



Constraining Formation Models of Binary Black Holes with Gravitational-wave Observations

Michael Zevin¹, Chris Pankow¹ , Carl L. Rodriguez², Laura Sampson¹,
 Eve Chase^{1,3}, Vassiliki Kalogera¹ , and Frederic A. Rasio¹

¹ Center for Interdisciplinary Exploration and Research in Astrophysics (CIERA) and Dept. of Physics and Astronomy,
 Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, USA

² MIT-Kavli Institute for Astrophysics and Space Research, 77 Massachusetts Avenue, 37-664H, Cambridge, MA 02139, USA

Received 2017 June 16; revised 2017 July 24; accepted 2017 August 1; published 2017 September 1

Abstract

Gravitational waves (GWs) from binary black hole (BBH) mergers provide a new probe of massive-star evolution and the formation channels of binary compact objects. By coupling the growing sample of BBH systems with population synthesis models, we can begin to constrain the parameters of such models and glean unprecedented knowledge about the inherent physical processes that underpin binary stellar evolution. In this study, we apply a hierarchical Bayesian model to mass measurements from a synthetic GW sample to constrain the physical prescriptions in population models and the relative fraction of systems generated from various channels. We employ population models of two canonical formation scenarios in our analysis—isolated binary evolution involving a common-envelope phase and dynamical formation within globular clusters—with model variations for different black hole natal kick prescriptions. We show that solely with chirp mass measurements, it is possible to constrain natal kick prescriptions and the relative fraction of systems originating from each formation channel with $\mathcal{O}(100)$ of confident detections. This framework can be extended to include additional formation scenarios, model parameters, and measured properties of the compact binary.

Key words: galaxies: star clusters: general – gravitational waves – methods: statistical – stars: black holes – supernovae: general

1. Introduction

Recent observations of gravitational waves (GWs) have launched a new branch of observational astronomy. The confident detections of binary black hole (BBH) mergers GW 150914, GW 151226, and GW 170104, as well as BBH candidate LVT 151012 by the advanced LIGO (aLIGO) detectors marked the discovery of BBH systems in our universe, and enticed deeper exploration of massive-star evolution (Aasi et al. 2013; Abbott et al. 2016a, 2016b, 2016c; The LIGO Scientific Collaboration et al. 2017). The final stages of BBH evolution enable the measurement of their physical properties, which connects us to their preceding history and can potentially constrain the environments responsible for facilitating BBH formation, the relative fraction of systems produced through various formation channels, and the physical processes underlying binary stellar evolution. With expected BBH merger rates ranging from 12 to 213 $\text{Gpc}^{-3} \text{yr}^{-1}$ (The LIGO Scientific Collaboration et al. 2016, 2017), it is an opportune time to develop methods for utilizing all existing and future observations to constrain and inform astrophysical models.

Two canonical formation channels are generally considered for contributing to the full population of BBHs: isolated binary evolution (i.e., “the field”) and dynamical formation (i.e., “clusters”). In the isolated evolution scenario, binaries are predicted to evolve and tighten through a common-envelope phase (e.g., Voss & Tauris 2003; Dominik et al. 2013; Belczynski et al. 2014), or through chemically homogeneous evolution of close binaries that attain rapid rotation (de Mink & Mandel 2016; Marchant et al. 2016; Mandel & de Mink 2016).

Alternatively, the dynamical channel predicts BBHs that become bound and tighten through three-body encounters in dense star clusters such as globular clusters (e.g., Portegies Zwart & Mcmillan 2000; Downing et al. 2010, 2011; Rodriguez et al. 2015, 2016a), galactic nuclei, and AGN disks (Antonini & Rasio 2016; Bartos et al. 2017; Stone et al. 2017), or young stellar clusters (Ziosi et al. 2014; Chatterjee et al. 2017). In addition to these canonical scenarios, more exotic formation channels have been suggested for facilitating BBH mergers, such as field triples tightened by Lidov–Kozai cycles (Antonini et al. 2017; Silsbee & Tremaine 2017), primordially formed black holes (Bird et al. 2016), or remnants of population III stars (Inayoshi et al. 2016). While these models maintain the ability to predict heavy BBHs such as GW 150914 (e.g., Belczynski et al. 2016; Rodriguez et al. 2016b; Stevenson et al. 2017b), the rates and property distributions predicted from population synthesis simulations are highly sensitive to the prescriptions chosen for uncertain physical processes such as black hole natal kick prescriptions, wind mass-loss, and common envelope physics (Dominik et al. 2012; Stevenson et al. 2015).

The analysis of compact binary populations through GW observations provides a unique and powerful mechanism for determining the models that describe the true underlying BBH population. By pairing measured BBH properties from the growing sample of GW observations with population synthesis models that account for various formation scenarios and physical prescriptions, constraints can be placed on the relative fraction of systems produced by each formation channel (i.e., “branching ratio”) and the inherent physical processes that underpin binary stellar evolution. As the merger rates predicted by various channels are highly uncertain and overlapping, this approach to astrophysical model selection in the context of

³ LSSTC Data Science Fellow.

GWs primarily explores the distributions of the BBH source parameters. Neglecting eccentricity and finite size effects, the masses and spins of component black holes primarily determine the GW signal during the evolution of the BBH through inspiral, merger, and ringdown (Abbott et al. 2016d), allowing for parameter estimation of these quantities by comparing a measured GW signal with template waveforms generated from a sample of the physical parameters (Cutler & Flanagan 1994; Poisson & Will 1995; Veitch et al. 2015).

One confounding aspect of this model selection problem is the uncertain physical mechanisms underlying population modeling of binary compact objects. To circumvent the intricacies and uncertainties of binary evolution modeling, studies such as Mandel et al. (2017) have taken an agnostic, model-independent approach toward model selection by developing methods for distinguishing populations through clustering of source parameters, such as black hole masses. Though such an approach can help to identify multiple populations, it lacks the ability to directly identify the physical processes inherent to stellar evolution models. To this end, Stevenson et al. (2015) assessed the potential of using GW observations for differentiating population synthesis models that have various prescriptions for common envelope binding energy, maximum neutron star mass, black hole natal kick prescriptions, and stellar winds, finding that certain models could be ruled out in the near-future given expected merger rates. However, the inclusion of alternative formation channels would complicate this process.

Several studies, such as Vitale et al. (2017) and Stevenson et al. (2017a), have performed model selection to infer branching ratios using BBH spin distributions, finding that one may converge upon the branching ratio between field and cluster formation channels with dozens to hundreds of detections. Though spin distributions for differing population models can be constructed purely by geometrical arguments, isolated binary evolution and dynamical formation rely on vastly different procedures for compact binary formation and evolution, and physically motivated modeling is therefore required for accurate and comparable distributions in mass and redshift. Using such models, Rodriguez et al. (2016c) found that certain combinations of masses and spins can be produced exclusively by dynamical formation channels, and the detection of such an outlier could be a clear indication of this formation process. Furthermore, Farr et al. (2017) demonstrated that the clustering of effective spin measurements in the current catalog of GW observations hints at an isotropic spin angle distribution rather than an aligned one. However, there is still much work to be done in utilizing population synthesis and catalogs of GW observations to infer properties of the true compact binary populations.

In this paper, we present an approach of hierarchical model selection that utilizes physically motivated models of BBH populations from multiple environments to infer underlying physical prescriptions and branching ratios between formation channels. In particular, we focus on the utility of chirp mass measurements for inferring black hole natal kick prescriptions and the branching ratio between isolated binary evolution and dynamical formation models. However, this approach can easily scale to include more measured BBH properties, additional population models, and submodels accounting for different uncertain physical prescriptions.

The outline for the paper is as follows. In Section 2, we discuss the BBH population models used in this analysis, which model field binaries (Section 2.1) and cluster binaries (Section 2.2), accounting for selection biases (Section 2.3). Section 3 outlines the algorithm for hierarchical model selection (Section 3.1), the mock observations and analytical approximations for measurement uncertainty (Section 3.2), and sampling procedure (Section 3.3). In Section 4, we discuss our inference on branching ratio (Section 4.1) and kick prescription (Section 4.2) through this methodology. We conclude in Section 5 with a discussion of the analysis and future prospects.

2. BBH Population Models

In this section, we describe the population models used in our analysis. These models are identical to those used in Rodriguez et al. (2016c), except that the various natal kick prescriptions used as submodels of the field population are also incorporated into the globular cluster models. Our inference relies on output parameter distributions of BBHs that merge as potential LIGO sources, such as component masses, spin-tilts, and redshifts.

2.1. Isolated Binary Populations

Populations of field binaries are generated using an upgraded version of the binary evolution code BSE (Hurley et al. 2002). This code rapidly models stellar populations using metallicity-dependent fits for single-star evolution, while also modeling binary interactions such as mass transfer and changes in orbital angular momentum from black hole natal kicks. Furthermore, BSE now uses a radius-dependent fitting formula for the common envelope binding energy parameter λ (Claeys et al. 2014). Our modifications to this code implement physical prescriptions from more recent work related to stellar winds and supernova prescriptions. These include mass-loss prescriptions for O and B stars known as the “Vink prescription” (Vink et al. 2001), metallicity dependence to the evolution of Wolf-Rayet stars (Vink & de Koter 2005), and prescriptions for the supernova mechanism developed in Fryer et al. (2012). Despite these upgrades, many of the physical mechanisms governing binary stellar evolution are still poorly constrained, and incorrect physical prescriptions may propagate inaccuracies to the physical parameter distributions of our populations. We evolve 10^5 binaries in 11 metallicity bins ranging from 0.005 to $1.5 Z_{\odot}$, with masses from 18 to $150 M_{\odot}$ drawn from an initial mass function of $p(m) dm \propto m^{-2.3} dm$ (see, e.g., Kroupa 2001). Mass ratios are drawn from a uniform distribution on the interval $[0, 1]$, and initial semimajor axes are drawn from a distribution flat in log space on the interval 10^{-10} – $10^5 R_{\odot}$. All binaries that evolve into BBHs are maintained as GW candidates, scaled appropriately by their merger time.

While the following scenarios may not cover all the physical uncertainties in population models, they do provide a representative sample of possibilities. As such, we consider 12 permutations of physical assumptions that all affect the final population parameters:

1. Three different natal kick prescription models, imparting different amounts of linear velocity to the newly formed black holes in the binary. One model, the *fallback* kick prescription, assumes that some fraction of the mass

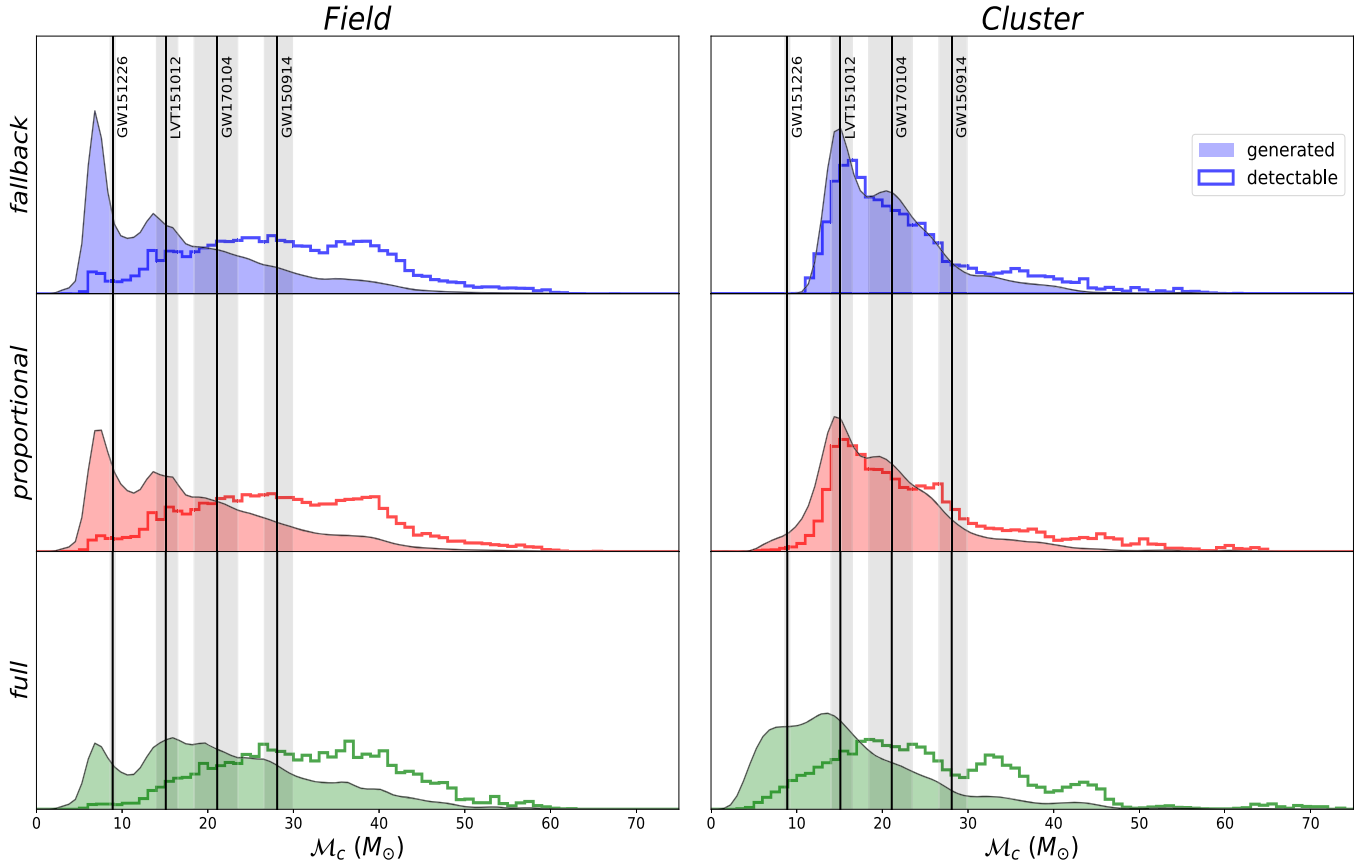


Figure 1. Chirp mass distributions for the field and cluster population models. Each panel shows the independently normalized distributions of sources generated (filled histogram) and sources weighted by detectability (unfilled histogram). For reference, the chirp masses of the four likely gravitational-wave events (GW 150914, GW 151226, GW 170104, and LVT 151012) are plotted, with the outer lines representing the 90% credible region. The top, middle, and bottom panels show the distributions for fallback, proportional, and full natal kick prescriptions, respectively. We construct each model using one kick magnitude prescription, comprised of equal abundances from the four submodels described in items 2 and 3 of Section 2.1.

ejected during core collapse will “fallback” on the black hole:

$$V_{\text{kick}}^{\text{BH}} = (1 - f_{\text{fallback}}) V_{\text{kick}}^{\text{NS}}. \quad (1)$$

The fraction of material that falls back is proportional to the core mass of the black hole progenitor. The second model, the *proportional* kick prescription, assumes that the kick imparted to the black hole is reduced by the ratio of the neutron star mass to the black hole mass:

$$V_{\text{kick}}^{\text{BH}} = \frac{m_{\text{NS}}}{m_{\text{BH}}} V_{\text{kick}}^{\text{NS}}, \quad (2)$$

where we assume $m_{\text{NS}} = 2.5 M_{\odot}$ for all systems, as this value represents the hypothetical “boundary” between neutron stars and black holes in most population synthesis codes. The final kick prescription, called the *full* kick prescription, assumes that the black hole kick is equal to the full kick velocity imparted on the neutron star:

$$V_{\text{kick}}^{\text{BH}} = V_{\text{kick}}^{\text{NS}}. \quad (3)$$

- Two differing kick directions. In one model we assume kicks are isotropically distributed in solid angle around the exploding star, which is the common assumption in population models. However, observations of pulsars have suggested a correlation between the kick direction and spin axis (Kaplan et al. 2008), motivating the inclusion of a polar kick prescription where the kicks are

confined to 10° cones about the rotational axis of the progenitor star.

- Two different methods of accounting for uncertainties in the realignment of the component spin axes after the first supernova. One model allows for realignment of the binary after the first kick, whereas the other model does not realign. Though this does not have an effect on the mass distributions of the field population models, it has a substantial effect on the spin distributions of the resultant BBHs.

All these variations in model assumptions largely affect the resultant spin-tilt distributions of the binaries. However, only kick magnitudes play a substantial role in the final distribution of BBH chirp masses. As seen in Figure 1, stronger kick prescriptions flatten out the relative abundance of low-mass binaries in field models; these systems acquire larger linear velocities from the kicks, allowing the kinetic energy of the binary component to more easily overcome the gravitational potential and become unbound. As this paper focuses on chirp mass measurements, we construct each model using one kick magnitude prescription, and equal abundances from the four submodels that are described in items 2 and 3 of Section 2.1. Furthermore, we expect only one kick magnitude to be true, whereas kick direction and binary realignment prescriptions may be dependent on processes such as stellar rotation. Future work will incorporate spin measurements in the inference and address these submodels with more detail.

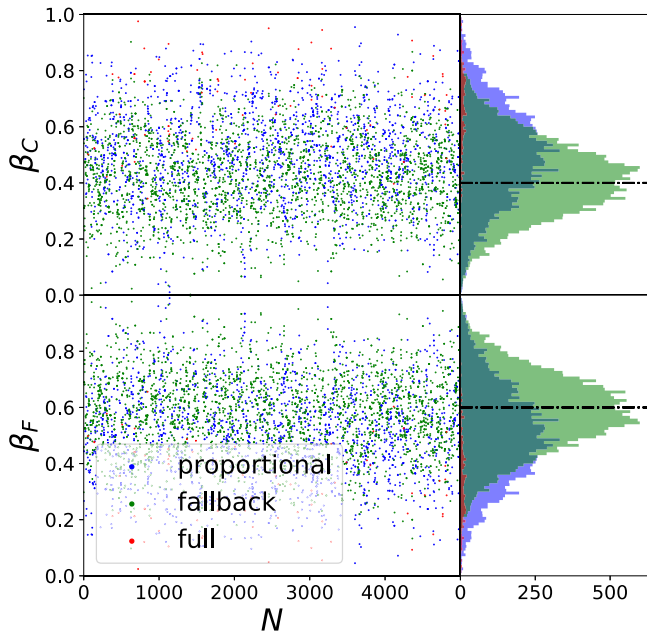


Figure 2. Example realization of the sampling, where 5000 samples are drawn from the RJMCMC chain. This particular realization is for 100 observations from the fallback model, with a cluster branching ratio of $\beta = 0.4$. β_C and β_F are the fraction of systems that are drawn from the cluster and field populations, respectively. The left panels show the value of β inferred for each step in the sampling, with colors indicating the model chosen. The right panels show the total binned histograms.

2.2. Cluster Binary Populations

In this study, we consider the “classical” channel of dynamical formation in old, metal-poor globular clusters. Cluster binaries are drawn from a few dozen globular cluster models generated using the Cluster Monte Carlo (CMC) code (see, e.g., Chatterjee et al. 2010). Black holes sink to the centers of globular clusters due to dynamical friction, fostering mass segregation through energy equipartition and turning the globular cluster cores into dynamical factories for forming heavy stellar-mass BBHs. Though these models are sensitive to initial conditions, dynamically formed binaries rely on single-star evolution and N -body dynamics rather than binary stellar evolution and therefore maintain fewer uncertainties in the physical processes (e.g., common envelope evolution, mass transfer) involved in generating BBHs (Rodríguez et al. 2016a). However, as the choice of kick magnitude prescription may alter the distribution of black holes within a cluster or eject black holes entirely, we include cluster submodels with the same variations in kick prescription presented for the field models in Section 2.1. This ensures a putative population mixing has a consistent kick prescription. As seen in Figure 1, the stronger kick magnitude prescriptions tend to flatten out peaks in the normalized chirp mass distributions for the cluster population as well as the field population. Note that in the cluster case the natal kicks do not act to disrupt individual binary systems, as the final partners are usually found long after the components have evolved into compact objects.

The globular cluster models generated in Rodríguez et al. (2016a) used the fallback prescription for black hole natal kicks described above. To create equivalent populations using the proportional and full kick prescriptions, we implement the following approximate procedure: for each of the 48 globular cluster models from Rodríguez et al. (2016a), we take the initial

($t = 0$) snapshot of the cluster, and evolve the massive stars (above $18 M_\odot$) forward with BSE until the stars have completed their evolution and the initial population of black holes has formed. We then record the velocity of the natal kick, and only retain those black holes for which $V_{\text{kick}} < \sqrt{-2\Phi(r)}$, where Φ is the gravitational potential of the cluster, and r is the initial radial position of the star in the cluster.⁴

Once we have an initial population of black holes for each cluster model, we proceed to eject black holes and BBHs from our synthetic population, assuming that the rate at which black holes are ejected is identical to that found with CMC.⁵ We select black hole masses without replacement from a list of $N_{\text{black holes}}$, according to

$$p(i)di \propto i^2 di \text{ for } 1 < i < N/3, \quad (4)$$

where i is the index of the list of black holes, sorted in order of decreasing mass. Equation (4) is physically motivated by the fact that globular clusters preferentially eject the most massive black holes first, continuing to eject black holes until depletion (Morscher et al. 2015). The functional form of $p(i)$ was found through trial-and-error to reproduce the masses and mass ratios of ejected BBHs from the CMC simulations. For every BBH, we also remove four single black holes from the list (see, e.g., Heggie & Hut 1993). The reasoning for this is that once a binary is nearing the hardness necessary to eject it from the cluster, the scatterings it undergoes will eject single objects with the same strong three-body encounters responsible for hardening and ejecting the BBH. However, since the binary is about twice as massive as the single objects that are scattered, there are a few scatterings where the velocity of the single object surpasses the cluster escape speed, while the velocity of the binary does not. Numerical tests indicate that on average 3–4 single black holes are ejected for each binary that is ejected. Finally, at the time of ejection, we set the eccentricity and semimajor axis of the binary using the half-mass-radius and mass of the cluster at the time of ejection, according to:

$$P(e)de = 2e de, \quad (5)$$

$$P(a|M_{GC}, R_h, \mu_{\text{bin}})da = \frac{1}{a\sigma\sqrt{2\pi}} \times \exp\left[-\frac{\left(\log \frac{\mu_{\text{bin}} R_h}{aM_{GC}} - a^*\right)^2}{2\sigma^2}\right] da, \quad (6)$$

where $\mu_{\text{bin}} = (m_1 m_2)/(m_1 + m_2)$ is the reduced mass of the binary, and a^* and σ are the parameters of a log-normal distribution with mean $a^* = 3.98$ and $\sigma = 0.59$ (see Rodríguez et al. 2016a, Equations (7) and (8)). The merger time of each binary is computed by adding the time each binary is ejected (assuming all globular clusters to be 12 Gyr old) to the GW

⁴ Because the black hole-formation timescale for massive stars (~ 5 Myr) is significantly smaller than the mass-segregation timescale (~ 100 Myr, see O’Leary et al. 2006), we can safely ignore the change in position of the star between birth and black hole-formation.

⁵ This assumption is well justified, as it is the total energy flux of the cluster, not of the black hole sub-system, that determines the ejection rate of black holes from a globular cluster (Breen & Heggie 2013). However, for clusters where there are not sufficiently many black holes to meet the energy requirements of the cluster (e.g., the full kick prescription), this assumption will overestimate the early ejection rate of black holes.

merger time from Peters (1964). The redshift assigned to each merger is the redshift at that cosmological lookback time, assuming a flat Λ CDM cosmology with $\Omega_M = 0.306$ and $H_0 = 67.9 \text{ km s}^{-1} \text{ Mpc}^{-1}$ (Planck Collaboration et al. 2016).

Once we have a population of ejected BBHs from clusters with different kick prescriptions, we resample the population of BBHs to better represent what we expect to see from globular clusters in the local universe. First, we draw a population of binaries from our effective globular cluster models by preferentially selecting binaries from globular clusters closer to the peak of the observed globular cluster mass function (i.e., massive globular cluster models that more-closely resemble the population of observed globular clusters in the local universe; see Rodriguez et al. 2016a). We then take this population of BBHs from globular clusters, where we have assumed a universal globular cluster mass function and constant spatial density of globular clusters, and create a 3D kernel-density estimate (KDE) of the binary mergers in m_1 , m_2 , and redshift. We then draw as many binaries as we want from this distribution using an MCMC (Foreman-mackey et al. 2013), with the KDE as our likelihood and a prior on the redshift, which is uniform in comoving volume.

The stronger kick prescriptions (i.e., full kicks) retain more low-mass binaries relative to their high-mass counterparts. This is because in the full kick case the natal kick velocity does not decrease with increasing black hole mass, contrary to fallback and proportional kicks (see Equations (1)–(3)). Therefore, the full kick prescription will kick out all black holes from the cluster with equal likelihood regardless of the black hole mass, whereas the velocity of the kick is stifled for higher-mass objects in the fallback and proportional cases, allowing more of these objects to be retained relative to their lower-mass analogs. However, as the stronger kick prescriptions will cause more newly formed black holes to be ejected from the cluster, it also results in a decrease to the overall merger rate. The above procedure is necessary to generate new BBH populations without having to generate new, computationally expensive models of massive globular clusters. We use this approximate method for all three populations, including the fallback prescription (for which we do have complete CMC models from Rodriguez et al. 2016a). This was done to avoid any systematic differences that our approximate technique may have introduced, and it was found that this method matched well with the true fallback population generated from CMC. Finally, we assume that the isotropic and polar kick models for clusters should be identical, since BH retention in clusters should be independent of kick angle.

2.3. Incorporation of Selection Biases

The distributions described above represent all BBH systems that are generated by these populations in the local universe. As the detectability of a given source is dependent on both its physical and orientation parameters (e.g., masses, spins, redshift, frequency content, detector network antenna pattern, inclination), the distribution of observed parameters will be different from the true source distribution. Therefore, we translate the raw source distributions into distributions of detectable sources by the expected design-sensitivity power spectrum and antenna pattern of a single detector assuming isotropic sky location and inclination distributions. A signal-to-

noise (S/N) threshold of 8 is applied, defined by

$$\rho^2 = 4\Re \int \frac{\tilde{h}^*(f)\tilde{h}(f)}{S_n(f)} df, \quad (7)$$

where $\tilde{h}(f)$ is the gravitational waveform in the frequency domain and $S_n(f)$ is the one-sided power spectral density of the noise.⁶ We then calculate the probability of a system with a given mass and redshift passing this threshold and then weight the distributions accordingly.⁷ As seen in the right panel of Figure 1, this tends to flatten out the low-mass peaks and amplify the number of higher-mass systems.

3. Model Selection

With population models in hand, we leverage BBH mass measurements to infer properties of the underlying distribution. The two questions we aim to address in this paper are as follows.

1. Given a catalog of N BBH chirp mass measurements from GW observations that come from a population made up of field and cluster binaries with a particular branching ratio (β), how well can one discern the inherent black hole natal kick prescription?
2. Assuming one prescription is correct, how many observations are required to confidently converge on the true value of this branching ratio?

We now describe the machinery behind this inference.

3.1. Hierarchical Modeling

As described by Mandel (2010) and Hogg et al. (2010), among many others, the objective of hierarchical modeling is to infer a set of model parameters λ given N observations $\{x_i\}$, which are characterized by a set of physical parameters $\{\theta_i\}$ and constrained by prior assumptions $\{\alpha_i\}$. The astrophysical model described by parameters λ gives a probability distribution for physical parameters, in our case chirp masses. By Bayes' theorem, the posterior on λ is

$$p(\lambda|\theta) = \frac{p(\theta|\lambda)p(\lambda)}{p(\theta)}, \quad (8)$$

where $p(\theta|\lambda)$ is the likelihood of observing a particular set of physical parameters, $p(\lambda)$ is the prior on the model parameters, and $p(\theta)$ is a normalization constant.

However, as we observe N independent GW signals rather than the physical parameters directly, we rewrite the likelihood as

$$p(\{x_i\}|\lambda) = \prod_{i=1}^N p(x_i|\lambda) = \prod_{i=1}^N \int d\theta p(x_i|\theta)p(\theta|\lambda). \quad (9)$$

⁶ For S/N calculations, we assume non-spinning component black holes and utilize the IMRPhenomPv2 waveform approximate (Khan et al. 2016).

⁷ Though the cluster models maintain redshift information, our field populations do not. We assume for simplicity that field binaries are distributed uniformly in comoving volume in the local universe and sample redshifts accordingly.

Again applying Bayes' Theorem, we write $p(x_i|\theta)$ as $p(\theta|x_i)p(x_i)/p(\theta)$ to get

$$p(\{x_i\}|\lambda) = \prod_{i=1}^N p(x_i) \int d\theta \frac{p(\theta|x_i)p(\theta|\lambda)}{p(\theta)}, \quad (10)$$

where $p(\theta)$ is the prior on the physical parameters that are used to generate the posterior samples.

We approximate the integral as a discrete sum over posterior samples

$$\int d\theta p(\theta|x_i)f(\theta) \approx \frac{1}{S} \sum_{k=1}^S f(\theta_k) \quad (11)$$

and ignore the multiplicative constant $p(x_i)$, as it is not a function of θ or λ and will not affect the sampling of the posterior. Therefore, the full expression for the likelihood that we wish to sample is

$$p(\{x_i\}|\lambda) = \prod_{i=1}^N \frac{1}{S} \sum_{k=1}^S \frac{p(\theta_k|\lambda)}{p(\theta_k)}, \quad (12)$$

where, again, N is the number of observed events, S is the number of posterior samples, θ_k are the astrophysical parameters, and λ are the model parameters.

We aim to do model selection between different kick prescriptions while simultaneously performing inference on branching ratios between field and cluster formation channels. The parameters of our astrophysical model are thus $\lambda = (\iota, \beta)$, where ι is an indexing parameter that indicates the kick prescription ($\iota \in [0, 1, 2]$ where 0, 1, and 2 designate the proportional, fallback, and full kick prescriptions, respectively) and β is the branching ratio parameter, defined as the fraction of observations that are drawn from cluster models ($0 \leq \beta \leq 1$). An illustrative example of this inference is shown in Figure 2.

As the branching ratios between the various formation channels are highly uncertain, we maintain minimal assumptions on our prior knowledge of the model parameters. For the prior on β , we use a Dirichlet distribution, which is a multivariate generalization of the beta distribution. This allows for minimal prior assumptions while ensuring the values of β for all channels sum to unity. Though X-ray binary observations (Repetto et al. 2017) and the current catalog of BBH observations (Belczynski et al. 2016) provide moderate evidence for certain black hole natal kick prescriptions, we use a uniform prior on the kick prescription, which for this discrete parameter puts equal weight on each prescription.

3.2. Mock Observations

We represent the chirp mass distributions for our populations with a Gaussian KDE and draw “observations” from this model. In practice, the observations themselves are manifested as a set of samples drawn from a posterior computed for each candidate event. Instead of employing very accurate, but computationally expensive, Markovian methods to estimate the parameter posteriors, we instead use the Fisher matrix as a proxy for the inverse covariance of a simpler Gaussian parameter distribution (see Cutler & Flanagan 1994). We are justified in this procedure in the case of chirp mass ($\mathcal{M}_c$), because it determines the leading-order evolution of GWs from compact binary coalescence and is therefore the best-measured physical property from a GW signal (see Abbott et al. 2016d, and references therein). However, this methodology is less

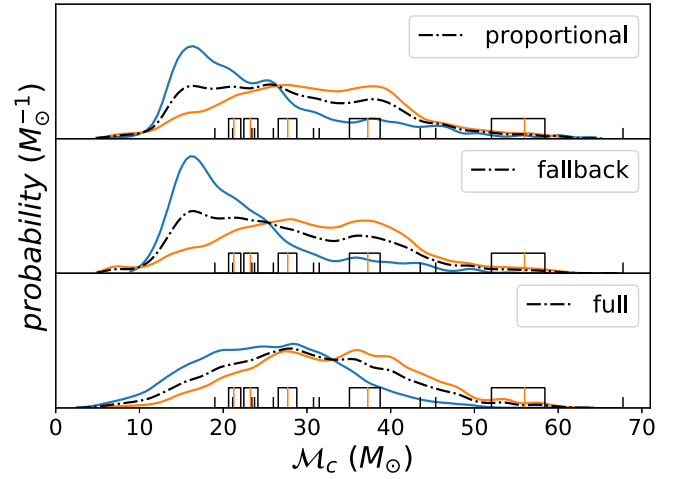


Figure 3. Mock observations drawn from a population specified by a kick prescription and branching ratio. This particular realization draws five observations from the fallback kick population model with $\beta = 0.4$. Each panel shows the distributions for a given kick prescription, where the blue and orange lines represent the field and cluster models, respectively, and the dashed black line shows the (normalized) combined population given the value of β . The box-and-whisker plots at the bottom of each panel show the median value of the posterior samples for an observation with an orange line, the upper and lower quartile of these samples as the edges of the box, and the maximum and minimum value of the posterior samples as whiskers. In this case, all observations are drawn from the fallback population model (dotted line in the middle panel).

accurate for parameters that are more correlated and less constrained, such as effective spin and symmetric mass ratio (see, e.g., Vallisneri 2008; Rodriguez et al. 2013; O’Shaughnessy et al. 2014).

As the Fisher matrix tends to overestimate the distributional width in $\mathcal{M}_c$, this also provides a conservative estimate for the true measurement uncertainty. However, as chirp mass measurements are highly constrained, we find that the inclusion of measurement uncertainty does not drastically affect our results. Nonetheless, we draw 100 mock posterior samples from a Gaussian distribution with a mean centered on the true value and standard deviation $\sigma_{\mathcal{M}_c}$ as each “observation”.

N smeared observations are drawn from a “true” distribution, described by a particular natal kick prescription and a value of the branching ratio β (such that $N_{\text{cluster}} = \beta N$ and $N_{\text{field}} = (1 - \beta)N$). Figure 3 shows one realization of this procedure. We then use these mock observations as the basis behind our statistical inference through hierarchical modeling.

3.3. Sampling

The technique we use for sampling the posterior on the model parameters is Reverse Jump Markov Chain Monte Carlo (RJMCMC; see e.g., Green 1995). In this method, the calculation of the Bayes factor between two models does not require the explicit calculation of the evidence integral. Rather, the model itself becomes a parameter of the chain. Depending on which value an indexing parameter takes, the likelihood and prior are evaluated using one of a set of models, which may or may not be of the same dimensionality. Said another way, the sampler first “jumps” in model index space, and then estimates the value for β within the particular model it lands. We assume that all models contain the same set of model parameters with the same meaning—in other words, the branching ratio β is the fraction between field and cluster in all models. The samples

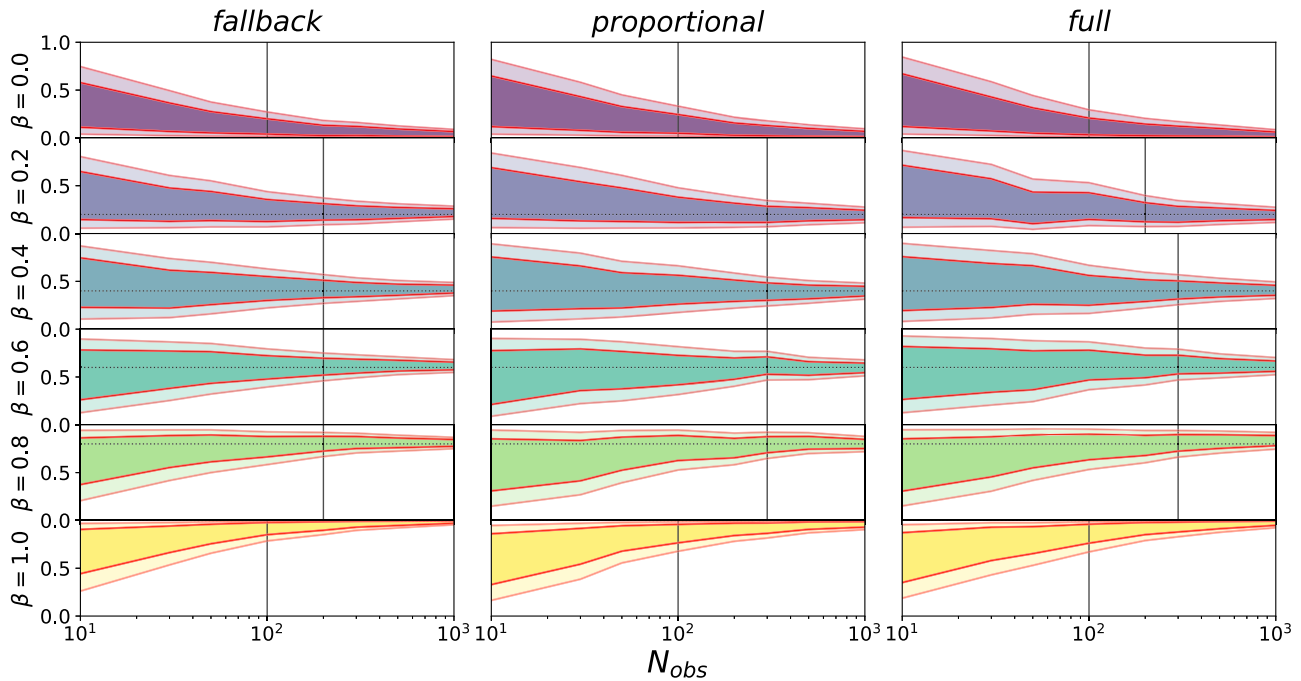


Figure 4. Convergence on the true value of β as a function of number of observations. The dark and light shadings represent the 68% and 90% credible intervals, respectively. The black vertical lines show the point in our discrete samplings of N_{obs} at which the 68% credible interval for β is constrained to less than 20% of the full range of β ; due to our discrete sampling this threshold is in fact reached before this point. The left, center, and right panels show the convergence on β when the injected kick prescription is fallback, proportional, and full, respectively. Note that the convergence rate varies depending both on the injected value of β and the injected kick prescription.

can then be sorted by model index at the end in order to generate posteriors for the individual kick prescriptions. Since our models have the same number of parameters with the same priors, the Bayes factor is then simply the ratio of the number of iterations that the chain spends in each model:

$$B_{ij} = \frac{\# \text{ of iterations in model } i}{\# \text{ of iterations in model } j}. \quad (13)$$

We use the open source MCMC python library `emcee` (Foreman-Mackey et al. 2013) for the implementation of this algorithm.

4. Astrophysical Inference with Chirp Mass Measurements

We now seek to utilize this inference for constraining branching ratios and properties of stellar evolution, given a catalog of BBH observations. Because the current number of BBH observations is likely too few to make any substantial claim about formation channels or physical prescriptions, we demonstrate the method using a mock catalog of BBH observations. We then apply one kick prescription model and formation channel branching ratio value to be the “correct” description of nature, and gauge how well one can converge on these injected values over an increasing number of observed systems. For the purposes of this study, we consider 100 realizations for each combination of the branching ratio β , kick prescription ι , and number of observations N_{obs} .

4.1. Branching Ratios

Convergence to the true branching ratio is a strong function of the number of observations drawn from the true population, and is also sensitive to the injected branching ratio itself. To summarize the convergence as a function of observations, we plot the marginalized posterior on β for different values of β in

Figure 4. To demonstrate this convergence in another way, Figure 5 shows the injected value of β for different combinations of β and N_{obs} , (i.e., many realizations of the sampling visualized in Figure 2), as well as the median value of the 100 realizations for each combination. As expected, there is rapid convergence on the true value of β as the number of observations increases, as well as smaller variance in the individual realizations. From Figure 5, it is also noted that the inference on β is unbiased.

Other model parameters aside, with only chirp mass measurements we converge on the true branching ratio to an accuracy of $\pm 10\%$ with $\mathcal{O}(100)$ observations of BBH systems. This convergence on branching ratio is similar to that found in Vitale et al. (2017) and Stevenson et al. (2017a) using population models with varied spin distributions; with dozens to a hundred observations, we will begin to see strong convergence on the true value of the branching ratio if BBH observations are dominated by the two canonical channels. Cumulative distribution functions of the searched intervals for β are shown in Figure 6. When only one mode in the distribution is present, the searched interval refers to the distance between the mode of the distribution and its true value. Quantitatively, when one channel dominates the overall event rate, convergence on the true value is noticeably better.

4.2. Natal Kick Model Selection

Our methodology also provides inference on the model index—that is, the underlying physical prescriptions assumed in the models. For the purposes of this study, the only physical prescription altered between models was the black hole natal kick magnitude. Since all models have the same parameters and prior ranges, the Bayes factor for one model compared to

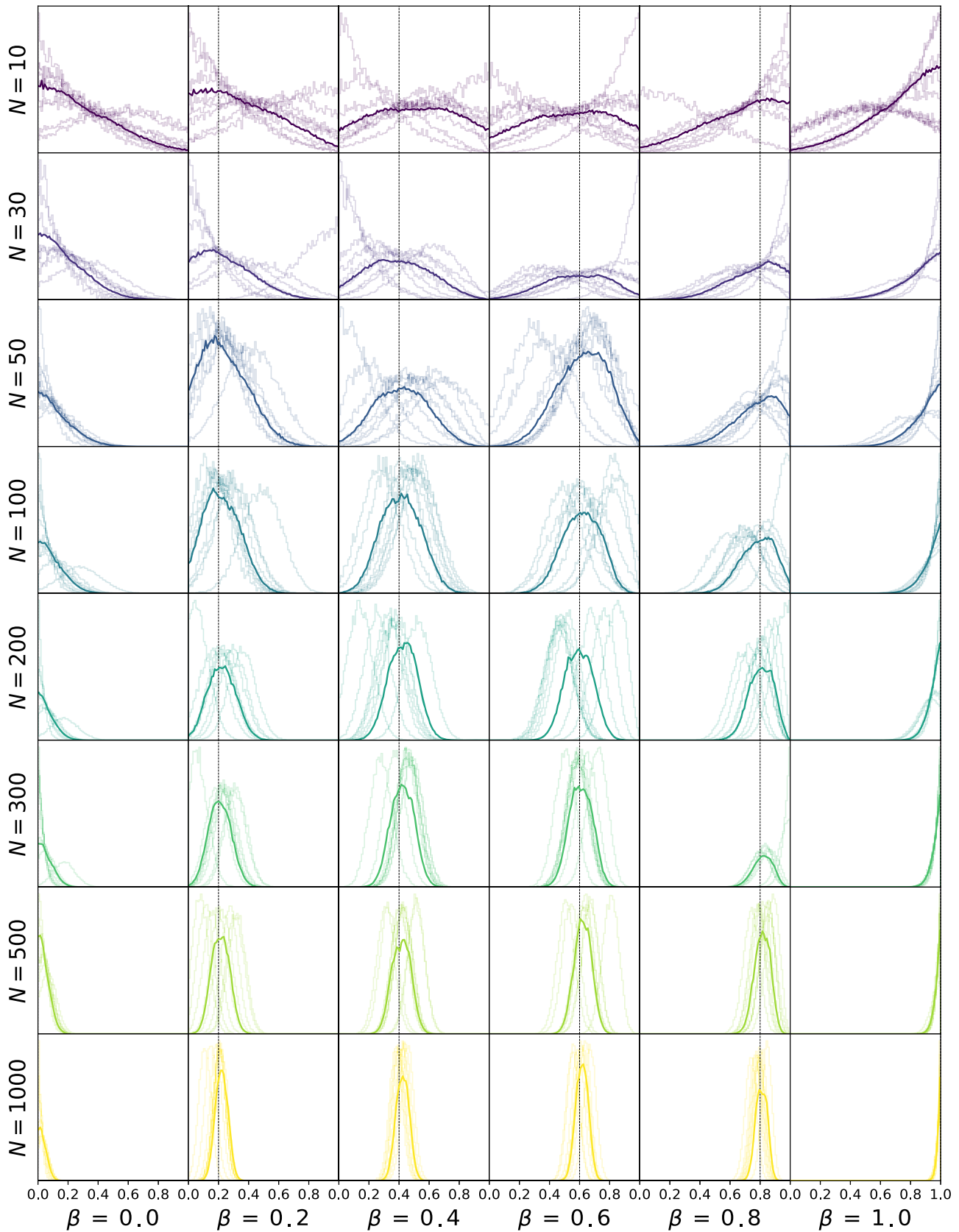


Figure 5. Convergence on the branching ratio β for various number of observations and injected values of β . Colors represent the number of observations, and are identical to the colors used in Figure 4. All observations are drawn from a population in which the “fallback” model is deemed true. In each panel, the 10 realizations of the inference on β are shown as faded lines, and the dark line is the median value of 100 realizations at each bin of the histogram. The dashed black vertical line marks the injected value of β .

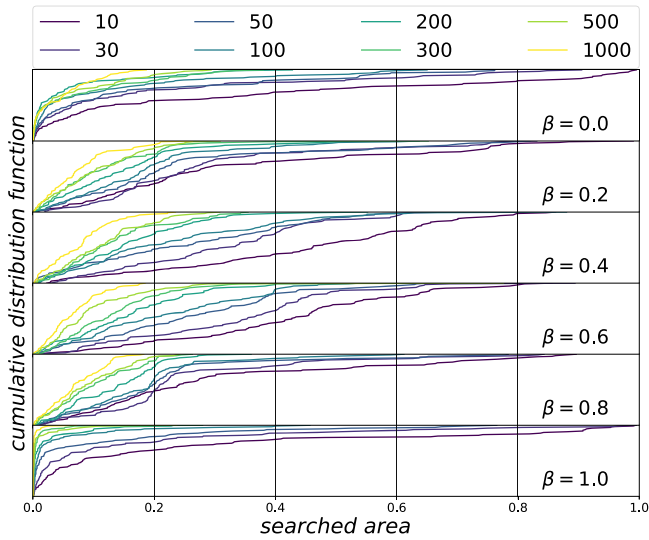


Figure 6. Cumulative distribution function on searched area for various injected values of β and number of observations (colored lines). The searched area represents the distance between the mode of the distribution and its true value. The searched area for significant N_{obs} is notably smaller when one channel dominates, e.g., $\beta \sim 1$ or $\beta \sim 0$.

another can be simply computed as the ratio of the number of iterations the chain spends in each model.

As the kick prescription has a noticeably different effect on the distribution of chirp masses in detectable cluster models relative to detectable field models (see Figure 1), the confidence for one prescription relative to another is a strong function of the branching ratio, as well as the number of observations. For example, branching ratios closer to $\beta = 1$ draw more observations from the cluster models, which have more distinctive features in the physical parameter distributions of detectable binaries compared to field models and allow for easier discrimination between populations. Furthermore, the growth of Bayes factors as a function of N_{obs} is expedited when comparing two kick prescriptions with dramatically different effects on the physical parameter distributions. This can also be seen in Figure 7: the Bayes factor between fallback kick and full kick increases much more rapidly than the Bayes factor between fallback kick and proportional kick.

We achieve Bayes factors between natal kick prescriptions of ~ 20 as the number of observations reaches $\mathcal{O}(100)$, though, as can be seen in Figure 7, the rate of increase of Bayes factors is extremely sensitive to the injected value of the branching ratio. Given the predicted discovery rates and projected interferometer sensitivity increases in the next few years, this indicates that we can begin to confidently infer the natal kick from supernovae within the lifetime of a design-sensitivity interferometer network. Even if the detection rate remains low, ruling out one physical model compared to others is within reach, especially given the conservative assumptions we have made. As population models are parameterized by many other discrete and continuous variables, we can expand this analysis to constrain other uncertain physical prescriptions of population synthesis using upcoming GW observations. Furthermore, this hierarchical approach does not require the models to be parameterized in the exact same way.

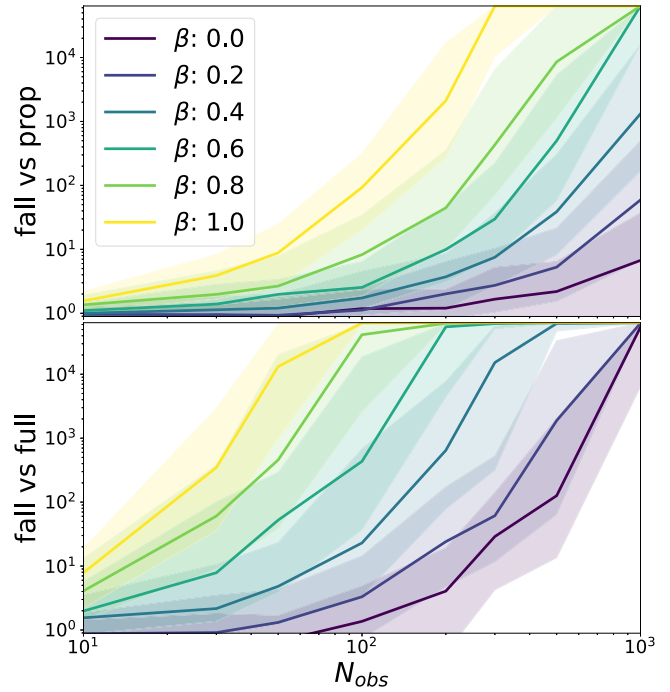


Figure 7. Bayes factors between different models as a function of the number of observations. The dark lines show the median value, and shaded regions a 68% credible interval derived from the 100 realizations. For this figure, observations were drawn from the “fallback” kick prescription, and various injected values of β are shown with different colors. The top and bottom panels show the Bayes factor between fallback kicks and proportional kicks, and the Bayes factor between fallback kicks and full neutron star kicks, respectively. The upper limit of the plot is the maximum Bayes factor calculable given our number of samples, and leads to the apparent flattening of the function near this maximal value. The Bayes factor between fallback and full kick models increases much more rapidly than the Bayes factor between fallback and proportional kick models as a function of N_{obs} , since the chirp mass distributions produced by fallback and full kick models are morphologically much more distinct.

5. Discussion and Conclusions

With the detection rates predicted for the advanced network of GW observatories, we can look forward to dozens to hundreds of BBH observations in the next decade. These systems provide a unique tool for studying massive-star evolution and the environments in which BBH systems arise, and by pairing a catalog of detections with detailed population models we can begin to constrain many of the uncertain processes driving binary stellar evolution.

This work investigates how hierarchical modeling can infer the parameters of binary stellar evolution from multiple formation channels using solely chirp mass measurements. We find that with $\mathcal{O}(100)$ observations, we will see convergence on the value for the branching ratio and the preferred natal kick prescription, provided the two channels considered dominate BBH rates. Furthermore, “extreme” values of β (i.e., domination by a single formation channel) facilitate quicker convergence on both the branching ratio and natal kick prescription.

Notably, we find a trade-off between inference on the branching ratio and inference on the natal kick prescription. This effect is dependent on the injected kick prescription itself; models with the fallback prescription shows the largest disparity between field and cluster models, thereby allowing quicker convergence on branching ratio, whereas models using

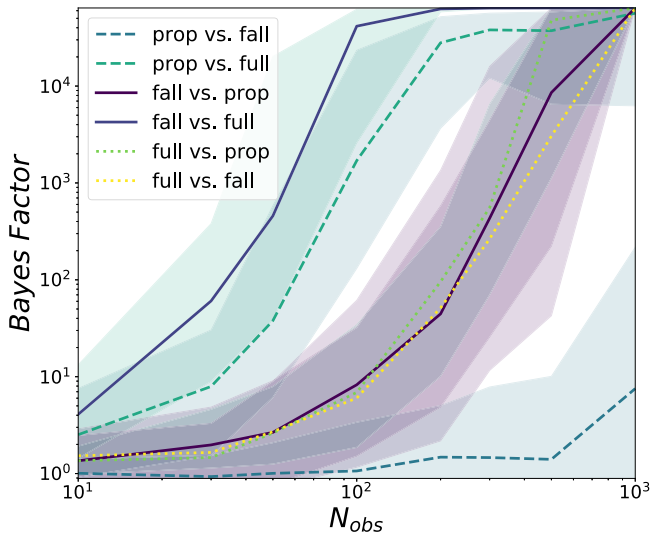


Figure 8. Bayes factors between different models as a function of N_{obs} , with observations drawn from models with different injected kick prescriptions. Dotted lines, dashed lines, and solid lines represent full kicks, proportional kicks, and fallback kicks as the correct distribution from which observations are drawn, respectively. All models in this plot have an injected value of $\beta = 0.8$. The median values and credible regions are indicated as in Figure 7. Interestingly, we find that when the proportional kick model is injected as the “true” model, our inference does not necessarily prefer proportional kicks as a function of N_{obs} . This is particularly apparent when comparing the Bayes factor between proportional kicks and fallback kicks, when observations are drawn from the proportional kick model. Though the populations are strikingly similar, we believe this issue arises from our conservative approximation of measurement uncertainty. Our approximations, which rely on Fisher matrix formalism for determining the spread of the posterior distribution, provide symmetric widths in our sampled posterior to both lower and higher values for $\mathcal{M}_c$. This may be unrealistically bolstering posterior samples in the low-mass peak of the fallback models (see Figure 3). Furthermore, as we limit to δ -function observations, this effect disappears.

the full kick prescription are less distinguishable and require more observations to converge on the branching ratio, as seen in Figure 4. However, as full kicks predict drastically different combined distributions of field and cluster populations relative to fallback and proportional kicks, comparisons with this model allow for accelerated inference on kick prescription, as we demonstrate in Figure 8.

Using our methodology, the current number of GW observations from BBH systems is far fewer than the number of observations needed to make any meaningful statement about kick prescription and branching ratio. Nonetheless, we inject the chirp masses of the three current GW events and one GW candidate from the first and second aLIGO observing runs of as our observations, using the 90% credible intervals for chirp mass measurements cited in Abbott et al. (2016a) and The LIGO Scientific Collaboration et al. (2017) to generate mock posterior samples. As expected, our analysis recovers the priors for both quantities and provides no discernment on branching ratio and true natal kick prescription.

The methodologies in this paper provide a framework for many extensions and refinements, both in the context of inferring additional parameters of population models and including more measured properties of BBH systems. One very simple extension, for example, is to measure the actual physical event rates, and compare with the rates derived from GW observations. Another possibility would be to constrain parameters that define a mass gap, either at low masses, between the maximum observed neutron star mass and

minimum observed black hole mass, or at high masses due to pair instability supernovae.

Multiple studies, such as Stevenson et al. (2017a) and Vitale et al. (2017), have also found that inference using spin distributions can converge on branching ratios with a similar number of GW observations. Though subject to higher measurement uncertainty and highly influenced by an unknown spin magnitude distribution, the inclusion of spin parameters in our analysis could help distinguish models, as dynamical formation predicts an isotropic distribution in spin-tilts, whereas isolated field binaries are believed to preserve the memory of their initial spin alignment and can further align their spins through mass transfer, common envelope evolution, and tidal torquing. Furthermore, kick prescriptions have a large effect on spin distributions and the inclusion of spin parameters could act to bolster the confidence of one kick model compared to another. Notably, the detection of a single outlier event, such as those described in Rodriguez et al. (2016c), could go a long way toward discriminating between models.

The inclusion of spins would also allow inference on other poorly constrained model parameters of population synthesis; though they have minimal impact on the mass distribution, the direction of the natal kick or partial realignment of the binary after the first supernova, for example, strongly affect the resultant spin distributions in population synthesis models. As the models used in this study are equipped with spin-tilts for the BBH systems, we plan to include spin measurements for purposes of model selection in future work. However, we note that other mechanisms unaccounted for in our models likely affect the spin-tilts of the binary components as well, and may act to contaminate the information we extract about the physical processes that are accounted for.

Besides taking a hierarchical approach to model selection using chirp mass distributions, this study provides the framework for determining Bayes factors between a discrete set of population models. In future work, we plan to expand this methodology to include other population synthesis model parameters, such as common envelope efficiency, natal kick direction, and the rate of binary coalescences as a function of redshift. Furthermore, the framework is extensible, allowing the inclusion of other proposed formation channel models, such as young stellar clusters, galactic nuclei, and chemically homogeneous evolution. As more BBHs are observed by the advanced network of GW detectors, this type of inference will evolve into a powerful tool for constraining the correct physical prescriptions in population synthesis models, thereby improving our understanding of the physical processes governing binary stellar evolution.

We would like to thank Simon Stevenson, Steven Reyes, and our anonymous referee for helpful suggestions on this manuscript, as well as Scott Coughlin for assistance in debugging code. This work was supported by NSF Grant AST-1312945, NSF Grant PHY-1607709, and NASA Grant NNX14AP92G at Northwestern University. M.Z. greatly appreciates financial support from the IDEAS Fellowship, a research traineeship program supported by the National Science Foundation under grant DGE-1450006. C.R. is grateful for the hospitality of the Kavli Institute for Theoretical Physics, supported by NSF Grant PHY11-25915, and is supported at MIT by a Pappalardo Fellowship in Physics. L.S. acknowledges support from the L’Oreal FWIS Fellowship program. E.C. thanks the LSSTC

Data Science Fellowship Program; her time as a fellow has benefited this work. V.K. and F.A.R. also acknowledge support from NSF Grant PHY-1066293 at the Aspen Center for Physics. The majority of our analysis was performed using the computational resources of the Quest high performance computing facility at Northwestern University, which is jointly supported by the Office of the Provost, the Office for Research, and Northwestern University Information Technology. This paper has been assigned LIGO document number ligo-P1700064.

ORCID iDs

Chris Pankow  <https://orcid.org/0000-0002-1128-3662>
 Vassiliki Kalogera  <https://orcid.org/0000-0001-9236-5469>
 Frederic A. Rasio  <https://orcid.org/0000-0002-7132-418X>

References

- Aasi, J., Abadie, J., Abbott, B. P., et al. 2013, *PhRvD*, **88**, 062001
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016a, *PhRvX*, **6**, 041015
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016b, *PhRvL*, **116**, 241103
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016c, *PhRvL*, **116**, 061102
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016d, *PhRvL*, **116**, 241102
- Antonini, F., & Rasio, F. A. 2016, *ApJ*, **831**, 187
- Antonini, F., Toonen, S., & Hamers, A. S. 2017, *ApJ*, **841**, 77
- Bartos, I., Kocsis, B., Haiman, Z., & Márka, S. 2017, *ApJ*, **835**, 165
- Belczynski, K., Buonanno, A., Cantiello, M., et al. 2014, *ApJ*, **789**, 120
- Belczynski, K., Holz, D. E., Bulik, T., & O’Shaughnessy, R. 2016, *Natur*, **534**, 512
- Bird, S., Cholis, I., Muñoz, J. B., et al. 2016, *PhRvL*, **116**, 201301
- Breen, P. G., & Heggie, D. C. 2013, *MNRAS*, **436**, 584
- Chatterjee, S., Fregeau, J. M., Umbreit, S., & Rasio, F. A. 2010, *ApJ*, **719**, 915
- Chatterjee, S., Rodriguez, C. L., Kalogera, V., & Rasio, F. A. 2017, *ApJL*, **836**, L26
- Claeys, J. S. W., Pols, O. R., Izzard, R. G., Vink, J., & Verbunt, F. W. M. 2014, *A&A*, **563**, A83
- Cutler, C., & Flanagan, É E. 1994, *PhRvD*, **49**, 2658
- de Mink, S. E., & Mandel, I. 2016, *MNRAS*, **460**, 3545
- Dominik, M., Belczynski, K., Fryer, C., et al. 2012, *ApJ*, **759**, 52
- Dominik, M., Belczynski, K., Fryer, C., et al. 2013, *ApJ*, **779**, 72
- Downing, J. M. B., Benacquista, M. J., Giersz, M., & Spurzem, R. 2010, *MNRAS*, **407**, 1946
- Downing, J. M. B., Benacquista, M. J., Giersz, M., & Spurzem, R. 2011, *MNRAS*, **416**, 133
- Farr, W. M., Stevenson, S., Miller, M. C., et al. 2017, arXiv:1706.01385
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, **125**, 306
- Fryer, C. L., Belczynski, K., Wiktorowicz, G., et al. 2012, *ApJ*, **749**, 91
- Green, P. J. 1995, *Biometrika*, **82**, 711
- Heggie, D. C., & Hut, P. 1993, *ApJS*, **85**, 347
- Hogg, D. W., Myers, A. D., & Bovy, J. 2010, *ApJ*, **725**, 2166
- Hurley, J. R., Tout, C. A., & Pols, O. R. 2002, *MNRAS*, **329**, 897
- Inayoshi, K., Kashiwama, K., Visbal, E., & Haiman, Z. 2016, *MNRAS*, **461**, 2722
- Kaplan, D. L., Chatterjee, S., Gaensler, B. M., & Anderson, J. 2008, *ApJ*, **677**, 1201
- Khan, S., Husa, S., Hannam, M., et al. 2016, *PhRvD*, **93**, 044007
- Kroupa, P. 2001, *MNRAS*, **322**, 231
- Mandel, I. 2010, *PhRvD*, **81**, 084029
- Mandel, I., & de Mink, S. E. 2016, *MNRAS*, **458**, 2634
- Mandel, I., Farr, W. M., Colonna, A., et al. 2017, *MNRAS*, **465**, 3254
- Marchant, P., Langer, N., Podsiadlowski, P., Tauris, T. M., & Moriya, T. J. 2016, *A&A*, **588**, A50
- Morscher, M., Pattabiraman, B., Rodriguez, C., Rasio, F. A., & Umbreit, S. 2015, *ApJ*, **800**, 9
- O’Leary, R. M., Rasio, F. A., Fregeau, J. M., Ivanova, N., & O’Shaughnessy, R. 2006, *ApJ*, **637**, 937
- O’Shaughnessy, R., Farr, B., Ochsner, E., et al. 2014, *PhRvD*, **89**, 064048
- Peters, P. C. 1964, *PhRv*, **136**, 1224
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, *A&A*, **594**, A13
- Poisson, E., & Will, C. M. 1995, *PhRvD*, **52**, 848
- Portegies Zwart, S. F., & McMillan, S. L. W. 2000, *ApJL*, **528**, L17
- Repetto, S., Igoshev, A. P., & Nelemans, G. 2017, *MNRAS*, **467**, 298
- Rodriguez, C. L., Chatterjee, S., & Rasio, F. A. 2016a, *PhRvD*, **93**, 084029
- Rodriguez, C. L., Farr, B., Farr, W. M., & Mandel, I. 2013, *PhRvD*, **88**, 084013
- Rodriguez, C. L., Haster, C.-J., Chatterjee, S., Kalogera, V., & Rasio, F. A. 2016b, *ApJL*, **824**, L8
- Rodriguez, C. L., Morscher, M., Pattabiraman, B., et al. 2015, *PhRvL*, **115**, 051101
- Rodriguez, C. L., Zevin, M., Pankow, C., Kalogera, V., & Rasio, F. A. 2016c, *ApJL*, **832**, L2
- Silsbee, K., & Tremaine, S. 2017, *ApJ*, **836**, 39
- Stevenson, S., Berry, C. P. L., & Mandel, I. 2017a, *MNRAS*, **471**, 2801
- Stevenson, S., Ohme, F., & Fairhurst, S. 2015, *ApJ*, **810**, 58
- Stevenson, S., Vigna-Gómez, A., Mandel, I., et al. 2017b, *NatCo*, **8**, 14906
- Stone, N. C., Metzger, B. D., & Haiman, Z. 2017, *MNRAS*, **464**, 946
- The LIGO Scientific Collaboration, the Virgo Collaboration, Abbott, B. P., et al. 2016, *ApJL*, **833**, L1
- The LIGO Scientific Collaboration, the Virgo Collaboration, Abbott, B. P., et al. 2017, *PhRvL*, **118**, 221101
- Vallisneri, M. 2008, *PhRvD*, **77**, 042001
- Veitch, J., Raymond, V., Farr, B., et al. 2015, *PhRvD*, **91**, 042003
- Vink, J. S., & de Koter, A. 2005, *A&A*, **442**, 587
- Vink, J. S., de Koter, A., & Lamers, H. J. G. L. M. 2001, *A&A*, **369**, 574
- Vitale, S., Lynch, R., Sturani, R., & Graff, P. 2017, *CQGra*, **34**, 03LT01
- Voss, R., & Tauris, T. M. 2003, *MNRAS*, **342**, 1169
- Ziosi, B. M., Mapelli, M., Branchesi, M., & Tormen, G. 2014, *MNRAS*, **441**, 3703