58L.

Isaac Legred, Katerina Chatziioannou, Reed Essick, Sophia Han, and Philippe Landry. Impact of the PSR $\text{J}0740+6620$ radius constraint on the properties of high-density matter. *Physical Review D*, 104(6):063003, September 2021. doi:10.1103/PhysRevD.104.063003. URL <https://link.aps.org/doi/10.1103/PhysRevD.104.063003>. Publisher: American Physical Society.

Ang Li, Zhiqiang Miao, Sophia Han, and Bing Zhang. Constraints on the Maximum Mass of Neutron Stars with a Quark Core from GW170817 and NICER PSR J0030+0451 Data. *The Astrophysical Journal*, 913(1):27, May 2021a. ISSN 0004-637X. doi:10.3847/1538-4357/abf355. URL <https://dx.doi.org/10.3847/1538-4357/abf355>. Publisher: The American Astronomical Society.

- Yin-Jie Li, Yuan-Zhu Wang, Ming-Zhe Han, Shao-Peng Tang, Qiang Yuan, Yi-Zhong Fan, and Da-Ming Wei. A Flexible Gaussian Process Reconstruction Method and the Mass Function of the Coalescing Binary Black Hole Systems. *The Astrophysical Journal*, 917 (1):33, August 2021b. ISSN 0004-637X, 1538-4357. doi:10.3847/1538-4357/ac0971. URL <https://iopscience.iop.org/article/10.3847/1538-4357/ac0971>.
- Tong Liu, Yun-Feng Wei, Li Xue, and Mou-Yuan Sun. Final Compact Remnants in Core-collapse Supernovae from 20 to 40 M?: The Lower Mass Gap. *The Astrophysical Journal*, 908:106, February 2021. ISSN 0004-637X. doi:10.3847/1538-4357/abd24e. URL <https://ui.adsabs.harvard.edu/abs/2021ApJ...908..106L>. ADS Bibcode: 2021ApJ...908..106L.
- Thomas Loredo. Handling Source Uncertainties in Surveys via Multilevel Modeling. In *Bulletin of the American Astronomical Society*, volume 41, page 283, January 2009. URL <https://ui.adsabs.harvard.edu/abs/2009AAS...21321104L>. Conference Name: American Astronomical Society Meeting Abstracts #213 ADS Bibcode: 2009AAS...21321104L.
- Thomas J. Loredo. Accounting for Source Uncertainties in Analyses of Astronomical Survey Data. 735:195–206, November 2004. doi:10.1063/1.1835214. URL <https://ui.adsabs.harvard.edu/abs/2004AIPC..735..195L>. Conference Name: Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering Place: eprint: arXiv:astro-ph/0409387 ADS Bibcode: 2004AIPC..735..195L.
- Thomas J. Loredo. *Bayesian Astrostatistics: A Backward Look to the Future*. eprint: arXiv:1208.3036, January 2013. doi:10.1007/978-1-4614-3508-2_2. URL <https://ui.adsabs.harvard.edu/abs/2013acna.conf...15L>. Conference Name: Astrostatistical Challenges for the New Astronomy Pages: 1013 ADS Bibcode: 2013acna.conf...15L.
- Rodrigo Luger, Megan Bedell, Daniel Foreman-Mackey, Ian J. M. Crossfield, Lily L. Zhao, and David W. Hogg. Mapping stellar surfaces III: An Efficient, Scalable, and Open-Source Doppler Imaging Model, October 2021. URL <https://ui.adsabs.harvard.edu/abs/2021arXiv211006271L>. ADS Bibcode: 2021arXiv211006271L.
- LVK. H1 Calibrated Sensitivity Spectra Oct 24 2015 (Representative for O1), 2015a. URL <https://dcc.ligo.org/LIGO-G1600150/public>.
- LVK. L1 Calibrated Sensitivity Spectra Oct 24 2015 (Representative for O1), 2015b. URL <https://dcc.ligo.org/LIGO-G1600151/public>.
- LVK. H1 Calibrated Sensitivity Spectra Jun 10 2017 (Representative Best of O2 – C02, With Cleaning/Subtraction), 2017a. URL <https://dcc.ligo.org/LIGO-G1801950/public>.
- LVK. L1 Calibrated Sensitivity Spectra Aug 06 2017 (Representative Best of O2 – C02, With Cleaning/Subtraction), 2017b. URL <https://dcc.ligo.org/LIGO-G1801952/public>.

- LVK. aLIGO, CAL, Official Advanced LIGO Sensitivity Plots, 2020. URL <https://dcc.ligo.org/LIGO-G1500623/public>.
- Ignacio Magaña Hernandez and Anarya Ray. Beyond Gaps and Bumps: Spectral Siren Cosmology with Non-Parametric Population Models, April 2024. URL <https://ui.adsabs.harvard.edu/abs/2024arXiv240402522M>. Publication Title: arXiv e-prints ADS Bibcode: 2024arXiv240402522M.
- Parthapratim Mahapatra, Anuradha Gupta, Marc Favata, K. G. Arun, and B. S. Sathyaprakash. Black hole hierarchical growth efficiency and mass spectrum predictions, September 2022. URL <http://arxiv.org/abs/2209.05766>. arXiv:2209.05766 [astro-ph, physics:gr-qc].
- K. G. Malmquist. On some relations in stellar statistics. *Meddelanden fran Lunds Astronomiska Observatorium Serie I*, 100:1–52, March 1922. URL <https://ui.adsabs.harvard.edu/abs/1922MeLuF.100....1M>. ADS Bibcode: 1922MeLuF.100....1M.
- K. G. Malmquist. A contribution to the problem of determining the distribution in space of the stars. *Meddelanden fran Lunds Astronomiska Observatorium Serie I*, 106:1–12, February 1925. URL <https://ui.adsabs.harvard.edu/abs/1925MeLuF.106....1M>. ADS Bibcode: 1925MeLuF.106....1M.
- Michele Mancarella and Davide Gerosa. Sampling the full hierarchical population posterior distribution in gravitational-wave astronomy, February 2025. URL <https://ui.adsabs.harvard.edu/abs/2025arXiv250212156M>. ADS Bibcode: 2025arXiv250212156M.
- Ilya Mandel and Alison Farmer. Merging stellar-mass binary black holes. *Physics Reports*, 955:1–24, April 2022. ISSN 0370-1573. doi:10.1016/j.physrep.2022.01.003. URL <https://ui.adsabs.harvard.edu/abs/2022PhR...955....1M>. ADS Bibcode: 2022PhR...955....1M.
- Ilya Mandel and Bernhard Müller. Simple recipes for compact remnant masses and natal kicks. *Monthly Notices of the Royal Astronomical Society*, 499:3214–3221, December 2020. ISSN 0035-8711. doi:10.1093/mnras/staa3043. URL <https://ui.adsabs.harvard.edu/abs/2020MNRAS.499.3214M>. ADS Bibcode: 2020MNRAS.499.3214M.
- Ilya Mandel, Will M. Farr, Andrea Colonna, Simon Stevenson, Peter Tiño, and John Veitch. Model-independent inference on compact-binary observations. *Monthly Notices of the Royal Astronomical Society*, 465(3):3254–3260, March 2017. ISSN 0035-8711. doi:10.1093/mnras/stw2883. URL <https://doi.org/10.1093/mnras/stw2883>.
- Ilya Mandel, Will M. Farr, and Jonathan R. Gair. Extracting distribution parameters from multiple uncertain observations with selection biases. *Monthly Notices of the Royal Astronomical Society*, 486:1086–1093, June 2019. ISSN 0035-8711. doi:10.1093/mnras/stz896. URL <https://ui.adsabs.harvard.edu/abs/2019MNRAS.486.1086M>. ADS Bibcode: 2019MNRAS.486.1086M.

- Ilya Mandel, Bernhard Müller, Jeff Riley, Selma E. de Mink, Alejandro Vigna-Gómez, and Debatri Chattopadhyay. Binary population synthesis with probabilistic remnant mass and kick prescriptions. *Monthly Notices of the Royal Astronomical Society*, 500:1380–1384, January 2021. ISSN 0035-8711. doi:10.1093/mnras/staa3390. URL <https://ui.adsabs.harvard.edu/abs/2021MNRAS.500.1380M>. ADS Bibcode: 2021MNRAS.500.1380M.
- Michela Mapelli. Binary black hole mergers: formation and populations. *Frontiers in Astronomy and Space Sciences*, 7:38, July 2020. doi:10.3389/fspas.2020.00038. URL <https://ui.adsabs.harvard.edu/abs/2020FrASS...7...38M>. ADS Bibcode: 2020FrASS...7...38M.
- Michela Mapelli and Nicola Giacobbo. The cosmic merger rate of neutron stars and black holes. *Monthly Notices of the Royal Astronomical Society*, 479:4391–4398, October 2018. ISSN 0035-8711. doi:10.1093/mnras/sty1613. URL <https://ui.adsabs.harvard.edu/abs/2018MNRAS.479.4391M>. Publisher: OUP ADS Bibcode: 2018MNRAS.479.4391M.
- Pablo Marchant and Julia Bodensteiner. The Evolution of Massive Binary Stars. *Annual Review of Astronomy and Astrophysics*, 62:21–61, September 2024. ISSN 0066-4146. doi:10.1146/annurev-astro-052722-105936. URL <https://ui.adsabs.harvard.edu/abs/2024ARA&A..62...21M>. ADS Bibcode: 2024ARA&A..62...21M.
- Pablo Marchant, Mathieu Renzo, Robert Farmer, Kaliroe M. W. Pappas, Ronald E. Taam, Selma de Mink, and Vassiliki Kalogera. Pulsational pair-instability supernovae in very close binaries. *The Astrophysical Journal*, 882(1):36, August 2019. ISSN 1538-4357. doi:10.3847/1538-4357/ab3426. URL <http://arxiv.org/abs/1810.13412>. arXiv:1810.13412 [astro-ph].
- Ben Margalit and Brian D. Metzger. Constraining the Maximum Mass of Neutron Stars from Multi-messenger Observations of GW170817. 850(2):L19, November 2017. ISSN 2041-8205. doi:10.3847/2041-8213/aa991c. URL <https://doi.org/10.3847/2041-8213/aa991c>. Publisher: American Astronomical Society.
- S. Mastrogiovanni, K. Leyde, C. Karathanasis, E. Chassande-Mottin, D. A. Steer, J. Gair, A. Ghosh, R. Gray, S. Mukherjee, and S. Rinaldi. On the importance of source population models for gravitational-wave cosmology. *Physical Review D*, 104:062009, September 2021. ISSN 1550-79980556-2821. doi:10.1103/PhysRevD.104.062009. URL <https://ui.adsabs.harvard.edu/abs/2021PhRvD.104f2009M>. ADS Bibcode: 2021PhRvD.104f2009M.
- Simone Mastrogiovanni, Danny Laghi, Rachel Gray, Giada Caneva Santoro, Archisman Ghosh, Christos Karathanasis, Konstantin Leyde, Danièle A. Steer, Stéphane Perriès, and Grégoire Pierra. Joint population and cosmological properties inference with gravitational waves standard sirens and galaxy surveys. *Physical Review D*, 108(4):042002, August 2023a. doi:10.1103/PhysRevD.108.042002. URL <https://link.aps.org/doi/10.1103/PhysRevD.108.042002>. Publisher: American Physical Society.

- Simone Mastrogiovanni, Grégoire Pierra, Stéphane Perriès, Danny Laghi, Giada Caneva Santoro, Archisman Ghosh, Rachel Gray, Christos Karathanasis, and Konstantin Leyde. ICAROGW: A python package for inference of astrophysical population properties of noisy, heterogeneous and incomplete observations, May 2023b. URL <https://ui.adsabs.harvard.edu/abs/2023arXiv230517973M>. Publication Title: arXiv e-prints ADS Bibcode: 2023arXiv230517973M.
- Wes McKinney. Data Structures for Statistical Computing in Python. pages 56–61, Austin, Texas, 2010. doi:10.25080/Majora-92bf1922-00a. URL <https://doi.curvenote.com/10.25080/Majora-92bf1922-00a>.
- Geoffrey McLachlan and David Peel. Multivariate Normal Mixtures. In *Finite Mixture Models*, pages 81–116. John Wiley & Sons, Ltd, 2000. ISBN 978-0-471-72118-5. doi:10.1002/0471721182.ch3. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/0471721182.ch3>. Section: 3 _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/0471721182.ch3>.
- C. Messenger and J. Read. Measuring a Cosmological Distance-Redshift Relationship Using Only Gravitational Wave Observations of Binary Neutron Star Coalescences. *Physical Review Letters*, 108:091101, March 2012. ISSN 0031-9007. doi:10.1103/PhysRevLett.108.091101. URL <https://ui.adsabs.harvard.edu/abs/2012PhRvL.108i1101M>. ADS Bibcode: 2012PhRvL.108i1101M.
- Jeffrey W. Miller and David B. Dunson. Robust Bayesian inference via coarsening. *arXiv e-prints*, page arXiv:1506.06101, June 2015. _eprint: 1506.06101.
- Matthew Mould, Davide Gerosa, Floor S. Broekgaarden, and Nathan Steinle. Which black hole formed first? Mass-ratio reversal in massive binary stars from gravitational-wave data. *Monthly Notices of the Royal Astronomical Society*, 517(2):2738–2745, October 2022. ISSN 0035-8711, 1365-2966. doi:10.1093/mnras/stac2859. URL <http://arxiv.org/abs/2205.12329>. arXiv:2205.12329 [astro-ph, physics:gr-qc].
- P. Mroz and L. Wyrzykowski. Measuring the mass function of isolated stellar remnants with gravitational microlensing. I. Revisiting the OGLE-III dark lens candidates. *Acta Astronomica*, 71(2):89–102, June 2021. ISSN 00015237. doi:10.32023/0001-5237/71.2.1. URL <http://arxiv.org/abs/2107.13701>. arXiv: 2107.13701.
- Bernhard Müller, Alexander Heger, David Liptai, and Joshua B. Cameron. A simple approach to the supernova progenitor-explosion connection. *Monthly Notices of the Royal Astronomical Society*, 460:742–764, July 2016. ISSN 0035-8711. doi:10.1093/mnras/stw1083. URL <https://ui.adsabs.harvard.edu/abs/2016MNRAS.460..742M>. ADS Bibcode: 2016MNRAS.460..742M.
- Coenraad J. Neijssel, Alejandro Vigna-Gómez, Simon Stevenson, Jim W. Barrett, Sebastian M. Gaebel, Floor S. Broekgaarden, Selma E. de Mink, Dorottya Szécsi, Serena Vinciguerra, and Ilya Mandel. The effect of the metallicity-specific star formation history

- on double compact object mergers. *Monthly Notices of the Royal Astronomical Society*, 490:3740–3759, December 2019. ISSN 0035-8711. doi:10.1093/mnras/stz2840. URL <https://ui.adsabs.harvard.edu/abs/2019MNRAS.490.3740N>. Publisher: OUP ADS Bibcode: 2019MNRAS.490.3740N.
- Ken K. Y. Ng, Salvatore Vitale, Will M. Farr, and Carl L. Rodriguez. Probing Multiple Populations of Compact Binaries with Third-generation Gravitational-wave Detectors. *The Astrophysical Journal*, 913:L5, May 2021. ISSN 0004-637X. doi:10.3847/2041-8213/abf8be. URL <https://ui.adsabs.harvard.edu/abs/2021ApJ...913L...5N>. Publisher: IOP ADS Bibcode: 2021ApJ...913L...5N.
- Alexander H. Nitz and Collin D. Capano. GW190521 May Be an Intermediate-mass Ratio Inspiral. *The Astrophysical Journal*, 907(1):L9, January 2021. doi:10.3847/2041-8213/abccc5. URL <https://doi.org/10.3847/2041-8213/abccc5>. Publisher: American Astronomical Society.
- Brendan O’Brien, Marek Szczepanczyk, V. Gayathri, Imre Bartos, Gabriele Vedovato, Giovanni Prodi, Guenakh Mitselmakher, and Sergey Klimenko. Detection of LIGO-Virgo binary black holes in the pair-instability mass gap. *arXiv e-prints*, page arXiv:2106.00605, June 2021. _eprint: 2106.00605.
- Madeline Oh, Maya Fishbach, Chase Kimball, Vicky Kalogera, and Christine Ye. The Role of Natal Kicks in Forming Asymmetric Compact Binary Mergers. *The Astrophysical Journal*, 953:152, August 2023. ISSN 0004-637X. doi:10.3847/1538-4357/ace349. URL <https://ui.adsabs.harvard.edu/abs/2023ApJ...953..152O>. ADS Bibcode: 2023ApJ...953..152O.
- A. Olejak and K. Belczynski. The Implications of High Black Hole Spins for the Origin of Binary Black Hole Mergers. *The Astrophysical Journal Letters*, 921(1):L2, October 2021. ISSN 2041-8205. doi:10.3847/2041-8213/ac2f48. URL <https://dx.doi.org/10.3847/2041-8213/ac2f48>. Publisher: The American Astronomical Society.
- Feryal Özel, Dimitrios Psaltis, Ramesh Narayan, and Jeffrey E. McClintock. THE BLACK HOLE MASS DISTRIBUTION IN THE GALAXY. 725(2):1918–1927, December 2010. ISSN 0004-637X. doi:10.1088/0004-637X/725/2/1918. URL <https://doi.org/10.1088/0004-637x/725/2/1918>. Publisher: American Astronomical Society.
- Rachel A Patton, Tuguldur Sukhbold, and J J Eldridge. Comparing compact object distributions from mass- and presupernova core structure-based prescriptions. *Monthly Notices of the Royal Astronomical Society*, 511(1):903–913, March 2022. ISSN 0035-8711. doi:10.1093/mnras/stab3797. URL <https://doi.org/10.1093/mnras/stab3797>.
- S. Perlmutter and B. P. Schmidt. Measuring Cosmology with Supernovae. *Supernovae and Gamma-Ray Bursters*, 598:195–217, January 2003. doi:10.1007/3-540-45863-8_11. URL <https://ui.adsabs.harvard.edu/abs/2003LNP...598..195P>. ADS Bibcode: 2003LNP...598..195P.

- Du Phan, Neeraj Pradhan, and Martin Jankowiak. Composable Effects for Flexible and Accelerated Probabilistic Programming in NumPyro, December 2019. URL <https://ui.adsabs.harvard.edu/abs/2019arXiv191211554P>. Publication Title: arXiv e-prints ADS Bibcode: 2019arXiv191211554P.
- Grégoire Pierra, Simone Mastrogiovanni, Stéphane Perriès, and Michela Mapelli. Study of systematics on the cosmological inference of the Hubble constant from gravitational wave standard sirens. *Physical Review D*, 109:083504, April 2024. ISSN 1550-79980556-2821. doi:10.1103/PhysRevD.109.083504. URL <https://ui.adsabs.harvard.edu/abs/2024PhRvD.109h3504P>. Publisher: APS ADS Bibcode: 2024PhRvD.109h3504P.
- Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian processes for machine learning*. Adaptive computation and machine learning. MIT Press, Cambridge, Mass, 2006. ISBN 978-0-262-18253-9. OCLC: ocm61285753.
- Anarya Ray, Ignacio Magaña Hernandez, Siddharth Mohite, Jolien Creighton, and Shasvath Kapadia. Nonparametric Inference of the Population of Compact Binaries from Gravitational-wave Observations Using Binned Gaussian Processes. *The Astrophysical Journal*, 957:37, November 2023. ISSN 0004-637X. doi:10.3847/1538-4357/acf452. URL <https://ui.adsabs.harvard.edu/abs/2023ApJ...957...37R>. Publisher: IOP ADS Bibcode: 2023ApJ...957...37R.
- Brian Reich. Chapter 6: Hierarchical models, March 2020. URL <https://st540.wordpress.ncsu.edu/files/2020/03/Hier.pdf>.
- M. Renzo, R. Farmer, S. Justham, Y. Götzberg, S. E. de Mink, E. Zapartas, P. Marchant, and N. Smith. Predictions for the hydrogen-free ejecta of pulsational pair-instability supernovae. *Astronomy & Astrophysics*, 640:A56, August 2020. ISSN 0004-6361, 1432-0746. doi:10.1051/0004-6361/202037710. URL <http://arxiv.org/abs/2002.05077>. arXiv:2002.05077 [astro-ph].
- Jeff Riley, Poojan Agrawal, Jim W. Barrett, Kristan N. K. Boyett, Floor S. Broekgaarden, Debatri Chattopadhyay, Sebastian M. Gaebel, Fabian Gittins, Ryosuke Hirai, George Howitt, Stephen Justham, Lokesh Khandelwal, Floris Kummer, Mike Y. M. Lau, Ilya Mandel, Selma E. de Mink, Coenraad Neijssel, Tim Riley, Lieke van Son, Simon Stevenson, Alejandro Vigna-Gómez, Serena Vinciguerra, Tom Wagg, Reinhold Willcox, and Team Compas. Rapid Stellar and Binary Population Synthesis with COMPAS. *The Astrophysical Journal Supplement Series*, 258:34, February 2022. ISSN 0067-0049. doi:10.3847/1538-4365/ac416c. URL <https://ui.adsabs.harvard.edu/abs/2022ApJS..258...34R>. ADS Bibcode: 2022ApJS..258...34R.
- Stefano Rinaldi and Walter Del Pozzo. (H)DPGMM: a hierarchy of Dirichlet process Gaussian mixture models for the inference of the black hole mass function. *Monthly Notices of the Royal Astronomical Society*, 509:5454–5466, February 2022. ISSN 0035-8711. doi:10.1093/mnras/stab3224. URL <https://ui.adsabs.harvard.edu/abs/2022MNRAS.509.5454R>. ADS Bibcode: 2022MNRAS.509.5454R.

- Carl L. Rodriguez, Michael Zevin, Pau Amaro-Seoane, Sourav Chatterjee, Kyle Kremer, Frederic A. Rasio, and Claire S. Ye. Black holes: The next generation—repeated mergers in dense star clusters and their gravitational-wave properties. *Physical Review D*, 100:043027, August 2019. ISSN 1550-7998/56-2821. doi:10.1103/PhysRevD.100.043027. URL <https://ui.adsabs.harvard.edu/abs/2019PhRvD.100d3027R>. ADS Bibcode: 2019PhRvD.100d3027R.
- Milton Ruiz, Stuart L. Shapiro, and Antonios Tsokaros. GW170817, general relativistic magnetohydrodynamic simulations, and the neutron star maximum mass. *Physical Review D*, 97(2):021501, January 2018. doi:10.1103/PhysRevD.97.021501. URL <https://link.aps.org/doi/10.1103/PhysRevD.97.021501>. Publisher: American Physical Society.
- Jam Sadiq, Thomas Dent, and Daniel Wysocki. Flexible and fast estimation of binary merger population distributions with an adaptive kernel density estimator. *Physical Review D*, 105:123014, June 2022. ISSN 1550-7998/56-2821. doi:10.1103/PhysRevD.105.123014. URL <https://ui.adsabs.harvard.edu/abs/2022PhRvD.10513014S>. ADS Bibcode: 2022PhRvD.10513014S.
- Jam Sadiq, Thomas Dent, and Mark Gieles. Binary vision: The merging black hole binary mass distribution via iterative density estimation, July 2023. URL <http://arxiv.org/abs/2307.12092>. arXiv:2307.12092 [astro-ph, physics:gr-qc].
- Fabian R. N. Schneider, Philipp Podsiadlowski, and Eva Laplace. Bimodal Black Hole Mass Distribution and Chirp Masses of Binary Black Hole Mergers. *The Astrophysical Journal*, 950:L9, June 2023. ISSN 0004-637X. doi:10.3847/2041-8213/acd77a. URL <https://ui.adsabs.harvard.edu/abs/2023ApJ...950L...9S>. ADS Bibcode: 2023ApJ...950L...9S.
- Bernard F. Schutz. Determining the Hubble constant from gravitational wave observations. *Nature*, 323(6086):310–311, September 1986. ISSN 1476-4687. doi:10.1038/323310a0. URL <https://www.nature.com/articles/323310a0>. Number: 6086 Publisher: Nature Publishing Group.
- Sohan Seth, Iain Murray, and Christopher K. I. Williams. Model Criticism in Latent Space. *Bayesian Analysis*, 14(3):703 – 725, 2019. doi:10.1214/18-BA1124. URL <https://doi.org/10.1214/18-BA1124>. Publisher: International Society for Bayesian Analysis.
- Yong Shao and Xiang-Dong Li. Population Synthesis of Black Hole Binaries with Compact Star Companions. *The Astrophysical Journal*, 920(2):81, October 2021. ISSN 0004-637X, 1538-4357. doi:10.3847/1538-4357/ac173e. URL <http://arxiv.org/abs/2107.03565>. arXiv: 2107.03565.
- Masaru Shibata, Enping Zhou, Kenta Kiuchi, and Sho Fujibayashi. Constraint on the maximum mass of neutron stars using GW170817 event. *Physical Review D*, 100(2):023015, July 2019. doi:10.1103/PhysRevD.100.023015. URL <https://link.aps.org/doi/10.1103/PhysRevD.100.023015>. Publisher: American Physical Society.

- Jared C. Siegel, Ilia Kiato, Vicky Kalogera, Christopher P. L. Berry, Thomas J. Maccarone, Katelyn Breivik, Jeff J. Andrews, Simone S. Bavera, Aaron Dotter, Tassos Fragos, Konstantinos Kovelakas, Devina Misra, Kyle A. Rocha, Philipp M. Srivastava, Meng Sun, Zepei Xing, and Emmanouil Zapartas. Investigating the Lower Mass Gap with Low Mass X-ray Binary Population Synthesis, September 2022. URL <http://arxiv.org/abs/2209.06844>. arXiv:2209.06844 [astro-ph].
- Dan Simpson. Un garçon pas comme les autres (Bayes) - Priors part 4: Specifying priors that appropriately penalise complexity, September 2022. URL <https://dansblog.netlify.app/posts/2022-08-29-priors4/priors4.html#the-dream-pc-priors-in-practice>.
- Daniel Simpson, Håvard Rue, Andrea Riebler, Thiago G. Martins, and Sigrunn H. Sørbye. Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors. *Statistical Science*, 32(1):1–28, February 2017. ISSN 0883-4237, 2168-8745. doi:10.1214/16-STS576. URL <https://projecteuclid.org/journals/statistical-science/volume-32/issue-1/Penalising-Model-Component-Complexity--A-Principled-Practical-Approach-to/10.1214/16-STS576.full>. Publisher: Institute of Mathematical Statistics.
- Sandip Sinharay and Hal S Stern. Posterior predictive model checking in hierarchical models. *Journal of Statistical Planning and Inference*, 111(1):209–221, February 2003. ISSN 0378-3758. doi:10.1016/S0378-3758(02)00303-8. URL <https://www.sciencedirect.com/science/article/pii/S0378375802003038>.
- M. Soares-Santos, A. Palmese, W. Hartley, J. Annis, et al. First Measurement of the Hubble Constant from a Dark Standard Siren using the Dark Energy Survey Galaxies and the LIGO/Virgo Binary–Black-hole Merger GW170814. *The Astrophysical Journal Letters*, 876(1):L7, April 2019. ISSN 2041-8205. doi:10.3847/2041-8213/ab14f1. URL <https://dx.doi.org/10.3847/2041-8213/ab14f1>. Publisher: The American Astronomical Society.
- Joshua S. Speagle. DYNESTY: a dynamic nested sampling package for estimating Bayesian posteriors and evidences. *Monthly Notices of the Royal Astronomical Society*, 493:3132–3158, April 2020. ISSN 0035-8711. doi:10.1093/mnras/staa278. URL <https://ui.adsabs.harvard.edu/abs/2020MNRAS.493.3132S>. ADS Bibcode: 2020MNRAS.493.3132S.
- D. Spergel, N. Gehrels, C. Baltay, D. Bennett, J. Breckinridge, M. Donahue, A. Dressler, B. S. Gaudi, T. Greene, O. Guyon, C. Hirata, J. Kalirai, N. J. Kasdin, B. Macintosh, W. Moos, S. Perlmutter, M. Postman, B. Rauscher, J. Rhodes, Y. Wang, D. Weinberg, D. Benford, M. Hudson, W.-S. Jeong, Y. Mellier, W. Traub, T. Yamada, P. Capak, J. Colbert, D. Masters, M. Penny, D. Savransky, D. Stern, N. Zimmerman, R. Barry, L. Bartusek, K. Carpenter, E. Cheng, D. Content, F. Dekens, R. Demers, K. Grady, C. Jackson, G. Kuan, J. Kruk, M. Melton, B. Nemati, B. Parvin, I. Poberezhskiy, C. Peddie, J. Ruffa, J. K. Wallace, A. Whipple, E. Wollack, and F. Zhao. Wide-Field Infrared Survey Telescope–Astrophysics Focused Telescope Assets WFIRST-AFTA 2015 Report, March 2015. URL <http://arxiv.org/abs/1503.03757>. arXiv:1503.03757 [astro-ph].

- Michael L. Stein. *Interpolation of Spatial Data*. Springer Series in Statistics. Springer, New York, NY, 1999. ISBN 978-1-4612-7166-6 978-1-4612-1494-6. doi:10.1007/978-1-4612-1494-6. URL <http://link.springer.com/10.1007/978-1-4612-1494-6>.
- Simon Stevenson, Frank Ohme, and Stephen Fairhurst. Distinguishing Compact Binary Population Synthesis Models Using Gravitational Wave Observations of Coalescing Binary Black Holes. *The Astrophysical Journal*, 810:58, September 2015. ISSN 0004-637X. doi:10.1088/0004-637X/810/1/58. URL <https://ui.adsabs.harvard.edu/abs/2015ApJ...810...58S>. ADS Bibcode: 2015ApJ...810...58S.
- Simon Stevenson, Matthew Sampson, Jade Powell, Alejandro Vigna-Gómez, Coenraad J. Neijssel, Dorottya Szécsi, and Ilya Mandel. The impact of pair-instability mass loss on the binary black hole mass distribution. *The Astrophysical Journal*, 882(2):121, September 2019. ISSN 1538-4357. doi:10.3847/1538-4357/ab3981. URL <http://arxiv.org/abs/1904.02821>. arXiv:1904.02821 [astro-ph, physics:gr-qc].
- Hiromichi Tagawa, Bence Kocsis, Zoltán Haiman, Imre Bartos, Kazuyuki Omukai, and Johan Samsing. Mass-gap Mergers in Active Galactic Nuclei. *The Astrophysical Journal*, 908(2):194, February 2021. doi:10.3847/1538-4357/abd555. URL <https://doi.org/10.3847/1538-4357/abd555>. Publisher: American Astronomical Society.
- Colm Talbot and Eric Thrane. Measuring the Binary Black Hole Mass Spectrum with an Astrophysically Motivated Parameterization. *The Astrophysical Journal*, 856:173, April 2018. ISSN 0004-637X. doi:10.3847/1538-4357/aab34c. URL <https://ui.adsabs.harvard.edu/abs/2018ApJ...856..173T>. ADS Bibcode: 2018ApJ...856..173T.
- Colm Talbot, Rory Smith, Eric Thrane, and Gregory B. Poole. Parallelized inference for gravitational-wave astronomy. *Physical Review D*, 100:043030, August 2019. ISSN 1550-7998 0556-2821. doi:10.1103/PhysRevD.100.043030. URL <https://ui.adsabs.harvard.edu/abs/2019PhRvD.100d3030T>. ADS Bibcode: 2019PhRvD.100d3030T.
- Colm Talbot, Amanda Farah, Shanika Galaudage, Jacob Golomb, and Hui Tong. GW-Population: Hardware agnostic population inference for compact binaries and beyond, September 2024. URL <https://ui.adsabs.harvard.edu/abs/2024arXiv240914143T>. Publication Title: arXiv e-prints ADS Bibcode: 2024arXiv240914143T.
- N. R. Tanvir, A. J. Levan, C. González-Fernández, O. Korobkin, I. Mandel, S. Rosswog, J. Hjorth, P. D’Avanzo, A. S. Fruchter, C. L. Fryer, T. Kangas, B. Milvang-Jensen, S. Rosetti, D. Steeghs, R. T. Wollaeger, Z. Cano, C. M. Copperwheat, S. Covino, V. D’Elia, A. de Ugarte Postigo, P. A. Evans, W. P. Even, S. Fairhurst, R. Figuera Jaimes, C. J. Fontes, Y. I. Fujii, J. P. U. Fynbo, B. P. Gompertz, J. Greiner, G. Hodosan, M. J. Irwin, P. Jakobsson, U. G. Jørgensen, D. A. Kann, J. D. Lyman, D. Malesani, R. G. McMahon, A. Melandri, P. T. O’Brien, J. P. Osborne, E. Palazzi, D. A. Perley, E. Pian, S. Piranomonte, M. Rabus, E. Rol, A. Rowlinson, S. Schulze, P. Sutton, C. C. Thöne, K. Ulaczyk, D. Watson, K. Wiersema, and R. A. M. J. Wijers. The Emergence of a Lanthanide-rich Kilonova Following the Merger of Two Neutron Stars. *The Astrophysical*

- Journal*, 848:L27, October 2017. ISSN 0004-637X. doi:10.3847/2041-8213/aa90b6. URL <https://ui.adsabs.harvard.edu/abs/2017ApJ...848L..27T>. Publisher: IOP ADS Bibcode: 2017ApJ...848L..27T.
- Stephen R. Taylor and Jonathan R. Gair. Cosmology with the lights off: Standard sirens in the Einstein Telescope era. *Physical Review D*, 86(2):023502, July 2012. doi:10.1103/PhysRevD.86.023502. URL <https://link.aps.org/doi/10.1103/PhysRevD.86.023502>. Publisher: American Physical Society.
- The LIGO Scientific Collaboration and The Virgo Collaboration. Data Release for “Population Properties of Compact Objects from the Second LIGO-Virgo Gravitational-Wave Transient Catalog, 2020a. URL <https://dcc.ligo.org/LIGO-P2000434/public>.
- The LIGO Scientific Collaboration and The Virgo Collaboration. GW190521 Data Release, 2020b. URL https://www.gw-openscience.org/eventapi/html/03_Discovery_Papers/GW190521/v2/.
- The LIGO Scientific Collaboration and The Virgo Collaboration. GW190814 Data Release, 2020c. URL https://www.gw-openscience.org/eventapi/html/03_Discovery_Papers/GW190814/v1/.
- Owen Thomas and Jukka Corander. Diagnosing model misspecification and performing generalized Bayes’ updates via probabilistic classifiers. *arXiv e-prints*, page arXiv:1912.05810, December 2019. _eprint: 1912.05810.
- Todd A. Thompson, Christopher S. Kochanek, Krzysztof Z. Stanek, Carles Badenes, Richard S. Post, Tharindu Jayasinghe, David W. Latham, Allyson Bieryla, Gilbert A. Esquerdo, Perry Berlind, Michael L. Calkins, Jamie Tayar, Lennart Lindegren, Jennifer A. Johnson, Thomas W.-S. Holoiën, Katie Auchettl, and Kevin Covey. A non-interacting low-mass black hole–giant star binary system. *Science*, November 2019. doi:10.1126/science.aau4005. URL <https://www.science.org/doi/abs/10.1126/science.aau4005>. Publisher: American Association for the Advancement of Science.
- Eric Thrane and Colm Talbot. An introduction to Bayesian inference in gravitational-wave astronomy: Parameter estimation, model selection, and hierarchical models. *Publications of the Astronomical Society of Australia*, 36:e010, March 2019. ISSN 1323-3580. doi:10.1017/pasa.2019.2. URL <https://ui.adsabs.harvard.edu/abs/2019PASA...36...10T>. ADS Bibcode: 2019PASA...36...10T.
- Vaibhav Tiwari. VAMANA: modeling binary black hole population with minimal assumptions. *Classical and Quantum Gravity*, 38(15):155007, July 2021. ISSN 0264-9381. doi:10.1088/1361-6382/ac0b54. URL <https://dx.doi.org/10.1088/1361-6382/ac0b54>. Publisher: IOP Publishing.
- Vaibhav Tiwari. Exploring Features in the Binary Black Hole Population. *The Astrophysical Journal*, 928:155, April 2022. ISSN 0004-637X. doi:10.3847/1538-4357/ac589a.

URL <https://ui.adsabs.harvard.edu/abs/2022ApJ...928..155T>. ADS Bibcode: 2022ApJ...928..155T.

Vaibhav Tiwari. What's in a binary black hole's mass parameter?, April 2023. URL <https://ui.adsabs.harvard.edu/abs/2023arXiv230403498T>. Publication Title: arXiv e-prints ADS Bibcode: 2023arXiv230403498T.

Stefano Torniamenti, Michela Mapelli, Carole P  rigois, Manuel Arca Sedda, Maria Celeste Artale, Marco Dall'Amico, and Maria Paola Vaccaro. Hierarchical binary black hole mergers in globular clusters: Mass function and evolution with redshift. *Astronomy and Astrophysics*, 688:A148, August 2024. ISSN 0004-6361. doi:10.1051/0004-6361/202449272. URL <https://ui.adsabs.harvard.edu/abs/2024A&A...688A.148T>. ADS Bibcode: 2024A&A...688A.148T.

Stefano Valenti, David J. Sand, Sheng Yang, Enrico Cappellaro, Leonardo Tartaglia, Alessandra Corsi, Saurabh W. Jha, Daniel E. Reichart, Joshua Haislip, and Vladimir Kouprianov. The Discovery of the Electromagnetic Counterpart of GW170817: Kilonova AT 2017gfo/DLT17ck. *The Astrophysical Journal*, 848:L24, October 2017. ISSN 0004-637X. doi:10.3847/2041-8213/aa8edf. URL <https://ui.adsabs.harvard.edu/abs/2017ApJ...848L..24V>. Publisher: IOP ADS Bibcode: 2017ApJ...848L..24V.

L. A. C. van Son, S. E. de Mink, T. Callister, S. Justham, M. Renzo, T. Wagg, F. S. Broekgaarden, F. Kummer, R. Pakmor, and I. Mandel. The Redshift Evolution of the Binary Black Hole Merger Rate: A Weighty Matter. *The Astrophysical Journal*, 931:17, May 2022a. ISSN 0004-637X. doi:10.3847/1538-4357/ac64a3. URL <https://ui.adsabs.harvard.edu/abs/2022ApJ...931...17V>. ADS Bibcode: 2022ApJ...931...17V.

L. A. C. van Son, S. E. de Mink, M. Renzo, S. Justham, E. Zapartas, K. Breivik, T. Callister, W. M. Farr, and C. Conroy. No Peaks without Valleys: The Stable Mass Transfer Channel for Gravitational-wave Sources in Light of the Neutron Star-Black Hole Mass Gap. *The Astrophysical Journal*, 940:184, December 2022b. ISSN 0004-637X. doi:10.3847/1538-4357/ac9b0a. URL <https://ui.adsabs.harvard.edu/abs/2022ApJ...940..184V>. Publisher: IOP ADS Bibcode: 2022ApJ...940..184V.

L. A. C. van Son, S. E. de Mink, M. Chru  li  nska, C. Conroy, R. Pakmor, and L. Hernquist. The Locations of Features in the Mass Distribution of Merging Binary Black Holes Are Robust against Uncertainties in the Metallicity-dependent Cosmic Star Formation History. *The Astrophysical Journal*, 948:105, May 2023. ISSN 0004-637X. doi:10.3847/1538-4357/acbf51. URL <https://ui.adsabs.harvard.edu/abs/2023ApJ...948..105V>. ADS Bibcode: 2023ApJ...948..105V.

Aki Vehtari, Andrew Gelman, and Jonah Gabry. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, (5):1413–1432, 2017. doi:10.1007/s11222-016-9696-4. URL <https://doi.org/10.1007/s11222-016-9696-4>.

- J. Veitch, V. Raymond, B. Farr, W. Farr, P. Graff, S. Vitale, B. Aylott, K. Blackburn, N. Christensen, M. Coughlin, W. Del Pozzo, F. Feroz, J. Gair, C. J. Haster, V. Kalogera, T. Littenberg, I. Mandel, R. O’Shaughnessy, M. Pitkin, C. Rodriguez, C. Röver, T. Sidery, R. Smith, M. Van Der Sluys, A. Vecchio, W. Voudsen, and L. Wade. Parameter estimation for compact binaries with ground-based gravitational-wave observations using the LALInference software library. *Physical Review D*, 91: 042003, February 2015. ISSN 1550-79980556-2821. doi:10.1103/PhysRevD.91.042003. URL <https://ui.adsabs.harvard.edu/abs/2015PhRvD..91d2003V>. Publisher: APS ADS Bibcode: 2015PhRvD..91d2003V.
- Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C. J. Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, and Paul van Mulbregt. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3): 261–272, March 2020. ISSN 1548-7105. doi:10.1038/s41592-019-0686-2. URL <https://www.nature.com/articles/s41592-019-0686-2>. Number: 3 Publisher: Nature Publishing Group.
- Salvatore Vitale, Davide Gerosa, Will M. Farr, and Stephen R. Taylor. Inferring the properties of a population of compact binaries in presence of selection effects. *arXiv e-prints*, page arXiv:2007.05579, July 2020. _eprint: 2007.05579.
- Eric-Jan Wagenmakers, Tom Lodewyckx, Himanshu Kuriyal, and Raoul Grasman. Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive Psychology*, 60(3):158–189, May 2010. ISSN 00100285. doi:10.1016/j.cogpsych.2009.12.001. URL <https://linkinghub.elsevier.com/retrieve/pii/S0010028509000826>.
- Kaze W. K. Wong and Miles Cranmer. Automated discovery of interpretable gravitational-wave population models, July 2022. URL <http://arxiv.org/abs/2207.12409>. arXiv:2207.12409 [astro-ph, physics:gr-qc].
- S. E. Woosley. Pulsational Pair-Instability Supernovae. *The Astrophysical Journal*, 836(2): 244, February 2017. ISSN 1538-4357. doi:10.3847/1538-4357/836/2/244. URL <http://arxiv.org/abs/1608.08939>. arXiv:1608.08939 [astro-ph].
- S. E. Woosley. The Evolution of Massive Helium Stars Including Mass Loss. *The Astrophysical Journal*, 878(1):49, June 2019. ISSN 1538-4357. doi:10.3847/1538-4357/ab1b41. URL <http://arxiv.org/abs/1901.00215>. arXiv:1901.00215 [astro-ph].
- Stan. E. Woosley and Alexander Heger. The Deaths of Very Massive Stars. 412:199, January 2015. doi:10.1007/978-3-319-09596-7_7. URL <https://ui.adsabs.harvard.edu/abs/>

2015ASSL..412..199W. Conference Name: Very Massive Stars in the Local Universe
Place: eprint: arXiv:1406.5657 ADS Bibcode: 2015ASSL..412..199W.

Łukasz Wyrzykowski and Ilya Mandel. Constraining the masses of microlensing black holes and the mass gap with Gaia DR2. *Astronomy & Astrophysics*, 636:A20, April 2020. ISSN 0004-6361, 1432-0746. doi:10.1051/0004-6361/201935842. URL <http://arxiv.org/abs/1904.07789>. arXiv: 1904.07789.

Christine Ye and Maya Fishbach. Inferring the Neutron Star Maximum Mass and Lower Mass Gap in Neutron Star-Black Hole Systems with Spin. *The Astrophysical Journal*, 937:73, October 2022. ISSN 0004-637X. doi:10.3847/1538-4357/ac7f99. URL <https://ui.adsabs.harvard.edu/abs/2022ApJ...937...73Y>. ADS Bibcode: 2022ApJ...937...73Y.

Claire S. Ye and Maya Fishbach. The Redshift Evolution of the Binary Black Hole Mass Distribution from Dense Star Clusters. *The Astrophysical Journal*, 967:62, May 2024. ISSN 0004-637X. doi:10.3847/1538-4357/ad3ba8. URL <https://ui.adsabs.harvard.edu/abs/2024ApJ...967...62Y>. Publisher: IOP ADS Bibcode: 2024ApJ...967...62Y.

Michael Zevin and Simone S. Bavera. Suspicious Siblings: The Distribution of Mass and Spin Across Component Black Holes in Isolated Binary Evolution. *The Astrophysical Journal*, 933(1):86, July 2022. ISSN 0004-637X, 1538-4357. doi:10.3847/1538-4357/ac6f5d. URL <http://arxiv.org/abs/2203.02515>. arXiv:2203.02515 [astro-ph, physics:gr-qc].

Michael Zevin and Daniel E. Holz. Avoiding a Cluster Catastrophe: Retention Efficiency and the Binary Black Hole Mass Spectrum. *The Astrophysical Journal*, 935:L20, August 2022. ISSN 0004-637X. doi:10.3847/2041-8213/ac853d. URL <https://ui.adsabs.harvard.edu/abs/2022ApJ...935L..20Z>. Publisher: IOP ADS Bibcode: 2022ApJ...935L..20Z.

Michael Zevin, Chris Pankow, Carl L. Rodriguez, Laura Sampson, Eve Chase, Vassiliki Kalogera, and Frederic A. Rasio. Constraining Formation Models of Binary Black Holes with Gravitational-wave Observations. *The Astrophysical Journal*, 846:82, September 2017. ISSN 0004-637X. doi:10.3847/1538-4357/aa8408. URL <https://ui.adsabs.harvard.edu/abs/2017ApJ...846...82Z>. ADS Bibcode: 2017ApJ...846...82Z.

Michael Zevin, Mario Spera, Christopher P. L. Berry, and Vicky Kalogera. Exploring the Lower Mass Gap and Unequal Mass Regime in Compact Binary Evolution. *The Astrophysical Journal*, 899:L1, August 2020. ISSN 0004-637X. doi:10.3847/2041-8213/aba74e. URL <https://ui.adsabs.harvard.edu/abs/2020ApJ...899L...1Z>. ADS Bibcode: 2020ApJ...899L...1Z.

Michael Zevin, Simone S. Bavera, Christopher P. L. Berry, Vicky Kalogera, Tassos Fragos, Pablo Marchant, Carl L. Rodriguez, Fabio Antonini, Daniel E. Holz, and Chris Pankow. One Channel to Rule Them All? Constraining the Origins of Binary Black Holes Using Multiple Formation Pathways. *The Astrophysical Journal*, 910(2):152, April 2021. doi:10.3847/1538-4357/abe40e. URL <https://doi.org/10.3847/1538-4357/abe40e>. Publisher: American Astronomical Society.