Probing Binary Black Hole Formation Channels with Gravitational Waves

by

Sangeet Paul

A dissertation accepted and approved in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in Physics

Dissertation Committee:
Ray Frey, Chair
Ben Farr, Advisor
Tien-Tien Yu, Core Member
Luca Mazzucato, Institutional Representative

University of Oregon
Summer 2025

DISSERTATION ABSTRACT

Sangeet Paul

Doctor of Philosophy in Physics

Title: Probing Binary Black Hole Formation Channels with Gravitational Waves

This dissertation presents the work I've done with gravitational waves using data from and as part of the LIGO-Virgo-KAGRA collaboration. The work in this dissertation showcases different approaches to analyzing the data in the third gravitational wave transient catalog. Since all of the detected signals in the catalog are consistent with compact binary coalescences, most of them binary black holes, it provides a great window to peek into the black hole demographics of the universe. The parametric noise modeling algorithm was introduced to conduct joint inference on signal parameters and noise parameters, enabling probes into hidden correlations between signal and noise models. Several clustering algorithms were developed to find patterns in gravitational-wave data which have significant uncertainties, using which, a lot of the patterns found in other population analyses were recovered and a few outliers were identified. The coagulation model for hierarchical mergers was refined and constrained with the catalog, providing evidence for the massive event GW190521 being a hierarchical merger very likely produced in an active galactic nucleus disk. All the techniques established in this dissertation will prove useful in analyzing the rapidly growing gravitational-wave census, thereby improving our understanding of black holes as well as the universe at large.

This dissertation includes previously unpublished co-authored material.

ACKNOWLEDGEMENTS

I thank my parents, my partner, and my friends for their unwavering love, support and kindness that keeps me going through the ups and downs of life. I thank my advisor Ben for always being by my side through this arduous journey with his brilliance, radiance and understanding. I thank my fellow graduate students Bruce, JD, Philippe, Matthew, Jaxen, Gino, and Ben for their mentorship and companionship that made my time in the program memorable. I thank Ray and Zoheyr for their wise and invaluable counsel. I thank everyone in the LVK collaboration for letting me be part of this grand endeavor. I thank all the scientists I've ever had the pleasure of meeting for the engaging and inspiring conversations. I thank GTFF and the university staff for making my academic career safe and easy. I thank every worker who has made the world a better place to live in. I thank the multitude of strangers who have knowingly or unknowingly aided me throughout my life. Finally, I thank the generations of thinkers and doers who came before me and built the pillars that hold up the world we see today.

To my loved ones.

I am but a mosaic of the pieces of your hearts you gave to me along the way.

TABLE OF CONTENTS

Chapter Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

CHAPTER I

THE STORY SO FAR

1.1 Introduction

In 1687, Isaac Newton published his law of universal gravitation, wherein every mass in the universe has its own gravitational field, and masses feel a net gravitational force due to the sum of all such fields in the universe. According to his theory, when a mass changes position, the entire gravitational field throughout the universe changes instantaneously, and the resultant gravitational forces are instantly changed accordingly. In 1915, Albert Einstein published his general theory of relativity, currently the most resilient description of gravity, wherein gravity is the result of the curvature of space-time. It asserts that no information can travel faster than the speed of light, including information on the positions of mass in the universe, which is communicated through the gravitational field. General relativity (GR) predicts that a change in gravitational field will travel through the universe at the speed of light. It is exactly these changes in gravitational field that are gravitational waves (GWs). They are created by moving masses just like electromagnetic waves are created by moving charges. But because gravity is the weakest of the four fundamental forces, even a strong gravitational wave will produce displacements on the order of 10^{-18} meters, a 1000 times smaller than the diameter of a proton. And these waves are produced by very massive systems undergoing large accelerations, like two orbiting black holes (BHs) that are about to merge into one. And since systems like these are rare, these sources will be light-years away. Einstein himself thought these tiny "ripples on space-time" were too small to detect, at least with the technology back then.

In 1974, Russell Alan Hulse and Joseph Hooton Taylor Jr observed the orbital decay of a binary pulsar which matched the decay predicted by GR as energy is lost to gravitational radiation, providing the first indirect evidence for the existence of gravitational waves (Hulse and Taylor (1975)). But the first direct observation of gravitational waves was made in 2015, when the two Laser Interferometer Gravitational-wave Observatory (LIGO) detectors received a signal generated by the merger of two black holes (The LIGO Scientific Collaboration and The Virgo Collaboration (2016)). Now, there's an international detector network consisting of LIGO-Hanford in Washington, LIGO-Livingston in Louisiana, Virgo in Italy, KAGRA in Japan, and GEO600 in Germany, with LIGO-Aundha in India set to join later this decade. The census of observed gravitational waves from the "stellar graveyard" has grown to almost a 100 compact binary coalescences (CBCs) (The LIGO Scientific Collaboration, The Virgo Collaboration, and The KAGRA Collaboration (2021a); The LIGO Scientific Collaboration et al. (2021b)), that is, the merger of two compact objects such as black holes or neutron stars. In addition to upgrades to currently existing detectors, more planned gravitational-wave detectors are on their way, including larger interferometers on the ground (Cosmic Explorer, Einstein Telescope) and in space (LISA), as well as four different pulsar timing arrays.

Gravitational waves interact with matter by compressing objects in one direction while stretching them in the perpendicular direction. Therefore, laser interferometric gravitational-wave detectors are either L-shaped or triangular, and detect gravitational waves by measuring the relative lengths of their "arms". Multiple interferometers are needed to confidently detect and locate the sources of transient gravitational waves. Since gravitational waves have a finite speed,

a detection delay between the detectors helps pinpoint the sky location of the gravitational-wave source. Multiple detectors also help sort out candidate gravitational-wave events caused by local sources by discarding candidate events observed at one detector but not the others within the light travel time between detectors.

The universe is soaked with gravitational waves, just as it is with electromagnetic radiation of various frequencies, neutrinos, and cosmic rays. Hence, the introduction of these new "messengers" of information has opened new windows into the universe. And with the wealth of new information come a plethora of fresh questions. The curiosities range from finding ways to improve the detectors even further, to investigating every single gravitational-wave event, to looking for patterns in the overall distribution of these events, to probing the cosmological history of the universe.

The work I have compiled in this thesis is an attempt to grapple with quite a few of these questions. In chapter II, we build a parametric model for the noise in gravitational-wave detectors. Being able to estimate the noise from only the data containing the gravitational-wave signal and being able to track how the noise is changing with time will become especially important as we keep detecting gravitational waves more frequently and for longer durations. Improving our noise models would lead to better estimation of gravitational-wave parameters, which brings us to the chapter III. While several techniques exist for analyzing gravitational-wave merger events, we introduce clustering algorithms that can look for patterns in the estimated parameters. This technique can deal with a large number of gravitational-wave events very quickly, making it an important exploratory tool for the future. All these efforts lead up to the question: why are

the gravitational-wave event parameters distributed in this specific way? Trying to answer that question itself opens up a whole menagerie of astrophysical questions. The one we intend to tackle is "Where are these mergers happening?" While various BBH formation scenarios have been proposed, we're specifically interested in hierarchical mergers in dynamical environments, such as globular clusters and AGN disks.

The rest of this chapter is a brief review of general relativity, gravitational waves, the basic physics of gravitational-wave detectors, and the observations that the LIGO-Virgo-KAGRA collaboration has studied so far.

1.2 General relativity

In general relativity, the bending and curving of spacetime is governed by the Einstein field equations. Deriving these and looking for their solutions should lead us to a theoretical understanding of gravitational waves. The general direction of the derivations has been guided by Tong (2025).

In classical dynamics, the Lagrangian L of a system specifies its dynamics. We can assign a functional to particle paths $x^i(t)$ called the action S that is defined as

$$S[x^i(t)] = \int_{t_1}^{t_2} dt\ L(x^i(t), \dot{x}^i(t)) \tag{1.1}$$

where the index i represents one of the three spatial dimensions: $x^i \in \{x, y, z\}$.

The principle of least action states that the true path taken by a particle is an extremum of S. It is equivalent to the Euler-Lagrange equations, which can be derived by setting $\delta S = 0$

$$\frac{\partial L}{\partial x^i} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}^i}\right) = 0 \tag{1.2}$$

The Lagrangian that describes motion in curved space is an extension of the kinetic energy

$$L = \frac{m}{2} g_{ij}(x) \dot{x}^i \dot{x}^j \tag{1.3}$$

where the matrix g_{ij} is called the metric which, by definition, satisfies $g^{ij} g_{ij} = \delta^i{}_k$.

The Euler-Lagrange equations for this Lagrangian are computed to be

$$\ddot{x}^i + \Gamma^i_{jk} \dot{x}^j \dot{x}^k = 0 \tag{1.4}$$

where the coefficients

$$\Gamma^i_{jk} = \frac{1}{2} g^{il} \left(\frac{\partial g_{lj}}{\partial x^k} + \frac{\partial g_{lk}}{\partial x^j} - \frac{\partial g_{jk}}{\partial x^l} \right) \tag{1.5}$$

are called the Christoffel symbols. These Euler-Lagrange equations are called the geodesic equations and their solutions are called geodesics.

We can build a few important quantities that will aid us along the way of our derivation. One of them is the Riemann curvature tensor,

$$R^\sigma{}_{\rho\mu\nu} = \partial_\mu \Gamma^\sigma_{\nu\rho} - \partial_\nu \Gamma^\sigma_{\mu\rho} + \Gamma^\lambda_{\nu\rho} \Gamma^\sigma_{\mu\lambda} - \Gamma^\lambda_{\mu\rho} \Gamma^\sigma_{\nu\lambda} \tag{1.6}$$

which appears in the geodesic deviation equation

$$\frac{D^2 S^\mu}{D\tau^2} = R^\mu{}_{\nu\rho\sigma} \frac{\partial x^\nu}{\partial \tau} \frac{\partial x^\rho}{\partial \tau} S^\sigma \tag{1.7}$$

where S^μ is called the deviation vector that takes us from one geodesic to a nearby geodesic with the same affine parameter τ. Note that we've moved from purely spatial coordinates (Latin indices) to space-time coordinates (Greek indices): $x^\mu \in \{t, x, y, z\}$. Thus, $\frac{D}{D\tau} = \frac{\partial x^\mu}{\partial \tau} \nabla_\mu$ is the covariant derivative. The equation tells us that the relative acceleration of neighboring geodesics is controlled by the Riemann tensor.

Another important quantity is the Ricci curvature tensor which can be constructed by contracting the Riemann tensor

$$R_{\mu\nu} = R^{\rho}{}_{\mu\rho\nu} \tag{1.8}$$

Going one step further, we can also create the Ricci scalar

$$R = g^{\mu\nu} R_{\mu\nu} \tag{1.9}$$

Now, the Riemann tensor obeys the Bianchi identity

$$\nabla_{[\lambda} R_{\sigma\rho]\mu\nu} = 0 \tag{1.10}$$

where $R_{\sigma\rho\mu\nu} = g_{\sigma\lambda} R^{\lambda}{}_{\rho\mu\nu}$ and the square brackets represent the anti-symmetrization operator. This implies that for the Ricci tensor,

$$\nabla^{\mu} R_{\mu\nu} = \frac{1}{2} \nabla_{\nu} R \tag{1.11}$$

Now we are motivated to build the Einstein tensor

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} \tag{1.12}$$

so that the Bianchi identity would imply

$$\nabla^{\mu} G_{\mu\nu} = 0 \tag{1.13}$$

This means that the Einstein tensor is covariantly constant.

We now introduce the simplest non-trivial action we can write for the spacetime manifold, the Einstein-Hilbert action

$$S = \int d^4x \sqrt{-g} R \tag{1.14}$$

The Euler-Lagrange equations arising from this are

$$G_{\mu\nu} = 0 \tag{1.15}$$

These are the Einstein field equations in the absence of any matter. The vacuum Einstein equations are equivalent to saying that the metric is Ricci flat

$$R_{\mu\nu} = 0 \tag{1.16}$$

Analogous to the Lagrangian being defined as the kinetic energy minus the potential energy, we introduce a constant, that we call the cosmological constant Λ, into the action

$$S = \frac{1}{16\pi G}\int d^4x\sqrt{-g}(R - 2\Lambda) \tag{1.17}$$

$$\implies G_{\mu\nu} = -\Lambda g_{\mu\nu} \tag{1.18}$$

$$\implies R_{\mu\nu} = \Lambda g_{\mu\nu} \tag{1.19}$$

Now we add an action for matter fields S_M that is dependent on both the fields and the metric.

$$S = \frac{1}{16\pi G}\int d^4x\sqrt{-g}(R - 2\Lambda) + S_M \tag{1.20}$$

And we define the energy-momentum tensor to be

$$T_{\mu\nu} = -\frac{2}{\sqrt{-g}}\frac{\delta S}{\delta g_{\mu\nu}} \tag{1.21}$$

Varying the full action with respect to the metric gives the equations of motion

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu} \tag{1.22}$$

These are the full Einstein equations, describing gravity coupled to matter.

The Einstein equations comprise ten, coupled partial differential equations. The general solution is a formidable challenge, but a number of highly-symmetric solutions are known.

When the metric is almost flat, assuming $\Lambda = 0$, the metric can be expressed as small perturbations $h_{\mu\nu} \ll 1$ on the Minkowski metric $\eta_{\mu\nu} =$

diag$(-1,+1,+1,+1)$, that is

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \tag{1.23}$$

We want to expand the Einstein equations to linear order in the small perturbation case, so we can think of gravity as a symmetric "spin 2" field $h_{\mu\nu}$ propagating in flat Minkowski space $\eta_{\mu\nu}$ that exhibits a Lorentz invariance.

Proceeding with the construction of various curvature tensors from the metric, while working in linear order in h, we eventually arrive at the Einstein tensor

$$G_{\mu\nu} = \frac{1}{2}\left(\partial^\rho\partial_\mu h_{\nu\rho} + \partial^\rho\partial_\nu h_{\mu\rho} - \Box h_{\mu\nu} - \partial_\mu\partial_\nu h - (\partial^\rho\partial^\sigma h_{\rho\sigma} - \Box h)\eta_{\mu\nu}\right) \tag{1.24}$$

For this linearized Einstein tensor, the Bianchi identity for the full Einstein tensor $\nabla^\mu G_{\mu\nu} = 0$ reduces to

$$\partial^\mu G_{\mu\nu} = 0 \tag{1.25}$$

The Einstein equations now become a linear, but still complicated, set of second order partial differential equations

$$\partial^\rho\partial_\mu h_{\nu\rho} + \partial^\rho\partial_\nu h_{\mu\rho} - \Box h_{\mu\nu} - \partial_\mu\partial_\nu h - (\partial^\rho\partial^\sigma h_{\rho\sigma} - \Box h)\eta_{\mu\nu} = 16\pi G T_{\mu\nu} \tag{1.26}$$

where the source $T_{\mu\nu}$ is assumed to be suitably small for consistency.

Linearized gravity has a nice gauge symmetry. So under an infinitesimal change of coordinates

$$x^\mu \to x^\mu - \xi^\mu(x) \tag{1.27}$$

the change in the linearized metric can be viewed as a transformation of the linearized field

$$h_{\mu\nu} \to h_{\mu\nu} + \partial_\mu\xi_\nu + \partial_\nu\xi_\mu \tag{1.28}$$

From comparison to the gauge transformation in electromagnetism, we can intuit that the linearized Riemann tensor must be gauge invariant.

Continuing the similarity with electromagnetism, we can then impose a gauge fixing condition in linearized gravity called the de Donder gauge

$$\partial^\mu h_{\mu\nu} - \frac{1}{2}\partial_\nu h = 0 \tag{1.29}$$

whose full non-linear version looks like $g^{\mu\nu}\Gamma^\rho_{\mu\nu} = 0$.

The linear de Donder gauge simplifies the Einstein equations to

$$\Box h_{\mu\nu} - \frac{1}{2}\Box h\eta_{\mu\nu} = -16\pi G T_{\mu\nu} \tag{1.30}$$

Defining

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}h\eta_{\mu\nu} \tag{1.31}$$

the Einstein equations reduce again to

$$\Box\bar{h}_{\mu\nu} = -16\pi G T_{\mu\nu} \tag{1.32}$$

1.3 Gravitational waves

Since gravitational waves propagate in vacuum, in the absence of any sources. This means that we need to solve the linearized equation

$$\Box\bar{h}_{\mu\nu} = 0 \tag{1.33}$$

One solution is provided by the gravitational wave

$$\bar{h}_{\mu\nu} = \mathrm{Re}(H_{\mu\nu}e^{ik_\rho x^\rho}) \tag{1.34}$$

where $H_{\mu\nu}$ is a complex, symmetric polarization matrix and the wavevector k_μ is a real 4-vector. Usually when writing these solutions we are lazy and drop the Re on the right-hand side, leaving it implicit that one takes the real part. This plane wave

ansatz solves the linearized Einstein equation provided that the wavevector is null,

$$k_\mu k^\mu = 0 \tag{1.35}$$

which tells us that gravitational waves, like light waves, travel at the speed of light. If we write the wavevector as $k^\mu = (\omega, \mathbf{k})$, with ω the frequency, then this condition becomes $\omega = \pm|\mathbf{k}|$.

Because the wave equation is linear, we may superpose as many different waves of the aforementioned form as we wish. In this way, we build up the most general solution to the wave equation.

Naively, the polarization matrix $H_{\mu\nu}$ has 10 components. But we still have to worry about gauge issues. Our ansatz satisfies the de Donder gauge condition $\partial^\mu \bar{h}_{\mu\nu} = 0$ only if

$$k^\mu H_{\mu\nu} = 0 \tag{1.36}$$

which tells us that the polarization is transverse to the direction of propagation. Furthermore, the choice of de Donder gauge does not exhaust our ability to make gauge transformations. If we make a further gauge transformation $h_{\mu\nu} \to h_{\mu\nu} + \partial_\mu \xi_\nu + \partial_\nu \xi_\mu$, then

$$\bar{h}_{\mu\nu} \to \bar{h}_{\mu\nu} + \partial_\mu \xi_\nu + \partial_\nu \xi_\mu - \partial^\rho \xi_\rho \eta_{\mu\nu} \tag{1.37}$$

This transformation leaves the solution in de Donder gauge $\partial^\mu \bar{h}_{\mu\nu} = 0$ provided that

$$\Box \xi_\nu = 0 \tag{1.38}$$

In particular, we can take

$$\xi_\mu = \lambda_\mu e^{ik_\rho x^\rho} \tag{1.39}$$

which obeys $\Box\xi_\mu = 0$ because $k_\rho k^\rho = 0$. A gauge transformation of this type shifts the polarization matrix to

$$H_{\mu\nu} \to H_{\mu\nu} + i(k_\mu\lambda_\nu + k_\nu\lambda_\mu - k^\rho\lambda_\rho\eta_{\mu\nu}) \tag{1.40}$$

Polarization matrices that differ in this way describe the same gravitational wave. We now choose the gauge transformation λ_μ in order to further set

$$H_{0\mu} = 0 \text{ and } H^\mu_\mu = 0 \tag{1.41}$$

These conditions are known as transverse traceless gauge. Because H is traceless, this choice of gauge has the advantage that $h_{\mu\nu} = \bar{h}_{\mu\nu}$.

At this stage we can do some counting. The polarization matrix $H_{\mu\nu}$ has 10 components. The de Donder condition gives 4 constraints, and there are 4 residual gauge transformations. The upshot is that there are just $10 - 4 - 4 = 2$ independent polarizations in $H_{\mu\nu}$.

For example, the wavevector of a gravitational wave propagating in the z direction would be

$$k^\mu = (\omega, 0, 0, \omega) \tag{1.42}$$

The constraints restrict the polarization matrix to be

$$H_{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & H_+ & H_\times & 0 \\ 0 & H_\times & H_+ & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{1.43}$$

where the elements H_+ and $H_\times$ can be complex, but only the real part constitutes the metric. H_+ and $H_\times$ are the two polarization states.

The relative physics of how gravitational waves affect neighboring objects can be captured by the geodesic deviation equation

$$\frac{D^2 S^\mu}{D\tau^2} = R^\mu{}_{\rho\sigma\nu} u^\rho u^\sigma S^\nu \tag{1.44}$$

where u^μ is the velocity 4-vector and S^μ is the displacement vector.

For a family of geodesics in a rest frame, $u^\mu = (1,0,0,0)$, a passing gravitational wave will change it as

$$u^\mu = (1,0,0,0) + \mathcal{O}(h) \tag{1.45}$$

Computing the deviation to leading order in the metric perturbation h, the geodesic equation becomes

$$\frac{d^2 S^\mu}{dt^2} = R^\mu{}_{00\nu} S^\nu \tag{1.46}$$

Plugging in the linearized Riemann tensor we had computed earlier,

$$\frac{d^2 S^\mu}{dt^2} = \frac{1}{2}\frac{d^2 h^\mu{}_\nu}{dt^2} S^\nu \tag{1.47}$$

Notice that a gravitational wave propagating in the z direction affects neither S^0 nor S^3. Its only effect on the geodesics is in the (x,y)-plane, transverse to the direction of propagation.

We can solve this equation in the $z = 0$ plane for the two polarizations. If we set $H_\times = 0$, then the geodesic deviation equations become

$$\frac{d^2 S^1}{dt^2} = -\frac{\omega^2}{2} H_+ e^{i\omega t} S^1 \tag{1.48}$$

$$\frac{d^2 S^2}{dt^2} = +\frac{\omega^2}{2} H_+ e^{i\omega t} S^2 \tag{1.49}$$

which can be solved perturbatively in H_+ to get

$$S^1(t) = S^1(0)\left(1 + \frac{1}{2} H_+ e^{i\omega t} + ...\right) \tag{1.50}$$

$$S^2(t) = S^2(0)\left(1 - \frac{1}{2} H_+ e^{i\omega t} + ...\right) \tag{1.51}$$

Imagine a collection of particles arranged around a circle of radius R. Then at $t = 0$, the displacement vectors S, the distances from the origin in this case, will satisfy $S^1(0)^2 + S^2(0)^2 = R^2$. The relative minus sign in the solutions tells us that when the geodesics move outwards in the x direction, they move inwards in the y direction, and vice versa. The overall picture, as shown in figure 1, is that the particles evolve from a circle to an ellipse, then back to a circle, then to an ellipse orthogonal to the previous one, then back to a circle, and so on. The movement traces out a soft "plus" sign in the plane, hence the name "plus polarization".

Similarly, setting $H_+ = 0$, the geodesic deviation equations become

$$\frac{d^2S^1}{dt^2} = -\frac{\omega^2}{2} H_\times e^{i\omega t} S^2 \tag{1.52}$$

$$\frac{d^2S^2}{dt^2} = -\frac{\omega^2}{2} H_\times e^{i\omega t} S^1 \tag{1.53}$$

which can again be solved perturbatively in $H_\times$ to get

$$S^1(t) = S^1(0) + \frac{1}{2} S^2(0) H_\times e^{i\omega t} + ... \tag{1.54}$$

$$S^2(t) = S^2(0) + \frac{1}{2} S^1(0) H_\times e^{i\omega t} + ... \tag{1.55}$$

Notice that the displacements $S^1(t) \pm S^2(t)$ have the same functional form as the plus polarization displacements, implying these are the same displacements rotated by 45°. This time, the circle to ellipse evolution traces out a soft "cross" sign in the plane, hence the name "cross polarization".

We can take linear combinations of these polarization states to form circular or elliptic polarizations.

The gravitational wave displacements being invariant under rotations by π is reflected in the fact that the hypothetical graviton has spin 2.

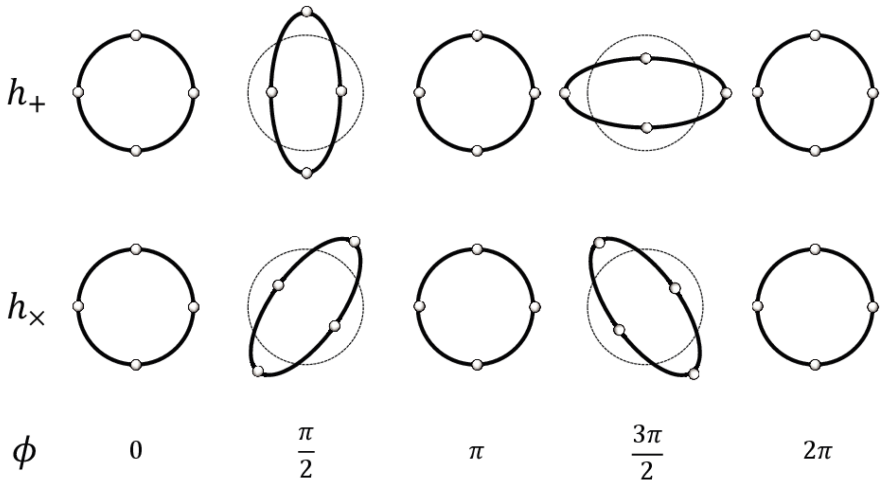

Figure 1. Gravitational wave polarizations.

1.4 Detectors

Ground-based gravitational wave detectors are essentially giant laser interferometers based on the Michelson design (Aasi et al. (2015)). Coherent light is halved with a beam splitter, bounced back and forth along the two arms with the mirrors at the two ends acting as the test masses. The light recombines at the beam splitter and the resulting interference pattern is fed into a detector. A change in this pattern translates to a difference in the arm lengths, which in the absence of noise, would be the result of a passing gravitational wave.

A detector would be the most sensitive to a gravitational wave that propagates perpendicular to the plane of the detector with a polarization maximally aligned with the arms, as it would produce the largest difference in the arm lengths. In this maximum sensitivity case, the difference in the arm lengths would be

$$L' = L\left(1 \pm \frac{H_+}{2}\right) \implies \frac{\delta L}{L} = \frac{H_+}{2} \tag{1.56}$$

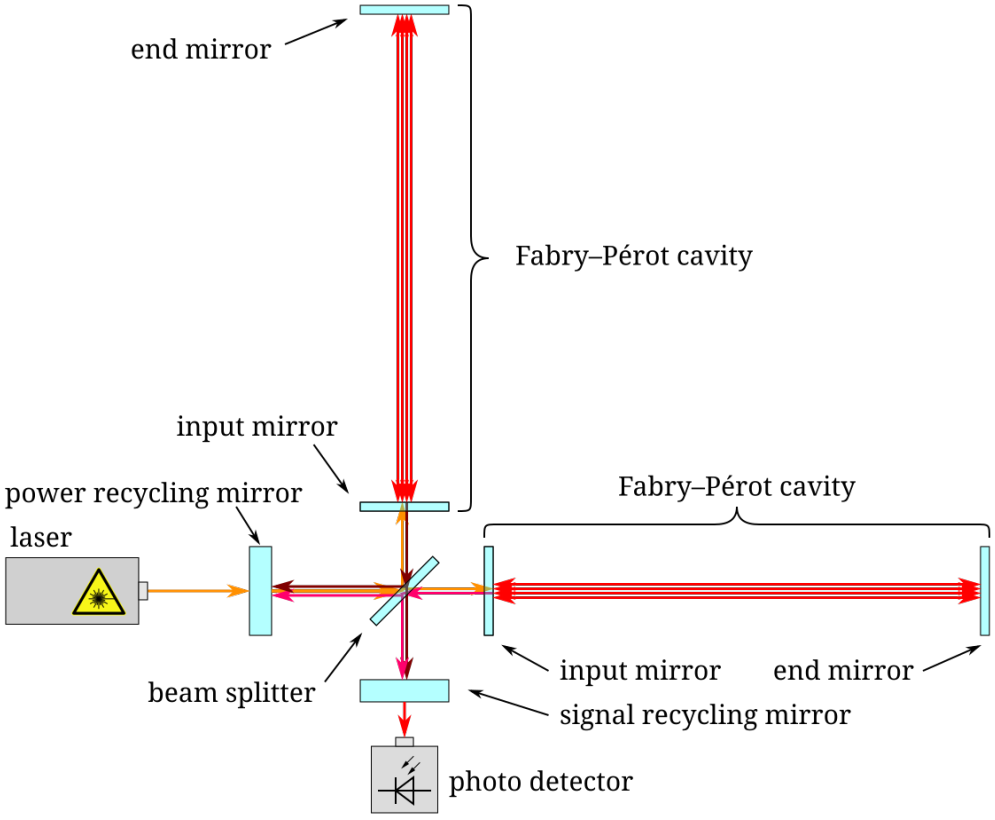

Figure 2. Simplified diagram of LIGO's interferometers. Source: Commons (2020).

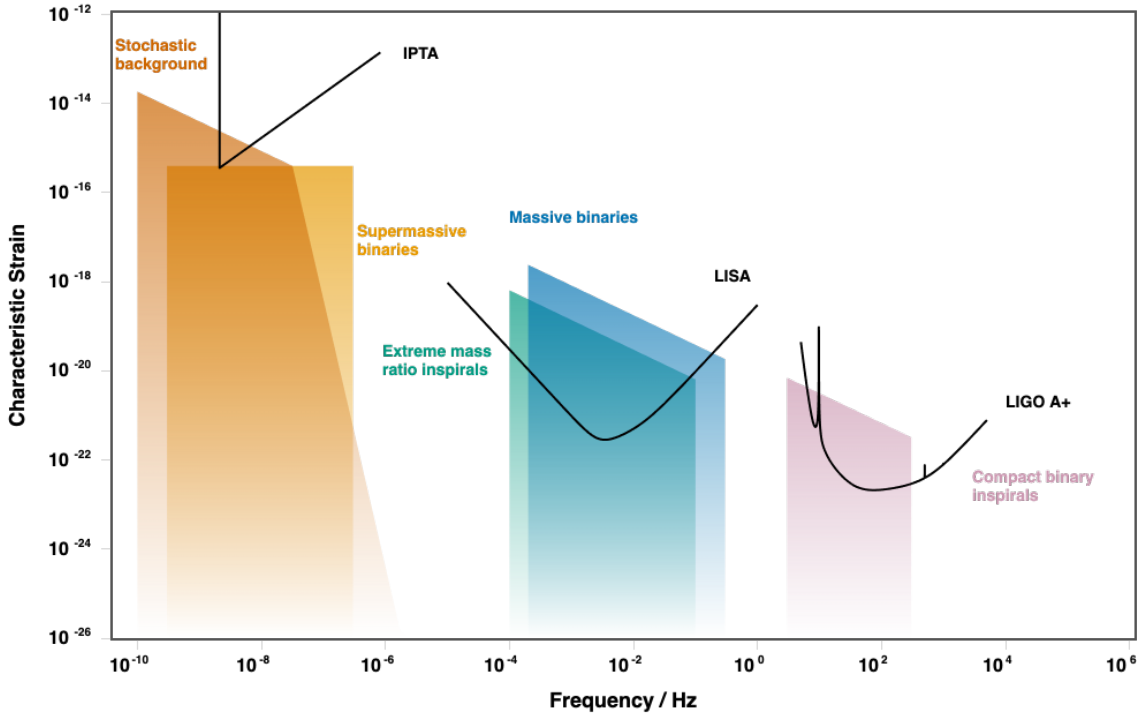

Figure 3. Strain sensitivity of LIGO, LISA and IPTA as a function of frequency. Also, characteristic ranges of strains from various astrophysical sources. Source: Moore et al. (2014).

For typical astrophysical sources, $H_+ \sim 10^{-21}$. For LIGO's 4km long arms, we can estimate $\delta L \sim 10^{-18}$ m, which is orders of magnitude smaller than the radius of a proton.

A range of optical and mechanical technologies, such as Fabry-Perot resonant cavities, power recycling, and frequency-dependent squeezing, are used to improve LIGO's sensitivity to the level required to observe such tiny strains. Figure 3 shows the sensitivity of LIGO, LISA and IPTA as a function of frequency, along with characteristic ranges of strains from various astrophysical sources.

A simplified schematic summarizing the main steps in LIGO-Virgo-KAGRA's data processing is shown in figure 4. Chapter II delves into the parameter estimation process in detail, but we'll briefly review it here.

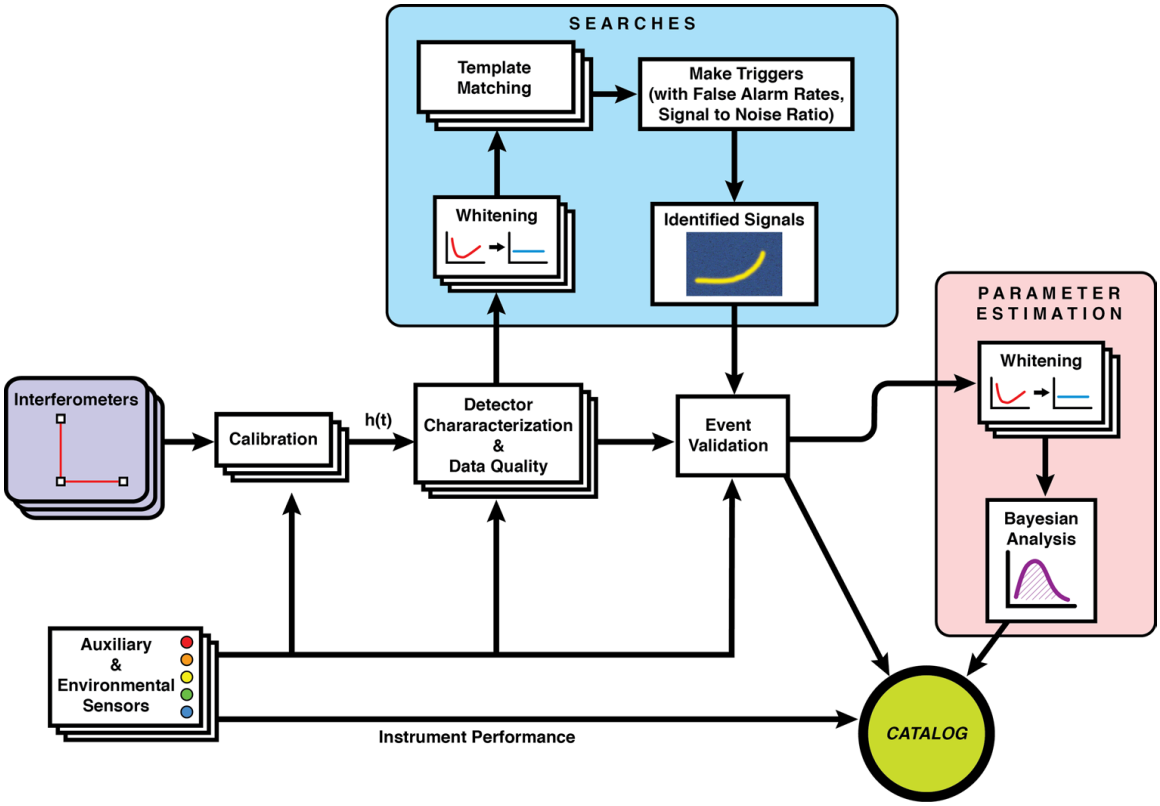

Figure 4. Simplified schematic summarizing the main steps in LIGO-Virgo-KAGRA's data processing, from the output of the data to the results reported in a catalog of transient events. Source: B. P. Abbott et al. (2020).

The timeseries data $\mathbf{d}$ observed by a gravitational-wave detector is its true response to a signal $\mathfrak{h}$ plus some true noise $\mathbf{n}$. For a good model $\mathbf{h}$ of the signal, the residual $\mathbf{r}$ should be consistent with the noise model.

$$\mathbf{d}(t) = \mathbf{n}(t) + \mathfrak{h}(t) = \mathbf{r}(t) + \mathbf{h}(t) \tag{1.57}$$

This means that the likelihood that the data $\mathbf{d}$ contains a possible signal $\mathfrak{h}$ is given by the probability that $\mathbf{r}$ is a realization of the noise model, that is, the likelihood function is essentially the noise model. So if the noise model is Gaussian, then the likelihood can be written as

$$p(\mathbf{d}|\mathbf{h}) = \frac{1}{\sqrt{\det(2\pi\mathbf{C})}} \exp\left(-\frac{1}{2}\langle\mathbf{r}|\mathbf{C}^{-1}|\mathbf{r}\rangle\right) \tag{1.58}$$

where $\mathbf{C}$ is the noise correlation matrix, and

$$\langle\mathbf{r}|\mathbf{C}^{-1}|\mathbf{r}\rangle = \int_0^\infty r_{Ik} C^{-1}_{(Ik)(Jm)} r_{Jm} df \tag{1.59}$$

where $C_{(Ik)(Jm)}$ is the correlation between the detector I's k-th data sample and the detector J's m-th data sample. If the noise between detectors is uncorrelated, then

$$C_{(Ik)(Jm)} = \delta_{IJ} C^I_{km} \tag{1.60}$$

and

$$p(\mathbf{d}|\mathbf{h}) = \prod_I^{N_{\rm IFO}} p(\mathbf{d}_I|\mathbf{h}) \tag{1.61}$$

Now, for a single detector,

$$\langle\mathbf{r}|\mathbf{C}^{-1}|\mathbf{r}\rangle = \int_0^\infty r(f_k) C^{-1}_{km} r(f_m) df \tag{1.62}$$

If the noise is stationary, then $\mathbf{C}$ will be diagonal in the Fourier domain

$$C_{km} = \delta_{km} S(f_k) \tag{1.63}$$

This $S(f_k)$ is the power spectral density (PSD). So, in the Fourier domain,

$$\langle\mathbf{r}|\mathbf{C}^{-1}|\mathbf{r}\rangle = \int_0^\infty \frac{r^*(f_k) r(f_k)}{S(f_k)} df = \frac{2}{T} \sum_k^{N_f} \frac{r^*_k r_k}{S_k} \tag{1.64}$$

Since $\mathbf{C}$ is now a diagonal matrix,

$$\det(\mathbf{C}) = \prod_{k}^{N_f} S(f_k) \tag{1.65}$$

Putting it all together,

$$p(\mathbf{d}|\mathbf{h}) = \prod_{k}^{N_f} \frac{2}{\pi T S_k} \times \prod_{k}^{N_f} \exp\left[-\frac{r_k^* r_k}{T S_k}\right] \tag{1.66}$$

$$= \prod_{k}^{N_f} \frac{2}{\pi T S_k} \exp\left[-\frac{d_k^* d_k - 2 d_k^* h_k + h_k^* h_k}{T S_k}\right] \tag{1.67}$$

which makes the log likelihood

$$\log p(\mathbf{d}|\mathbf{h}) = -\frac{2}{T} \sum_{k}^{N_f} \frac{d_k^* d_k - 2 d_k^* h_k + h_k^* h_k}{S_k} - \sum_{k}^{N_f} \log(S_k) + N_f \log\left(\frac{2}{\pi T}\right) \tag{1.68}$$

Note that in the first term, the sum cannot be factorized into a numerator sum and a denominator sum. Hence, it's called the noise-weighted inner product, as opposed to just an inner product with a factor. Since the signal model $\mathbf{h}$ is a deterministic function of signal parameters $\vec{\theta}_h$,

$$p(\mathbf{d}|\vec{\theta}_h) = \int p(\mathbf{d}|\mathbf{h}) p(\mathbf{h}|\vec{\theta}_h) d\mathbf{h} = p(\mathbf{d}|\mathbf{h}(\vec{\theta}_h)) \tag{1.69}$$

And the posterior distribution is

$$p(\vec{\theta}_h|\mathbf{d}) = \frac{p(\mathbf{d}|\vec{\theta}_h)\pi(\vec{\theta}_h)}{p(\mathbf{d})} \tag{1.70}$$

The model evidence $p(\mathbf{d})$ can be computed during the Bayesian inference process with techniques such as thermodynamic integration or stepping-stone approximation.

1.5 Observations

The first direct observation of gravitational waves occurred on September 14, 2015 with the detection of the signal GW150914 (The LIGO Scientific Collaboration and The Virgo Collaboration (2016)) by LIGO (Aasi et al. (2015); B. P. Abbott et al. (2016)). Since then, an intercontinental network comprising

LIGO in the United States, Virgo in Italy (Acernese et al. (2014)), and KAGRA in Japan (Akutsu et al. (2021)) has been operating with steadily improving sensitivities, leading to increasingly frequent gravitational-wave detections. Collectively, these terrestrial detectors have detected over 300 gravitational-wave signals to date (B. P. Abbott et al. (2019); R. Abbott et al. (2021, 2024); The LIGO Scientific Collaboration et al. (2021a)). Figure 5 illustrates the cumulative number of gravitational-wave detections as a function of time. These detectors operate in periods of coordinated observation called "observing runs", which are interspersed with phases of instrumental upgrades and commissioning. LIGO and Virgo have completed three such runs, with the fourth "O4" observing run currently in progress. All gravitational waves observed so far are believed to have arisen from stellar-mass compact binary mergers: the relativistic, gravitational-wave-driven collisions between stellar-mass black holes and/or neutron stars. These gravitational-wave detections are periodically compiled and published by the LIGO-Virgo-KAGRA collaboration as "gravitational-wave transient catalogs" (GWTCs).

Significant insights have been gained from the study of individual events, especially those with either unusual features or exceptionally well-measured properties. The very detection of GW150914 (The LIGO Scientific Collaboration and The Virgo Collaboration (2016)) and GW170817 (B. P. Abbott et al. (2017)) provided the first direct evidence for the existence of merging binary black holes and binary neutron stars, respectively. Events like GW190521 (R. Abbott et al. (2020a, 2020c)), a binary black hole with a primary component mass straddling the "pair-instability mass gap" may imply black hole formation processes beyond massive stellar evolution (S. Woosley (2017); S. Woosley and Heger (2021)). Similarly, events like GW190814 (R. Abbott et al. (2020b)) and GW230529_181500

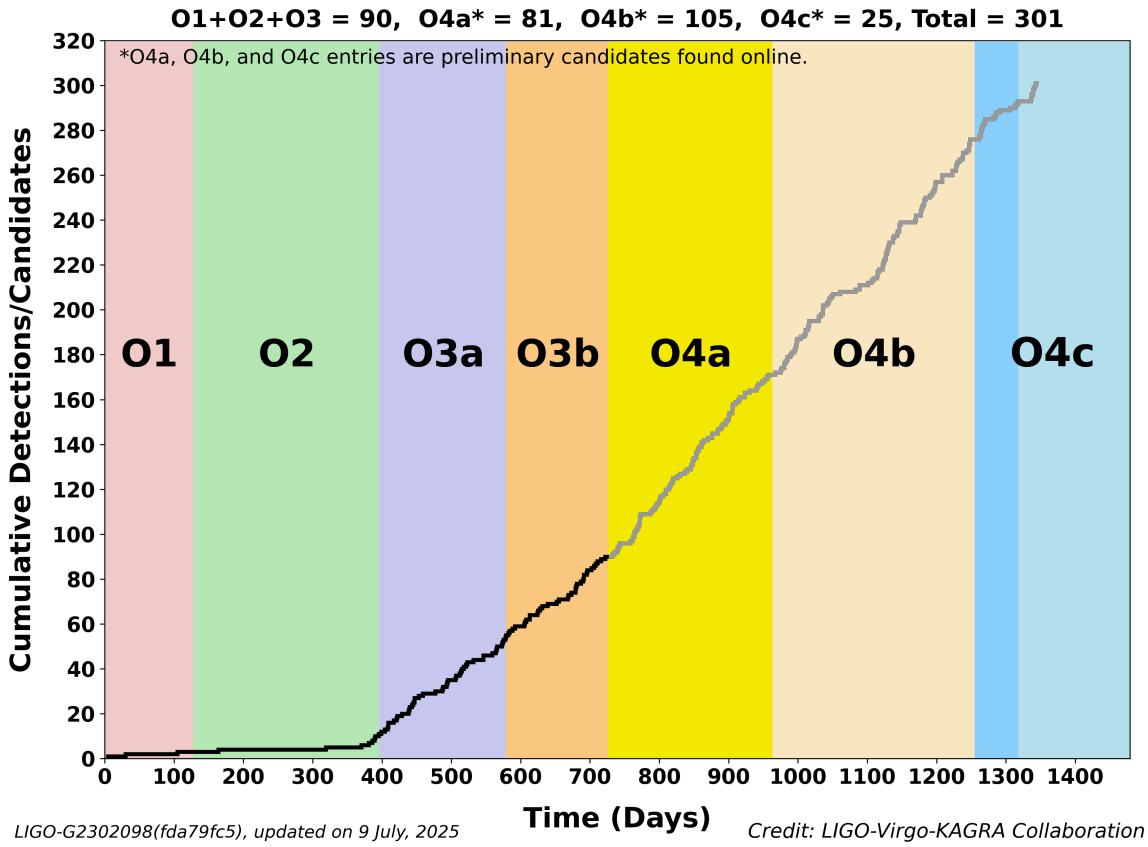

Figure 5. Cumulative number of gravitational-wave detections as a function of time. Observing runs are marked by vertical bands. Confident detections are in black, and candidate detections are in gray.

(Abac et al. (2024)) point to the existence of compact objects heavier than neutron stars but lighter than any other confirmed black holes. Despite this, drawing robust conclusions about the astrophysical population of merging compact binaries from individual exceptional events remains challenging, due to a variety of factors such as significant uncertainties on inferred parameters, and severe selection effects. Consequently, our current understanding of compact binary demographics is largely informed by statistical analyses of the overall observed population.

The multitude of analyses have taught us a lot about the astrophysical binary black hole population. The details can be boiled down to a few confident observations about binary black hole masses:

- The large majority of BBH mergers have primary masses in the $8 - 10\ M_\odot$ range.
- Above $\sim 10\ M_\odot$, the BBH merger rate decreases rapidly with primary mass.
- There's an approximately Gaussian excess of BBH mergers with primary masses $\approx 35\ M_\odot$.
- Black holes exist in the upper pair-instability mass gap.
- Binary black holes are preferentially symmetric in mass.

and about binary black hole spins:

- Most component black holes have small spins $0 < \chi \lesssim 0.4$.
- Black hole spins are preferentially aligned with the orbital angular momentum.
- Some BBHs exhibit significant spin-orbit misalignment.

Also, it has been observed that binary black holes with more unequal masses exhibit larger more positive effective spins.

1.6 Bayesian analysis

Almost all facets of gravitational-wave data analysis involve Bayesian statistics, so we'll briefly review it here.

Bayes' theorem provides a mathematical framework for inverting conditional probabilities, enabling the calculation of the probability of a cause given its observed effect. The theorem states that the "posterior" probability $p(H|E)$ of a hypothesis H given some observed data E is equal to the "likelihood" $p(E|H)$ of observing the data given the hypothesis, multiplied by the "prior" probability $p(H)$ of the hypothesis without the observed data, divided by $p(E)$, the total probability of observing the data given all possible hypotheses, also called the marginal likelihood or the model evidence.

$$p(H|E) = \frac{p(E|H) \cdot p(H)}{p(E)} \tag{1.71}$$

Bayesian inference is a statistical method that leverages the Bayes' theorem to update prior probability distributions of model parameters with new observed data to calculate their posterior probability distributions. Mathematically, the posterior distribution of parameters θ given data $\mathbf{d}$ and model hyperparameters α is calculated as

$$p(\theta|\mathbf{d}, \alpha) = \frac{p(\mathbf{d}|\theta, \alpha) \cdot p(\theta|\alpha)}{\int p(\mathbf{d}|\theta, \alpha) p(\theta|\alpha) d\theta} \tag{1.72}$$

In practice, for almost all complex Bayesian models, the posterior distribution is not obtained in a closed form distribution, but is instead computed in the form of samples. Numerous sampling algorithms now exist for efficient computation of

inferred posteriors, each using different techniques to optimize Bayesian inference for different problems.

1.7 Chapter notes

Co-authorship of chapters II, III, and IV is as follows:

- Chapter II is part of a manuscript in preparation co-authored by me and Ben Farr. I primarily developed the noise model, built the code extending the BILBY software, ran the analyses, and designed the validation studies.

- Chapter III is part of a manuscript in preparation co-authored by me, Zoheyr Doctor, and Ben Farr. Zoheyr Doctor had initiated this project with the deterministic k-means and probabilistic agglomerative methods. I developed the rest of the algorithms, modularized the code, and ran the analyses.

- Chapter IV is part of a manuscript in preparation co-authored by me, Zoheyr Doctor, Ben Farr, and Daniel Wysocki. It's based on the hierarchical merger model initially developed by Zoheyr Doctor, Daniel Wysocki, Richard O'Shaughnessy, Daniel E. Holz, and Ben Farr in Doctor et al. (2020). I primarily developed the astrophysically-motivated models, especially those for globular clusters and AGN disks, as well as designed the population mixture model inference technique. I also built the new population inference code from scratch, and ran the analyses.

Note that we've used the words "cluster" and "hierarchical" with two different meanings in two different contexts. In chapter III, a cluster is a grouping of data points, and hierarchical is a type of clustering method. In chapter IV, a cluster is a gravitationally-bound group of stars, and hierarchical is a hypothesized

model of repeated mergers. Also, neither usages refer to a computer cluster or Bayesian hierarchical modeling, both of which are important to my work in general.

CHAPTER II

UNDERSTANDING NOISE

The work in this chapter is part of a manuscript in preparation co-authored by me and Ben Farr. I primarily developed the noise model, built the code extending the BILBY software, ran the analyses, and designed the validation studies.

2.1 Introduction

Gravitational-wave detectors are time-domain instruments, recording gravitational-wave amplitude as a differential change in the lengths of each of the interferometer arms. The primary output of these detectors is a single time-stream of gravitational-wave "strain" - a one-dimensional real-valued time series. But the sensitivity of gravitational-wave detectors is frequency dependent and so is often measured as a power spectral density (PSD) over the range of interesting gravitational-wave frequencies. The PSD describes how the power of the signal or time series is distributed over frequency. It is often estimated using Welch's method, which computes the average of the modulus squared of the discrete Fourier transform of windowed overlapping time segments.

If we are to properly utilize the detectors to understand the universe, we have to keep them as noise-free as possible. But instrument noise dominates gravitational-wave data. Even with advanced current-gen interferometers, we need powerful analysis methods to squeeze out relevant information. All search pipelines looking for evidence of various kinds of gravitational-wave sources essentially ask the same question: "How statistically different is the detector output from our model for the data?" Template-based analyses model the data as a combination of noise and signals, and sometimes glitches. Thus, any errors in modeling the noise will lead to errors in inferring the template waveform parameters. Hence,

understanding the noise is just as crucial as understanding the template signals in order to do accurate parameter estimation.

Noise in ground-based interferometers has 3 broadband components:

- seismic noise at low frequencies,
- thermal noise at mid frequencies, and
- quantum photon shot noise at high frequencies.

On top of this lie various high-power, narrow-band, spectral lines due to resonant effects like the AC power supply, calibration lines, and vibrational modes of the suspension system. The noise is assumed to be approximately Gaussian and stationary for typical signal durations, allowing us to completely characterize it by its PSD. But the noise is not stationary on longer time scales, thereby motivating estimation of the noise from the on-source data itself. Both the issue and the motivation grow larger with improving detector sensitivity, especially towards lower frequencies, as the signals themselves remain "in band" for longer periods of time. The ultimate goal is to have parametrized models for both the noise and the gravitational-wave signal (or even a glitch) and infer them simultaneously. In a sampling framework, that implies our gravitational-wave detection inferences can be marginalized over all consistent PSDs, instead of assuming a fixed noise.

2.2 Noise model

Taking inspiration from the BayesLine (Littenberg and Cornish (2015)) algorithm, part of the BayesWave (Cornish and Littenberg (2015)) pipeline used for LIGO-Virgo-KAGRA analyses, we phenomenologically model the noise in the PSD vs frequency parameter space. The broadband noise is modeled with a spline and the spectral lines are modeled with Lorentzian functions. Thus, the noise PSD

$S_n(f)$ is modeled as a sum of a cubic spline and a linear combination of several Lorentzians:

$$S_n(f) = S_S(f) + S_L(f) \tag{2.1}$$

2.2.1 Splines. A spline is a piecewise polynomial function that can have a very simple local form, yet at the same time be globally flexible and smooth. A cubic spline can be represented in various ways, each with its own advantages and parameters (de Boor (1978)). For example, the piece-wise polynomial in the local power basis is parametrized by a set of N_S knots $\{f_i, S_i\}$ in frequency-PSD space as

$$S_S(f) = \sum_{m=0}^{k} c_{m,i}(f - f_i)^m \text{ for } f \in [f_i, f_{i+1}] \tag{2.2}$$

To compute these coefficients, we use functions from the Python sub-package `scipy.interpolate` (Virtanen et al. (2020)).

The B-spline basis is parametrized by a set of knots f_i and coefficients c_j as

$$S_S(f) = \sum_{j=0}^{n-1} c_j B_{j,k}(f) \tag{2.3}$$

where

$$B_{i,0}(f) = 1 \text{ if } f_i < f < f_{i+1} \text{ otherwise } 0 \tag{2.4}$$

and defined recursively as

$$B_{i,k}(f) = \frac{f - f_i}{f_{i+k} - f_i} B_{i,k-1}(f) + \frac{f_{i+k+1} - f}{f_{i+k+1} - f_{i+1}} B_{i+1,k-1}(f) \tag{2.5}$$

After having experimented with these splines, we chose to go with the Akima spline (Akima (1970)) for its smoothness and stability. It is parametrized by a set of N_S knots $\{f_i, S_i\}$ in frequency-PSD space as

$$S_S(f) = a_i + b_i(f - f_i) + c_i(f - f_i)^2 + d_i(f - f_i)^3 \text{ for } f \in [f_i, f_{i+1}] \tag{2.6}$$

where

$$m_i = \frac{S_{i+1} - S_i}{f_{i+1} - f_i} \tag{2.7}$$

$$\sigma_i = \frac{|m_{i+1} - m_i| m_{i-1} + |m_{i-1} - m_{i-2}| m_i}{|m_{i+1} - m_i| + |m_{i-1} - m_{i-2}|} \tag{2.8}$$

$$a_i = S_i \tag{2.9}$$

$$b_i = \sigma_i \tag{2.10}$$

$$c_i = \frac{3m_i - 2\sigma_i - \sigma_{i+1}}{f_{i+1} - f_i} \tag{2.11}$$

$$d_i = \frac{\sigma_i + \sigma_{i+1} - 2m_i}{(f_{i+1} - f_i)^2} \tag{2.12}$$

As evident in figure 6, Akima splines are less prone to the "ripple effect" found in other splines where a change in the PSD value of the spline at one frequency causes ripples that leak into the PSD value at other neighboring frequencies, which causes stability issues when sampling the spline's knot locations. Some time after we made the choice of Akima splines, the BayesWave team shared with us that they had independently decided to actively shift from using power-basis cubic splines to using Akima splines for the same reasons.

2.2.2 Lorentzians. Lorentzians are used to fit the spectral lines because a Lorentzian function is the characteristic PSD of a noise-driven damped harmonic oscillator which is a good physical model for some of our noise sources (Finn and Mukherjee (2001)). As for sources that are not so well-modeled by a Lorentzian, they're still reasonably fit by such a simple function. The Lorentzian line model can also be represented in various ways, such as in the location-amplitude-scale basis

$$S_L(f) = \sum_i^{N_L} \frac{a_i}{\pi} \frac{\gamma_i}{(f - f_i)^2 + \gamma_i^2} \tag{2.13}$$

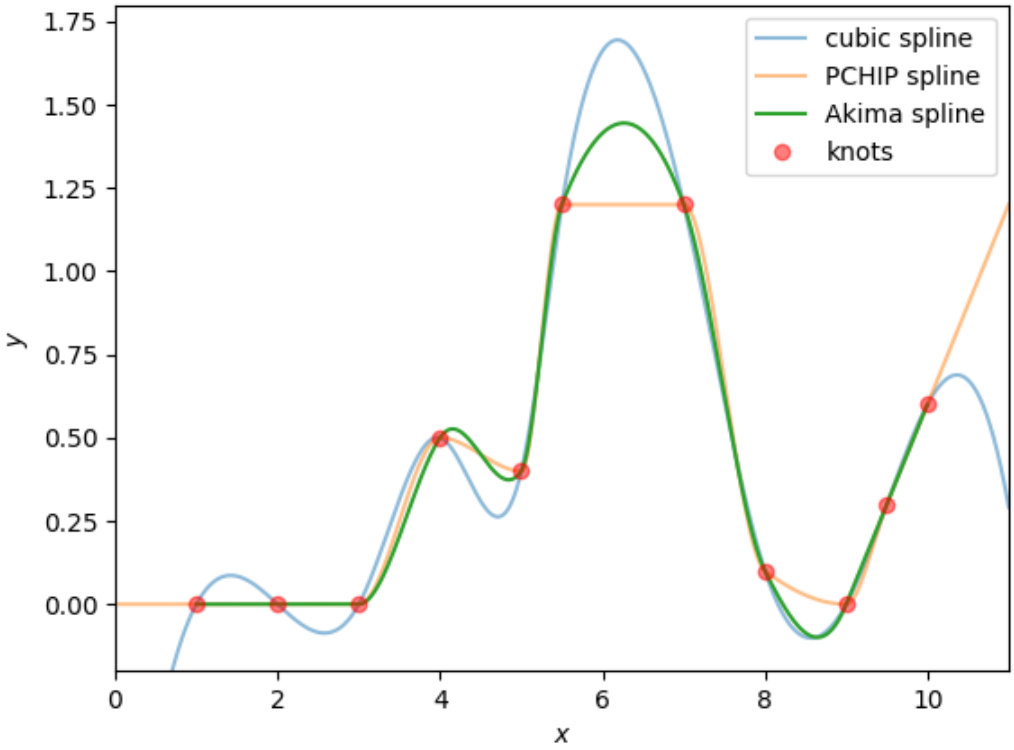

Figure 6. Comparison of different spline functions interpolating the same set of knots.

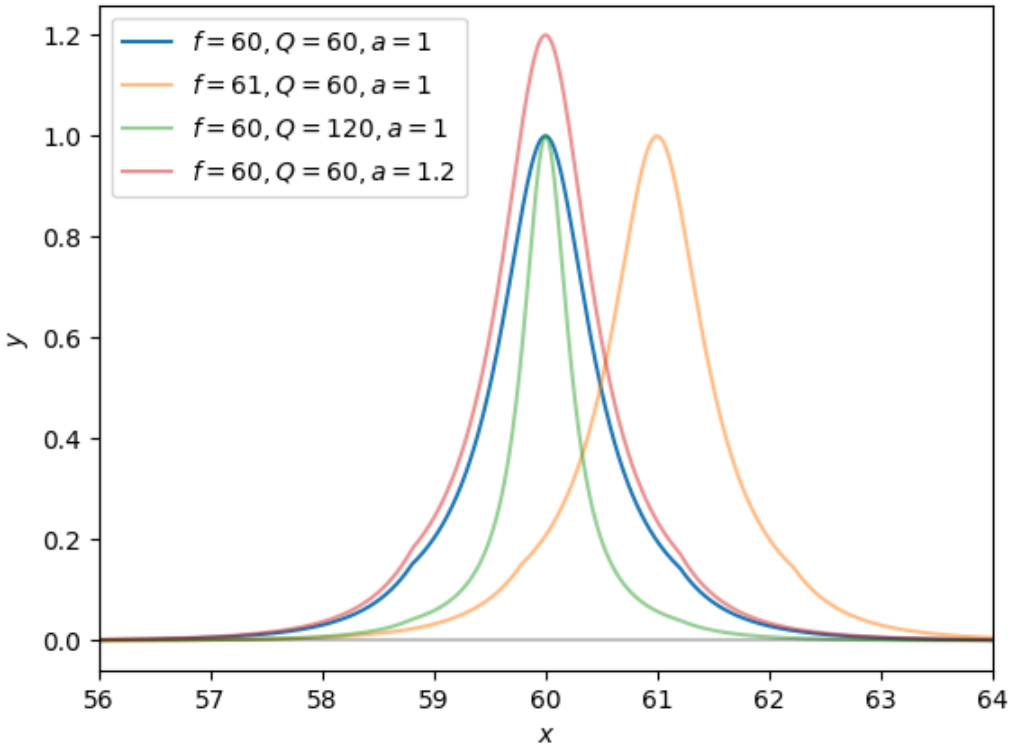

Figure 7. Lorentzian functions for different sets of parameters.

or in the location-height-width basis

$$S_L(f) = \sum_i^{N_L} \frac{h_i}{4} \frac{w_i^2}{(f - f_i)^2 + w_i^2} \tag{2.14}$$

We chose the location-quality-amplitude basis

$$S_L(f) = \sum_i^{N_L} a_i \frac{f_i^4}{f^2 f_i^2 + Q_i^2 (f^2 - f_i^2)^2} \tag{2.15}$$

because it provides the most stability for sampling the parameters. To be more precise, we sample in the location-log(quality)-log(amplitude) parameter space. The effect of varying these parameters is shown in figure 7

We can also multiply a window function to each Lorentzian to truncate its tails. We can choose a simple frequency-scaled rectangular window

$$w_i(f) = \begin{cases} 1 \text{ if } |f - f_i| < f_i/s \\ \\ 0 \text{ otherwise} \end{cases} \tag{2.16}$$

where s is some scaling factor that we usually choose to be 50. Or we can choose a smoother window that tapers the Lorentzian tails without creating a significant cliff where the ends join the splines.

$$w_i(f) = \begin{cases} 1 \text{ if } |f - f_i| < f_i/s \\ \\ \exp\left(-\dfrac{f - f_i/s}{f_i/s}\right) \text{ otherwise} \end{cases} \tag{2.17}$$

While more physically motivated models are possible, such as a broadband model parametrized by detector-related quantities, or coherent modeling and regression for spectral lines from electric sources, we'll only focus on our phenomenological model which fits the data reasonably well.

2.3 Inference

As shown in chapter I, the likelihood for parameter estimation is constructed by calculating the joint probability of the N complex Fourier coefficients of the data being realized from the PSD, assuming the whitened data should be consistent with a unit normal distribution.

$$\log p(\mathbf{d}|\theta) = -\frac{2}{T}\sum_{f=f_{\min}}^{f_{\max}} \frac{|d(f)|^2 - 2d(f)\cdot h(f) + |h(f)|^2}{S_n(f)} - \sum_{f=f_{\min}}^{f_{\max}} \log S_n(f) \tag{2.18}$$

where θ is the set of model parameters, $d(f)$ is the data in the frequency domain, $h(f)$ is the model signal in the frequency domain, $S_n(f)$ is the PSD, and T is the reciprocal of the frequency resolution. The constant normalization term has been dropped since we're only concerned with the relative likelihoods of points in the parameter space. Usual parameter estimation pipelines use only the first term since

they infer signal parameters while fixing the PSD equal to an off-source estimate. But our goal is to infer both signal and PSD parameters, so we'll have to compute the whole equation.

We extended the Bayesian inference library BILBY (Ashton et al. (2019); Ashton and Talbot (2021); Romero-Shaw et al. (2020)), to sample the $2N_S$ spline parameters and the $3N_L$ Lorentzian parameters, as well as the usual 15 BBH parameters, simultaneously. BILBY allows for a variety of samplers, among which we used the affine-invariant Markov chain Monte Carlo (MCMC) ensemble sampler EMCEE (Foreman-Mackey, Hogg, Lang, and Goodman (2013)), and the dynamic nested sampler DYNESTY (Koposov et al. (2024); Speagle (2020)). The EMCEE sampler was run on multiple parallel-tempered chains to deal with multimodality and avoid getting stuck in local optima. Additionally, we built a custom proposal selector that randomly jumps through a set of pseudo-Markovian proposals. Most of these are Gibbs sampling steps to handle the enormous dimensionality of the parameter space. Some of the proposals use previously calculated Fisher information matrices to make smarter jumps in heavily correlated parameter spaces such as spline knot locations.

The BBH parameters are initialized randomly as usual, but initializing the PSD parameters requires more ingenuity. The frequency locations of spline knots can be initialized either at uniformly separated frequencies, or based on equal quantile spacing which ensures more knots at places where the PSD changes rapidly. The PSD for a knot at a certain frequency is then set equal to the median power in a broadband filter centered at that frequency. To initialize Lorentzians, we scroll through the data with varying filters, putting in Lorentzians at frequencies where we see an excess of power in a narrowband filter compared to the median

power in a broadband filter centered at the same frequencies. Thus, the Lorentzian parameters are initialized based on where and how much of a power excess there is, as measured by its "prominence" crossing a threshold. Good initialization is crucial since the dimensionality of the parameter space is often large and stays constant while sampling. Well-informed priors and parallel tempering can be used to keep MCMC walkers in the extremely narrow useful sliver of the parameter space and improve time to convergence.

2.4 Application to GW150914

To test our algorithm, we ran both standard BILBY and our extended BILBY on the first detected gravitational wave event GW150914. We chose to use data from only one detector: LIGO-Hanford, to specifically explore scenarios where only one detector is online and modeling the noise accurately becomes even more important. A spectrogram of this data, is shown in figure 8; such spectrograms use constant-Q transforms and are often used to visualize signal data. For the actual likelihood calculation, we used 4 seconds of on-source data including the signal. For initializing the PSD parameters, we used 128 seconds of off-source data just before the signal, ending exactly where the on-source data begins. This is common to both versions of BILBY; the difference is that standard BILBY uses the off-source data to get a Welch estimate and uses it for all likelihood evaluations, whereas our extended BILBY uses it to initialize the PSD spline and Lorentzian parameters that will eventually be sampled. Figure 9 shows the initialization steps. At this stage, we're less concerned about how accurate the parameter estimates are, and more concerned with nudging the parameters into a good part of the very large-dimensional parameter space to save the sampler some time that it might otherwise waste on wandering the unreasonable parts.

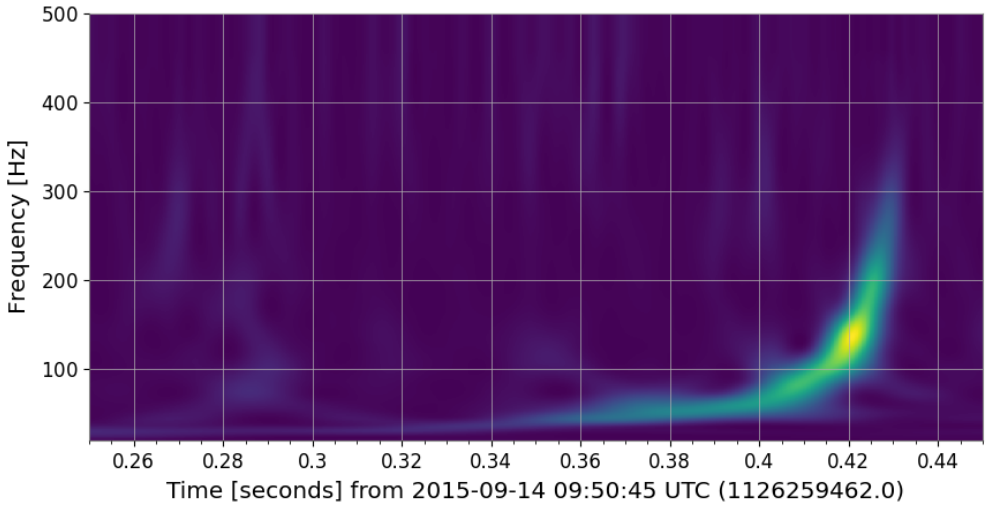

Figure 8. Spectrogram of LIGO-Hanford data including GW150914.

Both standard BILBY and our extended BILBY produce posteriors for the BBH parameters. Additionally, our extended BILBY produces posteriors for the PSD spline and Lorentzian parameters. Figure 10 shows PSDs built from these parameter posteriors. The PSD posteriors are in agreement with the PSD obtained via Welch's method on off-source data. The whitened frequency-domain strain data are consistent with the unit normal distribution $\mathcal{N}(0, 1)$. Most of the outliers come from the edge frequencies and "forests" of spectral lines that are tougher to fit.

Figure 11 compares the BBH parameter posteriors produced by standard BILBY and our extended BILBY. The extended code's BBH posteriors are obtained from marginalizing over the PSD posteriors. Most of the posterior distributions are in agreement. The bimodality in the $\phi_{JL} - \theta_{JN}$ space is lost in the extended version. Its cause can be traced to the correlation between θ_{JN} and the PSD, as can be seen in the correlation matrix in figure 12. The matrix also shows the expected heavy correlation between the chirp mass and the distance, and a light correlation between the chirp mass and the PSD. Intuitively, a higher PSD in

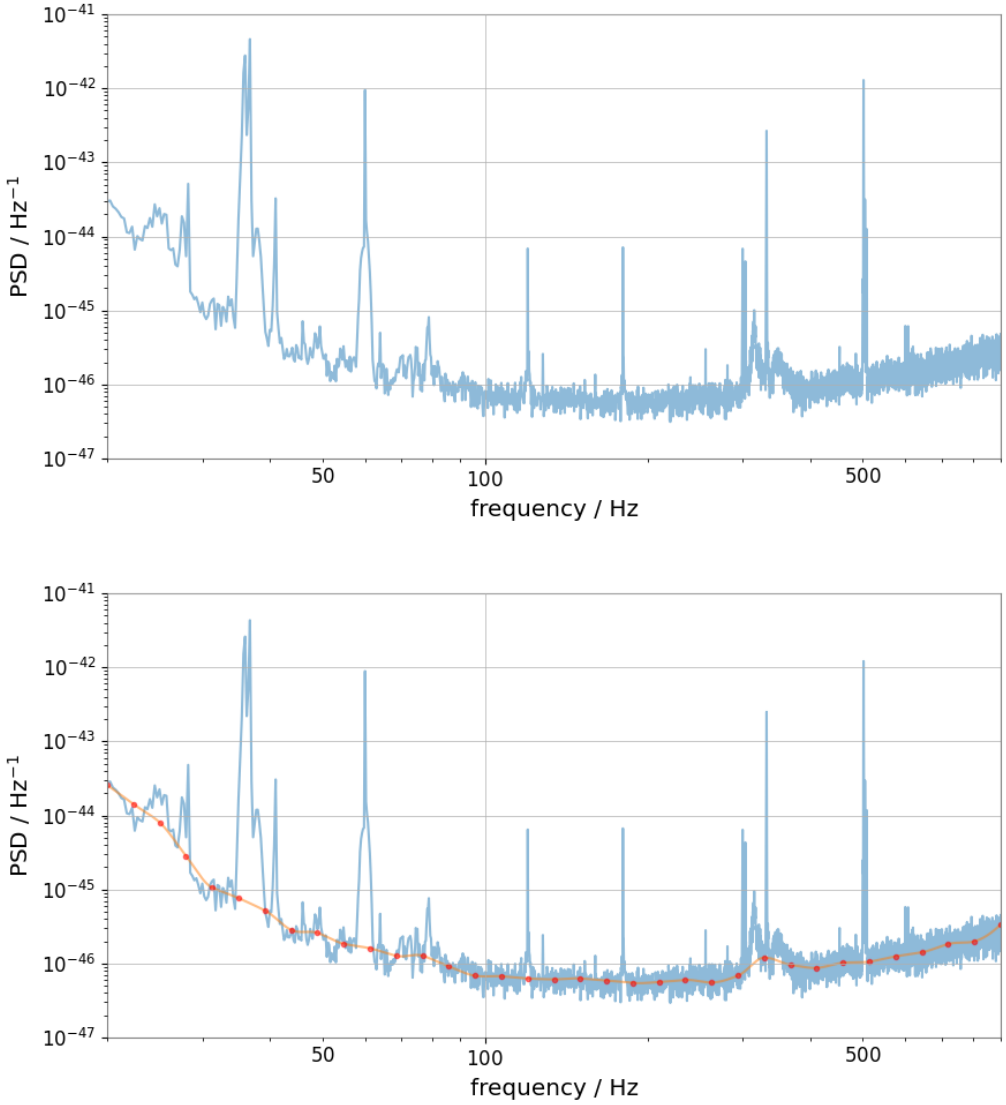

Figure 9. Off-source initialization steps on 128s of LIGO-Hanford data just before GW150914. Top: Welch estimate in blue. Vertical grid lines represent knot positions. Bottom: Estimated spline in orange. Knot locations in red.

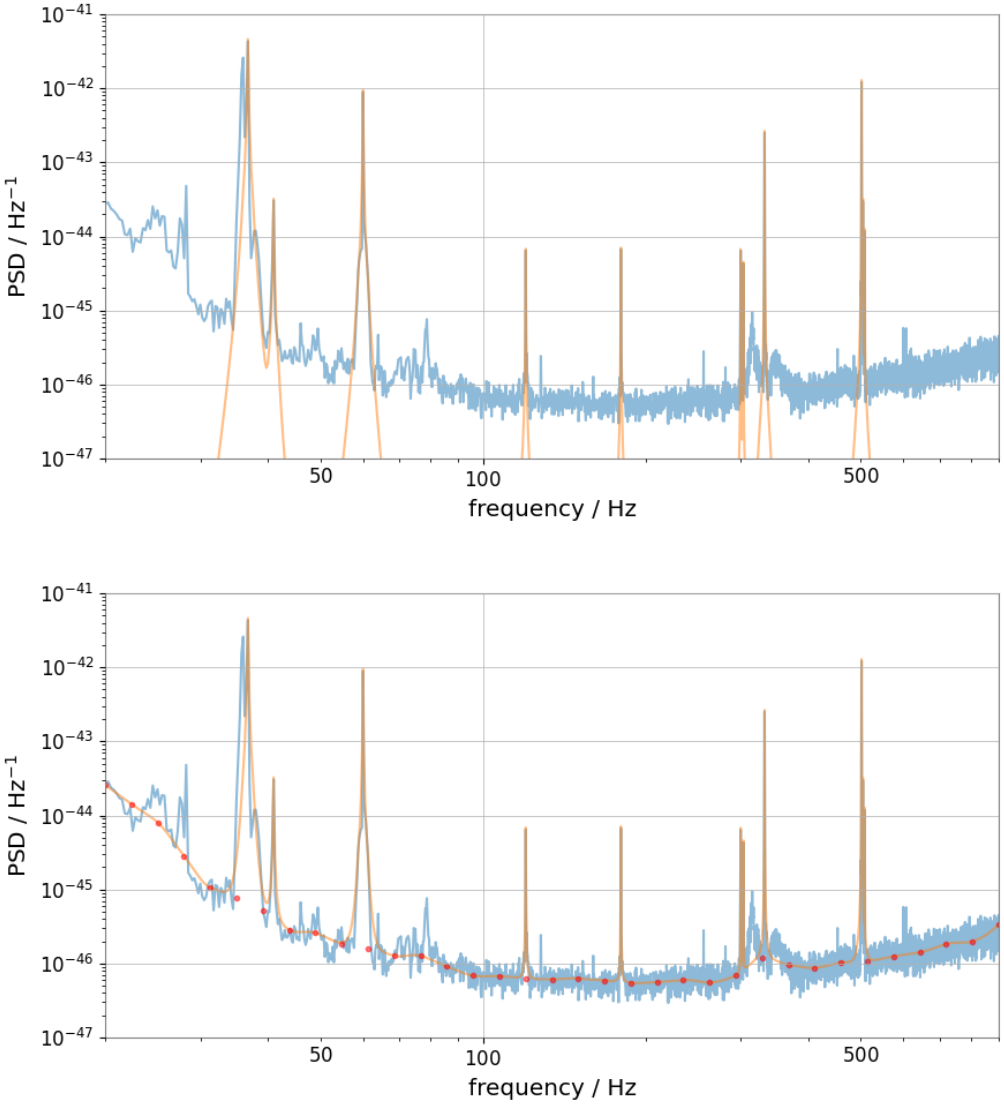

Figure 9. Continued: Off-source initialization steps on 128s of LIGO-Hanford data just before GW150914. Top: Estimated Lorentzians in orange. Bottom: Estimated PSD (spline + Lorentzians) in orange. Knot locations in red.

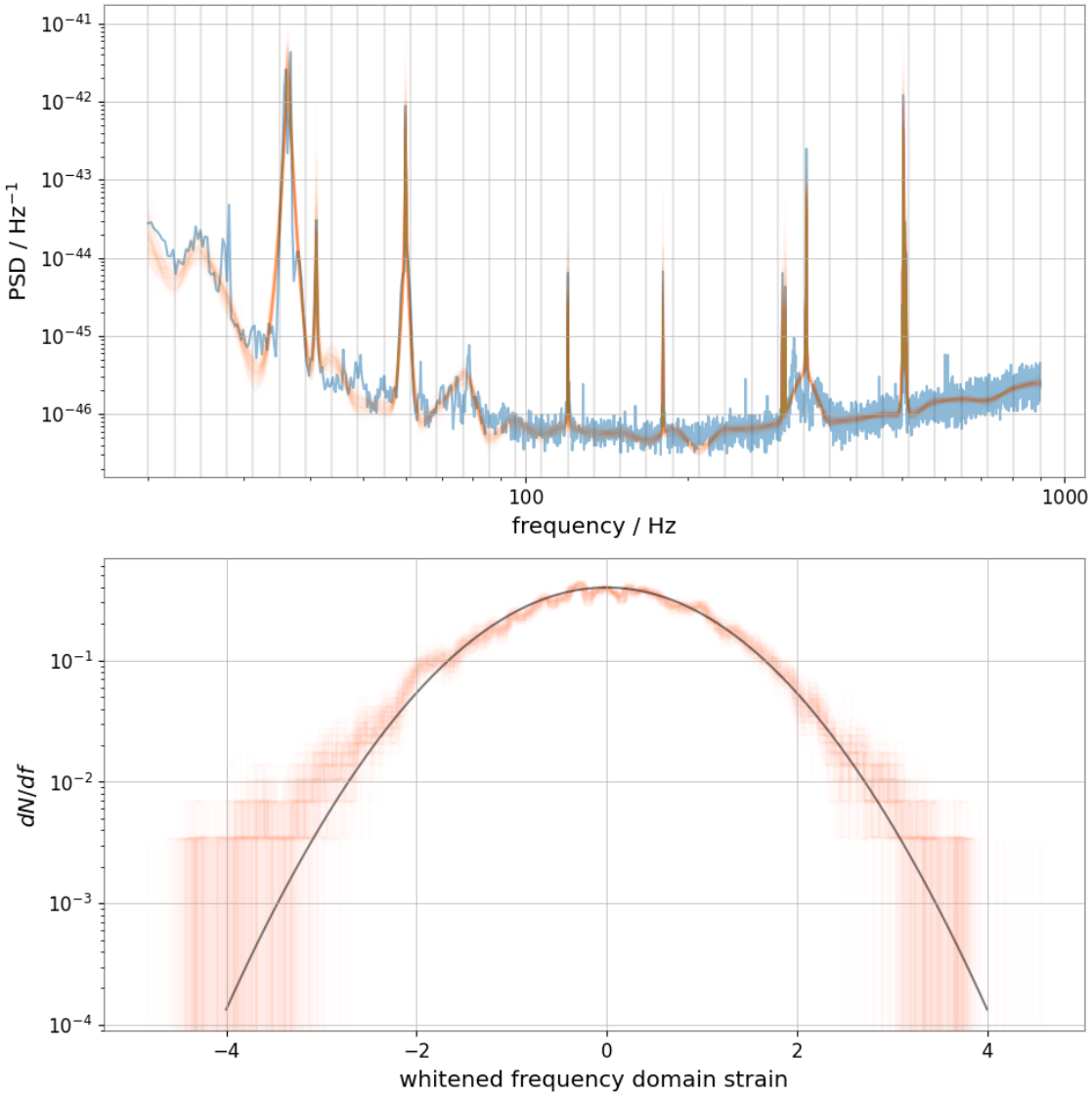

Figure 10. Posterior distribution of PSDs. Top: Welch estimate in blue. PSDs in orange. Bottom: Whitened strain data in frequency domain in orange. Unit normal distribution $\mathcal{N}(0, 1)$ in black.

a certain frequency range can be explained by either a heavier merger, a closer distance, or a louder noise. We can delve deeper with figure 13 that shows the correlation of BBH parameters with PSDs at specific frequencies. The $\theta_{JN}-$PSD correlation can be narrowed down to be occurring at around 55 Hz, so it's probably connected to the Lorentzian trying to fit the spectral line at 60 Hz due to power lines.

The waveform posteriors for GW150914 using standard BILBY, shown in figure 14, are not significantly different from those using our extended BILBY, shown in figure 15. The overall results should not be perceived as a preference towards one version of the parameter estimation over another. Instead, the conclusion is that the binary parameters' posterior distributions can be dependent on the noise model in unintuitive ways.

2.5 Discussion

In this work, we introduced an algorithm to conduct joint inference on CBC parameters and PSD parameters. It enables us to probe for hidden correlations between signal and noise models, that may otherwise be buried due to search pipelines using fixed estimates for PSDs. This may prove to be a good sanity check for cases where the noise is suspected to have muddied parameter estimation of the merger signal, especially when only one detector is online for observation. Its application to the event GW150914 showed minor differences in the posterior distributions of the parameters that can be connected to differences in the flexibility of noise models fitting the same data.

While CBC analyses use no a priori knowledge when fitting spectra, the posterior estimates from our method can be used to construct informed priors for future analyses. If the time series is long enough, we can even perform piece-wise

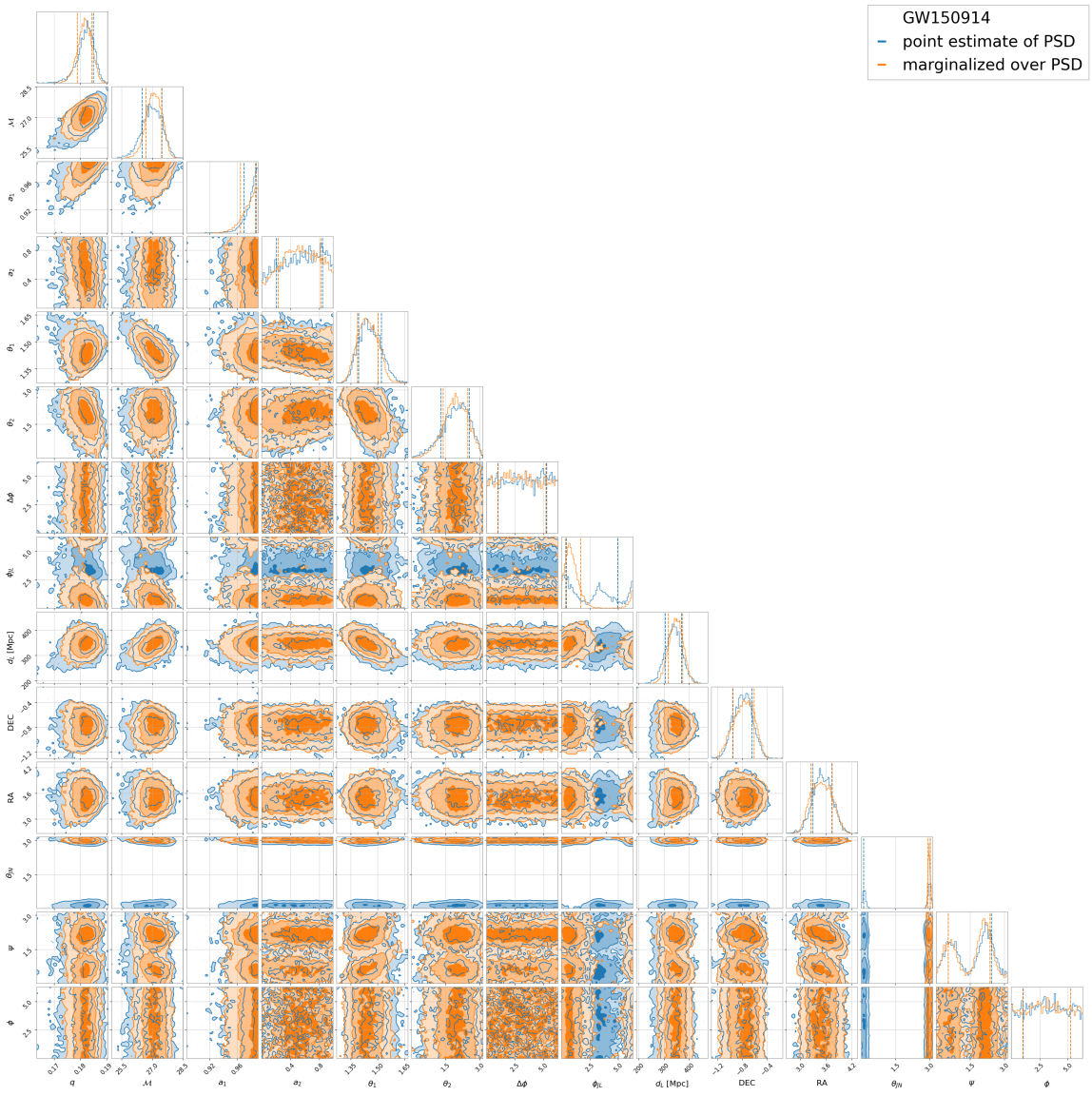

Figure 11. Posterior distribution of BBH parameters. Blue: Standard BILBY posteriors using a Welch estimate for PSD. Orange: Our extended BILBY posteriors marginalized over PSD posteriors.

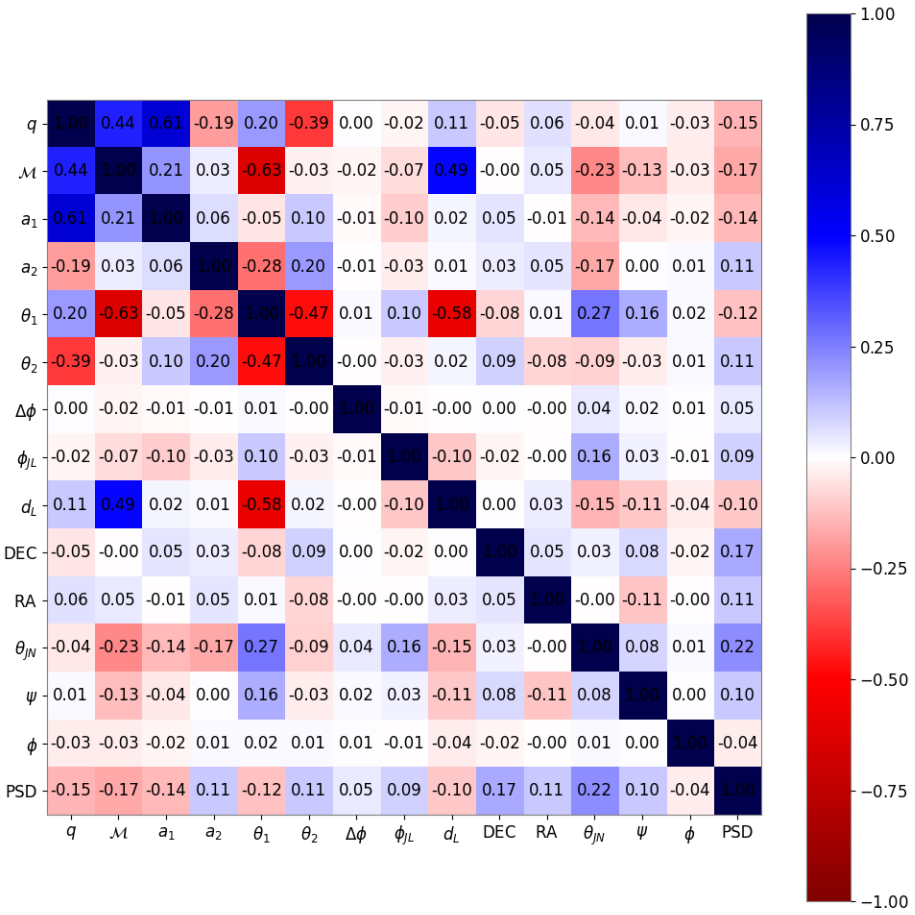

Figure 12. Correlation between BBH parameters and PSD.

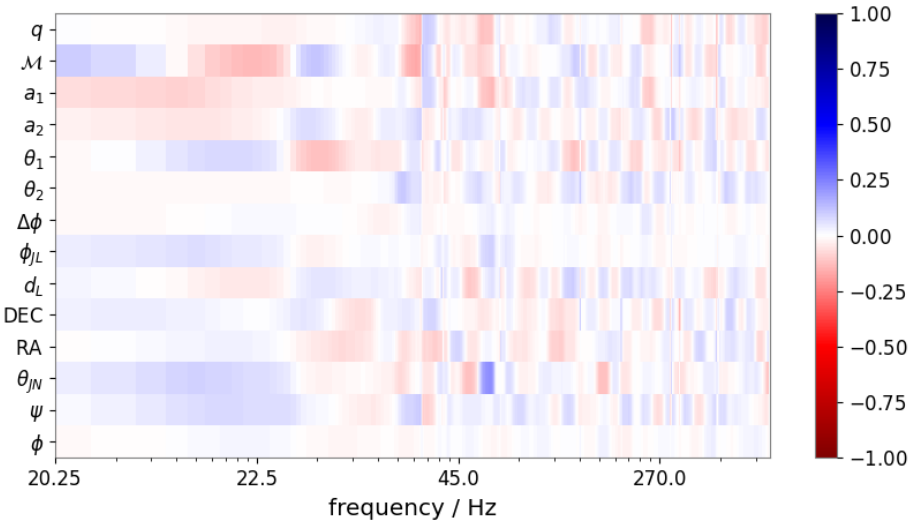

Figure 13. Correlation between BBH parameters and PSD frequencies.

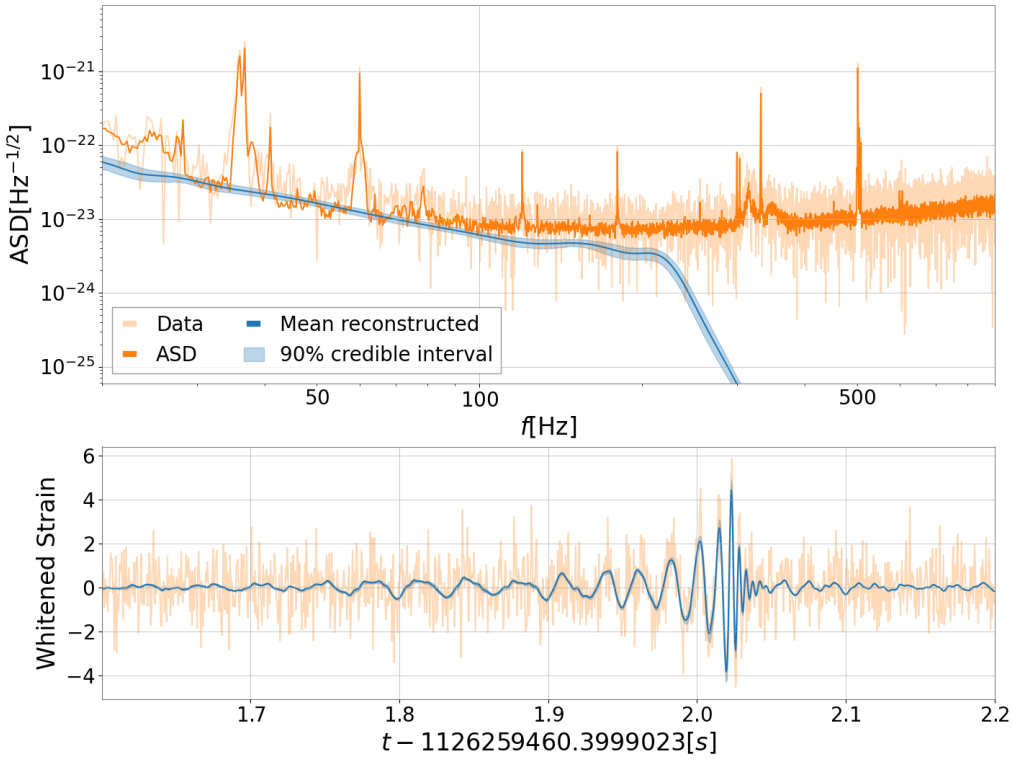

Figure 14. Waveform posteriors for GW150914 using standard BILBY.

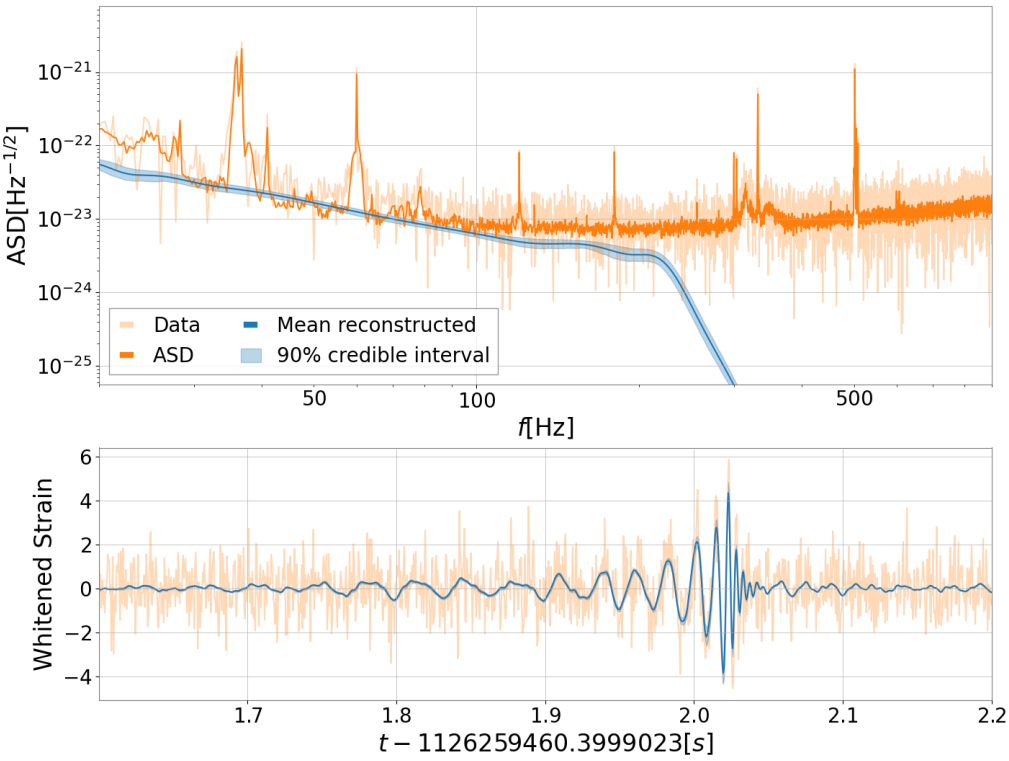

Figure 15. Waveform posteriors for GW150914 using extended BILBY.

estimation in time by linking the priors of neighboring time-chunks, and obtaining a 2d PSD in time-frequency space.

The parameter estimation process discussed in this chapter, when done for all confidently observed signals, provides us with a catalog of many events detected through time. While this chapter focused on analyzing individual events, the next will deal with looking for patterns in the collection of events as a whole.

CHAPTER III

LOOKING FOR CLUSTERS

The work in this chapter is part of a manuscript in preparation co-authored by me, Zoheyr Doctor, and Ben Farr. Zoheyr Doctor had initiated this project with the deterministic k-means and probabilistic agglomerative methods. I developed the rest of the algorithms, modularized the code, and ran the analyses.

3.1 Introduction

While parameter estimation informs us about the individual astrophysical sources that produced the gravitational waves (GWs), we have to look at all the gravitational-wave events together to dig into the overall rates of events and populations of compact binary merger sources. Even though we have no way of knowing for now how an individual binary was formed, we can still use the distribution of the binary parameters of the entire catalog of events to understand the statistical properties of various proposed formation channels, as we'll see in the next chapter. Although various sources of gravitational waves are possible, compact binary coalescences (CBCs) are the only sources we've detected so far; therefore we stick to analyzing only detected compact-object binary mergers in various parameter spaces, specifically binary black hole (BBH) mergers. The research community has come up with a few different approaches to analyze the sample of detected events. One approach is to fit different rate-density models to the data in order to characterize the underlying rate of BBH mergers as a function of component masses and spins. Another approach is to use population synthesis techniques to compare the measured BBH properties with the outputs of simulated BBH populations under different astrophysical prescriptions.

We present a third approach which is entirely empirical and does not attempt to reproduce the underlying rate distributions of binary mergers. Rather than asking about what astrophysical models fit the observed data, we instead ask whether there are interesting features in the data that may warrant theoretical explanation. In a field as observationally young as gravitational-wave astrophysics, clustering algorithms are often useful in exploratory data analysis, and can go on to motivate more mathematically involved models in statistical data analysis. These clustering algorithms are essentially unsupervised machine learning algorithms that perform pattern recognition by asking "Are some objects in a group more similar to each other than to others outside the group?"

3.2 Algorithms

In traditional clustering algorithms, each object is a point in some parameter space. However, in observational and computational science, an object is often not a single point but a distribution of points, owing to uncertainties, errors, or sampling methods. In such cases, the usual reaction is to convert the objects from distributions to point estimates and then clustering them. This works as a great approximation when the uncertainties are much smaller than the inter-cluster distances. However, gravitational-wave data doesn't have that leisure for now, and resorting to such popular tactics may give us false confidence in the results of our clustering analysis.

A better strategy is to build modified versions of these algorithms that take into account the distribution of each object, and result in either a probabilistic distribution of all possible clusterings, or a singular deterministic "expected" clustering. While our code allows any traditional algorithm to be extended to its probabilistic version, we've also constructed and explored the deterministic

versions of: k-means, x-means, agglomerative, and Bayesian hierarchical clustering algorithms.

Probabilistic clustering is an iterative procedure, where at each step, a set of points created by randomly choosing one sample from each event's posterior distribution undergoes a traditional clustering algorithm. The results are shown as a symmetric "co-cluster" matrix where the ij-th element is the probability of the i-th and the j-th events belonging to the same cluster. Now, we'll describe the 4 deterministic clustering algorithms.

3.2.1 k-means. k-means clustering (Bishop (1995); Duda and Hart (1973)) aims to partition objects into k clusters such that each object belongs to the cluster with the nearest cluster center, thereby minimizing the intra-cluster variances. The objective is to minimize the function

$$J[\{x\},\{\mu\}] = \sum_{k=1}^{K}\sum_{n=1}^{N} r_{nk}||\vec{x}_n - \vec{\mu}_k||^2 \tag{3.1}$$

where $\{\vec{x}_n\}$ are the N data vectors and $\{\vec{\mu}_k\}$ are the K cluster centers. r_{nk} is 1 if $\vec{x}_n$ is in cluster k and 0 otherwise. The trick to deal with the parameter uncertainties is that we instead minimize the *expected* objective function

$$\mathbb{E}[J] = \left[\prod_n \int d\vec{x}_n p(\vec{x}_n)\right] \sum_{k=1}^{K}\sum_{n=1}^{N} r_{nk}(\vec{x}_n - \vec{\mu}_k)^2 \tag{3.2}$$

$$= \sum_{k=1}^{K}\sum_{n=1}^{N} \int d\vec{x}_n p(\vec{x}_n) r_{nk}(\vec{x}_n - \vec{\mu}_k)^2 \tag{3.3}$$

$$= \sum_{k=1}^{K}\sum_{n=1}^{N} \langle r_{nk}(\vec{x}_n - \vec{\mu}_k)^2 \rangle_{\vec{x}_n} \tag{3.4}$$

thereby minimizing the *average* intra-cluster variances.

Minimizing $\mathbb{E}[J]$ with respect to $\vec{\mu}_k$ by taking the derivative and setting it to zero yields

$$\vec{\mu}_k = \frac{1}{N_k} \sum_{n=1}^{N} \int d\vec{x}_n p(\vec{x}_n) r_{nk} \vec{x}_n \tag{3.5}$$

$$= \frac{1}{N_k} \sum_{n=1}^{N} \langle r_{nk} \vec{x}_n \rangle_{\vec{x}_n} \tag{3.6}$$

Hence, for our algorithm, a cluster center is not simply the mean of the points in that cluster, but the mean of the means of the distributions in that cluster. In the original point estimate case, N_k would've been the number of data points assigned to the k-th cluster. But here,

$$N_k = \sum_{n} \int d\vec{x}_n r_{nk} p(\vec{x}_n) \tag{3.7}$$

where the integral is the probability that the n-th data point is associated with the k-th cluster.

This minimization problem is computationally difficult (NP-hard), so we use an efficient heuristic "expectation-maximization" algorithm that quickly converges to a local optimum. The standard Lloyd–Forgy algorithm, first proposed by Stuart Lloyd of Bell Labs in 1957, uses an iterative refinement technique that alternates between two steps:

- Expectation: Calculate cluster centers for objects assigned to each cluster.
- Maximization: Assign each object to the cluster with the nearest center.

until the assignments no longer change, i.e., it has converged.

3.2.2 x-means. This concept is also packaged in x-means clustering (Pelleg and Moore (2002)) which extends k-means to not require user input for the number of clusters k. This is done through model selection involving metrics related to the clustering, such as the Bayesian information criterion (BIC) (Kass

and Wasserman (1995)), also called the Schwarz information criterion. This algorithm also alternates between two operations:

- Improve Parameters: Run k-means to convergence.
- Improve Structure: For each cluster, decide whether to split the center in two via a model selection test based on local BIC improvement, essentially asking "Is there evidence that the two children are modeling real structure here, or would the original parent model the distribution equally well?" If there is a split, run a local 2-means and replace the single old center with the two new centers.

The loop runs until we hit a maximum allowed number of clusters $k > k_{\max}$, and then chooses the model with the best global BIC. The BIC deals with the trade-off between a model's goodness of fit and its simplicity, thus balancing between overfitting and underfitting, and is defined as

$$\mathrm{BIC}(M) = p \log n - 2 \log \hat{L}(D) \tag{3.8}$$

where $\hat{L}(D)$ is the likelihood of the data D according to the model M and taken at the maximum likelihood point, n is the number of data points in D, and p is a penalty term for the number of free parameters in M.

To be more transparent, we can list all ks with their corresponding BICs. A good estimate for the size parameter $\hat{\sigma}$ for a cluster is the root-mean-square distance of the cluster samples from the cluster centroid. While computing the BIC for the probabilistic version of x-means is straightforward, computing the BIC for the deterministic version is more involved and only an approximate version is computationally feasible. In this case, the *expected* BIC is calculated to be

$$\mathbb{E}[\text{BIC}] = (K + Kd)\log N - \frac{N - K^2}{2} + \sum_{k=1}^{K} N_k \log N_k \\ - \frac{Nd}{2}\mathbb{E}[\log \hat{\sigma}^2] - N \log N - \frac{N}{2}\log 2\pi \quad (3.9)$$

where K is number of clusters, d is the number of dimensions, N is the total number of objects, N_k is the number of objects assigned to the k-th cluster, and $\hat{\sigma}^2$ is the sum of the intra-cluster variances.

3.2.3 Bayesian GMM. A Gaussian mixture model (GMM) is a probabilistic model that assumes all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters.

$$p(\theta) = \sum_{i=1}^{K} z_i \mathcal{N}(\vec{\mu}_i, \Sigma_i) \quad (3.10)$$

where the probability density function of the multivariate normal distribution $\mathcal{N}(\vec{\mu}, \Sigma)$ is

$$f_X(\vec{x}) = \frac{\exp(-\frac{1}{2}\left((\vec{x} - \vec{\mu})^T \Sigma^{-1} (\vec{x} - \vec{\mu})\right))}{\sqrt{(2\pi)^k |\Sigma|}} \quad (3.11)$$

We can think of GMMs as generalizing k-means clustering to incorporate information about the covariance structure of the data along with the centers of the latent Gaussians.

Quite a few analyses of the GWTC-3 data have involved using GMMs, sometimes without but usually mixed with other distributions like power laws. These analyses use MCMC samplers to get posteriors for the parameters describing the mixture, such as means μ_i, covariances Σ_i, and mixture weights z_i. This is often taken a step further along the Bayesian chain so that the means, covariances, and mixture weights are themselves drawn from distributions whose parameters, called hyperparameters, are the ones being sampled, as shown in the Bayesian graph in

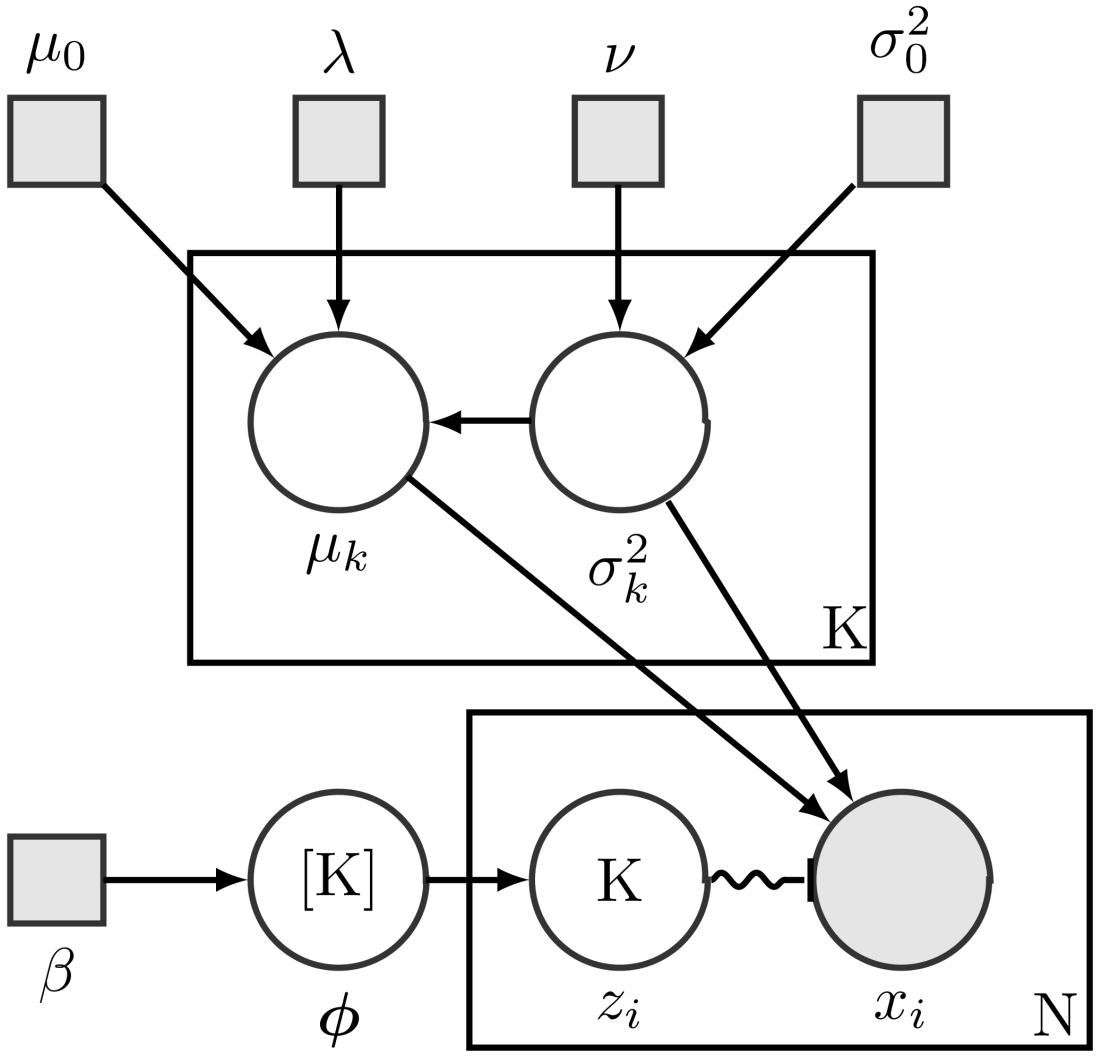

Figure 16. Bayesian Gaussian mixture model.

figure 16. In addition, the number of mixture components K is sampled from a distribution too, often a Dirichlet distribution.

As we're more interested in fast clustering algorithms instead of actual sampling, we implement an expectation-maximization algorithm for fitting the mixture model, just like we did for k-means. While this is more agnostic to cluster size and structure compared to k-means, it still retains some of the issues, such as singularities arising during covariance estimation for clusters with too few points,

or requiring information theoretical criteria to decide on the optimal number of components.

We extend the traditional expectation-maximization algorithm used for GMMs to account for uncertainty so that we can probabilistically cluster distributions instead of points. We first use k-means to choose the initial centers μs for the clusters. We can then estimate the initial covariances Σs from the data assigned to the initial clusters. Then we set the initial prior probabilities θs to be the reciprocal of the number of clusters, implying that each cluster is equally likely a priori. And then algorithm loops between two steps:

– Compute normalized mixture weights zs:

$$z_{n,k} = \frac{1}{Z_n}\text{Mult}(k|\theta)\mathcal{N}(x_n|\mu_k, \sigma_k^2) \tag{3.12}$$

– Re-estimate prior probabilities θs, means μs, and covariances Σs:

$$\theta_k = \frac{1}{N}\frac{\sum_n w_n z_{n,k}}{\sum_n w_n} \tag{3.13}$$

$$\mu_k = \frac{\sum_n w_n z_{n,k} x_n}{\sum_n w_n z_{n,k}} \tag{3.14}$$

$$\Sigma_k = \frac{\sum_n w_n z_{n,k} ||x_n - \mu_k||}{\sum_n w_n z_{n,k}} \tag{3.15}$$

where Mult is the multinomial distribution, Z_n is the normalizer to ensure that the zs sum to one, and ws are the relative weights of the events' posterior samples with respect to the entire catalog. The loop runs trying to maximize the likelihood computed at each step until it hits some convergence criteria signifying we've reached a local optimum.

We use the Bayesian information criterion again to efficiently select the number of components.

3.2.4 Agglomerative. Agglomerative clustering (Duda and Hart (1973)) seeks to build a hierarchy of clusters by merging them in a "greedy" manner. It decides which clusters to combine using a linkage criterion (D), which determines the dissimilarity between clusters (C_i) as a function of the pairwise distances (d) between points (x_i). Linkages include

- Single: $D(C_i, C_j) = \min_{i,j} d(x_i, x_j)$
- Complete: $D(C_i, C_j) = \max_{i,j} d(x_i, x_j)$
- Average: $D(C_i, C_j) = \frac{1}{|C_i||C_j|} \sum_{i,j} d(x_i, x_j)$
- Weighted: $D(C_i \cup C_j, C_k) = \frac{D(C_i,C_k)+D(C_j,C_k)}{2}$
- Centroid: $D(C_i, C_j) = d\left(\langle x_i \rangle, \langle x_j \rangle\right)$
- Median: $D(C_i, C_j) = \frac{\langle x_i \rangle + \langle x_j \rangle}{2}$
- Ward: $D(C_i \cup C_j, C_k)^2 = \frac{|C_i|+|C_k|}{T} D(C_i, C_k)^2 + \frac{|C_j|+|C_k|}{T} D(C_j, C_k)^2 - \frac{|C_k|}{T} D(C_i, C_j)^2$ where $T = |C_i| + |C_j| + |C_k|$

We modified the distance functions to accept distributions instead of points as input, to form an initial event-to-event distance matrix, which is then updated iteratively. All the events start off as their own clusters. Then at each step, the most similar pair of clusters are merged based on the linkage criterion, until all the events are in one cluster, thus creating a dendrogram.

3.2.5 BHC. Bayesian hierarchical clustering (Heller and Ghahramani (2005)) extends the agglomerative algorithm to use a probabilistic model-based statistical hypothesis test to decide cluster mergers instead of a simple distance metric. The result is a binary tree as before, but this tree is a Bayesian mixture

model with each tree node being a mixture component. Each node also carries a posterior probability r_k which can be interpreted as the probability that all the data under that k-th node belongs to the same cluster. Thus, we can prune the tree and obtain the optimal clustering by chopping off the nodes where $r_k < 0.5$. The trick here is more mathematically intensive but the gist is to calculate "expected" posterior probabilities, for lack of a better term.

3.3 Characteristics

3.3.1 Probabilistic vs Deterministic. Every clustering method can be implemented in a probabilistic way or a deterministic way.

The probabilistic scheme involves several iterations of the clustering method. In each iteration, a set of n_{events} samples is created by randomly choosing one sample from each event and then the clustering method is applied to this set and the resulting clustering is recorded. The result is a $n_{\text{events}} \times n_{\text{events}}$ symmetric matrix M with M_{ij} representing the probability of the i-th and j-th events belonging to the same cluster.

In the deterministic scheme, the clustering method is applied once to all events, each event being treated as the collection of its posterior samples, resulting in one clustering. This, of course, requires some mathematical extension to the algorithm so that an event is represented as its entire set of posterior samples instead of just one sample.

Note that one can obtain a probability matrix even with deterministic methods by changing hyper-parameters, e.g., changing initial cluster centers in k-means.

3.3.2 Parameters. The parameter spaces for the clustering algorithms are formed with combinations of the 15 parameters estimated in

analyses: 8 intrinsic parameters describing the 2 individual objects or the binary as a whole, and 7 extrinsic parameters describing the configuration of the merger in space-time relative to the observer. Two intrinsic parameters we've used extensively are the source-frame chirp mass $\mathcal{M}$ and the effective inspiral spin χ_{eff}, which are relatively well-measured because they appear explicitly in the CBC waveforms.

The chirp mass is defined in terms of the masses of the primary component m_1 and of the secondary component m_2 as

$$\mathcal{M} = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}} \tag{3.16}$$

The component masses are usually in the detector frame, so we shift them to the source frame using the relation

$$m_{\text{source}}(1 + z(d_L)) = m_{\text{det}} \tag{3.17}$$

where we obtain the redshift z as a function of luminosity distance d_L assuming Planck 2015's flat ΛCDM cosmology (Ade et al. (2016)).

The effective spin is defined as the mass-weighted sum of the component spins $\vec{\chi}_1$ and $\vec{\chi}_2$ projected onto the orbital angular momentum $\hat{L}$ of the system, thus

$$\chi_{\text{eff}} = \frac{(m_1 \vec{\chi}_1 + m_2 \vec{\chi}_2) \cdot \hat{L}}{m_1 + m_2} \tag{3.18}$$

The individual dimensions may be functions of these parameters, for example, we commonly use the log of the chirp mass $\log \mathcal{M}$ as a dimension instead of just the chirp mass.

3.3.3 Geometry. A Euclidean parameter space works for most parameter combinations and is what most algorithms have been designed for. 3-spherical parameter spaces like RA–Dec–d_L can be converted to Euclidean spaces with a change of variables, and then clustered. However, clustering in non-

Euclidean geometries such as 2-spherical parameter spaces like RA–Dec require modifications to the algorithms. For probabilistic schemes, a redefinition of the geodesic distance function is sufficient. But deterministic schemes need more involved mathematical extensions.

A subtle change for a 2-spherical space is using a von Mises-Fisher distribution instead of a bivariate normal in calculations. The probability density function of the von Mises-Fisher distribution on a 2-sphere is:

$$f(\mathbf{x}; \boldsymbol{\mu}, \kappa) = \frac{\kappa}{4\pi \sinh \kappa} \exp(\kappa \boldsymbol{\mu} \cdot \mathbf{x}) \tag{3.19}$$

where $||\mathbf{x}|| = ||\boldsymbol{\mu}|| = 1$ and $\kappa \geq 0$. $\boldsymbol{\mu}$ and κ are called the mean direction and the concentration parameter, respectively. The maximum likelihood estimates for these parameters are:

$$\hat{\boldsymbol{\mu}} = \frac{\langle \mathbf{x}_i \rangle}{\bar{R}} \tag{3.20}$$

$$\coth \hat{\kappa} - \frac{1}{\hat{\kappa}} = \bar{R} \implies \hat{\kappa} \approx \frac{\bar{R}(3 - \bar{R}^2)}{1 - \bar{R}^2} \tag{3.21}$$

where $\bar{R} = || \langle \mathbf{x}_i \rangle ||$.

3.3.4 Distance metric. For probabilistic methods, the sample-to-sample distance metric is usually the Euclidean distance or the spherical distance, both examples of geodesic distances. For deterministic methods, we have a few different metrics to choose from. For sample-to-event distances, we have:

- Mean: $D(x_i, C_j) = \frac{1}{|C_j|} \sum_j d(x_i, x_j)$
- Mahalanobis: $D(x_i, C_j) = \sqrt{(x_i - \mu_j)^T S_j^{-1} (x_i - \mu_j)}$

where μ and S are the mean and covariance matrix of the cluster C. For event-to-event distances we have:

- Mean: $D(C_i, C_j) = \frac{1}{|C_i||C_j|} \sum_{i,j} d(x_i, x_j)$

– Hellinger: $D(C_i, C_j) = \frac{1}{\sqrt{2}}\sqrt{\sum_x (\sqrt{p_i(x)} - \sqrt{p_j(x)})^2}$

– Bhattacharyya: $D(C_i, C_j) = -\ln(\sum_x \sqrt{p_i(x)p_j(x)})$

3.3.5 Evaluation. Evaluating a clustering result, i.e. judging the "goodness" of a clustering, is as difficult as the clustering algorithm itself. Since in our case the ground truth is not known, evaluations must be performed using the model itself, thus suffering from the problem that they represent functions that themselves can be seen as a clustering objective, thereby being biased towards algorithms that use the same cluster model. Nevertheless, evaluations can help us get a sense of how "clustery" the data is, especially for algorithms that require the number of clusters as input. In such "internal" evaulations, a clustering is assigned a score that rewards high intra-cluster similarity and low inter-cluster similarity. Such measures include the Davies-Bouldin index, the Dunn index, the Silhouette coefficient, the Calinski-Harabasz index, and the Hopkins statistic. A bad score may suggest that there are no clusters in the data to begin with, which we can utilize as a proxy for uniformity. For example, since we expect gravitational-wave sources to be uniformly distributed in sky location (at least for now), a k-means clustering in RA-Dec space should have a bad evaluation.

3.4 Application to GWTC-3

3.4.1 Deterministic. The third gravitational wave transient catalog (GWTC-3) contains almost a 100 compact binary coalescences, of which 69 are BBH detections with a false alarm rate of less than one per year, a reasonable confidence threshold. The results of running deterministic x-means and agglomerative clustering algorithms on these 69 BBH events are shown in figure 18. We compare this to two chirp mass distributions from LIGO-Virgo-KAGRA analyses (The LIGO Scientific Collaboration et al. (2021b)) shown in figure 17.

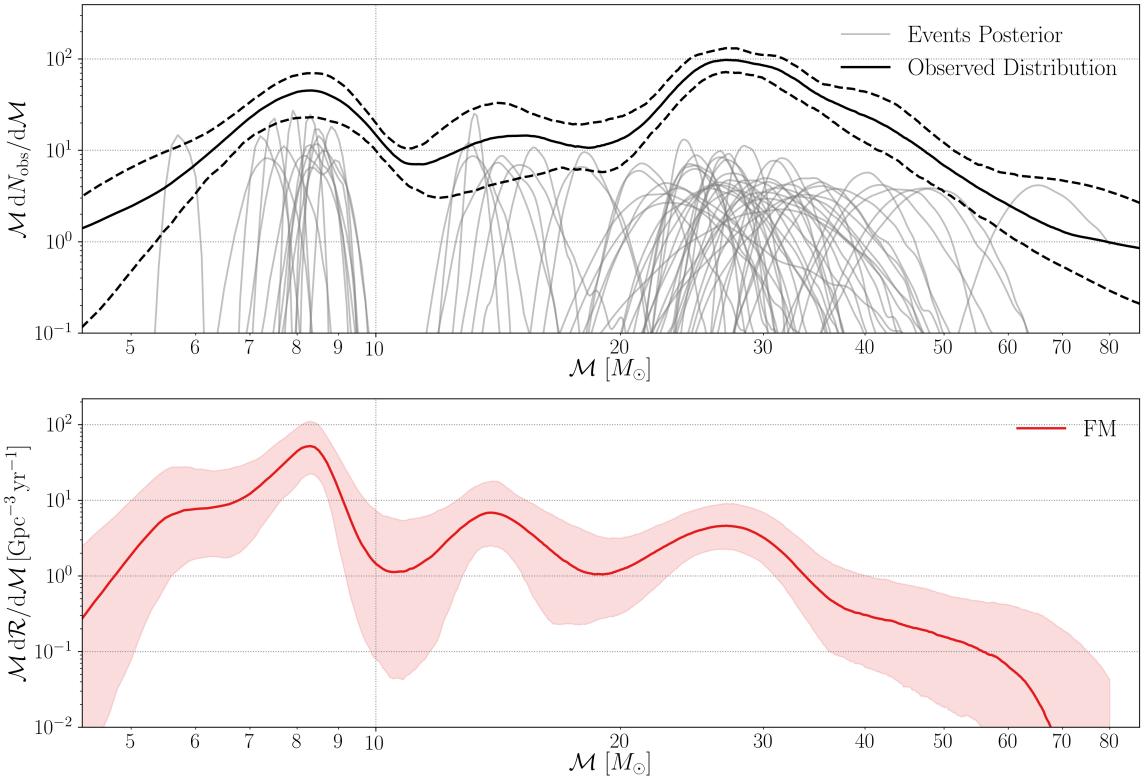

Figure 17. GWTC-3 chirp mass distributions from LIGO-Virgo-KAGRA population analyses. Source: The LIGO Scientific Collaboration et al. (2021b).

We recover the 3 clusters seen in both the inferred distribution using an adaptive KDE (Menne (2022); Sadiq, Dent, and Wysocki (2022); B. Wang and Wang (2007)) and the predicted distribution using the flexible mixture model framework (FM) (Tiwari (2021); Tiwari and Fairhurst (2021)). Agglomerative clustering with Ward's minimum variance linkage criterion results in a clustering almost similar to that from x-means.

3.4.2 Probabilistic. We have a variety of parameters to choose from to build the parameter spaces we cluster in. We chose the combinations we personally considered the most interesting, but also informed by what the work done by the wider collaboration is the most invested in. The parameter spaces we've presented here are:

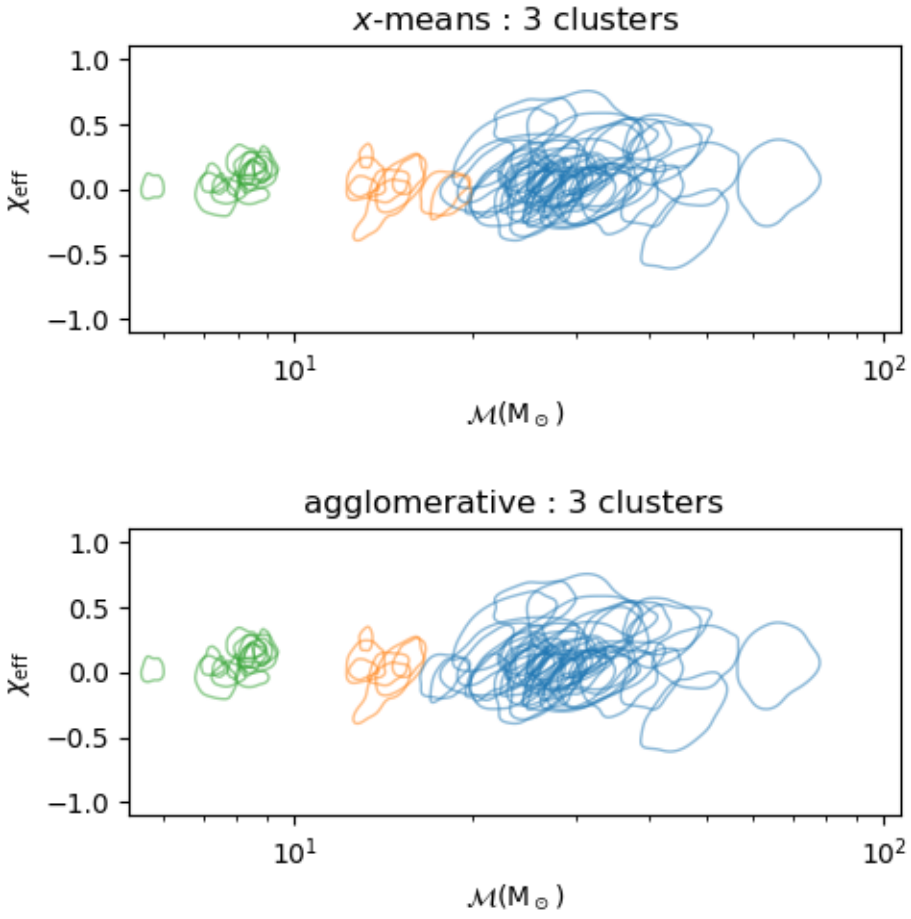

Figure 18. Results of deterministic x-means and agglomerative clustering algorithms on GWTC-3 BBHs in chirp mass – effective spin space.

– primary mass

– primary mass – mass ratio – effective spin

– chirp mass – effective spin

– mass ratio – effective spin – redshift

– RA – Dec

– RA – Dec – comoving distance

Visualizing deterministic clusterings is simply an extension of the standard clustering visualization technique, that is, assigning each cluster a unique color and coloring the clustered objects accordingly. Visualizing probabilistic clusterings, however, proved very challenging. There is no single perfectly informative visualization. Instead, we use different visualizations for the same clustering results to understand their various aspects from different perspectives. The visualizations we've shown here are:

– Co-cluster matrix: Each element p_{ij} in the matrix is the probability of the i-th and the j-th event belonging to the same cluster.

– Histogram of the number of events in each cluster over all clusterings.

– GMM Gaussians: Corner plot of 1d-marginalized and 2d-marginalized Gaussians, only for GMMs. 1d-marginalizations show the distribution of Gaussians associated with all clusters in all clusterings marginalized over all but one dimension. 2d-marginalizations show the same marginalized over all but two dimensions, but the individual Gaussians are plotted as ellipses representing slices at constant σ.

– Convex hulls: Corner plot of convex hulls projected in all 2d subspaces. The convex hull of a cluster is the smallest convex shape containing all the points in that cluster. This shows the cluster shapes more accurately than the Gaussian approximations, and can be used for non-GMM algorithms as well.

– Centroids: Corner plot of 2d histograms of the cluster centroid locations.

– Leave-one-out analysis: We leave out an event and run our clustering algorithms on the rest of the events in the catalog. Then we measure how much of an outlier the left-out event is compared to the models estimated by the clusterings, by computing the likelihood distribution of the event belonging to that clustering model. We also measure the change in the Bayesian evidence brought about by adding that particular event to the clustering.

The results of running probabilistic agglomerative clustering algorithms on GWTC-3 BBHs in two different parameter spaces are shown in figure 19. The 3 clusters expected in chirp mass ($\mathcal{M}$) – effective spin (χ_{eff}) space can be observed here too. But no obvious clusters can be seen in right ascension (α) – declination (δ) space which is consistent with the expectation of uniformity in sky location.

The results of running Bayesian GMM on GWTC-3 BBHs are shown for these parameter spaces:

– primary mass (m_1) in figure 20

– primary mass (m_1) – mass ratio (q) – effective spin (χ_{eff}) in figure 21

– mass ratio (q) – effective spin (χ_{eff}) – redshift (z) in figure 22

– RA – Dec – comoving distance in figure 23

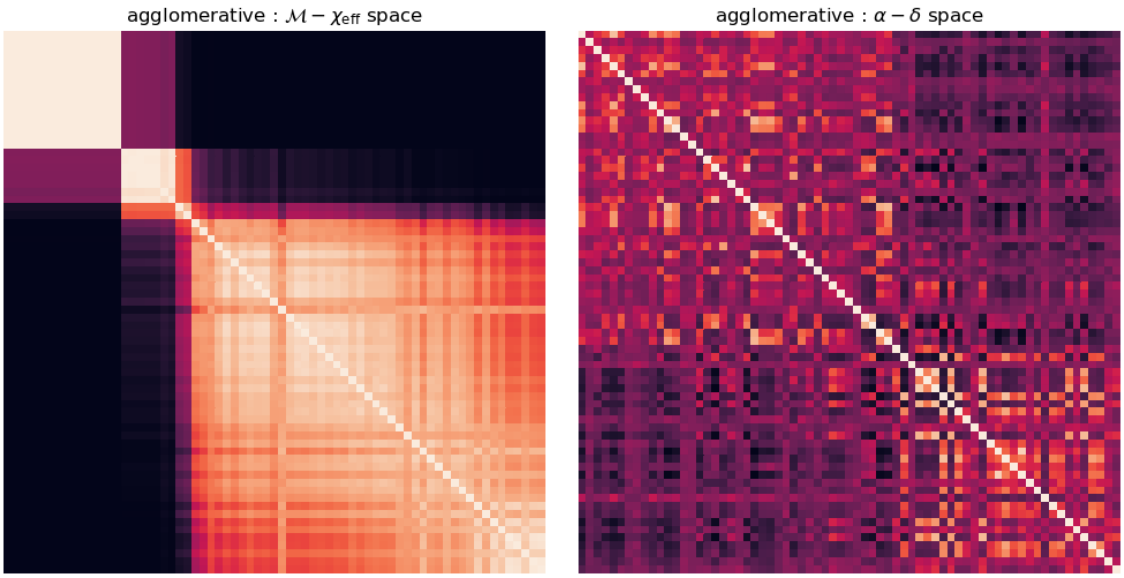

Figure 19. Results of agglomerative clustering algorithms on GWTC-3 BBHs in two different parameter spaces: chirp mass – effective spin, and right ascension – declination.

The leave-one-out analyses confirm that the massive GW190521 always stands out in primary mass, regardless of the clustering, implying that it is most probably an outlier to the rest of the catalog. GW190517_055101 stands out in $q - \chi_{\rm eff} - z$, probably due to its high effective spin of 0.49. Samples in the RA – Dec – comoving distance space were first transformed to those in a Euclidean space centered on the Earth, and then clustered. Some light clustering can be observed in this space, but no event particularly stands out in the leave-one-out analysis. This could be the result of there being no actual clusters except for the detectors' antenna pattern in the sky, or statistical artifacts of having a low number of events.

3.5 Discussion

In this work, we introduced several clustering algorithms to find patterns in data with significant uncertainty. We developed both deterministic methods that produce single clustering results in line with more traditional clustering algorithms,

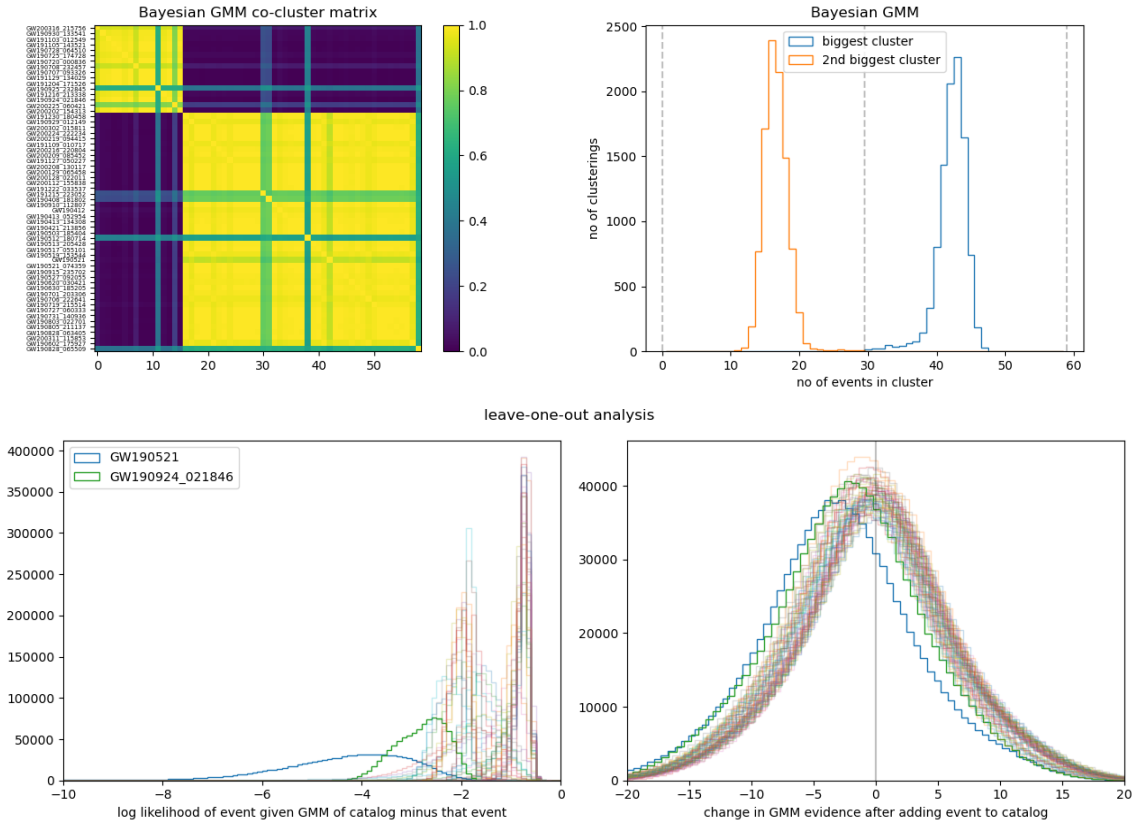

Figure 20. Results of Bayesian GMM on GWTC-3 BBHs in primary mass.

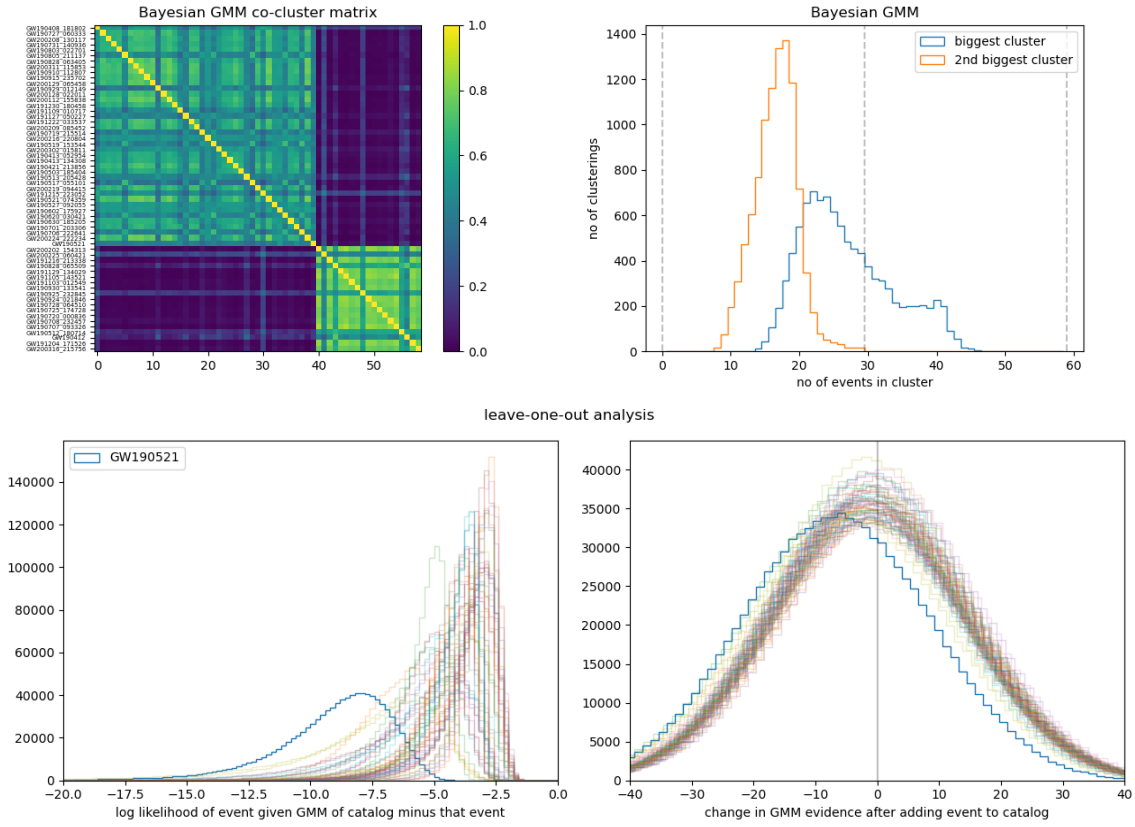

Figure 21. Results of Bayesian GMM on GWTC-3 BBHs in primary mass – mass ratio – effective spin space.

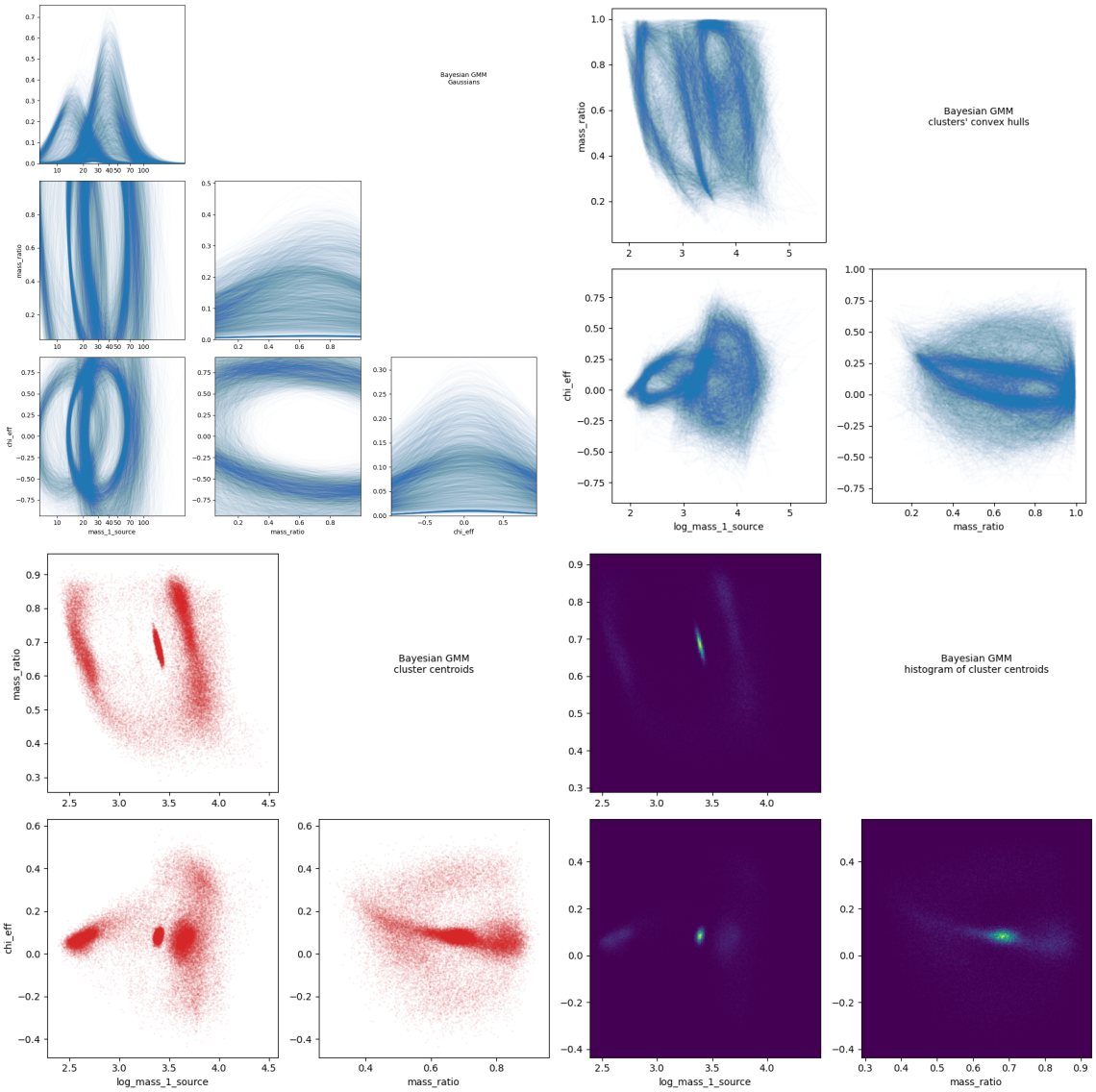

Figure 21. Continued: Results of Bayesian GMM on GWTC-3 BBHs in primary mass – mass ratio – effective spin space.

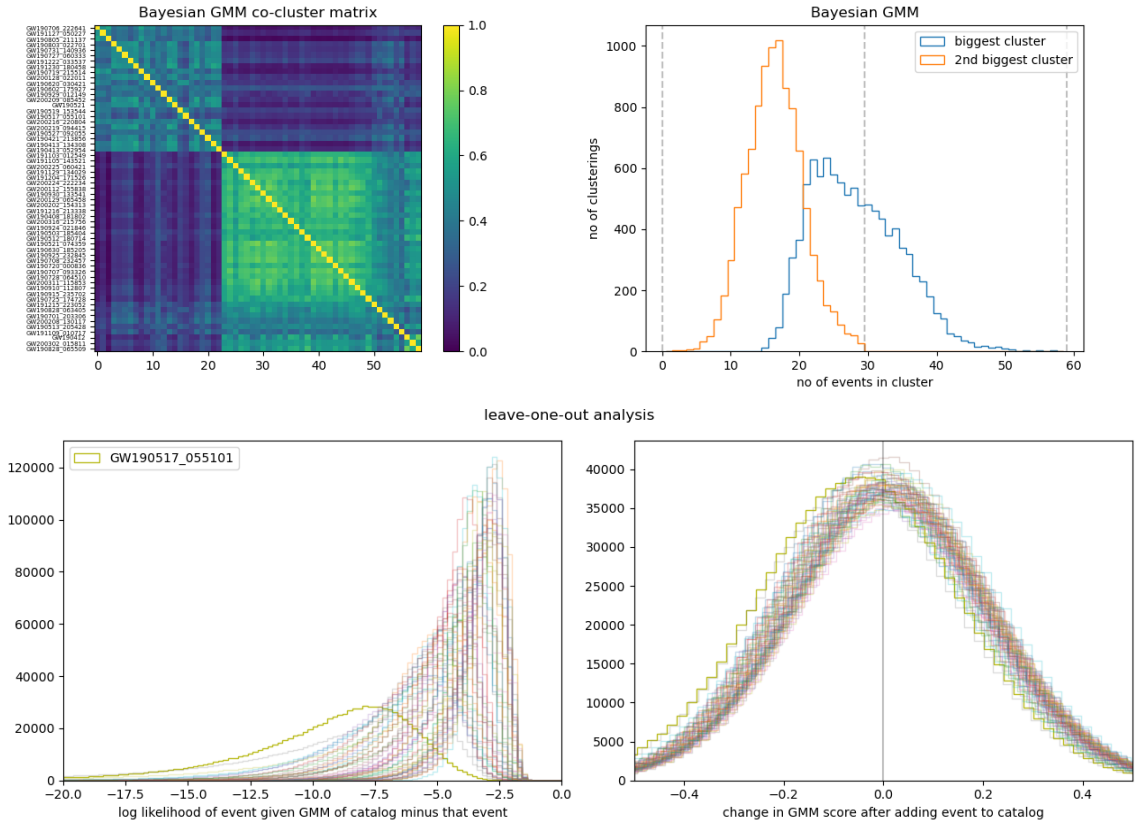

Figure 22. Results of Bayesian GMM on GWTC-3 BBHs in mass ratio – effective spin – redshift space.

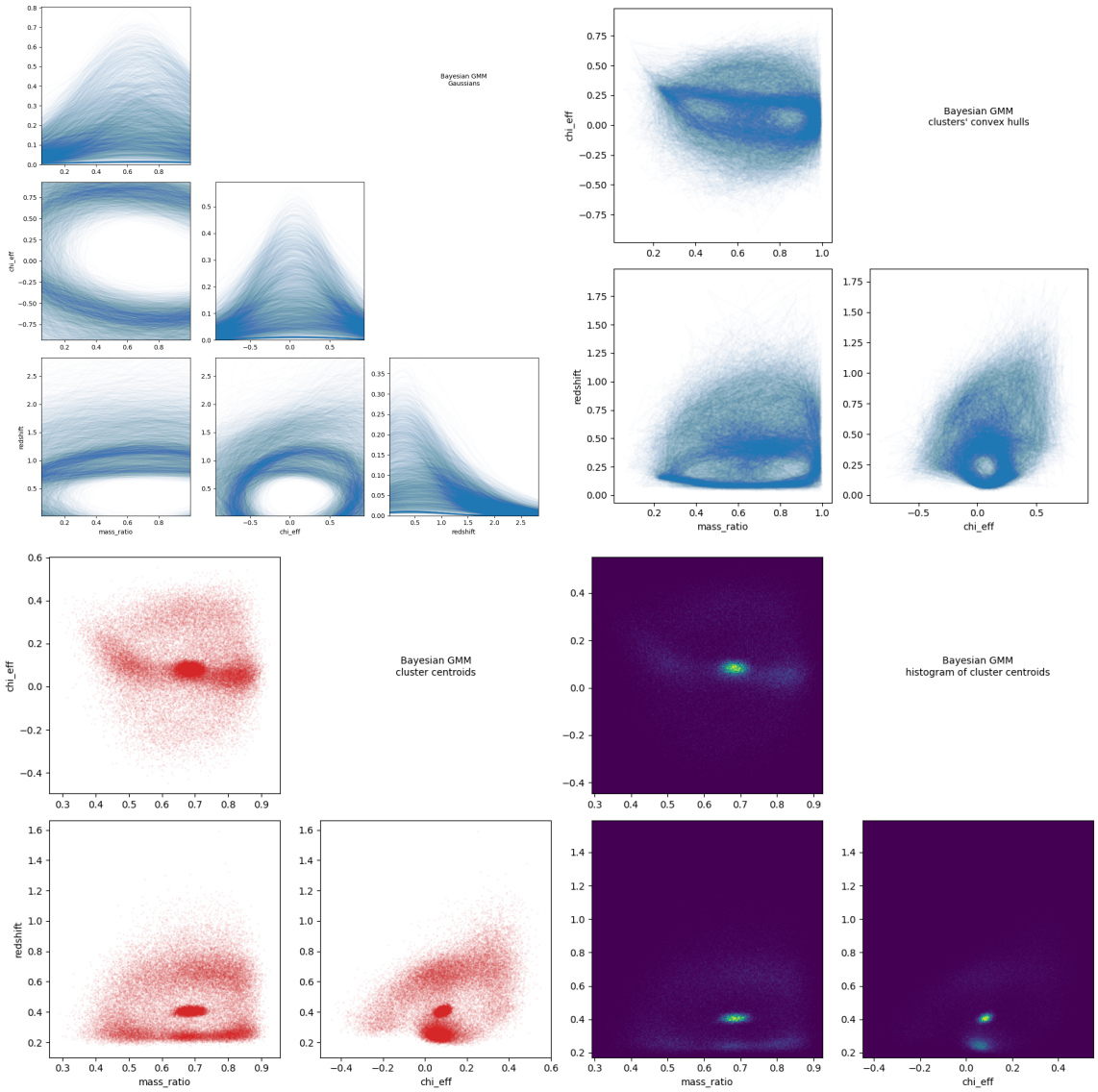

Figure 22. Continued: Results of Bayesian GMM on GWTC-3 BBHs in mass ratio – effective spin – redshift space.

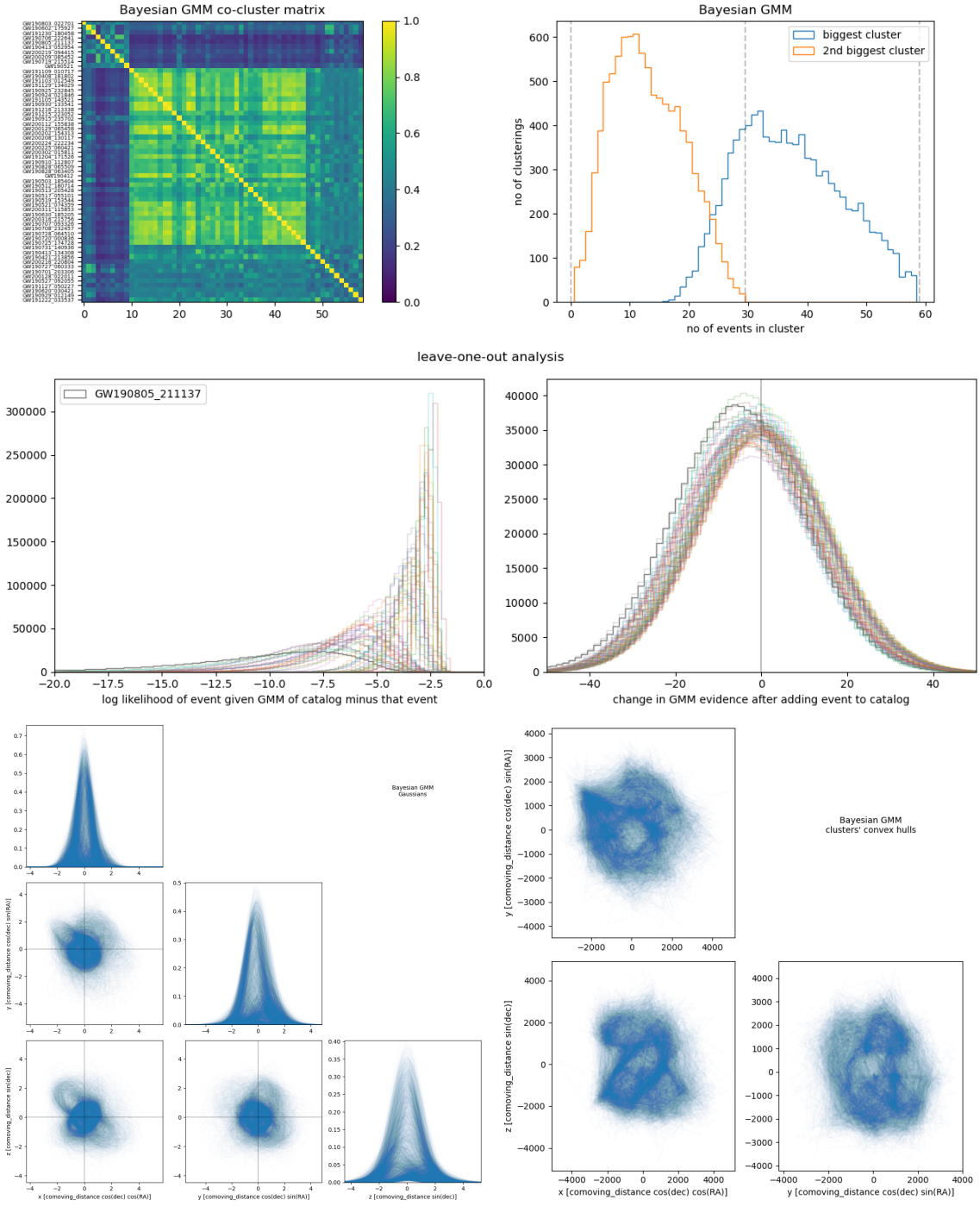

Figure 23. Results of Bayesian GMM on GWTC-3 BBHs in RA – Dec – comoving distance space.

as well as probabilistic algorithms that produce a distribution of clustering results to better reflect the uncertainty in the samples' parameter estimation. Applying this to gravitational-wave data from GWTC-3, we recovered a lot of the patterns found in other population analyses. We also noted the outlier nature of a few events like GW190521.

As the catalog of gravitational-wave events grows, and new detectors come online, clustering techniques will prove useful as a first exploratory approach to looking for patterns in the data. These algorithms can also aid in categorizing posterior samples of population model parameters from various population inference efforts, and thus point out important regions in the model space to direct future work.

The patterns revealed by the clustering algorithms provide a good springboard to dive into the gravitational-wave data with more sophisticated statistical tools and find the origins of those patterns. While this chapter dealt mainly with looking for clusters in the catalog with simple heuristic models, the next will deal with inferring astrophysically relevant parameters with physically motivated models of black hole formation channels.

CHAPTER IV

HIERARCHICAL MERGERS

The work in this chapter is part of a manuscript in preparation co-authored by me, Zoheyr Doctor, Ben Farr, and Daniel Wysocki. It's based on the hierarchical merger model initially developed by Zoheyr Doctor, Daniel Wysocki, Richard O'Shaughnessy, Daniel E. Holz, and Ben Farr in Doctor et al. (2020). I primarily developed the astrophysically-motivated models, especially those for globular clusters and AGN disks, as well as designed the population mixture model inference technique. I also built the new population inference code from scratch, and ran the analyses. Carl Rodriguez, Barry McKernan, and Saavik Ford shared their deep insights into astrophysical environments.

4.1 Introduction

As the population of observed gravitational-wave events grows, questions regarding where and how the merging compact binaries form become more prominent, as does our ability to answer them. The proposed formation scenarios can be broadly separated in two categories: isolated evolution in the galactic field, and dynamical assembly in dense environments such as star clusters and active galactic nucleus (AGN) disks. Isolated evolution (field) channels include the common envelope channel (Paczynski (1976); Tutukov and Yungelson (1993); Van Den Heuvel (1976)), the stable mass transfer channel (Neijssel et al. (2019); van den Heuvel, Portegies Zwart, and de Mink (2017)), and the chemically homogeneous evolution channel (De Mink and Mandel (2016); Mandel and De Mink (2016); Marchant, Langer, Podsiadlowski, Tauris, and Moriya (2016)). Dynamical channels include black holes merging through strong gravitational encounters in environments such as globular clusters (Rodriguez, Zevin, Pankow, Kalogera, and

Rasio (2016)), nuclear star clusters (Antonini, Gieles, and Gualandris (2019)), and AGN disks (Bartos, Kocsis, Haiman, and Márka (2017); McKernan, Ford, Kocsis, Lyra, and Winter (2014); Stone, Metzger, and Haiman (2017)). More esoteric formation channels like population III stars (Inayoshi, Hirai, Kinugawa, and Hotokezaka (2017); Kinugawa, Inayoshi, Hotokezaka, Nakauchi, and Nakamura (2014); Madau and Rees (2001)), primordial black holes (Clesse and Garcia-Bellido (2020); Sasaki, Suyama, Tanaka, and Yokoyama (2018)), and Proca stars (Bustillo et al. (2021); Sanchis-Gual et al. (2022)) have also been proposed.

We understand that stellar-mass black holes are usually formed by the gravitational collapse of stars undergoing supernova explosions. Models of the late-stage evolution of massive stars (Barkat, Rakavy, and Sack (1967); Bond, Arnett, and Carr (1984); Fowler and Hoyle (1964); Fraley (1968); Rakavy and Shaviv (1967); S. Woosley (2017); S. E. Woosley, Heger, and Weaver (2002)) predict that contraction of the core leads to electron-positron pair production that reduces internal pressure support, causing further contraction that powers explosive nuclear burning and a rebounding shock. For helium-core masses $\approx 32 - 64\ M_{\odot}$, multiple pulsational episodes can eject sufficient material to reduce the mass below the pair-instability regime, ending with a black hole remnant; the process leads to a supernova impostor phenomenon called pulsational pair-instability supernova (PPISN). A single pulse can entirely disrupt stars with larger helium cores, leaving behind no remnant in a pair-instability supernova (PISN). At even larger helium-core masses $\geq\ 135\ M_{\odot}$, this is avoided as the high core temperature results in photodisintegration that accelerates gravitational collapse to a massive black hole. This leads to the robust prediction from single-star evolution of the existence of a gap in the black hole mass distribution. Though this gap is broadly consistent with

the range $\approx 60 - 130\ M_{\odot}$, there are several theoretical uncertainties that affect both the lower edge and total extent of the gap.

Now one can see why the detection of GW190521 ($m_1 = 85^{+21}_{-14}\mathrm{M}_{\odot}$, $m_2 = 66^{+17}_{-18}\mathrm{M}_{\odot}$, $m_{\mathrm{rem}} = 142^{+28}_{-16}\mathrm{M}_{\odot}$) opened a new Pandora's box. Both the parent black holes straddle the mass gap, putting their own stellar origin into question. The PISN mass gap can be populated with some stellar and binary evolution processes, such as weaker stellar winds or core dredge-up episodes that allow a star to retain a hydrogen envelope, stellar mergers in dense clusters, multi-stellar systems, or fast-orbiting stellar binaries with rapidly rotating metal-poor stars that undergo chemically homogeneous evolution. But so far, the most popular formation channel for black holes in the mass gap has been hierarchical mergers, where one or both the parent black holes are themselves the result of previous BBH mergers (Antonini, Romero-Shaw, and Callister (2025); Gerosa and Fishbach (2021); Hussain, Isi, and Zimmerman (2024); Kimball et al. (2021); Li, Wang, Tang, and Fan (2024); Mould, Gerosa, and Taylor (2022); Pierra, Mastrogiovanni, and Perriès (2024); Y.-Z. Wang et al. (2022)).

Early in the development of the hierarchical merger hypothesis, it was believed that these mergers occur in several dynamical environments such as globular clusters, AGN disks, nuclear star clusters, or in the early universe. But the black hole remnants of binary mergers receive gravitational recoils with "kick" velocities as high as $10^2 - 10^4$ km/s. For these remnants to participate in further mergers, the BBH mergers have to occur in environments with high escape speeds. That cuts down the possible environments to very dense stellar clusters and AGN disks. We note that high-mass stellar mergers in the aforementioned environments

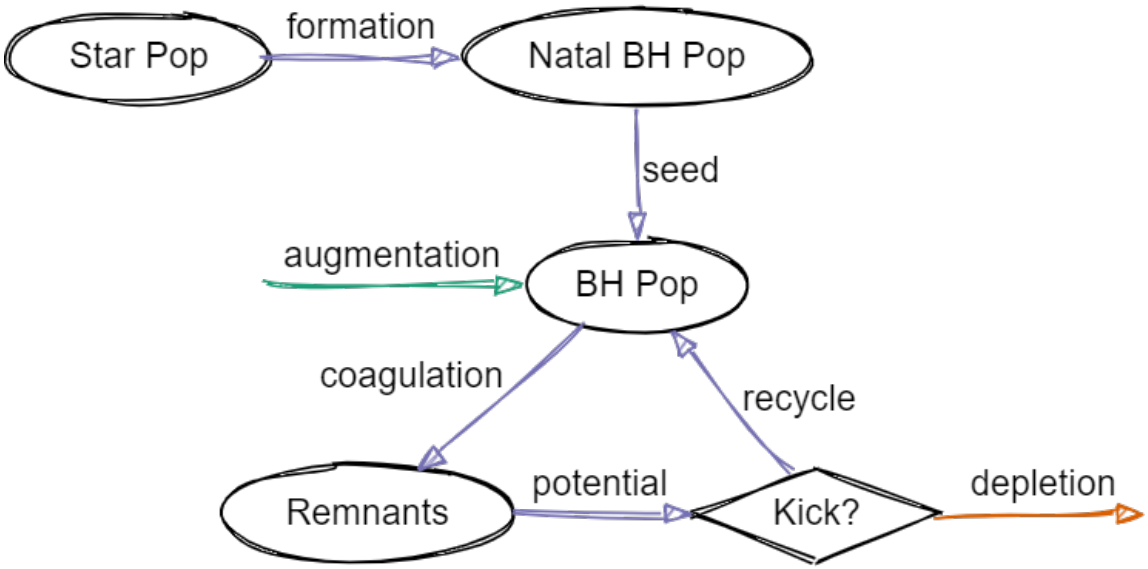

Figure 24. Framework of the coagulation hierarchical merger model.

can produce high-mass black holes with smaller recoils, and hence maybe a viable alternate formation channel.

4.2 Coagulation model

The coagulation model, introduced in Doctor et al. (2020), imagines black holes as sticky particles in a box undergoing collisions based on an effective cross section, without worrying about the complex dynamics of the individual stellar environments. The flexibility of this parameterized framework allows a wide range of submodels and prescriptions. This allows the creation of models that are purely phenomenological or based on astrophysical observations or simulations, which can then be followed by performing Bayesian inference using gravitational-wave observations.

In this framework, as shown in figure 24, an initial black hole population is evolved with a Monte Carlo procedure that approximates a continuous-time coagulation equation (Smoluchowski (1916))

$$\frac{\partial}{\partial_t} f(x;t) = \frac{1}{2}\int dx' dx'' f(x';t) f(x'';t) \Gamma(x', x'';t) \delta(x_{\rm rem}(x'', x') - x) \\ - \int dx' f(x;t) f(x';t) \Gamma(x, x';t) + r(x;t) - d(x;t) \quad (4.1)$$

where the first two integrals together represent coagulation (2 compact objects merging to form 1), r represents augmentation (new compact objects being introduced to the environment), and d represents depletion (objects leaving the environment never to merge again). The xs represent black hole parameters like mass, spin and kick velocity. In the iterative procedure, we start with a natal black hole population $f(x, t_0)$ and at each step, merge a set fraction w of black holes based on the coagulation coupling, compute the remnant parameters $x_{\rm rem}$, and determine whether they'll be retained in the environment based on their kick velocities. The merger fraction w, typically 5%, is a proxy for the merger timescale.

We use the coagulation process to only model mergers in dynamical environments, such as globular clusters and AGN disks. To model field mergers, we use the TRUNCATED mass model and a modified version of the DEFAULT spin model described in The LIGO Scientific Collaboration et al. (2021b). In the TRUNCATED mass model, the primary mass distribution follows a truncated power-law

$$p(m_1) \propto \begin{cases} m_1^{-\alpha}, & \text{for } m_{\rm min} \leq m_1 \leq m_{\rm max} \\ 0, & \text{otherwise.} \end{cases} \quad (4.2)$$

and the mass ratio follows a power-law truncated conditioned on the primary mass

$$p(q) \propto \begin{cases} q^{\beta}, & \text{for } m_{\rm min} \leq m_2 \leq m_1 \\ 0, & \text{otherwise.} \end{cases} \quad (4.3)$$

In our modified DEFAULT spin model, the dimensionless spin magnitude for each component black hole follows a Beta distribution

$$p(\chi_{1,2}) = \text{Beta}(\mathbb{E}[\chi], \text{Var}[\chi]) \tag{4.4}$$

and the cosine of the tilt angle between the component spin and the binary's orbital angular momentum follows a truncated half Gaussian distribution in $[-1, 1]$, with a peak at 1 and width σ_t

$$p(\cos\theta_{1,2}) \propto \begin{cases} \mathcal{N}(1, \sigma_t), & \text{for } -1 \leq \cos\theta_{1,2} \leq 1 \\ 0, & \text{otherwise.} \end{cases} \tag{4.5}$$

This reflects the assumption that the spins in field mergers are preferentially aligned.

We model the two dynamical environments slightly differently: globular clusters have extreme depletion due to their shallow gravitational potentials allowing low escape velocities, whereas AGN disks have negligible depletion due to the supermassive black holes' deep gravitational potential wells. Black hole spins in globular clusters are assumed to be isotropic, implying the tilt angles are spherically uniformly distributed, and we assign random tilts to black holes participating in any generation of hierarchical mergers. However, in AGN disks, since black hole spins and orbits are preferentially aligned to the disk, BBH component spins are also aligned to each other and the binary orbit. So we assign natal tilts from a truncated half Gaussian distribution similar to field mergers, but we keep track of and evolve the tilts as black holes progress through hierarchical mergers.

The most general models we can build are mixture models of the individual population models we discussed. In this work, we have focused on inferring the

parameters describing these mixture models. The following descriptions are about dynamical environmental models, unless explicitly stated otherwise.

4.2.1 Interaction. Γ in the coagulation equation denotes a volume- and time-averaged interaction rate, which we assume only depends on the total mass and the symmetric mass ratio η, as such

$$\Gamma_{m_1,m_2} \propto (m_1 + m_2)^a \eta^b \tag{4.6}$$

where

$$\eta = \frac{m_1 m_2}{(m_1 + m_2)^2} \tag{4.7}$$

ranges from 0 (perfectly asymmetric) to 0.25 (perfectly symmetric).

This single interaction term captures the average effect of diverse interactions, such as effective cross-sectional areas, dynamical friction, and mass segregation, the dependences of which have been swept under the exponents. Since we're focusing on interaction environments such as globular clusters and AGN disks where we can assume that the probability of interaction only depends on the component masses, we assume that spins do not influence Γ but do keep track of spin magnitudes to compute merger and remnant properties.

We use analytical fits to numerical relativity simulations to calculate remnant parameters, namely, mass and spin (Tichy and Marronetti (2008)), and recoil velocity (Zlochower and Lousto (2015)).

4.2.2 Depletion. We address two kinds of depletion: pre-merger and post-merger. Due to dynamical interactions, the vast majority of BBHs in globular clusters are expelled from their host globular clusters before they merge. Even though these BBHs merge outside globular clusters in the field, we still consider them globular cluster mergers because the stellar evolution pathways and dynamics

that formed these BBHs originated inside globular clusters. Observationally, we cannot locate where the BBHs merged; we can only infer where they were formed.

In our depletion model, we assume that a fraction $f_{\rm ret}$ of BBHs formed in globular clusters merge inside the globular clusters. $1 - f_{\rm ret}$ BBHs are first expelled from the globular cluster black hole population. For the remaining $f_{\rm ret}$ BBHs, we compute the remnant black hole parameters and decide whether to expel them based on their kick velocities as well as the globular cluster escape velocities.

If there is no post-merger depletion, like in an AGN disk, all merger remnants stay in the environment and may participate in further mergers. But we can turn on depletion using a prescription for the distribution of globular cluster escape velocities. Different star clusters would have different density profiles resulting in different central escape velocities. Globular clusters are low-binding energy environments that lead to high depletion. Shallower confining potentials are better at suppressing hierarchical mergers. We generate cluster escape potentials using the Plummer model (Plummer (1911)), assuming the black holes are located at the centers of clusters.

In the Plummer model of globular clusters, the gravitational potential goes as $-\frac{GM}{\sqrt{r^2+a^2}}$ where M is the total mass of the globular cluster and a is the Plummer radius, a scale parameter that sets the size of the globular cluster core. The 2D projected half-mass radius R is exactly a in this model. Assuming black holes are merging only at the center of globular clusters, the potential is then $-\frac{GM}{R}$. In our model, any remnant black hole with kinetic potential greater than the gravitational potential will escape the globular cluster, and will not otherwise, i.e., the probability of it escaping a particular globular cluster is a hard step function of

the kick velocity.

$$p(\text{escape}|v_{\text{kick}}, M, R) = \Theta\left(\frac{1}{2}v_{\text{kick}}^2 > \frac{GM}{R}\right) \tag{4.8}$$

Moving away from the simple globular cluster model used in Doctor et al. (2020), we used data from the FIRE-2 magnetohydrodynamics galaxy simulation (Grudić et al. (2022); Rodriguez et al. (2022)) to build an astrophysically motivated distribution of cluster potentials. We assume that the globular cluster masses follow a simple power law

$$p(M) \propto \begin{cases} M^{-\alpha}, & \text{for } 10^3 M_\odot \leq M \leq 10^7 M_\odot \\ 0, & \text{otherwise.} \end{cases} \tag{4.9}$$

and that the 2d projected half-mass radius R depends on the globular cluster mass as

$$R = k\left(\frac{M}{10^4 M_\odot}\right)^p \tag{4.10}$$

with a scatter of σ dex independent of globular cluster mass. The best-fit model from Grudić et al. (2022) gives $\alpha = 2.5, p = 0.25, k = 1.4\text{pc}, \sigma = 0.5\text{dex}$. A distribution of globular cluster masses and radii for these values is shown in figure 25.

Assuming the mergers are happening in random globular clusters, i.e. there is no intrinsic correlation between merger parameters and globular cluster parameters, the overall escape probability can be described as a function of kick velocity and globular cluster distribution parameters.

$$p(\text{escape}|v_{\text{kick}}, \alpha, k, p, \sigma) \propto \int dM dR\; p(\text{escape}|v_{\text{kick}}, M, R)\; p(M, R|\alpha, k, p, \sigma) \tag{4.11}$$

$$= F_1\left(\frac{1}{2}v_{\text{kick}}^2\right) F_2\left(\frac{1}{2}v_{\text{kick}}^2\right) + \frac{1}{2} \tag{4.12}$$

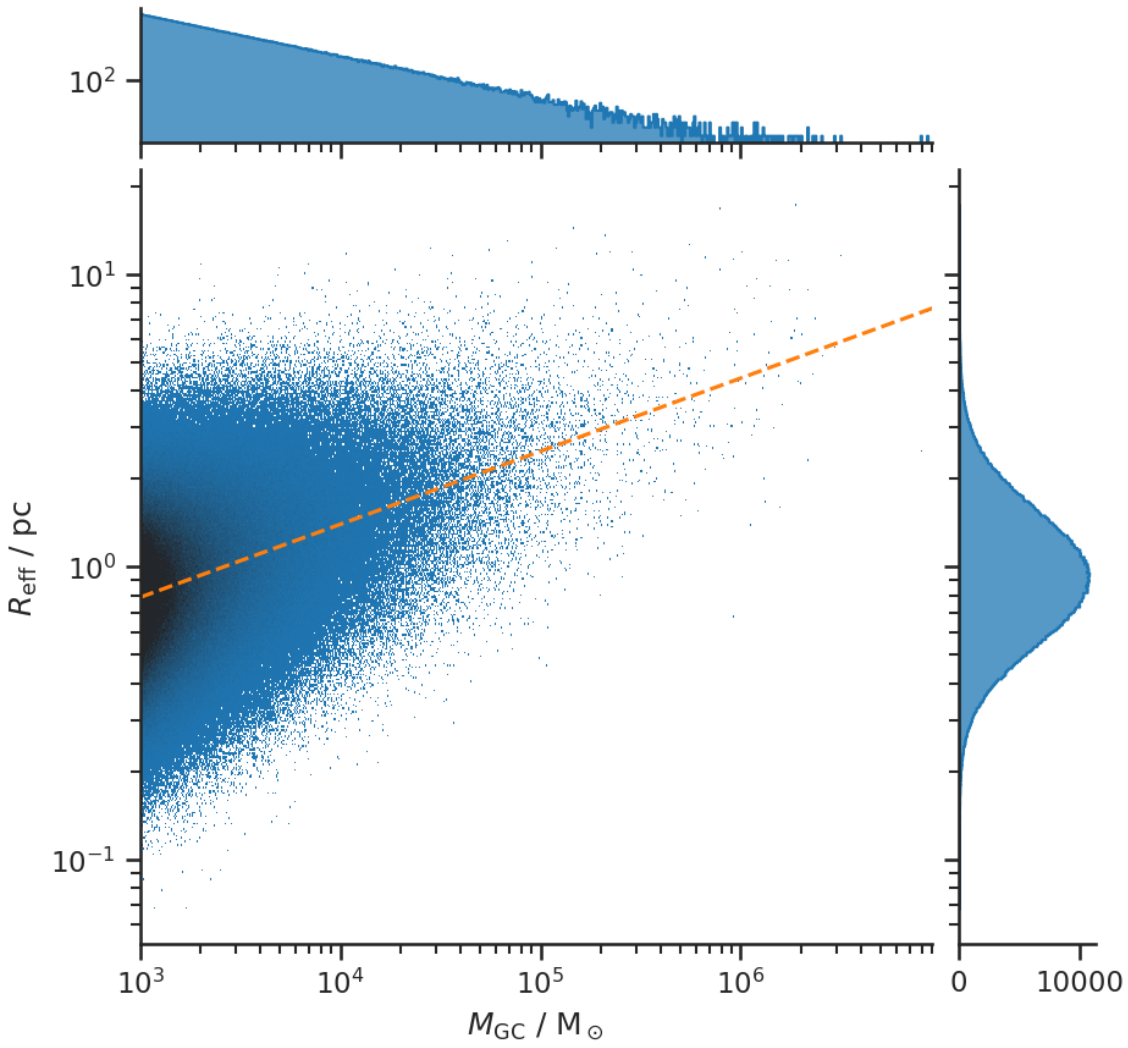

Figure 25. Globular cluster masses and radii for fiducial model parameters. Radius as a function of mass before introducing scatter is shown in orange.

where

$$F_1(\phi) = \frac{1}{2}\left(\frac{k}{10^{4p}G}\right)^{\beta}\frac{\phi^{\beta}}{M_{\rm max}^{1-\alpha} - M_{\rm min}^{1-\alpha}} \tag{4.13}$$

$$F_2(\phi) = F(\phi;\psi_{\rm max}) - F(\phi;\psi_{\rm min}) \tag{4.14}$$

$$F(\phi;\psi) = \exp\left(\frac{\beta^2\sigma^2}{2}\right)\mathrm{erf}\left(\frac{\ln(\psi/\phi) - \beta\sigma^2}{\sqrt{2}\sigma}\right) - \left(\frac{\psi}{\phi}\right)^{\beta}\mathrm{erf}\left(\frac{\ln(\psi/\phi)}{\sqrt{2}\sigma}\right) \tag{4.15}$$

$$\beta = \frac{1-\alpha}{1-p} \tag{4.16}$$

$$\psi = \frac{10^{4p}G}{k}M^{1-p} \tag{4.17}$$

If we simply pair the sampled mergers with the sampled globular clusters, the implication would be that merger rates are uniform across all globular clusters. However, merger rates do depend on globular cluster parameters. Assuming the number of mergers scales with the globular cluster mass as $M^{1.6}$ over a wide range of globular cluster masses from $10^2 - 10^8 M_\odot$ (Antonini and Gieles (2020)), we change $\alpha \to \alpha - 1.6$, making heavier globular clusters more likely to be drawn and paired with a merger. In general, the power law index for globular cluster masses $\alpha_{\mathcal{N}_{\rm GC},M}$ and the merger-to-GC-mass dependence $\alpha_{\mathcal{R},M}$ are degenerate and we can only infer the combined parameter $\alpha = \alpha_{\mathcal{N}_{\rm GC},M} - \alpha_{\mathcal{R},M}$.

4.2.3 Natal populations. We treat the initial distribution of black hole parameters as a separable distribution of masses, dimensionless spin magnitudes, and tilt angles (aka spin-orbit misalignments).

$$p(m, \boldsymbol{\chi}) = p(m)p(\chi)p(\cos\theta) \tag{4.18}$$

where $\cos\theta$ is the cosine of the tilt angle between component spin and a binary's orbital angular momentum. The natal mass distribution of component masses can be

– a simple power law

$$p(m) \propto \begin{cases} m^{-\alpha}, & \text{for } m_{\min} \leq m \leq m_{\max} \\ 0, & \text{otherwise.} \end{cases} \tag{4.19}$$

– a broken power law

$$p(m) \propto \begin{cases} m^{-\alpha_{\text{low}}}, & \text{for } m_{\min} \leq m \leq m_{\text{break}} \\ m^{-\alpha_{\text{high}}}, & \text{for } m_{\text{break}} \leq m \leq m_{\max} \\ 0, & \text{otherwise.} \end{cases} \tag{4.20}$$

– evolved from a stellar population following a simple power law.

A stellar initial mass function can be evolved using stellar evolution prescriptions, such as the "rapid" model in Fryer et al. (2012), introducing a new variable, the stellar metallicity Z_{metal}. We follow Golomb, Isi, and Farr (2024) and transform an initial mass distribution through a probabilistic function to a remnant mass distribution that automatically models the PISN mass gap as well as a PPISN "bump". A PPISN bump is the expected pile-up of black holes just below the mass gap created from progenitor stars that had undergone pair instability pulsations (Talbot and Thrane (2018)). Some studies use the PPISN mass bump, modeled as a Gaussian, to explain an overdensity in the BBHs' mass distribution found in LIGO-Virgo-KAGRA analyses. The evolution model introduces three new variables: the minimum mass for pulsations m_{PPISN}, the maximum black hole mass $m_{\text{BH,max}}$, and the standard deviation of the uncertainty σ. The effect of this function for different choices of these parameters is shown in figure 26. We use Farmer, Renzo, De Mink, Marchant, and Justham (2019) and Fryer, Olejak, and Belczynski (2022) to inform the possible range of the PISN mass gap parameters, that we bake into our priors.

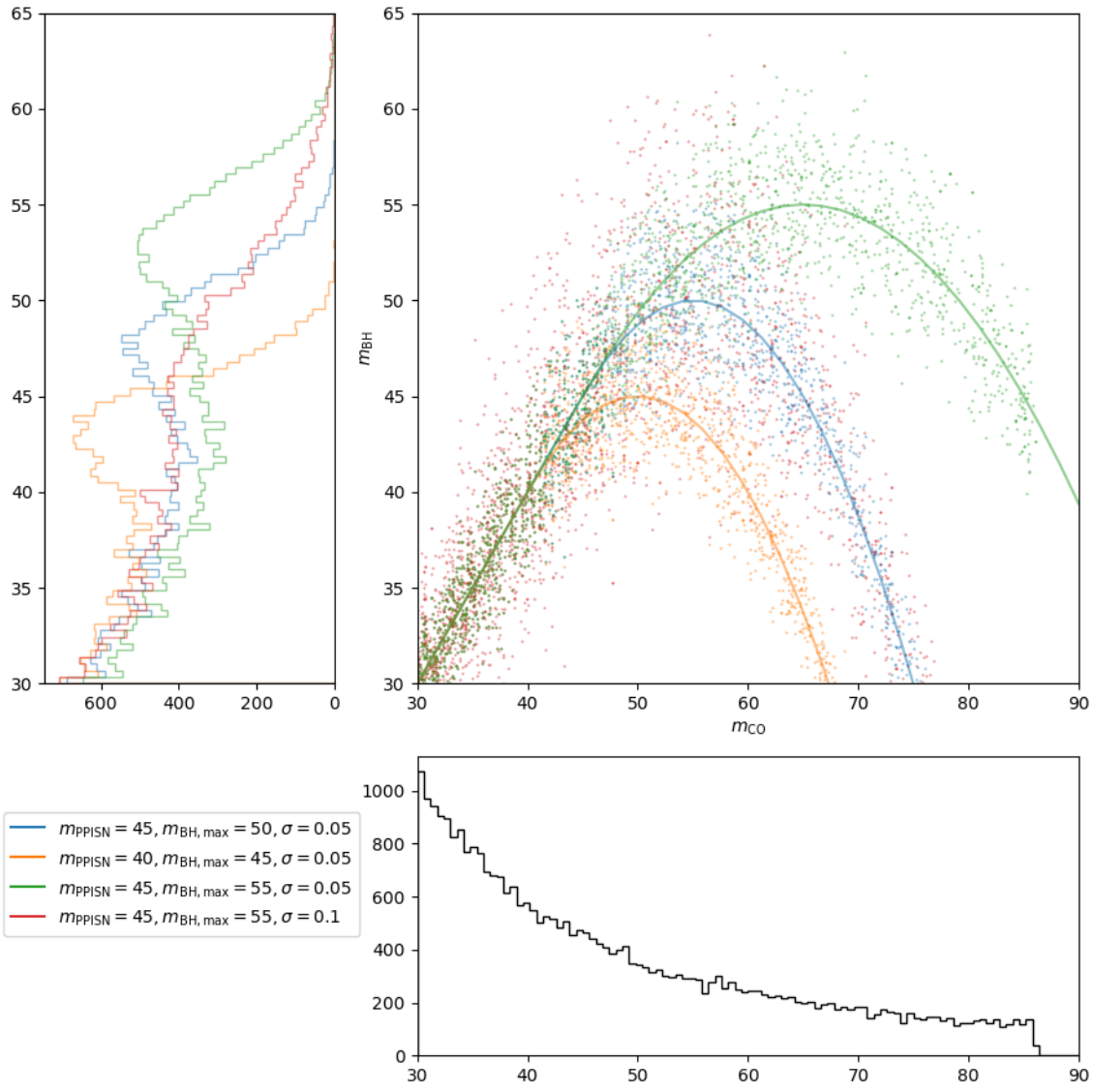

Figure 26. Black hole mass distributions for different choices of PISN mass gap parameters with a fiducial progenitor initial mass function.

The natal spin magnitudes χ follow a Beta distribution defined by its mean $\mathbb{E}[\chi]$ and variance $\mathrm{Var}[\chi]$, with fiducial values being $\mathcal{B}(\mathbb{E}[\chi] = 0.047, \mathrm{Var}[\chi] = 0.002)$. For BBHs in globular clusters, the natal spins are assumed to be isotropic, hence the tilt angles follow a uniform distribution in $[-1, 1]$.

4.2.4 Merger rates. A crucial implication of the finite set of black hole samples being a proxy for the entire population is that the overall merger rate is just a scaled value of those in our simulations. We infer the merger rate density while working within the stipulation that it is constant in co-moving volume. The BBH merger rate is expected to vary with redshift, and there's evidence to that end (Callister and Farr (2024); Edelman, Farr, and Doctor (2023); Fishbach, Holz, and Farr (2018); Payne and Thrane (2023); Ray, Hernandez, Mohite, Creighton, and Kapadia (2023)), but we do not include that dependence in our models.

In a mixture model, there are as many channel-wise merger rates as submodels in the mixture, with their sum being the overall merger rate $\sum_c \mathcal{R}_c = \mathcal{R}_{\mathrm{tot}}$.

4.3 Evolution

As expected and evident in figure 27, time evolution of a natal population without any depletion results in a secondary "island" of high mass, high spin black holes. Hierarchical mergers produce strong evolution of black hole spins due to remnants inheriting angular momentum from their parents and their orbit (Fishbach, Holz, and Farr (2017); Gerosa and Berti (2017)). The peak near $\chi \simeq 0.7$ is characteristic of comparable-mass binaries. Evolving with a high total mass index a and a high symmetric mass ratio index b results in multimodal mass distributions characterized by steps at multiples of the primordial maximum mass, that smooth out at high masses.

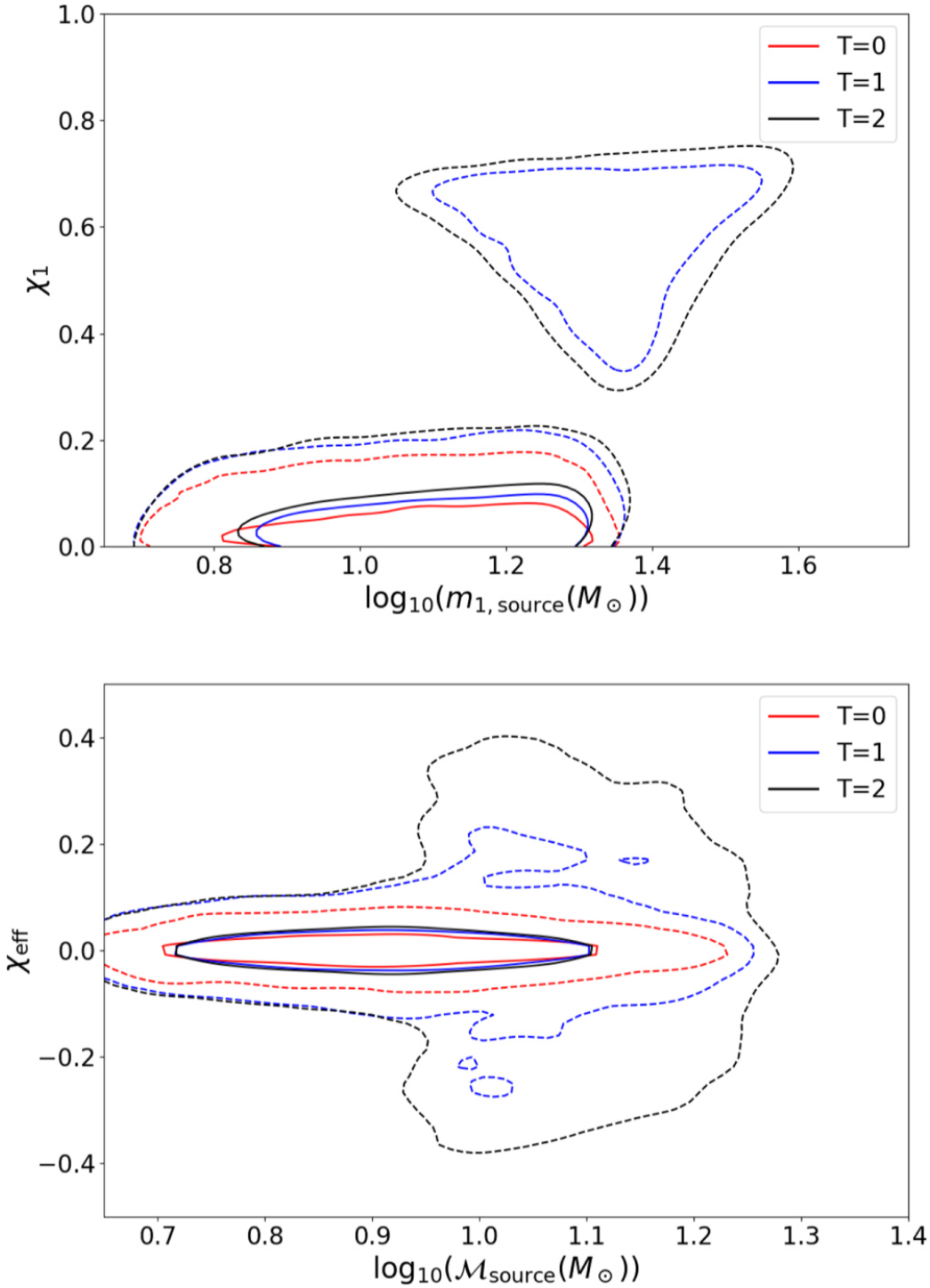

Figure 27. Joint mass-spin distributions for three successive time steps. Solid and dashed contours represent 90% and 99.9% confidence intervals. Source: Doctor et al. (2020).

We compare the BBH population characteristics of different formation channels in figure 28 and figure 29 by plotting the same number of events for each channel, leaving the channel-wise sub-populations unweighted by merger rates. The three channels are observed to dominate different parts of the mass space, with field mergers at lower masses, AGN disk mergers at higher masses, and globular cluster mergers nestled between the two. AGN disk mergers also exhibit the characteristic multimodality of hierarchical mergers. They're also extremely efficient at producing high effective spin mergers, compared to the other two channels. The $\mathcal{M} - \chi_{\rm eff}$ correlation imprinted by hierarchical mergers is a signature that we'll leverage to constrain our parameters with the catalog. The kick velocities are mostly higher than typical globular cluster escape velocities of $\sim$2$-$200 km/s, but lower than typical AGN disk escape velocities of $\sim 10^3 - 10^5$ km/s. This is in line with our expectations of hierarchical mergers being much more abundant in AGN disks than in globular clusters.

4.4 Inference

The code evaluates the inhomogeneous Poisson likelihood, that is the likelihood of observing N events with observed data $\{d_n\}$ given population parameters Λ

$$p(\{d_n\}|\Lambda) \propto e^{-\mu(\Lambda,\mathcal{R})} \prod_{n=1}^{N} \int p(d_n|\lambda)\rho(\lambda|\Lambda)d\lambda \tag{4.21}$$

where $p(d_n|\lambda)$ is the likelihood of data d_n given binary parameters λ (such as $\{\mathcal{M}, \eta, \chi_{\rm eff}\}$), $p(\lambda|\Lambda)$ is the probability of binary parameters given model parameters Λ.

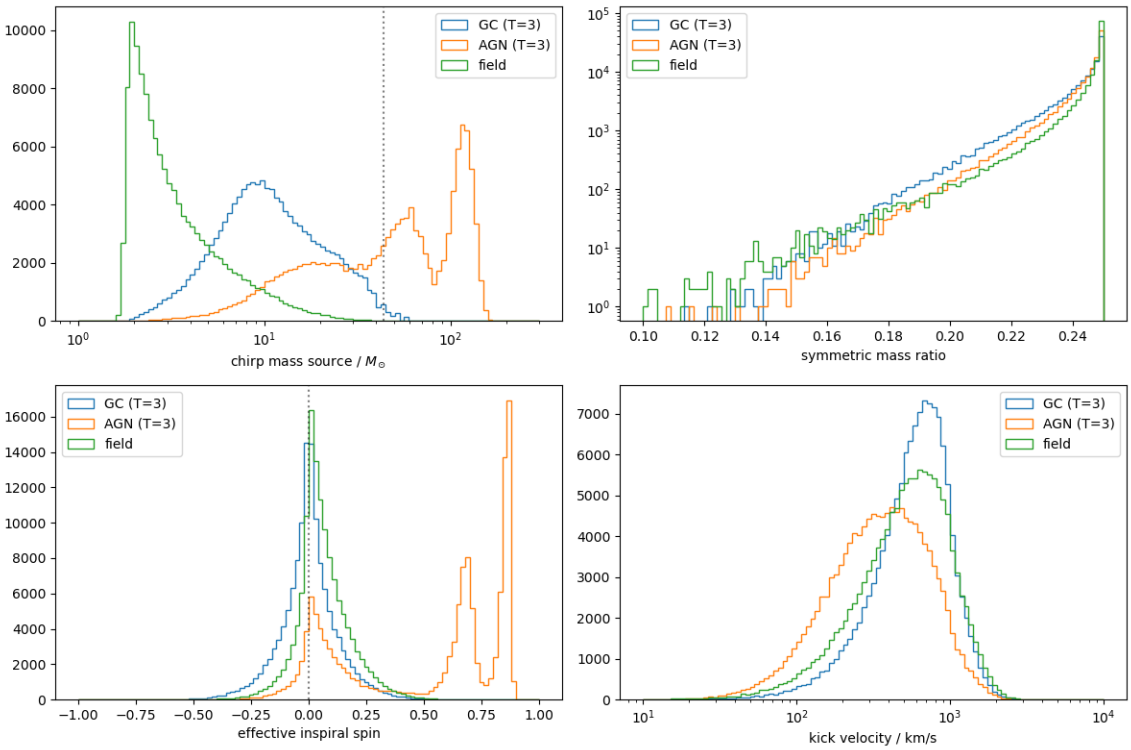

Figure 28. Comparison of parameter distributions for different formation channels with fiducial values.

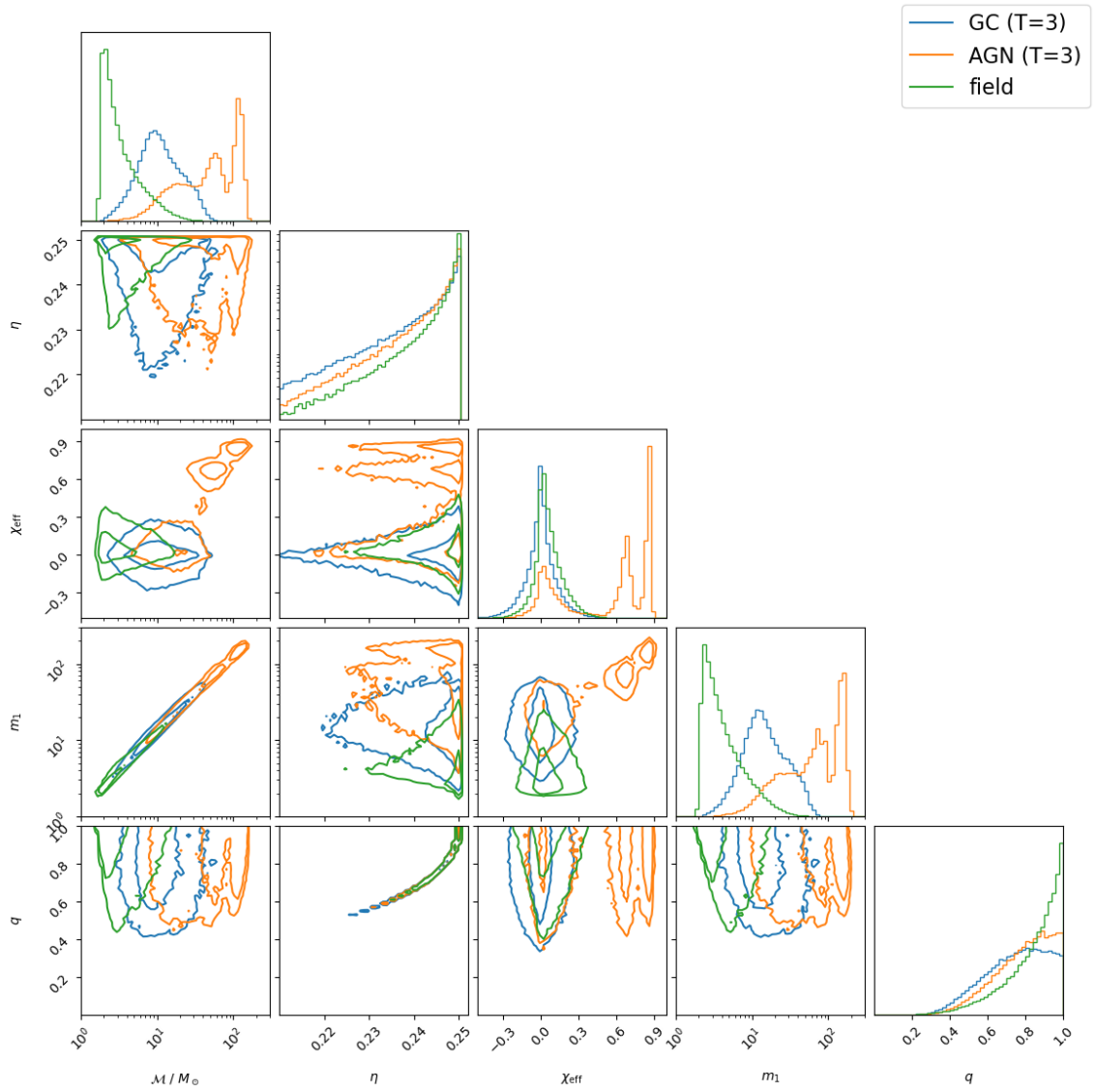

Figure 29. Comparison of parameter distributions for different formation channels with fiducial values.

$\mu(\Lambda, \mathcal{R})$ is the expected number of detections, also called the Poisson mean, calculated as

$$\begin{aligned}\mu(\Lambda, \mathcal{R}) &= \int \mathcal{V}(\lambda)\rho(\lambda|\Lambda)d\lambda \\ &= \int \mathcal{V}(\lambda)\mathcal{R}p(\lambda|\Lambda)d\lambda \\ &\approx \frac{\mathcal{R}}{S}\sum_{i=1}^{S}\mathcal{V}(\lambda_i) = \mathcal{R}\left\langle \mathcal{V}(\lambda_i)\right\rangle_i\end{aligned}$$

where $\mathcal{V}$ is the sensitive volume, $\mathcal{R}$ is the overall merger rate, and $\{\lambda_i\}$ are S draws from the simulated population.

The eventwise detection likelihood integral is approximated as

$$\begin{aligned}\int p(d_n|\lambda)\rho(\lambda|\Lambda)d\lambda &= \int p(d_n|\lambda)\mathcal{R}p(\lambda|\Lambda)d\lambda \\ &\approx \frac{\mathcal{R}}{S}\sum_{i=1}^{S}p(d_n|\lambda_i) = \mathcal{R}\left\langle p(d_n|\lambda_i)\right\rangle_i\end{aligned}$$

This casts the evaluated likelihood in the computable form

$$\log p(\{d_n\}|\Lambda) = -\mathcal{R}\left\langle \mathcal{V}(\lambda_i)\right\rangle_i + N\log\mathcal{R} + \sum_{n=1}^{N}\log\left\langle p(d_n|\lambda_i)\right\rangle_i \tag{4.22}$$

We can build several models by choosing different priors for inferred parameters or different prescriptions for constant parameters, with the parameters describing the natal population $(\alpha, m_{\min}, m_{\max}, \mathbb{E}[\chi], \mathrm{Var}[\chi], Z_{\mathrm{metal}})$, coagulation (a, b, T, w), and depletion $(\mu_M, \mu_{r_0}, \sigma_M, \sigma_{r_0})$. We can also build population models that are a mixture of our dynamical merger models and LIGO-Virgo-KAGRA's field merger models. To compute the likelihood for such mixture models, we modify our likelihood equation to

$$\log p(\{d_n\}|\Lambda) = -\sum_{c}^{k}\mathcal{R}_c\left\langle \mathcal{V}(\lambda_i)\right\rangle_i + \sum_{n=1}^{N}\log\sum_{c}^{k}\mathcal{R}_c\left\langle p(d_n|\lambda_{c,i})\right\rangle_i \tag{4.23}$$

This is equivalent to drawing the same number of events from the channel-wise sub-populations and weighing their contribution to the likelihood by their relative

merger rates, instead of drawing a different number of events from each sub-population.

We can use the summation constraint $\sum_c \mathcal{R}_c = \mathcal{R}_{\rm tot}$ to reparametrize the set of merger rates $\{\mathcal{R}_c\}$ as a set of merger fractions along with the total merger rate $\{\{f_c\}, \mathcal{R}\}$ where $f_c = \mathcal{R}_c/\mathcal{R}_{\rm tot}$. This allows us to use a single log uniform prior for the overall merger rate and a Dirichlet prior for the merger fractions, which makes sampling merger rates more efficient.

We also track the merger ancestries of black holes which helps us understand the generational composition of observed mergers, so that we can eventually calculate the probability of a particular BBH to have been the result of previous mergers.

Doctor et al. (2020) had used the population inference code PopModels (Wysocki, Lange, and O'Shaughnessy (2019)), specifically a version extended to model black hole coagulation, to compare their models to gravitational wave observations from the first gravitational wave transient catalog (GWTC-1) (B. P. Abbott et al. (2019)). Although we had initially followed this with an improved version of the same code to analyze the third gravitational wave transient catalog (GWTC-3) (The LIGO Scientific Collaboration et al. (2021a)), we later decided to build our own population inference software from scratch.

The new code is written in JAX (Bradbury et al. (2018)) and utilizes GPU acceleration and automatic differentiation to improve computing performance. This gives us more than 100x speedups compared to PopModels. The code performs Hamiltonian Monte Carlo (HMC) inference (Neal et al. (2011)), using the No U-Turn Sampler (NUTS) (Hoffman, Gelman, et al. (2014)) with adaptive path length and mass matrix adaptation (Betancourt (2017)).

For the binary parameters λ, we choose the chirp mass $\mathcal{M}$, the inverse symmetric mass ratio $1/\eta$, and the effective spin χ_{eff}, mainly because these are very well-estimated parameters, and many formation channel signatures tend to be imprinted in this space. The parameter estimation samples in these dimensions, especially η, are notoriously not Gaussians, whereas several sampling techniques that we use work best when sample distributions can be approximated as Gaussians. To resolve this, we enlist the help of normalizing flows. We use the package FlowJAX to train normalizing flows to convert Gaussians to parameter event samples. During sampling, we then use the flows' corresponding probability density functions to compute the likelihood per event. This procedure prevents a lot of the bias in posterior distributions that may otherwise appear due to bad approximations.

Once we have the posterior distributions, we can compute the posterior predictive distributions for the binary parameters as

$$p(\tilde{\lambda}|\{d_n\}) = \int p(\tilde{\lambda}|\Lambda)p(\Lambda|\{d_n\})d\Lambda \approx \left\langle p(\tilde{\lambda}|\Lambda_i)\right\rangle_i \tag{4.24}$$

For a mixture model, the overall posterior predictive distribution is a merger rate weighted sum of the channel-wise posterior predictive distributions.

$$p(\tilde{\lambda}|\{d_n\}) = \sum_c^k \frac{\mathcal{R}_c}{\mathcal{R}_{\text{tot}}} \cdot p(\tilde{\lambda}_c|\{d_n\}) \tag{4.25}$$

We can also compute the probability distribution of an observed event being produced by each of the formation channels as

$$\begin{aligned} p(d_n \in \text{Channel}_c|\{d_n\}) &= \int p(d_n \in \text{Channel}_c|\Lambda)p(\Lambda|\{d_n\})d\Lambda \\ &\approx \frac{1}{N_{\text{post}}} \sum_i^{N_{\text{post}}} p(d_n \in \text{Channel}_c|\Lambda_i) \end{aligned}$$

where

$$p(d_n \in \text{Channel}_c|\Lambda) = \frac{\mathcal{R}_c \cdot \int p(d_n|\lambda_c)p(\lambda_c|\Lambda)d\lambda_c}{\sum_c^k \mathcal{R}_c \cdot \int p(d_n|\lambda_c)p(\lambda_c|\Lambda)d\lambda_c} \approx \frac{\mathcal{R}_c \cdot \langle p(d_n|\lambda_{c,i})\rangle_i}{\sum_c^k \mathcal{R}_c \cdot \langle p(d_n|\lambda_{c,i})\rangle_i}$$

4.5 Application to GWTC-3

We used our code to perform Bayesian population inference on all the hierarchical merger models using the 69 confident BBH mergers in GWTC-3 (The LIGO Scientific Collaboration et al. (2021b)). The posterior distribution of the overall merger rate for the field+GC+AGN mixture model is shown in figure 30. The inferred merger rates of $\mathcal{R} = 13.62^{+12.77}_{-5.83}$ Gpc^{-3} yr^{-1} and $\mathcal{R} = 11.11^{+8.66}_{-5.21}$ Gpc^{-3} yr^{-1} straddle the lower end of the LIGO-Virgo-KAGRA results ($17.9 - 44$ Gpc^{-3} yr^{-1}) as reported in The LIGO Scientific Collaboration et al. (2021b).

Since the merger rate fractions of the formation channels are constrained together within a Dirichlet distribution, it's more logical to plot their posterior distributions in a triangular histogram, as shown in figure 31. We observe that field mergers constitute most of the events, followed by globular cluster mergers with most of the rest, and then AGN disk mergers with the rest.

We use triangular histograms again to show the probability of each event being produced by each of the formation channels in figure 32. Most events are categorized as field mergers, with some being split between field mergers and one other channel. Of particular note is GW190521, shown individually in figure 33, which is very confidently categorized as an AGN disk merger. This is expected as, in our models, AGN disks are the most likely source of hierarchical mergers that can best explain GW190521's large component masses.

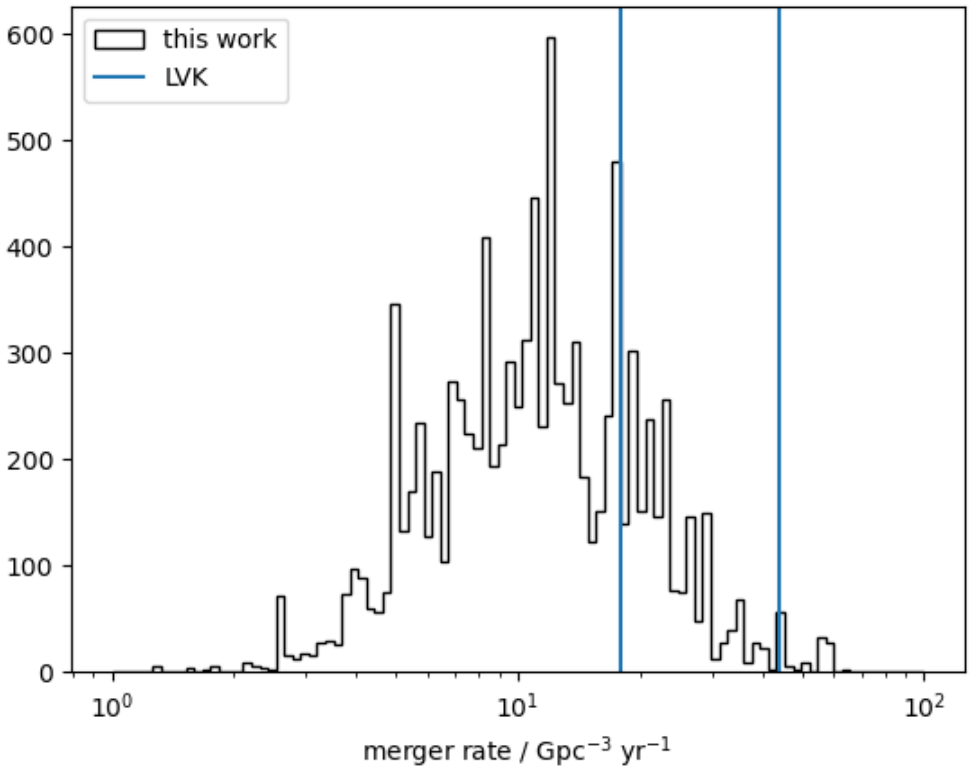

Figure 30. Black: Posterior distribution of the overall merger rate. Run on hierarchical BBH field+GC+AGN mixture model on GWTC-3. Blue: Range of LIGO-Virgo-KAGRA analyses' results.

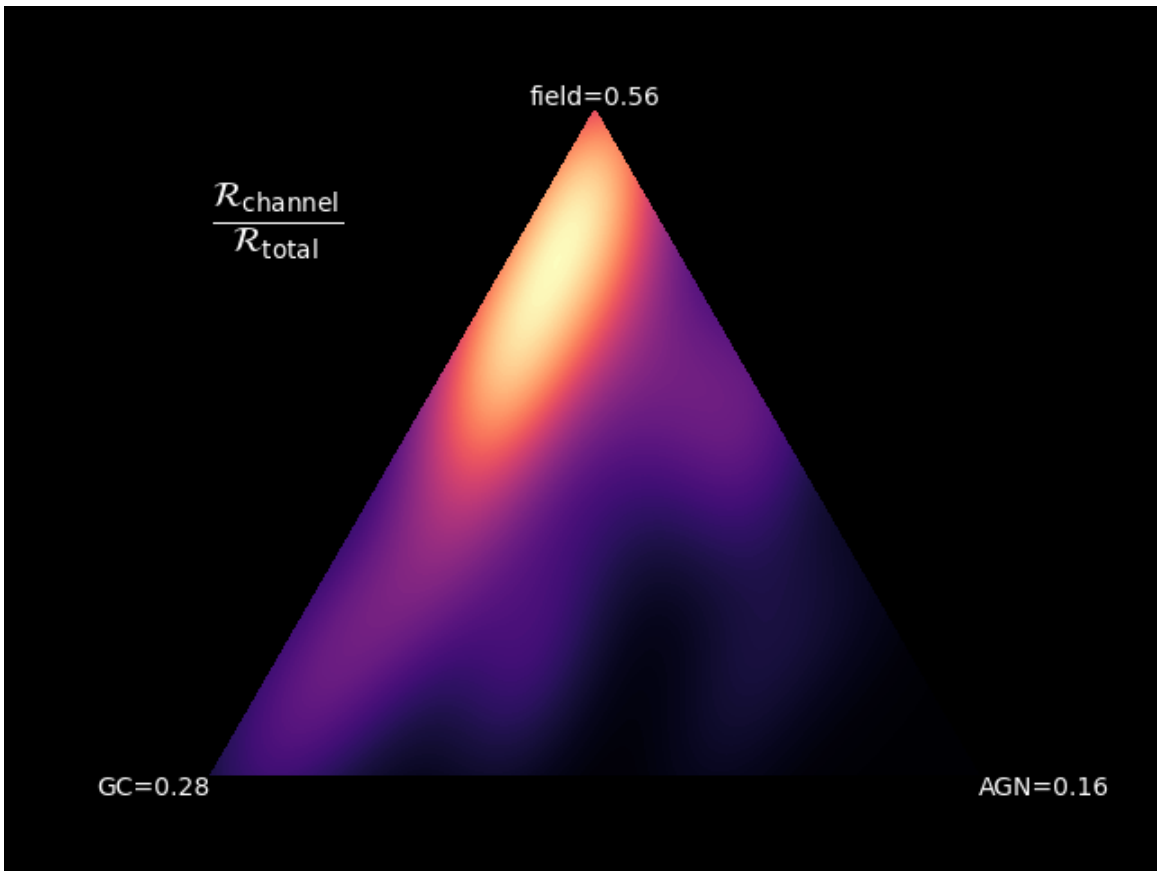

Figure 31. Posterior distribution of overall formation channel probabilities. Run on hierarchical BBH field+GC+AGN mixture model on GWTC-3. Brighter colors represent greater probabilities.

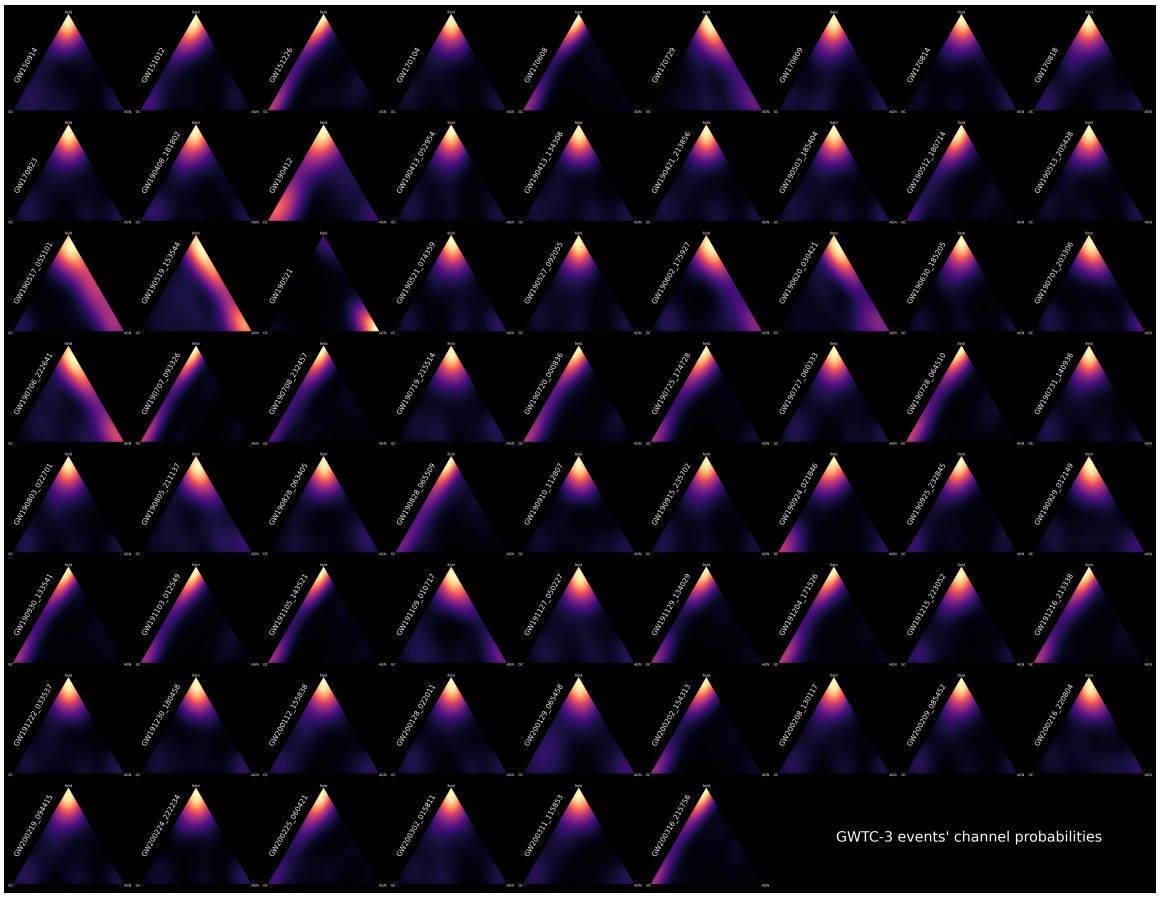

Figure 32. Probability of each event being produced by each formation channel. Run on hierarchical BBH field+GC+AGN mixture model on GWTC-3. Brighter colors represent greater probabilities. For every triangle, the top, left and right corners respresent field, GC, and AGN channels, respectively.

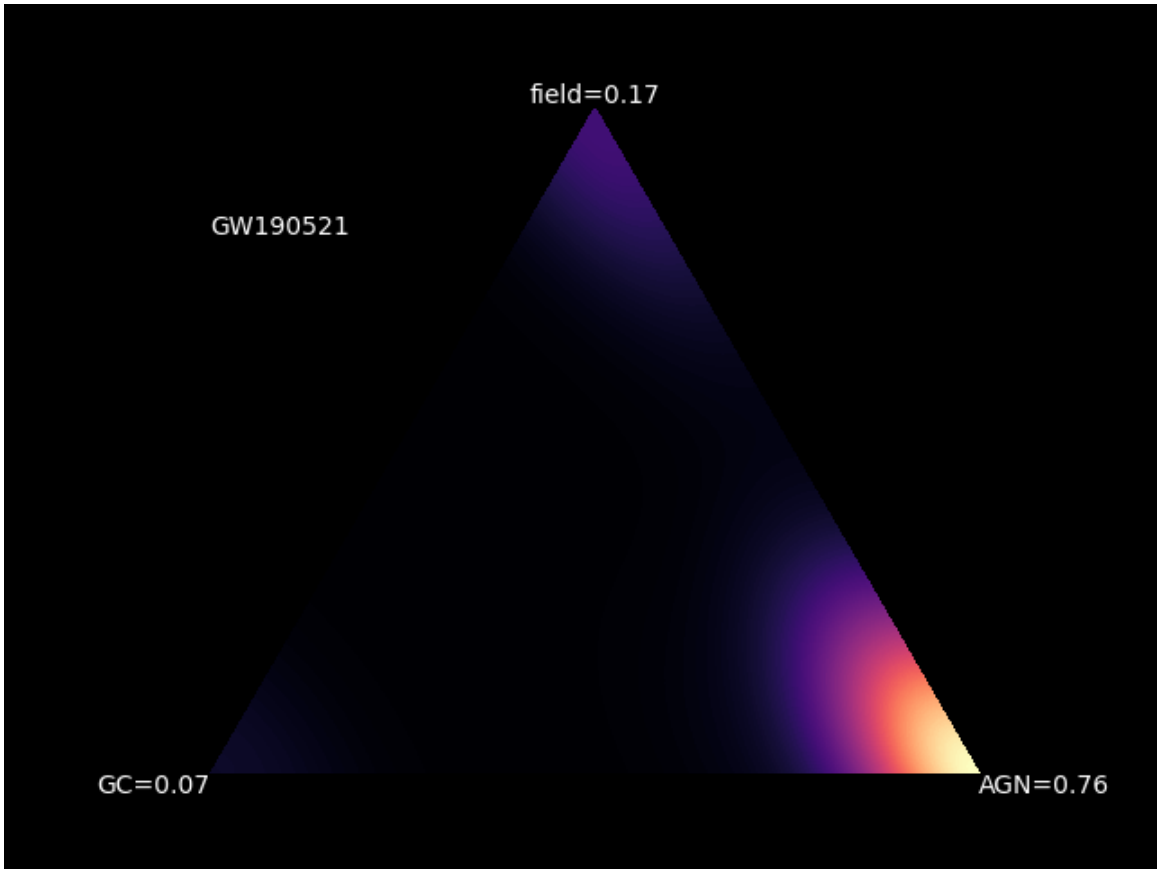

Figure 33. Probability of GW190521 being produced by each formation channel. Run on hierarchical BBH field+GC+AGN mixture model on GWTC-3. Brighter colors represent greater probabilities.

In figure 34, we compare posterior predictive distributions of several binary parameters for each formation channel, as well as for the full mixture model. We observe that field mergers dominate towards the lower masses, whereas AGN disk mergers constitute almost the entirety of the heaviest mergers, especially over the PISN mass gap. Dynamical channels seem more efficient at producing asymmetric mergers than the field channel. The effective spin distribution for globular clusters mergers is symmetric, as expected from an isotropic environment. While the field channel is only slightly biased towards positive χ_{eff}, the AGN disk channel is extremely biased towards positive χ_{eff}. The net result is the overall mixture model being biased towards positive χ_{eff} as well. Thus, the AGN disk model is very efficient at producing a subpopulation of high-mass high-spin mergers which is a strong signature for hierarchical mergers.

We compare the posterior predictive distributions for first generation mergers and higher generation mergers in figure 35 to explicitly showcase the effects of hierarchical mergers more. Evidently, hierarchical mergers are able to populate the PISN mass gap with heavy black holes. They're also efficient at producing high effective inspiral spins, and low mass ratio mergers. Interestingly, this is in line with the anti-correlation between mass ratios and effective spins observed in other analyses (Adamcewicz, Lasky, and Thrane (2023); Callister, Haster, Ng, Vitale, and Farr (2021); McKernan et al. (2022)). We can also calculate the probability of each observed event being a hierarchical merger, as shown in figure 36. While most events have very low probabilities, GW190521 stands out at 99.94%, reinforcing our belief that it is very likely a hierarchical merger.

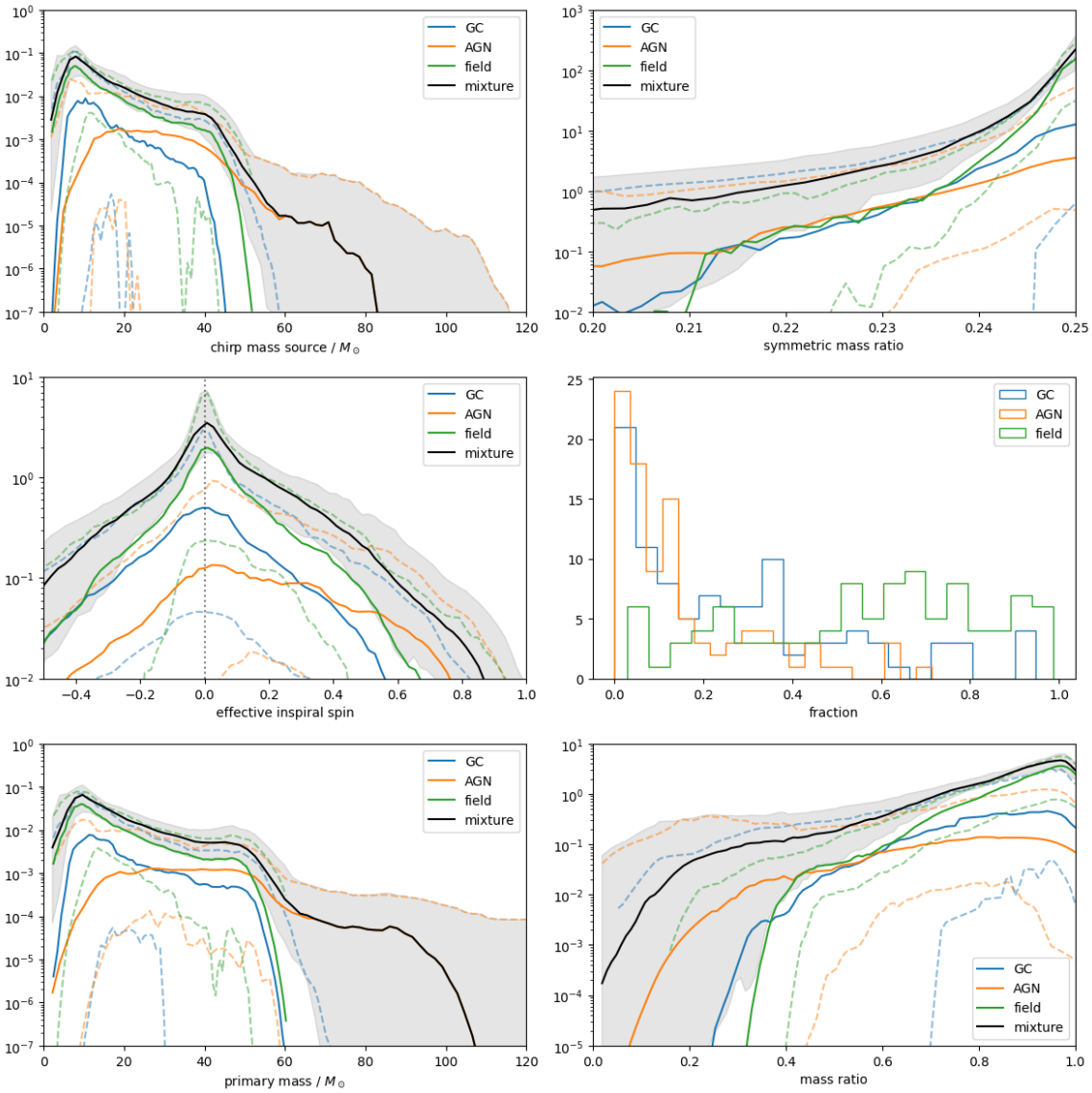

Figure 34. Posterior predictive distributions of hierarchical BBH field+GC+AGN mixture model inferred from GWTC-3, comparing formation channels.

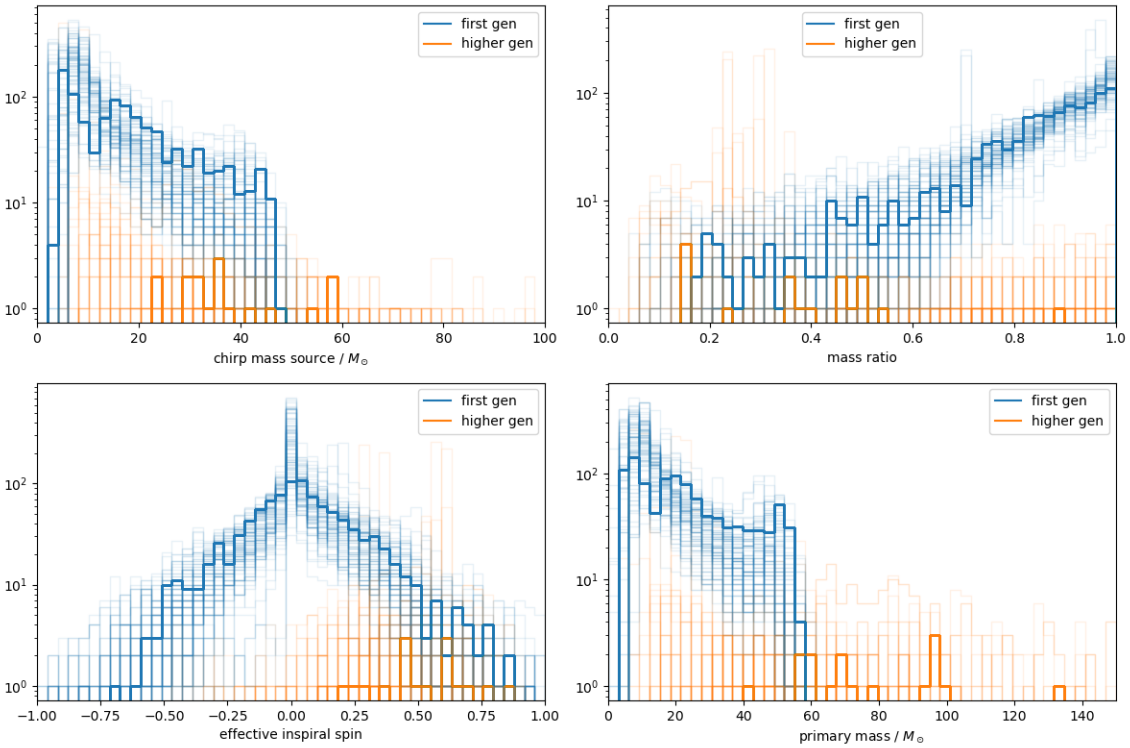

Figure 35. Posterior predictive distributions of hierarchical BBH field+GC+AGN mixture model inferred from GWTC-3, comparing merger generations.

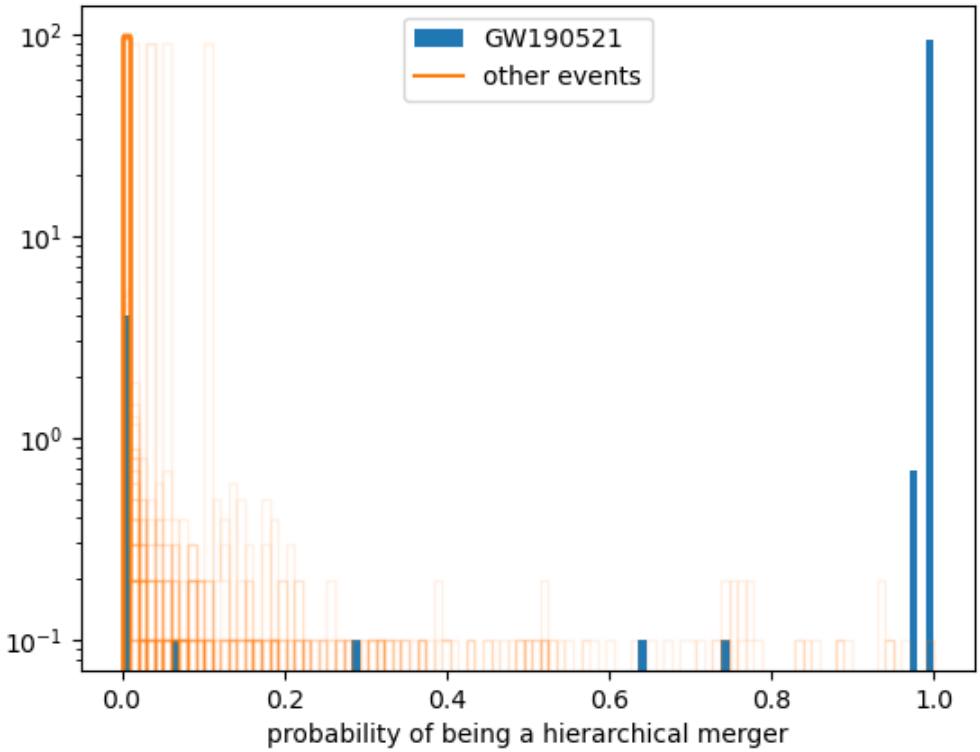

Figure 36. Probability of each event being a hierarchical merger. Run on hierarchical BBH field+GC+AGN mixture model on GWTC-3.

4.6 Discussion

In this work, we refined the coagulation model for hierarchical mergers introduced in Doctor et al. (2020), and developed new software to perform faster and more flexible population inference. Constraining our models on GWTC-3, we find that field mergers constitute the majority of events, followed by globular cluster mergers, and then AGN disk mergers. The AGN disk channel is the most efficient at producing high-mass high-spin mergers due to its ability to retain merger remnants that can go on to participate in hierarchical mergers. We observed the tendency of hierarchical mergers to form a high-mass high-spin sub-population of events that is also less preferentially symmetric than first-generation mergers. While most events are consistent with first-generation field or globular cluster mergers, we find that GW190521 is very likely a hierarchical merger from an AGN disk.

We plan to incorporate further extensions like modeling hierarchical neutron star mergers where the equation of state would have to be parametrized. A more immediate improvement for the environment models we're working on is addressing their redshift dependence. As astrophysical models of dynamical environments undergo improvement through various simulational and observational studies, we'll incorporate them in our models, and our inferences will be naturally evolving with time. Thus, gravitational waves can walk step-in-step with other messengers, refining each others' models with ever increasing precision.

CHAPTER V

LOOKING AHEAD

The fourth observing run (O4) is currently underway and has already detected more candidate gravitational-wave events than the previous three runs combined. Of particular note is GW231123 (The LIGO Scientific Collaboration, the Virgo Collaboration, and the KAGRA Collaboration (2025)), a loud binary black hole merger with total mass in the range $190-265\ M_\odot$. The two component black holes are estimated to have masses $\sim 140\ M_\odot$ and $\sim 100\ M_\odot$, and spins 0.9 and 0.8. That makes it the most massive black hole merger confidently identified so far, and an outlier to the third gravitational-wave transient catalog. While the final black hole is an intermediate-mass black hole well above the PISN mass gap, the heavier black hole straddles the upper boundary of the gap and the lighter lies well within the gap. The massive and rapidly spinning components make it difficult to form in isolated evolution scenarios, lending support to hierarchical origins. Staggeringly, the two heaviest grandparents (the heavier parent of each parent) are themselves probably in the mass gap.

The detection of GW231123 provides an abundance of use-cases for the work in this dissertation. Some non-stationary noise was observed in the hours around the signal, and there was a nearby glitch, both in LIGO-Hanford. While there's no evidence of any impact on the analyses, the disagreement within the different waveform models calls for some attention to the PSD modeling and looking for correlations with the inferred CBC posteriors, as explored in chapter II. The event being an outlier poses interesting questions for the clustering algorithms in chapter III and the hierarchical merger models in chapter IV, as well. While no confident comment can be made without a thorough statistical analysis including all the

gravitational-wave events observed in O4, one can imagine the event GW231123 driving the models in an interesting direction. It will likely lend even more credence to the idea of hierarchical mergers in AGN disks producing massive rapidly spinning black holes populating the PISN mass gap.

Modularity is a feature we have aspired to build into all the algorithms mentioned in our work. Each of the codes can accept a wide variety of data as input, making it easily usable in fields beyond gravitational-wave astronomy. The parametric PSD estimation algorithm in chapter II can accept any time series and generate a distribution of spectra, which is useful to analyze any data that contains a combination of broadband and narrowband noise. The clustering algorithms in chapter III can deal with any set of sample distributions, making them good exploratory tools for any data where the intra-distribution spread is comparable to the inter-distribution spread. The coagulation population inference algorithm in IV can fit any new compact object formation channel with a parametrized model and infer the model parameters as well as branching ratios in a mixture model. Thus these tools we have developed can prove to be very strong and versatile hammers even beyond the particular gravitational-wave-related nails we have focused on here.

With sensitivity upgrades, the gravitational-wave census is growing faster and richer every observing run. And with the advent of multi-messenger astronomy, gravitational-wave science is poised to become an inseparable component of astrophysics. Therefore it's necessary to keep developing tools to understand not only the gravitational-wave data we have but also the data we *will* have.

REFERENCES CITED

Aasi, J., Abbott, B., Abbott, R., Abbott, T., Abernathy, M., Ackley, K., ... others (2015). Advanced ligo. *Classical and quantum gravity*, *32*(7), 074001.

Abac, A., Abbott, R., Abouelfettouh, I., Acernese, F., Creighton, T., Diaz, M. C., ... Wang, W. (2024). Observation of gravitational waves from the coalescence of a 2.5–4.5 solar mass compact object and a neutron star. *The Astrophysical journal letters*.

Abbott, B. P., Abbott, R., Abbott, T., Abernathy, M., Acernese, F., Ackley, K., ... others (2016). Gw150914: The advanced ligo detectors in the era of first discoveries. *Physical review letters*, *116*(13), 131103.

Abbott, B. P., Abbott, R., Abbott, T., Abraham, S., Acernese, F., Ackley, K., ... others (2019). 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*(3), 031040.

Abbott, B. P., Abbott, R., Abbott, T., Acernese, F., Ackley, K., Adams, C., ... others (2017). Gw170817: observation of gravitational waves from a binary neutron star inspiral. *Physical review letters*, *119*(16), 161101.

Abbott, B. P., Abbott, R., Abbott, T. D., Abraham, S., Acernese, F., Ackley, K., ... others (2020). A guide to ligo–virgo detector noise and extraction of transient gravitational-wave signals. *Classical and Quantum Gravity*, *37*(5), 055002.

Abbott, R., Abbott, T., Abraham, S., Acernese, F., Ackley, K., Adams, A., ... others (2021). 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.

Abbott, R., Abbott, T., Abraham, S., Acernese, F., Ackley, K., Adams, C., ... others (2020a). Gw190521: a binary black hole merger with a total mass of 150 solar mass. *Physical review letters*, *125*(10), 101102.

Abbott, R., Abbott, T., Abraham, S., Acernese, F., Ackley, K., Adams, C., ... others (2020b). Gw190814: Gravitational waves from the coalescence of a 23 solar mass black hole with a 2.6 solar mass compact object. *The Astrophysical Journal Letters*, *896*(2), L44.

Abbott, R., Abbott, T., Abraham, S., Acernese, F., Ackley, K., Adams, C., ... others (2020c). Properties and astrophysical implications of the 150 solar mass binary black hole merger gw190521. *The Astrophysical Journal Letters*, *900*(1), L13.

Abbott, R., Abbott, T., Acernese, F., Ackley, K., Adams, C., Adhikari, N., ... others (2024). Gwtc-2.1: Deep extended catalog of compact binary coalescences observed by ligo and virgo during the first half of the third observing run. *Physical Review D*, *109*(2), 022001.

Acernese, F., Agathos, M., Agatsuma, K., Aisa, D., Allemandou, N., Allocca, A., ... others (2014). Advanced virgo: a second-generation interferometric gravitational wave detector. *Classical and Quantum Gravity*, *32*(2), 024001.

Adamcewicz, C., Lasky, P. D., & Thrane, E. (2023). Evidence for a correlation between binary black hole mass ratio and black hole spins. *The Astrophysical Journal*, *958*(1), 13.

Ade, P. A., Aghanim, N., Arnaud, M., Ashdown, M., Aumont, J., Baccigalupi, C., ... others (2016). Planck 2015 results-xiii. cosmological parameters. *Astronomy & Astrophysics*, *594*, A13.

Akima, H. (1970, October). A new method of interpolation and smooth curve fitting based on local procedures. *J. ACM*, *17*(4), 589–602. Retrieved from `https://doi.org/10.1145/321607.321609` doi: 10.1145/321607.321609

Akutsu, T., Ando, M., Arai, K., Arai, Y., Araki, S., Araya, A., ... others (2021). Overview of kagra: Detector design and construction history. *Progress of Theoretical and Experimental Physics*, *2021*(5), 05A101.

Antonini, F., & Gieles, M. (2020, February). Population synthesis of black hole binary mergers from star clusters. *MNRAS*, *492*(2), 2936-2954. doi: 10.1093/mnras/stz3584

Antonini, F., Gieles, M., & Gualandris, A. (2019). Black hole growth through hierarchical black hole mergers in dense star clusters: implications for gravitational wave detections. *Monthly Notices of the Royal Astronomical Society*, *486*(4), 5008–5021.

Antonini, F., Romero-Shaw, I. M., & Callister, T. (2025). Star cluster population of high mass black hole mergers in gravitational wave data. *Physical Review Letters*, *134*(1), 011401.

Ashton, G., Hübner, M., Lasky, P. D., Talbot, C., Ackley, K., Biscoveanu, S., ... Thrane, E. (2019, April). BILBY: A User-friendly Bayesian Inference Library for Gravitational-wave Astronomy. *ApJS*, *241*(2), 27. doi: 10.3847/1538-4365/ab06fc

Ashton, G., & Talbot, C. (2021, October). BILBY-MCMC: an MCMC sampler for gravitational-wave inference. *MNRAS*, *507*(2), 2037-2051. doi: 10.1093/mnras/stab2236

Barkat, Z., Rakavy, G., & Sack, N. (1967). Dynamics of supernova explosion resulting from pair formation. *Physical Review Letters*, *18*(10), 379.

Bartos, I., Kocsis, B., Haiman, Z., & Márka, S. (2017). Rapid and bright stellar-mass binary black hole mergers in active galactic nuclei. *The Astrophysical Journal*, *835*(2), 165.

Betancourt, M. (2017). A conceptual introduction to hamiltonian monte carlo. *arXiv preprint arXiv:1701.02434*.

Bishop, C. M. (1995). Neural networks for pattern recognition..

Bond, J., Arnett, W., & Carr, B. J. (1984). The evolution and fate of very massive objects. *Astrophysical Journal, Part 1 (ISSN 0004-637X), vol. 280, May 15, 1984, p. 825-847. Research supported by the Science and Engineering Research Council.*, *280*, 825–847.

Bradbury, J., Frostig, R., Hawkins, P., Johnson, M. J., Leary, C., Maclaurin, D., ... Zhang, Q. (2018). *JAX: composable transformations of Python+NumPy programs.* Retrieved from `http://github.com/jax-ml/jax`

Bustillo, J. C., Sanchis-Gual, N., Torres-Forné, A., Font, J. A., Vajpeyi, A., Smith, R., ... Leong, S. H. (2021). Gw190521 as a merger of proca stars: A potential new vector boson of 8.7× 10-13 ev. *Physical Review Letters*, *126*(8), 081101.

Callister, T. A., & Farr, W. M. (2024). Parameter-free tour of the binary black hole population. *Physical Review X*, *14*(2), 021005.

Callister, T. A., Haster, C.-J., Ng, K. K., Vitale, S., & Farr, W. M. (2021). Who ordered that? unequal-mass binary black hole mergers have larger effective spins. *The Astrophysical Journal Letters*, *922*(1), L5.

Clesse, S., & Garcia-Bellido, J. (2020). Gw190425 and gw190814: Two candidate mergers of primordial black holes from the qcd epoch. *arXiv preprint arXiv:2007.06481*.

Commons, W. (2020). *File:ligo simplified.svg — wikimedia commons, the free media repository.* Retrieved from `https://commons.wikimedia.org/w/index.php?title=File:LIGO_simplified.svg&oldid=490326135` ([Online; accessed 9-July-2025])

Cornish, N. J., & Littenberg, T. B. (2015, jun). Bayeswave: Bayesian inference for gravitational wave bursts and instrument glitches. *Classical and Quantum Gravity*, *32*(13), 135012. Retrieved from https://doi.org/10.1088%2F0264-9381%2F32%2F13%2F135012 doi: 10.1088/0264-9381/32/13/135012

de Boor, C. (1978). A practical guide to splines. In *Applied mathematical sciences.*

De Mink, S., & Mandel, I. (2016). The chemically homogeneous evolutionary channel for binary black hole mergers: rates and properties of gravitational-wave events detectable by advanced ligo. *Monthly Notices of the Royal Astronomical Society*, *460*(4), 3545–3553.

Doctor, Z., Wysocki, D., O'Shaughnessy, R., Holz, D. E., & Farr, B. (2020, apr). Black hole coagulation: Modeling hierarchical mergers in black hole populations. *The Astrophysical Journal*, *893*(1), 35. Retrieved from https://doi.org/10.3847%2F1538-4357%2Fab7fac doi: 10.3847/1538-4357/ab7fac

Duda, R. O., & Hart, P. E. (1973). Pattern classification and scene analysis. In *A wiley-interscience publication.*

Edelman, B., Farr, B., & Doctor, Z. (2023). Cover your basis: Comprehensive data-driven characterization of the binary black hole population. *The Astrophysical Journal*, *946*(1), 16.

Farmer, R., Renzo, M., De Mink, S., Marchant, P., & Justham, S. (2019). Mind the gap: the location of the lower edge of the pair-instability supernova black hole mass gap. *The Astrophysical Journal*, *887*(1), 53.

Finn, L. S., & Mukherjee, S. (2001, Feb). Data conditioning for gravitational wave detectors: A kalman filter for regressing suspension violin modes. *Phys. Rev. D*, *63*, 062004. Retrieved from https://link.aps.org/doi/10.1103/PhysRevD.63.062004 doi: 10.1103/PhysRevD.63.062004

Fishbach, M., Holz, D. E., & Farr, B. (2017, may). Are LIGO's black holes made from smaller black holes? *The Astrophysical Journal*, *840*(2), L24. Retrieved from https://doi.org/10.3847/2041-8213/aa7045 doi: 10.3847/2041-8213/aa7045

Fishbach, M., Holz, D. E., & Farr, W. M. (2018). Does the black hole merger rate evolve with redshift? *The Astrophysical Journal Letters*, *863*(2), L41.

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. (2013). emcee: The mcmc hammer. *PASP*, *125*, 306-312. doi: 10.1086/670067

Fowler, W. A., & Hoyle, F. (1964). Neutrino processes and pair formation in massive stars and supernovae. *Astrophysical Journal Supplement, vol. 9, p. 201 (1964)*, *9*, 201.

Fraley, G. S. (1968). Supernovae explosions induced by pair-production instability. *Astrophysics and Space Science*, *2*(1), 96–114.

Fryer, C. L., Belczynski, K., Wiktorowicz, G., Dominik, M., Kalogera, V., & Holz, D. E. (2012, mar). Compact remnant mass function: Dependence on the explosion mechanism and metallicity. *The Astrophysical Journal*, *749*(1), 91. Retrieved from `https://doi.org/10.1088/0004-637x/749/1/91` doi: 10.1088/0004-637x/749/1/91

Fryer, C. L., Olejak, A., & Belczynski, K. (2022). The effect of supernova convection on neutron star and black hole masses. *The Astrophysical Journal*, *931*(2), 94.

Gerosa, D., & Berti, E. (2017, Jun). Are merging black holes born from stellar collapse or previous mergers? *Phys. Rev. D*, *95*, 124046. Retrieved from `https://link.aps.org/doi/10.1103/PhysRevD.95.124046` doi: 10.1103/PhysRevD.95.124046

Gerosa, D., & Fishbach, M. (2021). Hierarchical mergers of stellar-mass black holes and their gravitational-wave signatures. *Nature Astronomy*, *5*(8), 749–760.

Golomb, J., Isi, M., & Farr, W. M. (2024). Physical models for the astrophysical population of black holes: Application to the bump in the mass distribution of gravitational-wave sources. *The Astrophysical Journal*, *976*(1), 121.

Grudić, M. Y., Hafen, Z., Rodriguez, C. L., Guszejnov, D., Lamberts, A., Wetzel, A., ... Faucher-Giguère, C.-A. (2022, December). Great balls of fire – i. the formation of star clusters across cosmic time in a milky way-mass galaxy. *Monthly Notices of the Royal Astronomical Society*, *519*(1), 1366–1380. Retrieved from `http://dx.doi.org/10.1093/mnras/stac3573` doi: 10.1093/mnras/stac3573

Heller, K. A., & Ghahramani, Z. (2005). Bayesian hierarchical clustering. *Proceedings of the 22nd international conference on Machine learning*.

Hoffman, M. D., Gelman, A., et al. (2014). The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *J. Mach. Learn. Res.*, *15*(1), 1593–1623.

Hulse, R. A., & Taylor, J. H. (1975, January). Discovery of a pulsar in a binary system. *The Astrophysical Journal Letters*, *195*, L51-L53. doi: 10.1086/181708

Hussain, A., Isi, M., & Zimmerman, A. (2024). Hints of spin-magnitude correlations and a rapidly spinning subpopulation of binary black holes. *arXiv preprint arXiv:2411.02252*.

Inayoshi, K., Hirai, R., Kinugawa, T., & Hotokezaka, K. (2017). Formation pathway of population iii coalescing binary black holes through stable mass transfer. *Monthly Notices of the Royal Astronomical Society*, *468*(4), 5020–5032.

Kass, R. E., & Wasserman, L. A. (1995). A reference bayesian test for nested hypotheses and its relationship to the schwarz criterion. *Journal of the American Statistical Association*, *90*, 928-934.

Kimball, C., Talbot, C., Berry, C. P., Zevin, M., Thrane, E., Kalogera, V., ... others (2021). Evidence for hierarchical black hole mergers in the second ligo–virgo gravitational wave catalog. *The Astrophysical Journal Letters*, *915*(2), L35.

Kinugawa, T., Inayoshi, K., Hotokezaka, K., Nakauchi, D., & Nakamura, T. (2014). Possible indirect confirmation of the existence of pop iii massive stars by gravitational wave. *Monthly Notices of The Royal Astronomical Society*, *442*(4), 2963–2992.

Koposov, S., Speagle, J., Barbary, K., Ashton, G., Bennett, E., Buchner, J., ... Goldstein, D. (2024, June). *joshspeagle/dynesty: v2.1.4.* Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.12537467` doi: 10.5281/zenodo.12537467

Li, Y.-J., Wang, Y.-Z., Tang, S.-P., & Fan, Y.-Z. (2024). Resolving the stellar-collapse and hierarchical-merger origins of the coalescing black holes. *Physical Review Letters*, *133*(5), 051401.

Littenberg, T. B., & Cornish, N. J. (2015, apr). Bayesian inference for spectral estimation of gravitational wave detector noise. *Physical Review D*, *91*(8). Retrieved from `https://doi.org/10.1103%2Fphysrevd.91.084034` doi: 10.1103/physrevd.91.084034

Madau, P., & Rees, M. J. (2001). Massive black holes as population iii remnants. *The Astrophysical Journal*, *551*(1), L27.

Mandel, I., & De Mink, S. E. (2016). Merging binary black holes formed through chemically homogeneous evolution in short-period stellar binaries. *Monthly Notices of the Royal Astronomical Society*, *458*(3), 2634–2647.

Marchant, P., Langer, N., Podsiadlowski, P., Tauris, T. M., & Moriya, T. J. (2016). A new route towards merging massive black holes. *Astronomy & Astrophysics*, *588*, A50.

McKernan, B., Ford, K., Callister, T., Farr, W., O'Shaughnessy, R., Smith, R., ... Vajpeyi, A. (2022). Ligo–virgo correlations between mass ratio and effective inspiral spin: testing the active galactic nuclei channel. *Monthly Notices of the Royal Astronomical Society*, *514*(3), 3886–3893.

McKernan, B., Ford, K., Kocsis, B., Lyra, W., & Winter, L. (2014). Intermediate-mass black holes in agn discs–ii. model predictions and observational constraints. *Monthly Notices of the Royal Astronomical Society*, *441*(1), 900–909.

Menne, T. (2022). *Adaptive width kde with gaussian kernels.* Retrieved from `https://github.com/mennthor/awkde`

Moore, C. J., Cole, R. H., & Berry, C. P. (2014). Gravitational-wave sensitivity curves. *Classical and Quantum Gravity*, *32*(1), 015014.

Mould, M., Gerosa, D., & Taylor, S. R. (2022). Deep learning and bayesian inference of gravitational-wave populations: Hierarchical black-hole mergers. *Physical Review D*, *106*(10), 103013.

Neal, R. M., et al. (2011). Mcmc using hamiltonian dynamics. *Handbook of markov chain monte carlo*, *2*(11), 2.

Neijssel, C. J., Vigna-Gómez, A., Stevenson, S., Barrett, J. W., Gaebel, S. M., Broekgaarden, F. S., ... Mandel, I. (2019). The effect of the metallicity-specific star formation history on double compact object mergers. *Monthly Notices of the Royal Astronomical Society*, *490*(3), 3740–3759.

Paczynski, B. (1976). Common envelope binaries. In *Symposium-international astronomical union* (Vol. 73, pp. 75–80).

Payne, E., & Thrane, E. (2023). Model exploration in gravitational-wave astronomy with the maximum population likelihood. *Physical Review Research*, *5*(2), 023013.

Pelleg, D., & Moore, A. (2002, 01). X-means: Extending k-means with efficient estimation of the number of clusters. *Machine Learning, p.*

Pierra, G., Mastrogiovanni, S., & Perriès, S. (2024). The spin magnitude of stellar-mass black holes evolves with the mass. *Astronomy & Astrophysics*, *692*, A80.

Plummer, H. C. (1911, March). On the problem of distribution in globular star clusters. *MNRAS*, *71*, 460-470. doi: 10.1093/mnras/71.5.460

Rakavy, G., & Shaviv, G. (1967). Instabilities in highly evolved stellar models. *Astrophysical Journal, vol. 148, p. 803*, *148*, 803.

Ray, A., Hernandez, I. M., Mohite, S., Creighton, J., & Kapadia, S. (2023). Nonparametric inference of the population of compact binaries from gravitational-wave observations using binned gaussian processes. *The Astrophysical Journal*, *957*(1), 37.

Rodriguez, C. L., Hafen, Z., Grudić, M. Y., Lamberts, A., Sharma, K., Faucher-Giguère, C.-A., & Wetzel, A. (2022). *Great Balls of FIRE II: The evolution and destruction of star clusters across cosmic time in a Milky Way-mass galaxy.* arXiv. Retrieved from `https://arxiv.org/abs/2203.16547` doi: 10.48550/ARXIV.2203.16547

Rodriguez, C. L., Zevin, M., Pankow, C., Kalogera, V., & Rasio, F. A. (2016). Illuminating black hole binary formation channels with spins in advanced ligo. *The Astrophysical Journal Letters*, *832*(1), L2.

Romero-Shaw, I. M., et al. (2020, December). Bayesian inference for compact binary coalescences with BILBY: validation and application to the first LIGO-Virgo gravitational-wave transient catalogue. *MNRAS*, *499*(3), 3295-3319. doi: 10.1093/mnras/staa2850

Sadiq, J., Dent, T., & Wysocki, D. (2022, jun). Flexible and fast estimation of binary merger population distributions with an adaptive kernel density estimator. *Physical Review D*, *105*(12). Retrieved from `https://doi.org/10.1103%2Fphysrevd.105.123014` doi: 10.1103/physrevd.105.123014

Sanchis-Gual, N., Bustillo, J. C., Herdeiro, C., Radu, E., Font, J. A., Leong, S. H., & Torres-Forné, A. (2022). Impact of the wavelike nature of proca stars on their gravitational-wave emission. *Physical Review D*, *106*(12), 124011.

Sasaki, M., Suyama, T., Tanaka, T., & Yokoyama, S. (2018). Primordial black holes—perspectives in gravitational wave astronomy. *Classical and Quantum Gravity*, *35*(6), 063001.

Smoluchowski, M. V. (1916, January). Drei Vortrage uber Diffusion, Brownsche Bewegung und Koagulation von Kolloidteilchen. *Zeitschrift fur Physik*, *17*, 557-585.

Speagle, J. S. (2020, April). DYNESTY: a dynamic nested sampling package for estimating Bayesian posteriors and evidences. *MNRAS*, *493*(3), 3132-3158. doi: 10.1093/mnras/staa278

Stone, N. C., Metzger, B. D., & Haiman, Z. (2017). Assisted inspirals of stellar mass black holes embedded in agn discs: solving the 'final au problem'. *Monthly Notices of the Royal Astronomical Society*, *464*(1), 946–954.

Talbot, C., & Thrane, E. (2018). Measuring the binary black hole mass spectrum with an astrophysically motivated parameterization. *The Astrophysical Journal*, *856*(2), 173.

The LIGO Scientific Collaboration, & The Virgo Collaboration. (2016, Feb). Observation of gravitational waves from a binary black hole merger. *Phys. Rev. Lett.*, *116*, 061102. Retrieved from https://link.aps.org/doi/10.1103/PhysRevLett.116.061102 doi: 10.1103/PhysRevLett.116.061102

The LIGO Scientific Collaboration, The Virgo Collaboration, & The KAGRA Collaboration. (2021a). *GWTC-3: Compact binary coalescences observed by ligo and virgo during the second part of the third observing run.* arXiv. Retrieved from https://arxiv.org/abs/2111.03606 doi: 10.48550/ARXIV.2111.03606

The LIGO Scientific Collaboration, The Virgo Collaboration, & The KAGRA Collaboration. (2021b). *The population of merging compact binaries inferred using gravitational waves through GWTC-3.* arXiv. Retrieved from https://arxiv.org/abs/2111.03634 doi: 10.48550/ARXIV.2111.03634

The LIGO Scientific Collaboration, the Virgo Collaboration, & the KAGRA Collaboration. (2025, July). GW231123: a Binary Black Hole Merger with Total Mass 190-265 $M_{\odot}$. *arXiv e-prints*, arXiv:2507.08219. doi: 10.48550/arXiv.2507.08219

Tichy, W., & Marronetti, P. (2008, Oct). Final mass and spin of black-hole mergers. *Phys. Rev. D*, *78*, 081501. Retrieved from https://link.aps.org/doi/10.1103/PhysRevD.78.081501 doi: 10.1103/PhysRevD.78.081501

Tiwari, V. (2021, jul). VAMANA: modeling binary black hole population with minimal assumptions. *Classical and Quantum Gravity*, *38*(15), 155007. Retrieved from https://doi.org/10.1088%2F1361-6382%2Fac0b54 doi: 10.1088/1361-6382/ac0b54

Tiwari, V., & Fairhurst, S. (2021, may). The emergence of structure in the binary black hole mass distribution. *The Astrophysical Journal Letters*, *913*(2), L19. Retrieved from https://doi.org/10.3847/2041-8213/abfbe7 doi: 10.3847/2041-8213/abfbe7

Tong, D. (2025). *General Relativity.* Retrieved from https://www.damtp.cam.ac.uk/user/tong/gr.html

Tutukov, A., & Yungelson, L. (1993). The merger rate of neutron star and black hole binaries. *Monthly Notices of the Royal Astronomical Society*, *260*(3), 675–678.

Van Den Heuvel, E. (1976). Late stages of close binary systems. In *Symposium-international astronomical union* (Vol. 73, pp. 35–61).

van den Heuvel, E. P., Portegies Zwart, S., & de Mink, S. E. (2017). Forming short-period wolf–rayet x-ray binaries and double black holes through stable mass transfer. *Monthly Notices of The Royal Astronomical Society*, *471*(4), 4256–4264.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., ... SciPy 1.0 Contributors (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, *17*, 261–272. doi: 10.1038/s41592-019-0686-2

Wang, B., & Wang, X. (2007). *Bandwidth selection for weighted kernel density estimation.* arXiv. Retrieved from `https://arxiv.org/abs/0709.1616` doi: 10.48550/ARXIV.0709.1616

Wang, Y.-Z., Li, Y.-J., Vink, J. S., Fan, Y.-Z., Tang, S.-P., Qin, Y., & Wei, D.-M. (2022). Potential subpopulations and assembling tendency of the merging black holes. *The Astrophysical Journal Letters*, *941*(2), L39.

Woosley, S. (2017). Pulsational pair-instability supernovae. *The Astrophysical Journal*, *836*(2), 244.

Woosley, S., & Heger, A. (2021). The pair-instability mass gap for black holes. *The Astrophysical Journal Letters*, *912*(2), L31.

Woosley, S. E., Heger, A., & Weaver, T. A. (2002). The evolution and explosion of massive stars. *Reviews of modern physics*, *74*(4), 1015.

Wysocki, D., Lange, J., & O'Shaughnessy, R. (2019, aug). Reconstructing phenomenological distributions of compact binaries via gravitational wave observations. *Physical Review D*, *100*(4). Retrieved from `https://doi.org/10.1103%2Fphysrevd.100.043012` doi: 10.1103/physrevd.100.043012

Zlochower, Y., & Lousto, C. O. (2015, Jul). Modeling the remnant mass, spin, and recoil from unequal-mass, precessing black-hole binaries: The intermediate mass ratio regime. *Phys. Rev. D*, *92*, 024022. Retrieved from `https://link.aps.org/doi/10.1103/PhysRevD.92.024022` doi: 10.1103/PhysRevD.92.024022