ASTROPHYSICS, COSMOLOGY, AND FUNDAMENTAL INTERACTIONS WITH GRAVITATIONAL WAVES

by

Anarya Ray

A Dissertation Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in Physics

at

The University of Wisconsin-Milwaukee

August 2024

ABSTRACT

ASTROPHYSICS, COSMOLOGY, AND FUNDAMENTAL INTERACTIONS WITH GRAVITATIONAL WAVES

by

Anarya Ray

The University of Wisconsin-Milwaukee, 2024
Under the Supervision of Professor Jolien Creighton, PhD

Several unresolved mysteries remain in our understanding of the universe that can be elucidated with a growing catalog of gravitational wave (GW) and multi-messenger observations of compact binary coalescences (CBCs). Current and future catalogs, assisted by robust modeling and scalable inference frameworks, will lead to new insights into the poorly understood physics of compact binary formation, independent measurements of the cosmological expansion, and unprecedented empirical probes into the very nature of fundamental interactions. The scientific potential of GW observations from CBCs is significantly enhanced through the detection of their multi-messenger counterparts. Similarly, multi-probe explorations of compact object universal properties that combine GW and electromagnetic observations of different systems comprising similar kinds of compact objects highly complement the prospects of GW observations in facilitating major scientific breakthroughs. With a large number of new observations expected shortly by current and next-generation detectors, discoveries capable of revolutionizing modern physics are imminent. In this thesis, I will discuss my contributions to GW astronomy which primarily involve extracting new observational insights into the astrophysics of compact binary formation, the cosmic expansion history, the properties of matter at extreme densities, and tests of general relativity, from growing GW catalogs. I will also discuss my contributions to multi-messenger astronomy, particularly in the field of low-latency categorization of GW candidates to assess the possibility of them being accompanied by multi-messenger counterparts.

To Ma, Baba, and Partha sir

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

EoS	Equation of State
GRB	Gamma-Ray Burst
GR	General Relativity
GW	Gravitational Wave
LIGO	Laser Interferometer Gravitational-Wave Observatory
LISA	Laser Interferometer Space Antenna
CBC	Compact Binary Coalescence
BBH	Binary Black Hole
BNS	Binary Neutron Star
NSBH	Neutron Star Black Hole
BH	Black Hole
NS	Neutron Star
O1	the first observing run of LIGO-Virgo
O2	the second observing run of LIGO-Virgo
O3	the third observing run of LIGO-Virgo-KAGRA
O4a	the first half of the fourth observing run of LIGO-Virgo-KAGRA
O4	the fourth observing run of LIGO-Virgo-KAGRA
O5	the fifth observing run of LIGO-Virgo-KAGRA
2G	second generation GW detectors
3G	third generation GW detectors
SME	Standard Model Extension
PE	Parameter Estimation
SNR	Signal to Noise Ratio

FAR False Alarm Rate

ACKNOWLEDGMENTS

I would like to thank my Ph.D. advisor Jolien Creighton for his guidance and extensive support over the past five years. Through your mentorship, I have matured significantly, both as a physicist and as a person. You have opened my eyes to the big picture and made me into an independent researcher. You have taught me to always search for the correct solution and never be satisfied with one that is barely good enough. I find that invaluable. You have given me an ideal to strive towards during my years to come in academia. I would also like to thank Patrick Brady for his deep insights into several of my projects. I am grateful to my grad school committee members David Kaplan, Phillip Chang, and Sarah Vigeland for their valuable input regarding my projects.

I want to thank Ignacio Magaña Hernandez for being a very supportive colleague, collaborator, friend, and well-wisher. Thank you for always being there, through thick and thin, and for pushing me to become a better and more productive version of myself. I am grateful to Caitlin Rose for being a valuable friend and an extremely supportive colleague and collaborator. I am also thankful to Christian, Aritra, Pratyusava, and Akash for their friendship and unhindered support during my times of need.

As part of the LVK, I have gotten the chance to know and work with the most amazing collaborators one could ask for, who have made invaluable contributions to all of the projects that comprise this thesis. I am particularly thankful to Shaon Ghosh, Jay Tasson, Chad Hanna, Leo Tsukada, Siddharth Mohite, Becca Ewing, Divya Singh, Victoria Niu, Shio Sakon, Shasvath Kapadia, Vinaya Valsan, Daniel Wysocki, and Soichiro Morisaki.

Finally, I would like to dedicate this thesis to Ma (Kaberi Ray), Baba (Gautam Ray), and Partha sir (Dr. Parthasarathi Majumdar) who have helped me find my way every time I lost it and continue to be my primary motivation and inspiration to carry on in life. I love you.

Chapter 1

Introduction

Up to now humanity has been deaf to the universe. Suddenly we know how to listen. The Universe has spoken and we have understood.

David Blair

Gravitational Waves (GWs) are ripples in the fabric of space-time caused by gravitating systems undergoing spherically asymmetric acceleration. Their generation, propagation, and interaction with matter, are governed by Einstein's general theory of Relativity (GR), which remains to date the most accurate description of large-scale dynamics in our Universe. GR depicts gravitation as the distortion of spacetime caused by energy and matter. The movement of celestial bodies in a gravitational field is thus interpreted as following geodesics in curved spacetime. Similarly, the tidal acceleration between different points in space is quantified in GR as the relative deviation between freely falling geodesics caused by spacetime curvature. Therefore, a passing GW induces oscillating deviations between test particle geodesics, that are in principle measurable. Additionally, the tidal stretching and squeezing of objects caused by a GW passing through them perform work, indicating that GWs carry energy away from their system of origin.

The first indirect observation of GWs was made by Hulse and Taylor in 1976 when they performed timing studies on a system of binary pulsars (Hulse & Taylor, 1975). The observed decrease in the binary system's rotational period could only be explained by energy loss via GW emission. The Einstein Quadrupole formula (Einstein, 1918), which quantifies the energy loss of a gravitationally radiating system in terms of the system's time-varying Quadrupole moment, was shown to match the observed decrease in rotational period accurately (Hulse & Taylor, 1975). This discovery was awarded the Nobel Prize in 1993. However, direct observation of the oscillatory

deviations between the space-time trajectories of test particles caused by GWs passing through the Earth imposed a difficult technological challenge. Even for GWs caused by violent and highly energetic astrophysical phenomena such as compact stellar remnants crashing into each other, the fractional change in the relative separation of test particles located cosmological distances away from the source would be of the order of 10^{-21}, making them extremely hard to detect.

Precision laser interferometry has long been proposed as a viable mechanism for measuring geodesic deviations of such order (Weiss, 2022) and hence for detecting gravitational wave signals from compact binary coalescences (CBCs). The advanced laser interferometric gravitational wave observatory (aLIGO, J. Aasi et al., 2015) is an enormous Michaelson interferometer with perpendicular arms stretching four kilometers in each direction, that was designed to detect tiny changes in its arm lengths resulting from the squeezing and stretching of spacetime by a passing GW. In principle, a passing GW that causes light to travel slightly different path lengths along the X and Y arms of the interferometer can be detected by measuring the resultant changes in the interference pattern. Relying on Fabry Perot cavities to build up the laser light and increase the effective distance traveled along each arm, power-recycling mirrors to sharpen the interference fringes, and sophisticated vibration-dampening technology to suppress terrestrial sources of noise, the aLIGO detector configuration can measure changes in length as small as 10^{-19}m, comparable to the strain caused by passing GWs from typical CBCs (J. Aasi et al., 2015).

On September 14th, 2015, the twin interferometers of aLIGO at Hanford and Livingston recorded the first direct observation of a GW signal from a binary black hole merger several hundred megaparsecs away (Abbott et al., 2016a). Since then the LIGO-Virgo-KAGRA (LVK, J. Aasi et al., 2015; Acernese et al., 2014; Akutsu et al., 2021) detector network has observed around a hundred compact binary mergers including binary neutron star and neutron star black hole mergers (Abbott et al., 2023b). These observations offer a new perspective on the observable Universe, in complement to and often beyond the reach of terrestrial experiments and other

astrophysical messengers. With the LVK's ongoing and upcoming observing runs likely to unveil a higher number of new detections, GW astronomy promises revolutionary discoveries in several fields of modern physics ranging from astrophysics and cosmology to the very nature of fundamental interactions (e.g., Abbott et al., 2023d, 2021b,h, 2018a).

The discovery of multi-messenger counterparts to GW sources can greatly magnify their scientific potential (Mészáros et al., 2019). For example, the observation of electromagnetic transients from the GW source GW170817(Abbott et al., 2017c; Goldstein et al., 2017; Savchenko et al., 2017; Abbott et al., 2017b; Coulter et al., 2017; Abbott et al., 2017d), later identified to be a BNS merger in the galaxy NGC 4993, revealed a plethora of new information beyond the scope of what either messenger could have provided individually. Interpretation of the complementary information carried by photons, and GWs, originating from a single source has revealed new insights into unsolved mysteries like the origins of short gamma-ray bursts and the production of heavy elements in the Universe (Kasen et al., 2017; Abbott et al., 2017b). This has also enabled independent probes of cosmological expansion and high-precision tests of fundamental physics (Abbott et al., 2017a, 2019e; Radice et al., 2018). The prospects of more such detections further add to the potential of GW observations from CBCs to be unique and powerful probes of fundamental physics (Abbott et al., 2018b; Kiendrebeogo et al., 2023).

To extract new physics from GW observations, it is necessary to model their generation, propagation, and interaction with matter. Throughout the rest of this chapter, I will briefly introduce the fundamentals of GW generation, and propagation, the scientific potential of growing CBC catalogs, the role of GWs in multi-messenger astronomy, and the exciting future of GW science. I will then conclude this chapter with a summary of how this thesis is organized and an emphasis on the primary motivation behind this research.

1.1 GRAVITATIONAL WAVES FROM COMPACT BINARY MERGERS

As mentioned before, the propagation and generation of GWs as well as their interaction with matter are governed by Einstein's GR which provides a geometric description of the gravitational interaction. In this section, following the treatment of the subject in Creighton & Anderson (2011), I will briefly summarize the dynamics of GWs, starting from the basics of GR and ending with the Newtonian order CBC waveform as well as its effects on an interferometric detector. This summary lays the foundation upon which the new analysis methods that I have developed for my thesis rely, to implement physical inference with detected GW signals. For rigorous derivations and insights relevant to the summary presented in this section, please see Creighton & Anderson (2011).

1.1.1 The general theory of relativity

In GR, spacetime is considered to be a pseudo-Riemnanian manifold with a metric ($g_{\mu\nu}$) of Lorentzian signature. The physical effects of a non-zero gravitational field manifest in the form of space-time curvature, causing nearby geodesics to bend towards or away from each other, and is encoded in the components of the Riemann curvature tensor. At any point in space-time, it is possible to choose a coordinate system in which the metric tensor is flat. For example, weightlessness in a freely falling elevator indicates that in the reference frame of the elevator, the metric tensor is the same as that of Minkowski spacetime. However, tidal acceleration between two particles in the elevator, an inevitable consequence of a gravitational field, will be non-zero even in that reference frame. Therefore, components of the Riemann curvature tensor, which govern the tidal acceleration between nearby geodesics, cannot be made to vanish by coordinate transformation. In a freely falling cartesian coordinate system, the Riemann tensor can be expressed in terms of the second derivatives of the metric, such as:

$$R_{\alpha\beta\gamma\delta} = \frac{1}{2}\left\{\partial_\alpha\partial_\gamma g_{\beta\delta} - \partial_\beta\partial_\gamma g_{\alpha\delta} + \partial_\beta\partial_\delta g_{\alpha\gamma} - \partial_\alpha\partial_\delta g_{\beta\gamma}\right\}. \tag{1.1}$$

This expression of the Riemann tensor in freely falling cartesian co-ordinates, and the fact that the covariant derivative reduces to the normal partial derivative in the same coordinate system ($\nabla_\mu = \partial_\mu$) can be used to derive the Bianchi identity:

$$\nabla_{[\mu} R_{\alpha\beta]\gamma\delta} = 0, \tag{1.2}$$

where $A_{[\mu\nu\alpha]}$ denotes a sum over cyclic permutations of the indices $\mu\nu\alpha$. The Bianchi identity is an important property of spacetime curvature and plays a crucial role in the description of the gravitational field and its coupling with matter.

If the Riemann tensor is non-zero, the deviation between nearby geodesics will evolve when parallelly transported along them. In other words, initially parallel lines may move towards or away from each other in curved spacetime. This tidal acceleration between freely falling geodesics is governed by the equation of geodesic deviation, which takes the following form:

$$\frac{d^2\xi^\mu}{d\lambda^2} = R^\mu_{\nu\alpha\beta} u^\nu u^\alpha \xi^\beta, \tag{1.3}$$

where, $u^\mu(\lambda, s)$ represents a family of nearby geodesics with affine parameter λ, $\xi^\mu = \frac{du^\mu}{ds}|_{s=0}$ is the deviation between geodesics, and $\frac{d}{d\lambda} = u^\alpha \nabla_\alpha$ is the covariant derivative along u. While Eq. (1.3) determines the tidal acceleration between objects, the geodesic equation: $u^\mu \nabla_\mu u^\nu = 0$ governs the motion of a point particle in a gravitational field. Since both of these equations depend on the metric, given a particular geometry of spacetime, it is possible to determine how matter moves in it.

On the other hand, the distribution of matter itself determines the gravitational field. The dynamics of the gravitational field and its coupling with matter are described by the Einstein field equations which relate spacetime curvature to the energy and momentum density of matter in the

following way,

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = \frac{8\pi G}{c^4} T_{\mu\nu} \tag{1.4}$$

where, $R_{\mu\nu}$ is the Ricci curvature tensor, R the Ricci scalar, and $T_{\mu\nu}$ the stress-energy tensor that represents the distribution of matter in spacetime. Together, the field equations and the conservation equations they imply ($\nabla_\mu T^{\mu\nu} = 0$), offer a self-consistent description of the gravitational field and dynamics of matter in spacetime. This justifies the famous quote by John Wheeler: "Matter tells spacetime how to curve. Spacetime tells matter how to move".

1.1.2 Linearized gravity and the wave equation

Far away from the gravitating object, spacetime can be thought of as *nearly* flat, with its metric tensor expressible in the following way:

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \tag{1.5}$$

where $\eta_{\mu\nu}$ is the flat background metric and $h_{\mu\nu}$ is a small perturbation around it. Upto linear order in h, it can be shown that the Riemann tensor and the Bianchi identity reduce to the following forms in any cartesian coordinate system:

$$R_{\alpha\beta\gamma\delta} = \frac{1}{2} \left\{ \partial_\alpha \partial_\gamma h_{\beta\delta} - \partial_\beta \partial_\gamma h_{\alpha\delta} + \partial_\beta \partial_\delta h_{\alpha\gamma} - \partial_\alpha \partial_\delta h_{\beta\gamma} \right\}, \tag{1.6}$$

$$\partial_{[\mu} R_{\alpha\beta]\gamma\delta} = 0. \tag{1.7}$$

Contracting Eq. (1.7) with two of its indices (such as μ and α) and using the Einstein field equations, it can be shown that the Riemann tensor is transverse in all of its indices, i.e., $\partial_\delta R_{\alpha\beta\gamma}{}^{\delta} = 0$, in vaccum. Taking the divergence of the Eq. (1.7) further shows that the Riemann tensor follows

the wave equation:

$$\Box R_{\mu\nu\gamma\delta} = 0, \tag{1.8}$$

where $\Box = \eta^{\mu\nu}\partial_\mu\partial_\nu$ is the wave operator in flat spacetime. Given that the Riemann tensor encodes the physical effects of the gravitational field, Eq. (1.8) demonstrates that those effects propagate at the speed of light through vacuum.

1.1.3 Propagation of gravitational waves

The linearized metric perturbation $h_{\mu\nu}$ is a symmetric tensor in four dimensions and hence has 10 independent components, not all of which represent independent and physical degrees of freedom. In order to understand how GWs propagate and their effects on matter, it is necessary to extract the physical degrees of freedom of h from the Riemann tensor.

The general solution to Eq. (1.8) is a monochromatic plane wave which makes the Riemann tensor a function of $k_\mu x^\mu$ only, where k_μ is a null vector. For such a functional form of the Riemann tensor, the Bianchi identity reduces to $k_{[\mu}R_{\alpha\beta]\gamma\delta} = 0$. Using this identity, all components of the Riemann tensor can be expressed in terms of only $R_{0\alpha0\beta}$, in the following way:

$$R_{\alpha\beta\gamma\delta} = \frac{1}{(k^0)^2}\left(k_\alpha k_\gamma R_{0\beta0\delta} - k_\alpha k_\delta R_{0\beta0\gamma} - k_\beta k_\gamma R_{0\alpha0\delta} + k_\beta k_\delta R_{0\alpha0\gamma}\right). \tag{1.9}$$

Therefore, all the independent physical degrees of freedom of a gravitational wave are contained $R_{\beta0\delta0}$. Given the symmetries of the Riemann tensor, it will be non-zero only for spatial β and δ making R_{i0j0} the only nonzero independent components of the linearized Riemann tensor. Furthermore, using the field equations in vacuum ($R_{\alpha\beta} = 0$) it can be shown that $k^i R_{i0j0} = 0 = \delta^{ij} R_{i0j0}$. This demonstrates that for GWs traveling through vacuum, GR predicts that there are only two degrees of freedom (or polarizations), both of which are spatial, transverse, and traceless.

Utilizing the invariance of the Einstein field equations under infinitesimal coordinate transfor-

mations, it is possible to choose a reference frame in which the metric perturbation corresponding to a GW traveling through vacuum is spatial, transverse, and traceless. In such a reference frame the Einstein field equations take the following form:

$$\Box h_{ij}^{TT} = -\frac{16\pi G}{c^4} T_{ij}^{TT}, \tag{1.10}$$

where TT represents the transverse traceless gauge. For a GW propagating along an arbitrary direction $\hat{n}$, the spatial components of any tensor can be projected into the transverse traceless gauge in the following way:

$$A_{ij}^{TT} = P_i^k P_j^l A_{lm} - \frac{1}{2} P_{ij} P^{lm} A_{lm}, \tag{1.11}$$

where $P_{ij} = \delta_{ij} - \hat{n}_i \hat{n}_j$. The wave equation in Eq. (1.10) can be solved to obtain the waveforms corresponding to accelerating systems such as compact object binaries. Similarly, the tidal acceleration due to a GW can be represented in terms of the transverse traceless metric perturbation using the following expression of the linearized Riemann tensor:

$$R_{0i0j} = -\frac{1}{2} \partial_t^2 h_{ij}^{TT} \tag{1.12}$$

1.1.4 The generation of gravitational waves

Eq. (1.10) has a well-known solution in terms of the retarded Green's function, which takes the following form:

$$h_{ij}^{TT}(t, \mathbf{r}) = \frac{2\pi G}{c^4} \int d^3\mathbf{r}' \frac{T_{ij}^{TT}(t - \frac{|\mathbf{r}-\mathbf{r}'|}{c}, \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}. \tag{1.13}$$

In the spacetime of linearized gravity, the energy-momentum tensor of matter obeys the linearized conservation law:

$$\partial_\mu T^{\mu\nu} = 0. \tag{1.14}$$

Using this conservation relation, one can obtain the following identity for the spatial component of the energy-momentum tensor:

$$T^{ij} = \frac{1}{2}\partial_0^2(r^i r^j T^{00}) + \partial_k(T^{ik}r^j + T^{jk}r^j) - \frac{1}{2}\partial_k\partial_l(T^{kl}r^i r^j). \tag{1.15}$$

We intend to evaluate the integral in Eq. (1.13) for $\mathbf{r}$ in the far field region, which is determined by:

$$r >> \text{wavelength of GW} >> r'_{max}, \tag{1.16}$$

where r'_{max} is the size of the source. In this far field limit, up to first order $\frac{r'_{max}}{r}$, the last two terms of Eq. (1.15) only contribute at the surface of the volume being integrated over in Eq. (1.13). These contributions vanish given that far away from the source, $T_{ij} = 0$. Under such considerations and the assumption that the source is slowly moving ($t - \frac{|\mathbf{r}-\mathbf{r}'|}{c} \approx t - \frac{r}{c}$), we arrive at the following expression for the transverse traceless metric perturbation caused by a time-varying mass-energy distribution:

$$h_{ij}^{TT} = \frac{2G}{c^4 r}\frac{d^2 Q_{ij}^{TT}}{dt^2}, \tag{1.17}$$

$$\text{where } Q_{ij} = \int d^3\mathbf{r}' T^{00}(\mathbf{r}') r'_i r'_j \tag{1.18}$$

is the quadrupole tensor of the gravitating source, and Q_{ij}^{TT} can be obtained from Q_{ij} using the projection operators of Eq. (1.11). Eq. (1.17) is known as the Einstein quadrupole formula, which quantifies the metric perturbation corresponding to a gravitational wave in terms of the time-varying quadrupole moment of its source.

1.1.5 Energy and momentum carried by gravitational waves

As mentioned before, a passing GW will cause oscillatory tidal fields and objects moving through such a tidal field will have work done upon them. For example, a tidal field $R_{0i0j} = -(1/2)\partial_t^2 h_{ij}^{TT}$ will impart energy into a rigid body with quadrupole moment Q^{ij} at a rate $dW/dt = -(1/2)R_{0i0j}\partial_t Q^{ij}$ $= (1/4)\partial_t^2 h_{ij}^{TT}\partial_t Q^{ij}$. Since the objects experiencing the tidal field can only be at a time-like or null separation from the source, it is agnostic of their existence. This implies that gravitational waves carry energy and momentum away from their system of origin. In order to compute the energy carried by gravitational waves we have to look at $O(h^2)$ terms in the Einstein equation, which would then look like:

$$\Box \bar{h}_{ij}^{TT} = -\frac{16\pi G}{c^4}\left(T_{ij}^{TT} + T_{ij}^{GW,TT}\right) + O(h^3), \tag{1.19}$$

$$\text{where } T_{\mu\nu}^{GW} = \frac{c^4}{16\pi G}\left[R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R\right]^{(2)}, \tag{1.20}$$

and $[B]^{(2)}$ represents only $O(h^2)$ terms in B. It can be shown that the energy-momentum tensor of a gravitational wave is given by:

$$T_{\mu\nu}^{GW} = \frac{c^4}{32G}\left\langle \partial_\mu h_{ij}^{TT}\partial_\nu h_{TT}^{ij}\right\rangle, \tag{1.21}$$

where $\langle ...\rangle$ represents an average over many wavelengths which is necessary because the unaveraged energy-momentum tensor of the gravitational field can be made to vanish at any point by choosing a locally inertial frame. Using the Quadrupole formula, it is possible to express the loss

of energy via gravitational radiation, in terms of the quadrupole moments of the radiating source in the following way:

$$\frac{dE^{GW}}{dt} = -\frac{G}{5c^5}\left\langle \frac{d^3Q_{ij}^{TT}}{dt^3}\frac{d^3Q_{TT}^{ij}}{dt^3}\right\rangle \tag{1.22}$$

1.1.6 The compact binary waveforms: the Newtonian chirp

For two point particles with masses m_1, m_2 orbiting each other in a circular orbit about their center of mass, on a plane whose normal makes an angle ι with the positive z-axis, the quadrupole tensor can be expressed as:

$$Q_{ij} = \frac{1}{2}\mu a^2 \begin{pmatrix} 1+\cos 2\phi(t) & \cos\iota \sin 2\phi(t) & -\sin\iota \sin 2\phi(t) \\ \cos\iota \sin 2\phi(t) & \cos^2\iota(1-\cos 2\phi(t)) & -\cos\iota \sin\iota(1-\cos 2\phi(t)) \\ -\sin\iota \sin 2\phi(t) & -\sin\iota\cos\iota(1-\cos 2\phi(t)) & \sin^2\iota(1-\cos 2\phi(t)) \end{pmatrix}, \tag{1.23}$$

where, $\mu = \frac{m_1 m_2}{m_1+m_2}$ is the reduced mass of the system, a the orbital separation, and $\phi(t)$ is the orbital phase as a function of time. For Keplerian motion, the orbital separation and phase can be expressed in terms of the orbital speed (v) and the binary masses, such as:

$$\phi(t) = \omega t = \left(\frac{v^3}{c^3}\right)\frac{c^3 t}{G(m_1+m_2)}, \tag{1.24}$$

$$\text{and } v = \sqrt{\frac{G(m_1+m_2)}{a}}, \tag{1.25}$$

where ω is the orbital angular frequency. Using these expressions and the Quadrupole formula, one can obtain the transverse-traceless metric perturbation along the z-axis to be:

$$h_{ij}^{TT} \equiv -\frac{4G\mu a^2\omega^2}{c^4 r}\begin{pmatrix} (1+\cos^2\iota)cos2\phi(t) & \cos\iota\sin 2\phi(t) & 0 \\ \cos\iota\sin 2\phi(t) & -(1+\cos^2\iota)cos2\phi(t) & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{1.26}$$

where r is the distance from the source. It can be seen that Eq. (1.26) is a monochromatic plane wave for a Keplerian orbit. However, given that GWs carry energy away from the system, we expect the orbit to evolve with time. This evolution can be incorporated into the waveform by exploiting the fact that the GW luminosity given by the Quadrupole formula should be equal to the rate at which the binary system is losing orbital energy. According to the virial theorem, a gravitationally bound binary system has orbital energy $E^{orbit} = -\frac{1}{2}\mu v^2$. Setting $\frac{dE^{orb}}{dt} = -\frac{dE^{GW}}{dt}$ and using the quadrupole formula and Kepler's laws, it can be shown that the GW frequency (f, which is twice the orbital frequency) evolves in the following way:

$$\frac{df}{dt} = \frac{96}{5}\pi^{8/3}\left(\frac{G\mathcal{M}}{c^3}\right)^{5/3} f^{11/3}, \tag{1.27}$$

where $\mathcal{M} = \frac{(m_1 m_2)^{3/5}}{(m_1+m_2)^{1/5}}$ is the chirp mass of the system. Eq. (1.27) indicates that the waveform's amplitude and frequency increase with time, ultimately leading to merger which occurs when $f \to \infty$. Given a reference frequency/ starting point (f_0), the time to merger (t_c) can be calculated from Eq. (1.27). In terms of t_c and the phase of the orbit during merger (ϕ_c), the metric perturbation corresponding to this Newtonian chirp waveform from CBCs can be expressed in the following

way:

$$h_{ij}^{TT} = \begin{pmatrix} h_+ & h_\times & 0 \\ h_\times & -h_+ & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{1.28}$$

$$\text{where, } h_+ = -\frac{G\mathcal{M}}{c^2 r}\frac{1+\cos^2\iota}{2}\left(\frac{c^3(t_c-t)}{5G\mathcal{M}}\right)^{-1/4}\cos\left\{2\phi_c - 2\left(\frac{c^3(t_c-t)}{5G\mathcal{M}}\right)^{5/8}\right\}, \tag{1.29}$$

$$\text{and } h_\times = -\frac{G\mathcal{M}}{c^2 r}\cos\iota\left(\frac{c^3(t_c-t)}{5G\mathcal{M}}\right)^{-1/4}\sin\left\{2\phi_c - 2\left(\frac{c^3(t_c-t)}{5G\mathcal{M}}\right)^{5/8}\right\}. \tag{1.30}$$

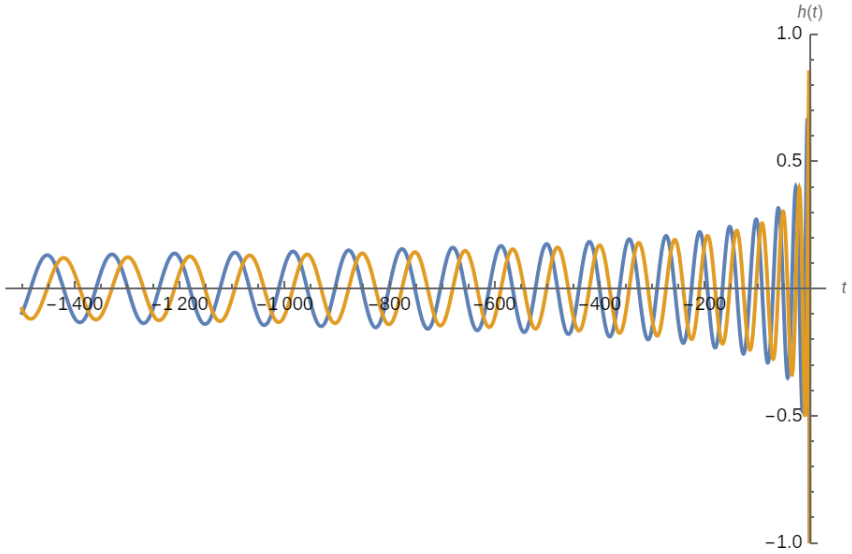


Figure 1.1: A visual representation of the Newtonian Chirp waveform. The amplitudes are presented in arbitrary units with blue representing the $+$ polarization and orange representing the $\times$ polarization. Credit: Leo C. Stein and Wolfram Mathworld.

For a visual representation of the Newtonican chirp waveform see figure 1.1. Given a waveform such as this, it is possible in principle to study its effects on the space-time trajectories of particles and thereby measure the parameters $\mathcal{M}, r, \iota, t_c, \phi_c$ that characterize the compact binary system.

Before exploring the effects of GWs on test particle geodesics, it is necessary to note that the actual merger waveform is more complicated than Eqs. (1.29), (1.30). As the binary components

get closer and speed up due to radiation reaction, Kepler's laws (which hold only in the weak field and slow-motion limit of GR) will no longer be valid, and Post-Newtonian corrections to the equations of motion of the binary system and the quadrupole radiation formula will both start to have non-negligible effects on the waveform's phase evolution (Cutler & Flanagan, 1994). Accounting for those corrections to the phase evolution of the compact binary waveform will introduce additional terms in Eqs. (1.27) and (1.29),(1.30) which scale as higher powers of $\frac{v}{c}$ than the ones considered in the case of a Newtonian chirp. Eventually, as the binary orbit shrinks to a size comparable to the inner-most stable circular orbit of the system, $\frac{v}{c}$ will approach 1 and the Post-Newtonian description of the waveform will break down entirely. In such a scenario, numerical relativity (NR) simulations which solve Einstein's equations numerically, will have to be invoked to describe the evolution of the binary system as well as the corresponding waveform.

1.1.7 Motion of test particles in space-time with GWs

Spacetime curvature, be it due to a passing GW or a nearby massive body, will cause parallel geodesics to bend towards or away from each other causing a relative acceleration between test particles. This acceleration, also known as tidal acceleration, is determined by the Riemann curvature tensor and is governed by the equation of geodesic deviation. In the transverse traceless gauge, up to linear order in h, the independent components of the Riemann curvature tensor will be given by:

$$R_{0i0j} = -\frac{1}{2}\partial_t^2 h_{ij}^{TT}. \tag{1.31}$$

Using this form of the Riemann tensor and the equation of geodesic deviation, it is possible to compute the relative acceleration of test particles in the context of various GW detectors (such as resonant bars, etc.). I will however sketch the detector response of only interferometric detectors since my thesis work primarily relies on GW data obtained from such detectors.

For two timelike geodesics (such as the beam splitter and the X-arm mirror of an interferometric detector) spatially separated by $\vec{\xi} = \xi\hat{p}$, the relative tidal acceleration is given by:

$$\ddot{\xi} = \xi R_{0i0j}\hat{p}^i\hat{p}^j = -\frac{1}{2}\xi\partial_t^2 h_{ij}^{TT}\hat{p}^i\hat{p}^j. \tag{1.32}$$

Assuming ξ to be a constant (L) plus some order h deviation δL, Eq. (1.32) can be integrated to obtain:

$$\delta L(t) = -\frac{L}{2}h_{ij}^{TT}\hat{p}^i\hat{p}^j. \tag{1.33}$$

For an interferometric detector with two equal length arms along $\hat{p}$ and $\hat{q}$ respectively, the fractional strain between the two arms is therefore given by:

$$h(t) = \frac{1}{2}\left(\hat{p}^i\hat{p}^j - \hat{q}^i\hat{q}^j\right)h_{ij}^{TT}. \tag{1.34}$$

For low-frequency GWs, photons going back and forth along the x and y arms will therefore traverse different path lengths, with the fractional change in length given by Eq. (1.34).

For a GW traveling in an arbitrary direction $\hat{n} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$, the metric perturbation in the TT-gauge can be expressed as:

$$h_{ij}^{TT}(t) = h_+(t)(e_i^1e_j^1 - e_i^2e_j^2) + h_x(t)(e_i^1e_j^2 + e_j^1e_i^2) \tag{1.35}$$

where, $e^1 = (\cos\psi\sin\phi - \sin\psi\cos\phi\cos\theta, -\cos\psi\cos\phi - \sin\psi\sin\phi\cos\theta, \sin\psi\sin\theta)$, and $e^2 = (-\sin\psi\sin\phi - \cos\psi\cos\phi\cos\theta, \sin\psi\cos\phi - \cos\psi\sin\phi\cos\theta, \cos\psi\sin\theta)$ are orthogonal basis vectors perpendicular to $\hat{n}$. Here ψ is the angle between e^1 and the line of nodes, i.e. $\hat{N}\times\hat{n}$, where $\hat{N}$ is the direction towards the celestial north pole. ψ defines the two polarization states and is hence known as the polarization angle. In this co-ordinate system, choosing $\hat{p} = (1,0,0)$, and $\hat{q} = (0,1,0)$, the fractional strain between the arms of an interferometric detector can be expressed

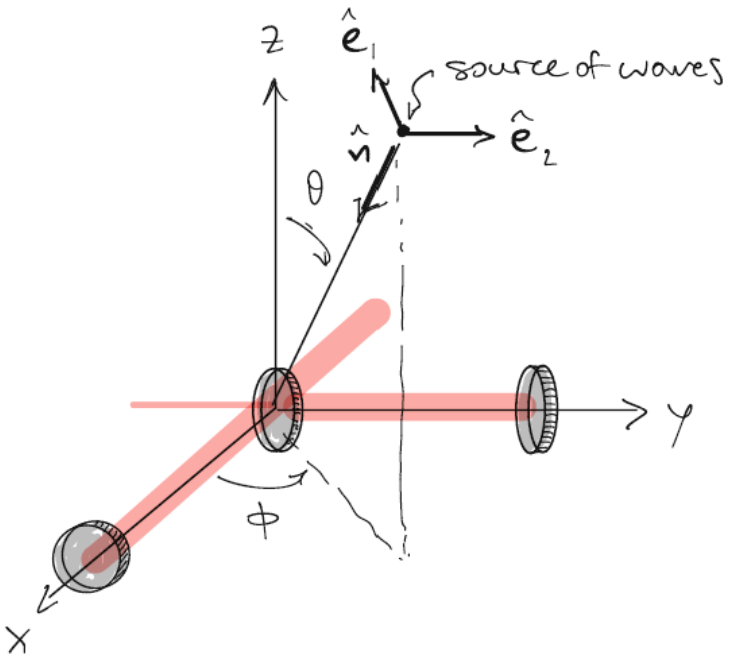


Figure 1.2: Diagram representing the co-ordinate system and basis vectors for a GW incident on an interferometer. Credit: Jolien Creighton.

as:

$$h(t) = h_+(t)F_+(\theta,\phi,\psi) + h_\times(t)F_\times(\theta,\phi,\psi), \tag{1.36}$$

$$\text{where } F_+ = -\frac{1}{2}\cos^2\theta\cos 2\phi\cos 2\psi - \cos\theta\sin 2\phi\sin 2\psi, \tag{1.37}$$

$$\text{and } F_\times = \frac{1}{2}\cos^2\theta\cos 2\phi\sin 2\psi - \cos\theta\sin 2\phi\cos 2\psi \tag{1.38}$$

are the antenna patterns of an interferometric detector's response to a GW coming from the sky position θ, ϕ (see figure 1.2 for a diagram representing the co-ordinate system and the orthogonal basis vectors). A simple Michaelson interferometer will therefore demonstrate a change in interference pattern in response to a strain of the form of Eq. (1.36) which can be induced by a

passing GW. When using a laser of wavelength λ_{laser}, the power read out of a simple Michaelson interferometer's photodiode would be the following function of the GW-induced strain:

$$I = I_0 \sin^2 \left(\frac{2\pi L h(t)}{\lambda_{laser}} \right), \tag{1.39}$$

again for low-frequency GWs.

However, as mentioned before, for typical CBC sources, this fractional strain is of the order of 10^{-21}, making Eq. (1.39) very difficult to measure. An order of magnitude estimate can be obtained using Eqs. (1.29) and (1.30), by substituting $m_1, m_2 \sim 30 M_\odot$, $r \sim 1 Gpc$, and $f \sim 100 Hz$ into them. More sophisticated interferometer configurations such as Fabry Perot cavities, power-recycling mirrors, and high-quality vibration dampening are required to directly measure such minuscule strains. However, indirect evidence of GW emission from CBC systems is obtainable by observing them through other means.

1.2 DETECTION OF GWS FROM COMPACT BINARY MERGERS

In the previous section, I presented a brief summary of the compact binary waveform, the information it encodes and the effect it has on interferometric detectors located far away from the source. In this section, I briefly recapitulate how GWs from compact binary systems were first detected, both indirectly and directly.

1.2.1 The first indirect observation of gravitational waves

As shown in the previous section, a gravitationally bound compact binary system will radiate away orbital energy and gradually speed up. If the binary components are observable by other means, tracking this orbital evolution is possible. For example, the orbital evolution of a system of binary neutron stars, one of which is a Pulsar, can be tracked by timing the observed radio pulses over a long period of time. The decay in orbital period for a circular orbit due to GW emission can be

obtained from Eq. (1.27) by substituting the orbital frequency with the orbital period $P = \frac{1}{f}$. For an elliptical orbit with eccentricity e, the decay in period due to GW emission is instead given by:

$$\frac{dP}{dt} = \frac{192\pi}{5} \frac{m_1 m_2}{(m_1+m_2)^2} \left(\frac{2\pi G(m_1+m_2)}{c^3 P}\right)^{5/3} \frac{1+\frac{73}{4}e^2+\frac{17}{96}e^4}{(1-e^2)^{7/2}}. \tag{1.40}$$

By timing the binary pulsar star system `PSR 1913+16` Hulse and Taylor were able to measure $\frac{dP}{dt}$ as well as the masses and eccentricity of the system, which agreed with Eq. (1.40) within a factor of 0.2 percent (Hulse & Taylor, 1975). This discovery therefore served as the first indirect observation of GWs from a compact binary system and was awarded the Nobel prize in 1993. However, the direct observation of GWs using ground based interferometric detectors had to wait two more decades as technology gradually caught up to the level of precession required to make such a detection.

1.2.2 The direct detection of gravitational waves

A Michaelson interferometer such as aLIGO will record changes in its interference pattern when the detector undergoes a strain similar to the RHS of Eq. (1.34). Even though the presence of Fabry Perot cavities and power recycling mirrors make the readout at the photodiode a more complicated and sensitive function of the GW-induced strain than Eq. (1.39), GW data from aLIGO detectors will ultimately comprise the strain time-series of Eq. (1.36). For a compact binary merger occurring somewhere in the Universe, the Newtonian chirp waveform of Eqs. (1.29) and (1.30) will induce a strain on aLIGO-like detectors that can be expressed as a function of time and the parameters of the CBC system in the following way:

$$h(t) = \frac{G\mathcal{M}}{c^2 r_{eff}} \left(\frac{c^3(t_0-t)}{5G\mathcal{M}}\right)^{1/4} \cos\left(2\phi_0 - 2\phi(t)\right), \tag{1.41}$$

$$\text{where }, r_{eff} = r\sqrt{F_+^2\left(\frac{1+\cos^2\iota}{2}\right)^2 + F_\times^2\cos^2\iota}, \tag{1.42}$$

$$2\phi_0 = 2\phi_c - \tan^{-1}\left(\frac{F_\times}{F_+}\frac{2\cos\iota}{1+\cos^2\iota}\right), \tag{1.43}$$

and $\phi(t) = \left(\frac{c^3(t_c-t)}{5G\mathcal{M}}\right)^{5/8}$ for the Newtonian chirp. Using the stationary phase approximation, one can take a Fourier transform of Eq. (1.41), to obtain the following frequency domain detector response:

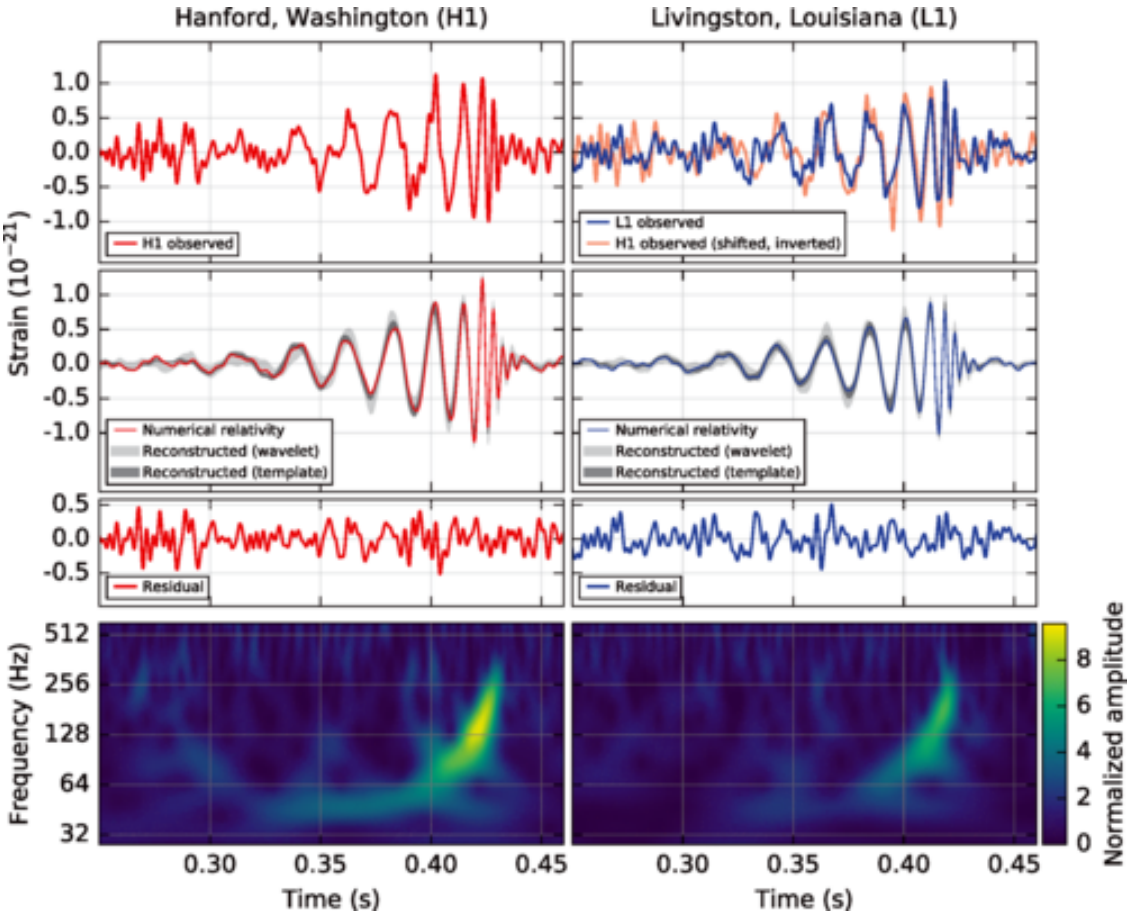


Figure 1.3: The strain data and frequency evolution of GW150914. Credit: the LIGO Scientific Collaboration, 2015 (Abbott et al., 2016a)

$$\tilde{h}(f) = \frac{G\mathcal{M}^{5/6}}{c^2 r_{eff}} f^{7/6}\exp\{2\pi i f t_0 - 2i\phi_0 - i\pi/4 + i\frac{3}{128}\left(\frac{c^3 f}{G\mathcal{M}}\right)^{-5/3}\}, \tag{1.44}$$

where the phase inside exponential corresponds to the Newtonian chirp signal and will generally contain Post-Newtonian corrections for more realistic signals. Such a signal can be searched for in detectors like aLIGO, aVirgo, and KAGRA using matched filtering (Allen et al., 2012). Once a signal is detected, Bayesian parameter estimation can be used to infer the compact binary properties (Thrane & Talbot, 2019), by computing the likelihood of obtaining strain data that contains a signal similar to Eq. (1.44), given the hypothesis that the true parameters have some particular value, with prior probabilities associated to each alternate hypotheses.

The twin aLIGO interferometers at Hanford and Livingston both span four kilometers along each perpendicular arm length (J. Aasi et al., 2015). On September 14th 2015, they recorded a GW signal from a BBH merger, a first-of-its-kind discovery (Abbott et al., 2016a). The chirping signal extracted from aLIGO data is displayed in figure 1.3. The source parameters of this system such as masses, inclination, and distance from the Earth, etc. were measured by analyzing the strain data recorded by the interferometers (figure 1.4, Abbott et al., 2016b).

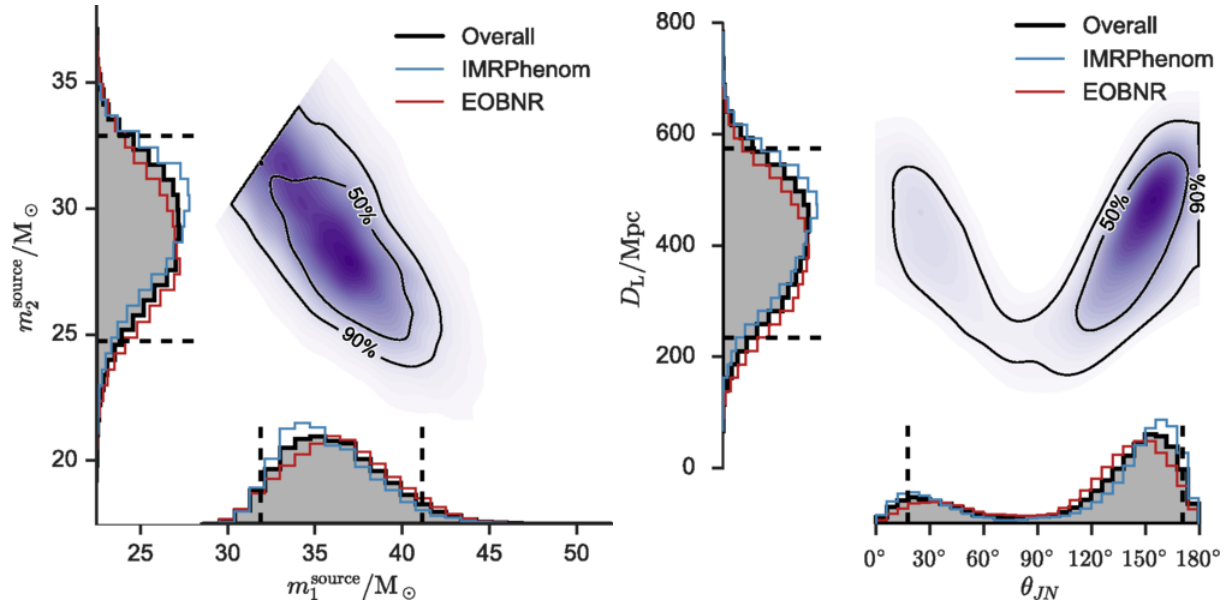


Figure 1.4: Estimating the source properties of GW150914 from GW data. Here, d_L is the luminosity distance of the source and $\theta_{JN} = \pi/2 - \iota$. Credit: Abbott et al. (2016b)

In addition to these two observatories, the LVK detector network comprises the Virgo interfer-

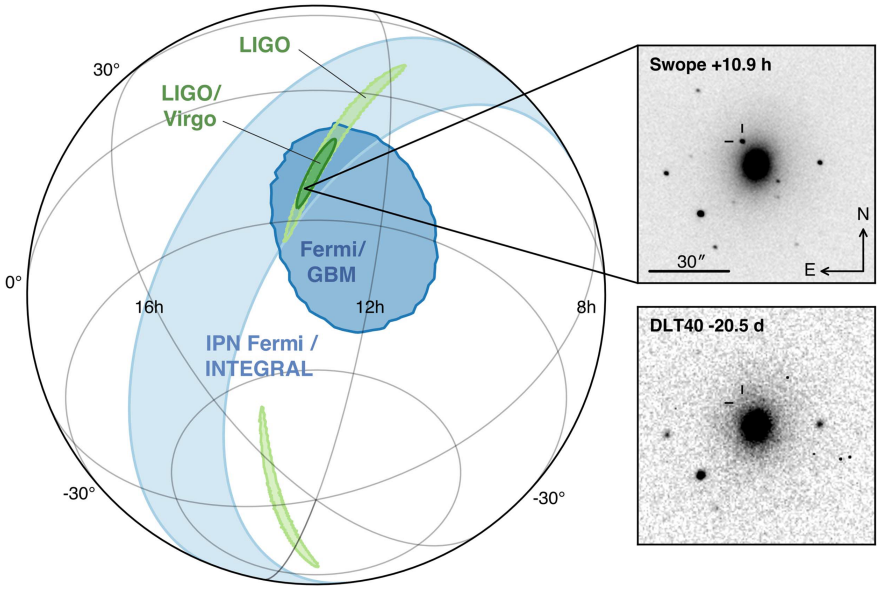


Figure 1.5: Sky localization of GW170817. Credit: Abbott et al. (2019c)

ometer located near Pisa which spans three kilometers along both arms (Acernese et al., 2014), and the KAGRA interferometer located near the city of Hida, which is of similar dimensions (Akutsu et al., 2021). A GW passing through the earth will induce a frequency domain strain in each of these detectors, similar to Eq. (1.44). Combining data from multiple detectors adds extra information to the analysis of GW data, which results in improved measurements. For example, multiple detector observations break the degeneracy between r and r_{eff} leading to accurate sky localization, which is highly relevant in the context of discovering the multi-messenger counterparts of GW candidates.

On 17th August, 2017, the LIGO-Virgo detector network observed, for the first time, GWs from a BNS merger (Abbott et al., 2017c). Around the same time, the Fermi, and INTEGRAL observatories recorded an SGRB from a sky location that had significant overlap with the multi-detector GW-based sky position measurement (figure 1.5, Savchenko et al., 2017; Goldstein et al., 2017). The accurate sky localization provided by GW data was used to optimize the search for other EM transients from the same BNS, which were consequently discovered (Abbott et al., 2019c). These

detections ushered in a new era of multi-messenger astronomy.

To date the LVK network has gone through a total of three observing runs with the fourth run currently ongoing. During Each observing run, the LVK observed at a higher sensitivity than the previous ones, leading to a larger number of detections. While data from the first half of the LVK's fourth observing run is still being analyzed, the cumulative catalog of CBCs obtained at the end of the third observing run has already facilitated revolutionizing discoveries.

1.3 THE SCIENTIFIC POTENTIAL OF GWS FROM CBCS

In the previous section, I briefly highlighted the detection of GWs from compact binary sources and the role played by the LVK detector network in detecting both GWs and their multimessenger counterparts. In this section, I will summarize the ground breaking discoveries that has already been enabled by GW observations from CBCs as well as what the future upholds for the same. In doing so I will try to emphasize the main motivation for my thesis work.

1.3.1 The third gravitational wave transient catalog

At the end of its third observing run, the LIGO-Virgo-KAGRA detector network unveiled a catalog of GW observations that comprised 90 astrophysically significant CBCs (figure 1.6, Abbott et al., 2023b). In addition to detecting these systems and measuring their source properties, we were also able to use their detected waveforms to test GR in the strong gravity regime (Abbott et al., 2021h). Such investigations using the third GW transient catalog (GWTC-3) revealed new information about the observable Universe that led to several advancements in almost every major field of contemporary physics.

The majority of CBCs found throughout GWTC-3 were identified to be stellar-mass binary black holes (Abbott et al., 2023b). The existence of these systems, while postulated, had never been observationally verified prior to LIGO's first detection on 14th September 2015. Similarly,

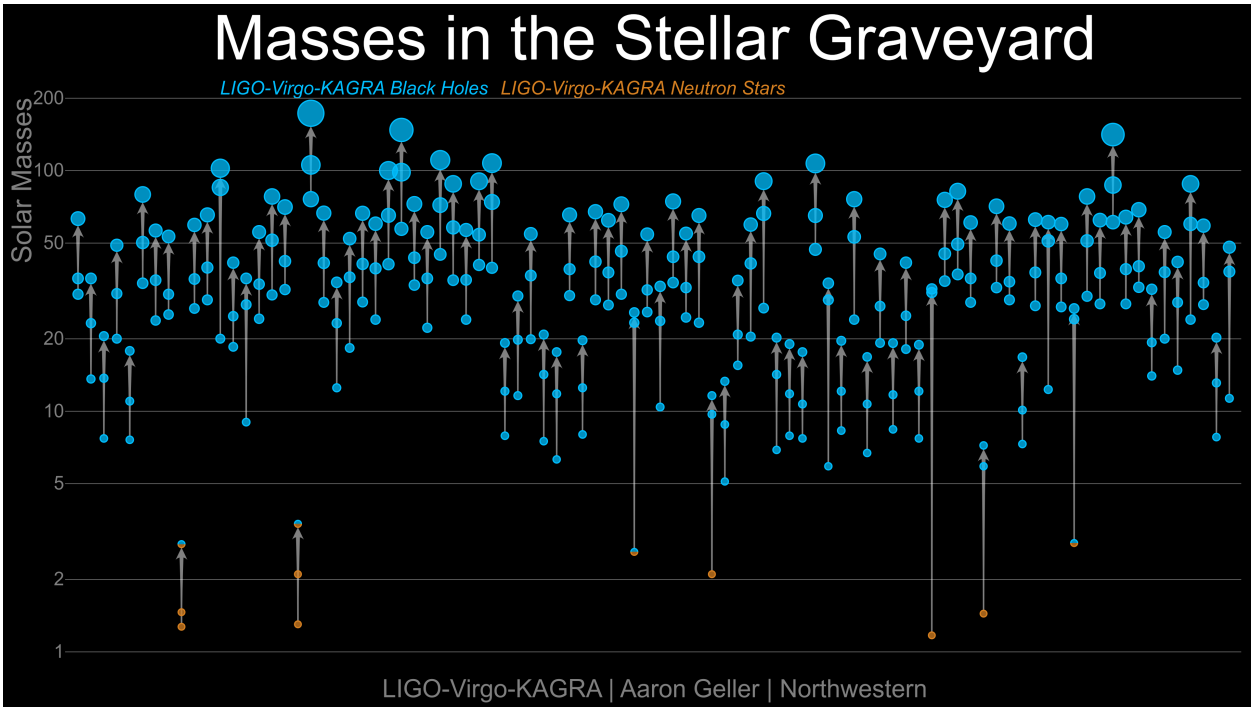


Figure 1.6: Credit: Aron Geller

their proposed formation mechanisms had also remained uncertain and unconstrained due to the lack of observations. The highly complex astrophysical processes thought to be responsible for the birth of these systems can be parameterized by several uncertain quantities which are expected to leave unique imprints on their observable population properties (Mandel & Farmer, 2022; Mapelli, 2020a,b). By studying the ensemble properties of the 69 high-confidence BBH systems recorded in GWTC-3, we were able to extract previously inaccessible information relevant to several fields of modern astrophysics ranging from the evolution of massive stellar binaries to the star formation history of the Universe (Abbott et al., 2023d).

In addition to BBHs, GWTC-3 observations comprised a handful of CBCs whose components were Neutron stars (NSs, Abbott et al., 2017c, 2020b, 2021e). These highly compact stellar remnants provide a natural laboratory for studying the properties of matter in extreme environments that are irreproducible in terrestrial laboratories. GWs from binary neutron stars (BNSs) encode

information on finite-size effects in the binary's evolution and hence can facilitate systematics-free mass-radius measurements for NSs (Özel & Freire, 2016). Exploiting the one-to-one mapping between an NSs mass-radius relationship and the nuclear equation of state (Lindblom, 1992), we were able to constrain several theories describing the strong interaction and hence, significantly improve our understanding of the same (Abbott et al., 2018a, 2020b).

Furthermore, the merger of all kinds of compact binary systems observed through GWTC-3, namely BBH, BNS, and neutron star black holes (NSBH) systems, occurs in the strong-gravity regime of GR. Analyzing the detected waveforms from GWTC-3 led to unprecedented strong field tests of GR. Several of its predictions, ranging from the values of Post-Newtonian coefficients in the inspiral part of the CBC waveform to the absence of GW dispersion and birefringence, were constrained using GWTC-3 data (Abbott et al., 2021h). As a result, our confidence in the predictions of GR, as well as our expectations of alternative theories of gravity underwent significant changes in light of the new data from GWTC-3.

Similarly, all CBC sources of GWs are standard sirens in the sense that they provide a direct measurement of their luminosity distances independent of the cosmic distance ladder (Schutz, 1986). While the measurement of the cosmological redshifts of these sources is degenerate with that of CBC masses[1], they can be extracted using various methods such as cross-correlation of the sources' sky positions with galaxy catalogs (Del Pozzo, 2012), the detection of multi-messenger counterparts (Holz & Hughes, 2005) and CBC universal properties such as features in the mass-spectrum (Farr et al., 2019) or the unique mass-radius relationship of NSs (Del Pozzo et al., 2017). Employing some of these methods on GWTC-3 data, we were able to obtain independent measurements of the Hubble parameter, which demonstrated the potential of GW catalogs in resolving the Hubble tension and other inadequacies in our understanding of the cosmic expansion history (Abbott et al., 2021b).

[1] See section 2.2.1.1 for more details regarding this degeneracy.

In short, catalogs of GW observations such as GWTC-3 present unprecedented observational probes of the mysteries of compact binary formation, the unknown properties of matter at extreme densities, the validity of GR in the strong field regime, and the cosmic expansion history. With the ongoing and upcoming observing runs of the LVK promising a large number of new observations (Abbott et al., 2018b; Kiendrebeogo et al., 2023), growing GW catalogs will provide increasingly precise constraints on previously unexplored physics and hence, continue to enable revolutionary discoveries.

1.3.2 Multimessenger astronomy and low-latency alerts

The detection of multi-messenger counterparts to CBC sources, such as the observation of electromagnetic (EM) transients from the BNS merger GW170817, greatly increases their potential as unique probes of physical phenomena far beyond the scope of what either photons or GWs could have probed individually (Abbott et al., 2017d; Mészáros et al., 2019). For example, as GW observations confirmed the source to be a BNS merger, EM observations captured imprints of r-process nucleosynthesis, leading to the discovery that BNS mergers contribute significantly to the production of heavy elements in the Universe (Kasen et al., 2017). Similarly, the association of GRB170817A with GW170817 provided direct evidence that BNS mergers can power short gamma-ray bursts (SGRBs Abbott et al., 2017b). This scenario had long been postulated and yet never before been verified observationally. Furthermore, timing the arrival of gamma-rays and GWs from GW170817 led to measurements of the deviation of the speed of GWs from that of light with extreme precision (Abbott et al., 2019e). EM observations further added to the EoS information that could be extracted from GW data alone (Radice et al., 2018), while also facilitating measurements of the Hubble parameter by identifying the redshift of 170817's host galaxy in complement to the independent distance measurements provided by the GW data (Abbott et al., 2017a).

To facilitate the prompt discovery of multi-messenger counterparts to GW sources, the LVK sends out real-time alerts to astronomers during its observing runs (Abbott et al., 2018b; Kiendrebeogo et al., 2023), so that they can potentially follow-up on GW candidates of interest. To maximize the amount of EM data obtained and to minimize the efforts wasted on following up CBC sources that are not expected to be electromagnetically bright, these alerts include information about the nature of the CBC sources in addition to the sky location and false alarm rates (FARs) of the candidates. Sophisticated low-latency algorithms that rely on our current knowledge of the compact binary population, merger rates (Kapadia et al., 2020), and the NS EoS (Chatterjee et al., 2020) to assess GW candidates in real-time and the resulting classification estimates are sent out as part of GW alerts to the public. CBC candidates that have a high probability of being BNS or NSBH, and have the potential to be electromagnetically bright sources can be followed up by telescopes which can be made to sweep regions of the sky wherein the candidate was localized using GW data. As the sensitivity of GW detectors continues to increase, reliable estimates of trigger classification will continue to optimize the search and discovery of mutimessenger counterparts to GWs from CBCs.

1.3.3 The big picture

The advanced LIGO and Virgo detector configurations are expected to reach design sensitivity by the LVK's 5th observing run which is scheduled to start in a couple of years. Given current estimates for the merger rates of different kinds of CBCs, the LVK will observe, on an average 10s, 100s, and 1000s of BNSs, NSBHs, and BBHs respectively, per every year of observing time through O5 and beyond (see figure 1.7 for a pictorgraphic representations). The planned addition of new detectors such as LIGO-India to the network of ground-based observatories will highly improve the prospects of source-localization in addition to detection. In the era of third-generation ground-based detectors such as the Einstein Telescope (Maggiore et al., 2020) and the Cosmic

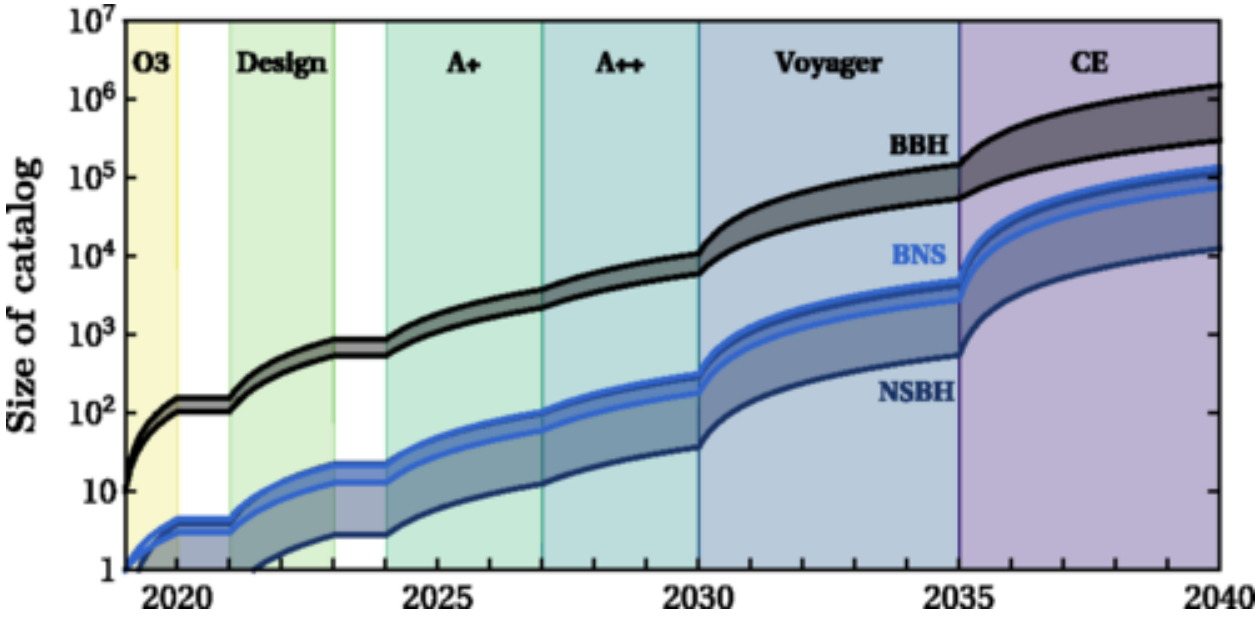


Figure 1.7: Growing catalog of GWs from CBCs using both current and future detectors. Credit: Baibhav et al. (2023).

Explorer (Reitze et al., 2019), the number of expected detections per year will grow by roughly 3 orders of magnitude (Baibhav et al., 2023). Hierarchical inference of the compact binary population properties, the NS EoS, the cosmological parameters, and the deviations from GR are expected to become increasingly informative with the number of observations analyzed. In particular, the widths of the credible intervals of inferred quantities will likely decrease as one over the square-root of the number of events. Therefore, possibly by O5 and definitely by the 3G era, these analyses will yield robust conclusions on several unresolved mysteries of modern physics.

Pinning down the CBC population distribution, the cosmological parameters, the NS EoS, and the parameters quantifying deviations from GR to sub-percent level precision will require robust statistical modeling and associated inference frameworks that scale tractably with the size of the analyzed datasets. On the other hand, highly accurate trigger classification algorithms will be necessary to optimize the search for multi-messenger counterparts in the O5 era and beyond. This necessity of developing rigorous and scalable analysis methods that utilize the rapidly growing

gravitational catalogs to extract new physical insights and enable fast discovery of multimessenger counterparts comprises the primary motivation for this thesis which is organized as follows.

1.4 INTRODUCTION TO THIS THESIS

In the previous section, I summarized how GW observations of CBCs, in particular growing GW catalogs, promise revolutionary discoveries in several fields of contemporary physics, and how some of those promises have already been fulfilled. Motivated by the enormous potential of GWs, I have worked throughout the past five years to develop new analysis methods for extracting physics from growing CBC catalogs, and facilitating the prompt discovery of multi-messenger counterparts to CBCs. I have also analyzed contemporary datasets using these methods to gain new observational insights. Specifically, my thesis aims to exploit GW observations from CBCs to answer the following unsolved questions:

- How do compact binaries capable of merging via GW emission form?
- What is the nature of the cosmological expansion?
- What are the universal properties of degenerate matter at supra nuclear densities?
- How accurate are the predictions of GR?

To this end, as part of my thesis work, I have contributed to the development of tractable, unbiased and robust analysis frameworks that can ingest GW data and attempt to extract valuable information pertaining to these mysteries in our current understanding of the observable Universe. The rest of this thesis is organized as follows.

In chapter two, I discuss the fundamentals of the Bayesian inference methods, which I later use as a cornerstone to develop novel algorithms that implement accurate trigger classification and GW-based physical inference. In chapter three, I summarize my contributions to the development

of improved trigger classification algorithms, which were later on used by the LVK collaboration to categorize GW triggers both in real-time during O4 and in post-processing at the end of O3.

In chapters four and five, I list my contributions to the development of data-driven population models for compact binary systems, and demonstrate their potential to conduct model-agnostic astrophysical and cosmological inference from GW data that is free of our uncertainties regarding compact binary formation scenarios. In addition to obtaining the first data-driven constraints on the redshift evolution of the astrophysical BBH mass distribution, I also found evidence of a new kind of mass-spin correlation in the GWTC-3 BBH population which has highly interesting astrophysical implications. I further contributed to obtaining the first non-parametric spectral siren constraint on the Hubble parameter from GWTC-3 data. It is a GW-based cosmological measurement that is free of the distance ladder, the limitations of EM telescopes, and the uncertainties regarding the shape of the BBH population distribution.

In chapter six, I summarize my contributions to the development of a rapid EoS-inference method that can constrain phenomenologically parameterized models for the nuclear EoS from multiple BNS and NSBH observations. By relying on previous works to bypass the problem of dimensional reduction, I was able to implement a framework that can infer the EoS from 16 events in less than 20 hours. Armed with this rapid analysis method, I did a systematics study of how the EoS constraints vary with the mass of the NSs and the SNR of the observed signal, by simulating different populations of BNS mergers. This inference framework was later used by the LVK collaboration to obtain EoS constraints from a newly discovered NSBH-like CBC during O4.

In chapter seven, I list my contributions to obtaining constraints on alternative theories of gravity by measuring the speed of GWs from GWTC-3 observations. By relying on an effective field theory framework for studying small deviations from the predictions of GR and the standard model in low-energy experiments that can be caused by the violation of fundamental symmetries such as Lorentz invariance at the level of the action, we were able to constrain simultaneously, for the first

time, all nine coefficients of the Lorentz violating GW propagation, using GW data alone. In addition to developing a fully Bayesian framework for measuring these coefficients which, again for the first time, accounted for selection effects in speed of gravity measurements, I compared it with other popular methods and demonstrated why it is superior to them, in terms of both the reliability and the interpretation of the yielded uncertainty estimates.

In chapter eight, I conclude with a summary of the aforementioned findings and developments, a list of follow-up investigations, and the future applications of these works given the expected number of new detections promised by LVK's current and future observing runs, as well as by the planned third generation ground-based detectors.

1.5 DECLARATION

This thesis represents the culmination of five years of research, during which various themes and topics related to gravitational wave science have been investigated. Most of the findings presented in this work have already been published or made available as preprints elsewhere. As of May 2024, a comprehensive list of publications (and one article in preparation) relevant to this dissertation is,

[1] *Searching for binary black hole sub-populations in gravitational wave data using binned Gaussian processes*

Ray, A., Magaña, I., Breivik, K., and Creighton, J.

Submitted to Nat. Astron. (arxiv:2404.02522) 1 citation

[2] *Beyond Gaps and Bumps: Spectral Siren Cosmology with Non-Parametric Population Models*

Magaña, I., and **Ray, A.**

Submitted to PRL (arxiv:2404.02522)

[3] *Updating gravitational-wave source classification estimates with rapid parameter estimation*

Rose, C., **Ray, A.**, Valsan, V., Brady, P., *et. al.*

(In preparation)

[4] *Non-parametric inference of the population of compact binaries from gravitational wave observations using binned Gaussian processes*

Ray, A., Magaña, I., Mohite, S., Crehighton, J., Kapadia, S.

ApJ 957 37 (2023) (arXiv:2304.08046) 11 citations

[5] *Measuring Gravitational Wave Speed and Lorentz Violation with the First Three Gravitational-Wave Catalogs*

Ray, A., Fan, P., He, V. F., Bloom, M. *et. al.*

Submitted to PRD (arXiv:2307.13099) 1 citation

[6] *When to Point Your Telescopes: Gravitational Wave Trigger Classification for Real-Time Multi-Messenger Followup Observations*

Ray, A., Niu, W., Sakon, S., Ewing R., *et. al.*

To be submitted to CQG (arXiv:2304.08046v1) 2 citations

[7] *Rapid Hierarchical Inference of Neutron Star Equation of State from multiple Gravitational Wave Observations of Binary Neutron Star Coalescences*

Ray, A., Camilo, M., Creighton, J., Ghosh, S., Morisaki, S.

Phys. Rev. D 107, 043035 (2023) (arXiv:2211.06435) 2 citations

[8] *Observation of Gravitational Waves from the Coalescence of a 2.5-4.5 $M_\odot$ Compact Object and a Neutron Star*

The LIGO Scientific, Virgo and KAGRA Collaborations (including **Ray, A.**)

Submitted to APJL (arXiv:2404.04248)

[9] *GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run*

The LIGO Scientific, Virgo and KAGRA Collaborations (including **Ray, A.**)

Phys. Rev. X 13, 041039 (2023) (arXiv:2111.03606) 382 citations

[10] *The population of merging compact binaries inferred using gravitational waves through GWTC-3*

The LIGO Scientific, Virgo and KAGRA Collaborations (including **Ray, A.**)

Phys. Rev. X 13, 011048 (2023) (arXiv:2111.03634) 186 citations

[11] *GWTC-2.1: Deep Extended Catalog of Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run*

The LIGO Scientific, Virgo and KAGRA Collaborations (including **Ray, A.**)

Phys. Rev. D 109, 022001 (2023) (arXiv:2111.03634) 108 citations

[12] *Observation of gravitational waves from two neutron star-black hole coalescences*

The LIGO Scientific, Virgo and KAGRA Collaborations (including **Ray, A.**)

ApJL, 915, L5 (2021) (arXiv:2106.15163) 562 citations

Each of these works has contributed differently to the content of the thesis. A brief summary of the relevant publications with each chapter of this dissertation is as follows:

- The results and/or methods of trigger classification and rate estimation presented chapter 3 have been communicated through [6], and the LVK collaboration papers [8, 9, 10, 11, 12], with the contents of the second-last section currently being prepared for submission [3].
- The population inference methods of chapter 4 and the results therein constitute [1, 4]
- The cosmological probe discussed in chapter 5 is communicated through [2]

- The equation of state inference research summarized in chapter 6 is published through [7] and the LVK collaboration paper [8].

- The research involving testing general relativity described in chapter 7 is communicated through [5]

This work was supported by NSF Award No. PHY-2207728. I am also grateful for computational resources provided by the Leonard E Parker Center for Gravitation, Cosmology and Astrophysics at the University of Wisconsin-Milwaukee. I thank the LIGO and Virgo Collaboration for providing the data for this work. This research has made use of data, software and/or web tools obtained from the Gravitational Wave Open Science Center (https://www.gw-openscience.org/), a service of LIGO Laboratory, the LIGO Scientific Collaboration, and the Virgo Collaboration. I am grateful for computational resources provided by the LIGO Laboratory and supported by National Science Foundation Grants PHY-0757058 and PHY-0823459, PHY-0823459, PHY-1626190, and PHY-1700765. This material is based upon work supported by NSF's LIGO Laboratory which is a major facility fully funded by the National Science Foundation

Part I

Statistical Framework

Chapter 2

Bayesian Hierarchical Inference

But astrophysics is an observational, not an experimental science. The phenomena under study are nearly always inaccessible to direct manipulation, and must be observed from great distances, and often indirectly. As a consequence, the scientific inferences of astrophysicists are fraught with uncertainty. To realize the scientific potential of astrophysics thus demands an understanding, not only of the necessary physics, but also of the principles of inference that dictate how information can be optimally extracted from observational data and how theoretical predictions can be rigorously compared with such data.

T.J. Loredo, 1992

2.1 INTRODUCTION

As emphasized in the previous chapter, the primary motivation of this thesis work was to implement the inference of physical models and the classification of CBC candidates using growing CBC catalogs, both of which rely on the measurement of individual event source properties. Prior to extracting any information regarding the source of a GW signal, it must first be identified and distinguished from noise transients lurking within the interferometric strain data.

Given that the expected CBC signals are extremely weak and amidst noise transients of comparable strength, robust inference frameworks for candidate identification, parameter estimation, and hierarchical modeling are necessary to implement the above-mentioned analyses accurately. In this chapter, I will summarize the basics of searching for GW signals in noisy strain data and derive a self-consistent framework that performs trigger classification and physical inference using a cumulative collection of GW candidates that were found by search pipelines to be significant above some threshold.

2.1.1 The identification and ranking of CBC candidates in strain data

Given models for the temporal variation of CBC signals and the distribution of random noise present in the data, one can construct a detection statistic that can be used to estimate whether or not the data prefers the hypothesis that a signal is present in a particular data segment, over its complement. The Neyman-Pearson lemma dictates that the optimal detection statistic is the ratio of the likelihood of the signal hypothesis to that of the noise hypothesis.

For stationary Gaussian noise, the likelihood ratio is a monotonic function of the matched filter signal-to-noise ratio (SNR), which can be expressed as the peak of the following time series:

$$z(t) = 4\mathcal{R}e \int df \frac{\tilde{d}(f)\tilde{h}^*(\theta, f)}{S_n(f)} e^{2\pi i f t}, \tag{2.1}$$

where $\tilde{d}(f)$ and $\tilde{h}(f, \theta)$ are the Fourier transforms of the time series data and the CBC waveform respectively, and $S_n(f)$ is the power spectral density of the random noise time-series (Allen et al., 2012). Given the computational cost of evaluating the waveform at a large number of frequency points, the CBC parameter space is discretized leading to a bank of template waveforms. GW candidates can potentially be identified as templates being "rung up" with a certain SNR $\rho = z(t_0)$, where t_0 is the time of arrival of the signal. For further details regarding how the noise power spectral density is measured and the numerical implementation of matched filtering, see Allen et al. (2012).

Due to the presence of non-stationary noise artifacts or "glitches" in the data, SNR alone cannot be used to rank GW candidate events optimally. To assess whether or not a glitch has caused a high SNR candidate several consistency tests have been developed. The χ^2 consistency check (Allen et al., 2012; Allen, 2005), for example, is a form of time-frequency decomposition that tests for expected signal accumulation in different frequency sub-bands. For a total of p bands, the minimum and maximum frequencies can be uniquely chosen by ensuring that each band aggregates

an equal amount of power from the template waveform (h). The scaled mean square deviation of the matched filter SNR across all bands then yields the χ^2 statistic, which is expected to follow a χ^2 distribution for Gaussian noise. Since noise transients often produce high SNR at only a few frequency bands, they can cause large χ^2 values that are unlikely to be due to a perfectly matched CBC signal in additive Gaussian noise.

Another form of signal consistency test is achieved by comparing the SNR time series with the auto-correlation of the template waveform time series (Messick et al., 2017). The absolute deviation of this time series with the SNR can be normalized with respect to its value in the absence of a signal and averaged over some time window containing the trigger to obtain the ξ^2 statistic (Messick et al., 2017). This quantity is essentially representative of how similar the data is to the rung-up template waveform and is expected to be zero for an exact match with some GW candidate and 1 for a noise transient (Messick et al., 2017).

Using ρ, χ^2 or ξ^2, and other characteristics of the signal such as arrival time and phase at different detectors, various search statistics can be constructed for ranking GW candidates and potentially identifying astrophysical sources. For example, the `PyCBC` search pipeline (Davies et al., 2020; Nitz et al., 2018; Canton et al., 2021) reweights SNR by some function of χ^2 to construct their their ranking statistic while `MBTA` (Adams et al., 2016; Aubin et al., 2021) uses SNR weighted by a function of ξ^2 instead. On the other hand, the `GstLAL` search pipeline (Cannon et al., 2021; Messick et al., 2017; Sachdev et al., 2019; Cannon et al., 2015) uses the likelihood ratio as a function of ρ, ξ^2, arrival time, coalescence phase, etc to rank CBC candidates and distringuish them from noise transients.

During ranking, triggers from different detectors that ring up the same template are clustered across some time window for coincidence formation (Messick et al., 2017; Davies et al., 2020). Single detector triggers are often down-ranked as compared to coincident candidates (Messick et al., 2017; Sachdev et al., 2019) since it is much less likely for two glitches in two different

detectors to match with the same CBC template at the same time. The probability distribution of the ranking statistics conditioned on the noise hypothesis can be used to compute the false alarm rate (FAR) of each ranked candidate, which assesses the rate at which noise artifacts can produce a trigger of the same significance/rank (e.g. Cannon et al., 2015).

In low-latency (real-time/online) searches, a high-ranked candidate will often need to be categorized into particular classes of CBCs so as to alert astronomers of their existence and potential to be electromagnetically bright. In offline searches, the collection of ranked and classified triggers above some established threshold is used to generate CBC catalogs which comprise accurate measurements of individual event source properties. Combining individual event measurements from growing catalogs leads to the inference of highly interesting physical models that constrain the population-level/universal properties of CBC systems and signals.

2.1.2 Individual and universal source properties: the need for Bayesian inference

Once a significant CBC candidate is identified, it is necessary to ascertain the nature and parameters of its system of origin. Afterward, it is of interest to combine data from multiple such events and conduct physical inference. While implementing such analyses, the so-called frequentist viewpoint of statistical inference that is employed by search pipelines has mostly been avoided in favor of Bayesian methods, for several reasons. I shall try to highlight some of them using examples relevant to my thesis work.

Firstly in the context of asserting the origin of a GW candidate, FAR alone is often not sufficient to decide whether or not it is astrophysical. In addition to FAR, one must account for the rate of CBC signals that are expected to be present in the data at the same level of significance. For example, different kinds of CBC candidates at the same FAR will be assigned different values for the probability of astrophysical origin depending on how their rate of occurrence compares to that of noise transients. Therefore, the degree to which we can ascertain the origin of a GW can-

didate given data is reliant not only on its significance but also on our prior expectations regarding the signal and noise hypotheses, which in turn encode rate information. It is difficult to compute this degree of posterior support in the frequentist viewpoint that hinges upon averaging over alternate data rather than alternate hypotheses and is hence unable to assign prior probabilities to the latter (Loredo, 1992).

Secondly, in the case of individual event parameter estimation, the frequentist approach would involve constructing a function of the data known as an estimator and then computing its sampling distribution from the likelihood. The sampling distribution can then be used to obtain the average behavior of the estimator over realizations of alternate data. This implies that the frequentist $P\%$ confidence interval of the estimate would contain the true value of the parameter $P\%$ of the times the exact same observation is repeated. However, to ensure a long-run frequency of success for the estimated value, this approach completely neglects prior information about the true values of the parameter that is necessary for reliable single-case inference (Cornfield, 1969; Loredo, 1992). Given that we will never again observe the same CBC in the same detector configuration, it is instead desirable to obtain a measure of the degree to which we believe the true values of the parameter to be in some range, given data and relevant a priori expectations (Finn, 1997).

Similarly, while estimating a universal quantity that constrains a subset of CBC observables from multiple uncertain observations, it is necessary to marginalize over the uncertainties of all individual event CBC parameters including the ones that are agnostic to the model being inferred. The inclusion of new observations therefore amounts to a build-up of nuisance parameters in the inference problem. In such a scenario, the *profile* likelihood of the universal quantity maximized over the nuisance parameters can lead to biased estimates (Loredo, 2004). Furthermore, accounting for physically informed priors on the universal parameters themselves greatly improves their inference (Read et al., 2009). However, frequentist methods disregard such information altogether despite their robustness and availability.

Furthermore, frequentist approaches toward rate estimation, trigger classification, parameter estimation, and hierarchical inference have several additional problems to the ones discussed thus far (Loredo, 1992). While some of these hurdles can be bypassed through tactful maneuvers, others might be insurmountable. On the other hand, the Bayesian approach offers a straightforward solution to the mentioned inference problems while being free of the challenges faced by its frequentist counterpart (Loredo, 1992).

In particular, Bayesian inference instinctively accounts for all available information on various hypotheses by assigning prior probabilities to each of them. Furthermore, nuisance parameters can be straightforwardly marginalized to yield unbiased constraints on hierarchical models. Throughout the rest of this chapter, I will outline the derivation of a self-consistent framework of hierarchical Bayesian inference, that averages over alternative hypotheses regarding the physical scenarios we want to elucidate using GW catalogs and yields robust and unbiased estimates of the degree to which we can believe such hypotheses given GW data.

2.2 THE HIERARCHICAL LIKELIHOOD

Strain data from interferometric detectors constitutes a noisy time series spanning over the detector's time of operation, with GW signals potentially lurking within its various segments. In addition to astrophysical signals, the data can contain noise transients or glitches in some segments. Therefore, it is necessary, in general, to self-consistently account for the possibility of GW candidates being glitches while attempting to extract physics from multiple GW observations, and while assessing the significance of candidates for generating low-latency alerts and GW transient catalogs. In other words, for each trigger, we intend to average over the alternative hypotheses regarding their origin while assigning physically informed priors to each hypothesis.

To develop this self-consistent Bayesian framework, we start by partitioning the time series data ($\vec{x}$) into N time chunks $\vec{d} = \{x_1, x_2, ..., x_N\}$ of width $\delta T = \frac{T}{N}$, that is small enough to support

the assumption that at most one candidate is found above threshold in each chunk (Farr et al., 2015; Vitale et al., 2020). Under these assumptions, the possible scenarios to count for each time chunk will depend on the set of possible hypotheses regarding the sources of triggers that may or may not be above the detection threshold. We assume the following four cases determined by two hypotheses:

- Hypotheses: $\eta \in \{F, B\}$. Candidates in the data are assumed to be of either astrophysical or terrestrial in origin
- Detection flags: $g_i = \begin{cases} 1 & \text{for } \mathcal{F}(d_i) \geq \mathcal{F}_{th} \\ 0 & o.w. \end{cases}$

With these hypotheses and detection flags defined we can proceed to construct the likelihood functions conditional on different hypotheses, by first implementing Bayesian inference on a single chunk.

2.2.1 Bayesian Inference of a single chunk

Strain data in each chunk can be thought to comprise a random noise time series $n(t)$ and possibly a modeled signal $h(t, \theta)$ characterized by CBC system parameters θ, such as:

$$d_i(t) = \begin{cases} h(t,\theta) + n(t) & \text{for } \eta = \mathrm{F} \\ n(t) & \text{for } \eta = \mathrm{B} \end{cases}. \tag{2.2}$$

Here we further divide the time-series data in a single chunk into n_s points where n_s is determined by the sampling frequency. Using, Eq. 2.2 and a noise model, i.e. a joint distribution function for the random noise at each time point in the series $p_n(\{n(t_a)\}_{a=1}^{n_s})$, we can construct the likelihood

function of obtaining data d_i given signal parameters θ, for both the signal and noise hypotheses:

$$p(d|\theta, F) = p_n(\{d(t_a) - h(t_a, \theta)\}_{a=1}^{n_s}) \tag{2.3}$$

$$p(d|\theta, B) = p_n(\{d(t_a)\}_{a=1}^{n_s}) \tag{2.4}$$

Note that these likelihoods are normalized distributions of the data in the following sense:

$$\int ddp(d|\theta, \eta) = 1, \tag{2.5}$$

where the integral is over all possible values of all data points $\{d(t_a)\}_{a=1}^{n_s}$. Combining these likelihood functions with a priori expectations regarding signal parameter values, one can compute the posterior probability distribution of these parameters. According to Bayes' theorem, the posterior distribution of CBC parameters under the signal hypothesis ($p(\theta|d, F)$ takes the following form:

$$p(\theta|d, F) = \frac{p(d|\theta, F)p(\theta|F)}{p(d|F)}, \tag{2.6}$$

where $p(\theta|F)$ is the prior probability of signal parameters conditioned on the signal hypothesis and $p(d|F)$ is known as the Bayesian evidence for the signal hypothesis. For stationary Gaussian noise, the likelihood corresponding to the modeled waveform $h(\theta, t)$ is given by:

$$\log p(d|\theta, F) = -\frac{1}{2}\sum_{i,j}(C_n^{-1})^{ij}\left\{d(t_i) - h(\theta, t_i)\right\}\left\{d(t_j) - h(\theta, t_j)\right\} + const \tag{2.7}$$

where $C_n^{ij} = f_n(|t_i - t_j|)$ is the covariance matrix of the stationary Gaussian process whose realizations constitute the detector noise. The RHS of Eq. (2.7) can also be expressed in the frequency domain such as:

$$\log p(d|\theta, F) = -\frac{1}{2}\left\langle d - h(\theta)|d - h(\theta)\right\rangle + const, \tag{2.8}$$

where I've defined the noise-averaged product of two time series in the following way:

$$\langle A|B\rangle = 4\mathrm{Re}\left[\int_0^\infty df \frac{\tilde{A}(f)\tilde{B}^*(f)}{S_n(f)}\right], \tag{2.9}$$

with

$$S_n(f) = \int f_n(\tau) e^{-2\pi i f\tau} d\tau \tag{2.10}$$

being the noise power spectral density and $\tilde{A}(f)$ is the Fourier transform of the time-series $A(f)$. See Finn (1998) for a derivation of Eq. (2.9) from Eq. (2.7).

Given a waveform model such as Eq. (1.44), and strain data corresponding to a chunk that was flagged by search pipelines for containing a signal above detection threshold, Eq. (2.8) can be evaluated as a function of CBC parameters $\theta = \{m_1, m_2, r, ...\}$ by using fast Fourier transforms (FFTs) on the data. Choosing suitable priors for the parameters, one can use stochastic sampling-based Bayesian inference engines such as Markov Chain Monte Carlo (MCMC) or nested sampling (Skilling, 2006) to sample their posterior distributions and obtain Bayesian credible intervals which will contain the true values of the parameters with certain posterior probability (Thrane & Talbot, 2019).

2.2.1.1 Observable CBC parameters

The parameters of the compact binary system that can modulate the detected waveform can be constrained through Bayesian inference of a single chunk. The Newtonian chirp waveform of Eq. (1.44), for example, can be used to estimate the effective distance r_{eff} and the chirp mass ($\mathcal{M}$), since these two parameters directly modulate the amplitude and phase evolution of the signal at Newtonian order. Due to the cosmological expansion, the observed frequency/time-evolution and flux of the signal get redshifted, which leads to a scaling of the measured parameter values by the cosmological redshift of the source. In other words, the observables from a Newtonian chirp

waveform are the effective luminosity distance and redshifted (detector frame) chirp mass which are defined as follows:

$$\mathcal{M}_z = \mathcal{M}(1+z), \tag{2.11}$$

$$D_{eff} = r_{eff}(1+z). \tag{2.12}$$

A single GW observation cannot break the degeneracy between mass and redshift of the source.

In addition to these two parameters, the direct observables from a Newtonian chirp signal are the arrival time and coalescence phase of the signal. For Gaussian noise, the likelihood function can be efficiently maximized or marginalized over time, phase, and amplitude/effective distance, which makes template bank construction by searches unnecessary for these parameters. Such parameters that can be efficiently marginalized without requiring costly waveform evaluations are known as extrinsic parameters. For multi-detector observations, the arrival time at different detectors can be used to triangulate the sky location of the source which also comprises a couple of extrinsic parameters.

Unlike the extrinsic parameters, measurements of the detector frame chirp-mass require full waveform evaluation for each value of the sampled parameter, making it an intrinsic parameter. During searches, it would therefore be necessary to construct a template bank of waveforms corresponding to different values of chirp mass. Post-Newtonian corrections to the waveform are sensitive to more intrinsic parameters such as component spins of the binaries and tidal deformations of binary components in the case of NS-containing binaries. All of these quantities can potentially be measured by running Bayesian PE on de-glitched data segments that are likely to contain a GW signal.

2.2.2 Inferring the population-level properties of CBCs

As described in the previous subsection, Bayesian parameter estimation of a single CBC detection will yield measurements of the compact binary system parameters. A single event, however, cannot tell us about the population-level properties of CBCs, which encode highly interesting physics. Of particular interest is the differential merger rate density of compact binaries per unit θ, which takes the following form:

$$\frac{dN_F}{d\theta}(\lambda) = N_F p(\theta|\lambda_F), \tag{2.13}$$

where N_F is the total number of CBCs that have occurred in the Universe during the detector's observation time, λ_F are hyper-parameters which either determine the population-level distribution of θ or some universal relation on θ for every CBC, and $p(\theta|\lambda_F)$ is the hierarchical prior distribution of θ conditioned on λ_F.

The particular form of the hierarchical prior $p(\theta|\lambda_F)$ will depend upon the physics we aim to extract. For example, when attempting to constrain uncertain astrophysical parameters such as the carbon-oxygen reaction rate of massive stars, the mass-distribution of BBHs can be modeled $p(m_1, m_2|\lambda_F)$ based on known formation scenarios and inferring λ_F from a collection of BBH observations will yield measurements of the reaction rate in question (Golomb et al., 2023). In this example, $p(\theta|\lambda)$ is a population-level probability distribution. On the other hand, while attempting to measure the nuclear EoS from BNS observations, the unique relationship imposed by an EoS model between BNS observables (such as masses and tidal deformabilities $m_1, m_2, \Lambda_1, \Lambda_2$), can be used to constrain parameters of the EoS model (λ_F) by choosing a deterministic hierarchical prior: $p(\Lambda_1, \Lambda_2, m_1, m_2|\lambda_F) = \delta(\Lambda_1 - f(m_1, \lambda_F))\delta(\Lambda_2 - f(m_2, \lambda_F))p(m_1, m_2)$.

In order to infer the population-level quantities/hyperparameters λ from a catalog of CBC observations, one can combine information from the Bayesian analyses of each individual chunk by treating every CBC event as an independent draw from some large population (Vitale et al., 2020;

Mandel et al., 2019; Thrane & Talbot, 2019). Imposing priors on the hyperparameters themselves and on the event-specific nuisance parameters that are agnostic of the population-level model characterized by the hyperparameters, it is possible to implement a hierarchical Bayesian inference of the hyperparameters. However, given that only triggers above the detection threshold are analyzed for population inference and that non-detections are likely *always* censored (Mandel et al., 2019; Farr et al., 2015), one needs to account for Malmquist biases by correctly normalizing the likelihoods after conditioning them on detection.

2.2.3 Astrophysical and Detectable populations

If we choose to analyze only the chunks that contain a trigger ranked above the detection threshold then our likelihood functions are implicitly conditioned on detection and need to be *re-normalized*:

$$p(d|g,\theta,\eta) = \frac{p(g|d)p(d|\theta,\eta)}{p(g|\theta,\eta)} \tag{2.14}$$

$$p(g|d) = \delta^{g,1}\Theta(\mathcal{F}(d)-\mathcal{F}_{th}) + \delta^{g,0}\Theta(-\mathcal{F}(d)+\mathcal{F}_{th}) \tag{2.15}$$

$$p(g|\theta,\eta) = \int ddp(g|d)p(d|\theta,\eta). \tag{2.16}$$

Note that our $p(g=1|\theta,F)$ is what is commonly labeled in literature as $P_{det}(\theta)$ and holds the same meaning of detection probability conditioned on parameters θ under the foreground hypotheses (Vitale et al., 2020; Mandel et al., 2019). We can use these likelihoods conditioned on detection to relate the true and detectable populations which are different due to Malmquist biases. If astrophysical and terrestrial sources comprise different populations, each characterized by some distribution function of parameters θ, we can represent both the astrophysical and detected populations as follows.

- Astrophysical and Terrestrial populations: $\frac{dN_F}{d\theta}(\theta|\lambda_F)$, $\frac{dN_B(\theta|\lambda_B)}{d\theta}$

- Fraction of detectable events:

$$N_{\eta,det} = \int d\theta p(g=1|\theta,\eta)\frac{dN_\eta}{d\theta} = N_\eta - N_{\eta,nodet} = N_\eta - \int d\theta p(g=0|\theta,\eta)\frac{dN_\eta}{d\theta} \quad (2.17)$$

- Detectable population:

$$\frac{dN_{\eta,det}}{d\theta} \quad = \quad p(g=1|\theta,\eta)\frac{dN_\eta}{d\theta} \quad (2.18)$$

We finally have all the pieces necessary to construct our joint hierarchical likelihood of GW data containing triggers that are realizations of mixed inhomogeneous Poisson processes. But before that let us take a quick look at mixtures of homogeneous Poisson processes to help transition the jump.

2.2.4 Aside: mixture of Poisson processes

If we have two independent Poisson-distributed random variables such as $n_1 \sim Poisson(N_{1,det}(\lambda_1))$, and $n_2 \sim Poisson(N_{2,det}(\lambda_2))$, with count parameters $N_{1,det}$, and $N_{2,det}$, then the probability distribution of their sum $N = n_1 + n_2$ can be expressed in the following way:

$$\begin{aligned} p(N|\lambda_1,\lambda_2) &= \sum_k^N p(n_1 = N-k|\lambda_1)p(n_2=k|\lambda_2) \\ &= \frac{e^{-N_{1,det}(\lambda)-N_{2,det}(\lambda_2)}}{N!}\sum_k N_{1,det}^{N-k}(\lambda_1)N_{2,det}^k(\lambda_2) \times \frac{N!}{k!(N-k)!} \end{aligned} \quad (2.19)$$

In other words, if we have two independent Poisson processes characterized by count parameters $N_{1,det}(\lambda_1)$ and $N_{2,det}(\lambda_2)$, that are contributing to the occurrence of some observable phenomena,

then the total number of events detected will have the following likelihood function:

$$p(N|\lambda_1,\lambda_2) \quad = \quad \frac{1}{N!}e^{-N_{1,det}(\lambda_1)-N_{2,det}(\lambda_2)}\left\{N_{1,det}(\lambda_1)+N_{2,det}(\lambda_2)\right\}^N \tag{2.20}$$

which when summed over N leads to 1. Now if the two Poisson processes were **inhomogeneous, i.e., if the count parameters were more preferential towards some values of data x than others** (Farr et al., 2015), we would have:

$$p(\{x\},N|\lambda_1,\lambda_2) = \frac{1}{N!}e^{-N_{1,det}(\lambda_1)-N_{2,det}(\lambda_2)}\prod_i^N\left\{\frac{dN_{1,det}}{dx}(x_i,\lambda_1)+\frac{dN_{2,det}}{dx}(x_i,\lambda_2)\right\} \tag{2.21}$$

Note again that $p(\{x\},N|\lambda_1,\lambda_2)$ is normalized when integrated over the x's and summed over N. In the next section, we will apply this to GW data by doing the following in the back of our heads:

$$x \quad \rightarrow \quad d \tag{2.22}$$

$$\frac{dN_{\eta,det}}{dd} \quad \rightarrow \quad \int d\theta p(d|g=1,\theta,\eta)\frac{dN_{\eta,det}}{d\theta}. \tag{2.23}$$

For a derivation of a similar expression to Eq. (2.21) from first principles, see Creighton (2017).

2.2.5 GW triggers as realizations of mixed inhomogeneous Poisson processes

One can construct the likelihood of obtaining GW data in N chunks, each with one trigger above threshold given some astrophysical and terrestrial populations characterized by hyper-parameters λ_F,λ_B, which takes the following form:

$$p(\vec{d},N|\lambda_F,\lambda_B,\{g=1\}) = \frac{1}{N!}e^{-N_{B,det}(\lambda_B)-N_{F,det}(\lambda_F)}$$

$$\times \prod_i^N \left\{ \int d\theta p(d_i|g=1,\theta,F) \frac{dN_{F,det}}{d\theta}(\theta,\lambda_F) + \int d\theta p(d_i|g=1,\theta,B) \frac{dN_{B,det}}{d\theta}(\theta,\lambda_B) \right\}. \tag{2.24}$$

From here on we will drop the condition $\{g=1\}$, which we will assume to be implicit since we only analyze "detected triggers". Substituting Eqs (2.18) and (2.15) into Eq. (2.24), we get:

$$\begin{aligned} p(\vec{d}, N|\lambda_F, \lambda_B) &= \frac{1}{N!} e^{-N_{B,det}(\lambda_B) - N_{F,det}(\lambda_F)} \\ &\times \prod_i^N \left\{ \int d\theta p(d_i|\theta,F) \frac{dN_F}{d\theta}(\theta,\lambda_F) + \int d\theta p(d_i|\theta,B) \frac{dN_B}{d\theta}(\theta,\lambda_B) \right\} p(g=1|d_i) \end{aligned} \tag{2.25}$$

Since $p(d_i|\theta,B)$ is independent of θ, we get:

$$\begin{aligned} p(\vec{d}, N|\lambda_F, \lambda_B) &= \frac{1}{N!} e^{-N_{B,det}(\lambda_B) - N_{F,det}(\lambda_F)} \\ &\prod_i^N \left\{ \int d\theta p(d_i|\theta,F) \frac{dN_F}{d\theta}(\theta,\lambda_F) + p(d_i|B) N_B(\lambda_B) \right\} p(g=1|d_i) \end{aligned} \tag{2.26}$$

Eqs (2.25), and (2.26) both represent the hierarchical likelihood of obtaining data corresponding to a collection of GW candidates conditioned on the population-level characteristics of the astrophysical and terrestrial sources. It can be used to conduct trigger classification, rate estimation, and the inference of population-level quantities and models given data from a collection of ranked CBC candidates.

2.3 TRIGGER CLASSIFICATION AND RATE ESTIMATION USING A FIXED POPULATION

As mentioned before, search pipelines provide estimates of FARs corresponding to each GW candidate. While FARs are useful in assessing the significance of candidate events, they do not account for merger rate information. To classify triggers into astrophysical and terrestrial categories, it is necessary to incorporate information on the merger rate density of CBCs, since depending on the rate of occurrence, different regions of the CBC parameter space will have different likelihood of obtaining astrophysical triggers at the same FAR. While it is necessary in general to self-consistently infer the merger rate densities with these probabilities of astrophysical origin, when classifying triggers for generating catalogs and low-latency alerts, it often suffices to use a fixed straw-person distribution of CBC parameters for each astrophysical population considered in the analysis (Farr et al., 2015; Kapadia et al., 2020).

For fixed hyper-parameters $\lambda_{F,0}, \lambda_{B,0}$ and likelihoods corresponding to the best matched template reported by the search pipeline, the integrals in Eq. (2.25) can be approximated in the following way:

$$\int d\theta p(d_k|g=1,\theta,\eta)\frac{dN_{\eta,det}}{d\theta} = p(x_k|g=1,\eta)N_{\eta,det}, \tag{2.27}$$

where x_i is the ranking data computed by the search pipeline, including the ranking statistics, the best-matched template t_k, SNR ρ_k etc., all for the kth trigger. The hierarchical likelihood conditioned on the detected counts therefore becomes:

$$p(\{x\},N|N_{F,det},N_{B,det}) \propto \frac{1}{N!}e^{-N_{F,det}-N_{B,det}}\prod_i \{p(x_i|g=1,F)N_{F,det}+p(x_i|g=1,B)N_{B,det}\}\,. \tag{2.28}$$

Imposing uninformative priors such as Jeffrey's prior on the detected counts (Farr et al., 2015;

Kapadia et al., 2020), one can obtain their posterior distribution $p(N_{F,det}, N_{B,det}|\{x\}, N)$. Using this count's posterior, it is possible to self-consistently compute the probability of astrophysical origin ($\mathrm{P_{astro}}$) of a particular GW candidate:

$$P_i(F|\{x,t\}) = \int dN_{F,det} dN_{B,det} p(N_{F,det}, N_{B,det}|\{x\}, N) \frac{p(x_i|g=1,F)N_{F,\mathrm{det}}}{\sum_\eta p(x_i|g=1,\eta)N_{\eta,\mathrm{det}}}. \tag{2.29}$$

Similarly, the astrophysical rate of CBCs can be estimated from the counts posterior and the spacetime volume to which the detector network is sensitive to:

$$p(\mathcal{R}_F|\{x\}) = \int dN_{B,det} p(N_F = \mathcal{R}_F \langle VT\rangle, N_{B,det}|\{x\}) \langle VT\rangle, \tag{2.30}$$

where $\langle VT\rangle$ is the sensitive hypervolume within which CBCs from the fixed astrophysical population are expected to be detectable. It can be computed from the probability of detection in the following way:

$$\langle VT\rangle = \int dz d\theta p(g=1|\theta,F) p(\theta|F) \frac{dV}{dz} \frac{T_{obs}}{1+z} \tag{2.31}$$

where $\frac{dV}{dz}$ is the differential co-moving volume. Note that a fixed cosmological model needs to be assumed to obtain the differential volume and the redshift as a function of the observable luminosity distance. Furthermore, throughout this analysis, all distributions conditioned on η are implicitly also conditioned on the fixed hyperparameters λ_η.

2.3.1 Multi-component classification and rate estimation

In rate estimation and trigger classification, it is often necessary to split up the astrophysical category into several sub-categories of CBCs (Kapadia et al., 2020), due to the following reasons. Firstly, during the generation of real-time alerts to GW candidates, the nature of the CBC is highly

valuable information since BNSs and NSBHs are far more likely to be electromagnetically bright than BBHs (Metzger, 2019; Perna et al., 2019). Furthermore, estimating the rate of CBC sub-categories individually is often necessary to explore the astrophysical implications of new CBC detections.

Sub-dividing the astrophysical category amounts to expressing the foreground population, and hence Eq. (2.27) as a mixture of multiple components, in the following way:

$$p(x_k|F, g = 1)N_{F,det} = \sum_{\alpha} p(x_k|g = 1, F_\alpha)N_{F_\alpha,det}, \tag{2.32}$$

where $\alpha \in \{BBH, BNS, NSBH\}$. The new multi-component counts' posterior then becomes:

$$p(\vec{N}_{F,det}, N_B|\{x\}) \propto p(\vec{N}_{F,\mathrm{det}}, N_{B,det})\frac{1}{N!}e^{-N_{B,det}-\sum_\alpha N_{F_\alpha,det}} \prod_i \left\{\sum_\alpha p(x_i|g = 1, F_\alpha)N_{F_\alpha,det} + p(x_i|g = 1, B)N_{B,det}\right\}, \tag{2.33}$$

which can be used to compute the category-specific probabilities of astrophysical origin ($P_i(F_\alpha|\{x\})$):

$$P_i(F_\alpha|\{x\}) = \int d\vec{N}_{F,det}dN_{B,det}\left\{p(\vec{N}_{F,det}, N_{B,det}|\{x\}, N)\times \frac{p(x_i|g = 1, F_\alpha)N_{F_\alpha,\mathrm{det}}}{\sum_\beta p(x_i|g = 1, F_\beta)N_{F_\alpha,\mathrm{det}} + p(x_i|g = 1, B)N_{B,det}}\right\}. \tag{2.34}$$

These classification probabilities are crucial to low-latency alerts since astronomers can optimize their follow-up campaigns based on say whether or not $P_i(BNS|\{x, t\}$ is high enough. The Rate of each subpopulation of CBCs can also be computed using the multi-component counts-posterior of Eq. (2.33) and the category-specific sensitive volumes:

$$\langle VT\rangle_\alpha = \int dzd\theta p(g = 1|\theta, F)p(\theta|F_\alpha)\frac{dV}{dz}\frac{T_{obs}}{1+z}. \tag{2.35}$$

2.4 SIMULTANEOUS INFERENCE OF POPULATION PROPERTIES

While the fixed population hierarchical likelihood suffices to implement trigger classification and rate estimation, the primary goal of hierarchical inference is to constrain the population level characteristics themselves from the data. To implement such an analysis, Bayesian inference of single-event parameters should be used to evaluate the integrals of Eq. (2.26), instead of the point estimates returned by search pipelines since the latter are inaccurate for several parameters and hence might lead to biases.

To incorporate information from the Bayesian parameter estimation of single event candidates into the integrals in Eq. (2.26), one can use Bayes theorem to express the foreground likelihood as a function of the parameter estimation posterior:

$$p(d_i|\theta, F) = \frac{p(\theta|d_i, F)}{p(\theta|F)} p(d_i|F), \tag{2.36}$$

to get:

$$\begin{aligned} p(\vec{d}, N|\lambda_F, \lambda_B) = N_B^N \frac{1}{N!} e^{-N_{B,det}(\lambda_B) - N_{F,det}(\lambda_F)} \\ \times \prod_i^N \left\{ 1 + \frac{p(d_i|F)}{p(d_i|B)} \frac{1}{N_B(\lambda_B)} \int d\theta \frac{p(\theta|d_i, F)}{p(\theta|F)} \frac{dN_F}{d\theta}(\lambda_F) \right\} p(g = 1|d_i). \end{aligned} \tag{2.37}$$

While in parameter estimation the single event posterior $p(\theta|d_i, F)$ is constructed from the stationary Gaussian noise assumption, we must account for the possibility of noise transients in the data. This is achieved by estimating the Bayes factor $\frac{p(d_i|F)}{p(d_i|B)}$ from search-pipeline reported quantities which account for non-stationarity and non-Gaussianity in the noise.

2.4.1 Using pipeline $P_{\rm astro}$

As mentioned in the previous section, the search pipelines report the probability of the astrophysical origin of each trigger which is computed using a fixed straw-person model for the astrophysical and background populations. Those can be expressed in the following way:

$$P_{astro,i}(\lambda_{F,0},\lambda_{B,0}) = \frac{p(d_i|F)p(F|\lambda_{F,0})}{p(d_i|F)p(F|\lambda_{F,0}) + p(d_i|B)p(B|\lambda_B)}. \tag{2.38}$$

$$\implies \frac{p(d_i|F)}{p(d_i|B)} = \frac{P_{astro,i}(\lambda_{F,0},\lambda_{B,0})}{1-P_{astro,i}(\lambda_{F,0},\lambda_{B,0})} \times \frac{p(B|\lambda_{B,0})}{p(F|\lambda_{F,0})}, \tag{2.39}$$

$$\text{where } \frac{p(B|\lambda_{B,0})}{p(F|\lambda_{F,0})} = \frac{N_B(\lambda_{B,0})}{N_F(\lambda_{F,0})}, \tag{2.40}$$

and $\lambda_{F,0}, \lambda_{B,0}$ are fixed hyper-parameters characterizing the straw person population models used to calculate pipeline $\rm P_{astro}$s. At this point, it is interesting to note that Eqs. (2.38)- (2.40) are not in contradiction with Eq. (2.29), and instead represent the exact same expression. This is because the ratio of likelihoods conditioned on detection times that of detectable counts is the same as the ratio of unconditioned likelihoods times that of detectable counts, such as:

$$\frac{p(d_i|F)}{p(d_i|B)}\frac{N_F(\lambda_{F,0})}{N_B(\lambda_{B,0})} = \frac{p(d_i|g=1,F)}{p(d_i|g=1,B)}\frac{N_{F,det}(\lambda_{F,0})}{N_{B,det}(\lambda_{B,0})} = \frac{P_{astro,i}(\lambda_{F,0},\lambda_{B,0})}{1-P_{astro,i}(\lambda_{F,0},\lambda_{B,0})}. \tag{2.41}$$

This can be derived from Eqs. (2.14) and (2.18) by marginalizing over θ. Another point to note is that the threshold used to compute the counts posterior for pipeline P_{astro}s can in general be different from the threshold used in constructing the hierarchical likelihood. However, due to Eq. (2.41), this does not affect our results.

Substituting Eqs. (2.39) and (2.40) into Eq. (2.37), we get:

$$p(\vec{d}, N|\lambda_F,\lambda_B) \propto N_B^N(\lambda_B)\frac{1}{N!}e^{-N_{B,det}(\lambda_B)-N_{F,det}(\lambda_F)}$$

$$\times \prod_i^N \left\{ 1 + \frac{N_B(\lambda_{B,0})}{N_B(\lambda_B)} \frac{P_{astro,i}(\lambda_{F,0}, \lambda_{B,0})}{1 - P_{astro,i}(\lambda_{F,0}, \lambda_{B,0})} \frac{1}{N_F(\lambda_{F,0})} \int d\theta \frac{p(\theta|d_i, F)}{p(\theta|F)} \frac{dN_F}{d\theta}(\lambda_F) \right\} p(g = 1|d_i) \tag{2.42}$$

Assuming the same background model as the pipeline:

$$N_{B,det}(\lambda_{B,0}) \approx N_B(\lambda_B), \tag{2.43}$$

we get:

$$\begin{aligned} p(\vec{d}, N|\lambda_F, \lambda_B) &\propto N_B^N(\lambda_B) \frac{1}{N!} e^{-N_{B,det}(\lambda_B) - N_{F,det}(\lambda_F)} \\ &\times \prod_i^N \left\{ 1 + \frac{P_{astro,i}(\lambda_{F,0}, \lambda_{B,0})}{1 - P_{astro,i}(\lambda_{F,0}, \lambda_{B,0})} \frac{1}{N_F(\lambda_{F,0})} \int d\theta \frac{p(\theta|d_i, F)}{p(\theta|F)} \frac{dN_F}{d\theta}(\lambda_F) \right\} p(g = 1|d_i) \end{aligned} \tag{2.44}$$

Using suitable priors on the hyper-parameters, one can construct their posterior distribution:

$$p(\lambda_F, \lambda_B|\vec{d}, N) \propto p(\lambda_F, \lambda_B) p(\vec{d}, N|\lambda_F, \lambda_B), \tag{2.45}$$

which can be sampled stochastically using Bayesian inference engines to obtain credible intervals for the physically informative quantities λ_F. In the case of population inference, the hyper-posterior can be used to update the pipeline P_{astro}'s into quantities that are now marginalized over the uncertainties regarding the CBC population, such as:

$$P_{astro,i}^{marg} = \int d\lambda_F d\lambda_B p(\lambda_F, \lambda_B|\vec{d}, N) \frac{K_i P_i(\lambda_F)}{1 + K_i P_i(\lambda_F)}, \tag{2.46}$$

where

$$K_i = \frac{P_{astro,i}(\lambda_{F,0}, \lambda_{B,0})}{1 - P_{astro,i}(\lambda_{F,0}, \lambda_{B,0})}, \tag{2.47}$$

and

$$P_i(\lambda_F) = \frac{1}{N_{F,det_0}(\lambda_{F,0})} \int d\theta \frac{p(\theta|d_i,F)}{p(\theta|F)} \frac{dN_F}{d\theta}(\lambda_F) \tag{2.48}$$

.

While updating P_{astro}'s makes sense only in the scenario where the hierarchical model being constrained is a population distribution of CBC parameters, Eq. (2.42) can be used to infer any hierarchical model such as those of the cosmological expansion, the nuclear EoS, and alternative theories of gravity. However, in several such cases it is often desirable to only include very low FAR candidates and restrict to inferring from a catalog of high significance events only that are almost certainly astrophysical.

2.5 RESTRICTING TO SIGNIFICANT CATALOGS

Choosing a very low FAR threshold implies $P_{astro,i} \sim 1$ for all events that are being hierarchically combined. In that limit, the hierarchical likelihood of Eq. (2.42) marginalized over the background hyperparameters λ_B take the following form:

$$p(\vec{d}, N|\lambda_F) \propto e^{-N_{F,det}(\lambda_F)} \prod_i^N \int d\theta \frac{p(\theta|d_i,F)}{p(\theta|F)} \frac{dN_F}{d\theta}(\lambda_F). \tag{2.49}$$

It is also often desirable to marginalize this likelihood over the total astrophysical count $N_F(\lambda_F)$. This is achieved by re-defining N_F to be a hyperparameter: $\lambda_F \equiv \{N_F, \lambda'_F\}$ and expressing the population distribution as $\frac{dN_F}{d\theta} = N_F p(\theta|\lambda'_F)$. Choosing a suitable prior on N_F, we get:

$$p(\vec{d}, N|\lambda'_F) \propto \int dN_F p(N_F) e^{-N_F \beta_{F,det}(\lambda'_F)} (N_F)^N \prod_i^N Z_i(\lambda'_F), \tag{2.50}$$

where,

$$\beta_{F,det}(\lambda'_F) = \frac{N_{F,det}(\lambda'_F)}{N_F} = \int d\theta p(g=1|\theta,F) p(\theta|\lambda'_F) \tag{2.51}$$

is the detection fraction as a function of population-level quantities, and

$$Z_i(\lambda'_F) = \int d\theta \frac{p(\theta|d_i, F)}{p(\theta|F)} p(\theta|\lambda'_F) \tag{2.52}$$

is proportional to the Bayesian evidence of the hierarchical model in data from the ith event, again as a function of population-level quantities. Choosing $p(N_F) = \frac{1}{N_F}$, we get the familiar expression of the hierarchical likelihood:

$$p(\vec{d}|\lambda) \propto \prod_i^N \frac{Z_i(\lambda)}{\beta_{det}(\lambda)}, \tag{2.53}$$

where we have removed the explicit dependence on N, dropped the subscript 'F', and renamed $\lambda' \to \lambda$. Imposing a suitable prior on λ, we can construct their posterior distribution and sample it stochastically to obtain credible intervals that encode information on the underlying physical assumptions of the hierarchical model.

2.6 IMPLEMENTATION USING AVAILABLE DATA PRODUCTS

To conduct trigger classification, rate estimation, and physical inference using the framework presented thus far, it is necessary to evaluate the hierarchical likelihoods of Eqs. (2.33), (2.42), and (2.53) from the data products generated during an observing run. The relevant ones are listed as follows.

1. Posterior samples of CBC parameters for each event above some threshold obtained using Bayesian parameter estimation.

2. The ranking data of all GW candidates above some threshold. For pipelines such as `GstLAL`, these products comprise the FAR, the log-likelihood ratio ranking statistic, and the observed SNR.

3. The ranking data of simulated signals injected into detector noise realizations.

Throughout the rest of this section, I will sketch how these are used to conduct hierarchical inference.

2.6.1 Trigger classification using GstLAL data products

For implementing the fixed population trigger classification framework described in Sec. 2.3.1, it is necessary to construct the counts posterior which relies on the marginalized likelihoods $p(x_k, t_K|\eta)$ for each trigger. In fact, it can be shown that up to a normalization constant, the counts posterior only depends on the ratio $p(x_k, t_K|F_\alpha)/p(x_k, t_K|B)$ (Kapadia et al., 2020). For the `GstLAL` search pipeline (Cannon et al., 2021; Messick et al., 2017; Sachdev et al., 2019), this ratio can be expressed in terms of search data products in the following way:

$$\frac{p(x_k, t_K|F_\alpha)}{p(x_k, t_K|B)} \propto A e^{\mathcal{L}_k} \frac{P(t_k, \rho_k|F_\alpha)}{\sum_\beta P(t_k, \rho_k|F_\alpha)}, \tag{2.54}$$

where $\mathcal{L}_k$ is the log-likelihood ratio ranking statistics computed by `GstLAL` (Cannon et al., 2015), and $P(t_k, \rho_k|F_\alpha)$ are semi analytic template weights. The constant A can be estimated from `GstLAL`'s background distribution $p(\mathcal{L}_k|B)$. Similarly, the template weights as a function of SNR, for each template in the bank, can be computed assuming some noise model (Fong, 2018). See chapter and appendix for further details regarding how exactly these quantities are computed. Once the counts posterior is constructed, it can be used to compute the multi-component P_{astro}'s of triggers, using Eq. (2.34).

2.6.2 Using posterior samples to compute the evidence

Unlike the fixed population analysis, inferring a hierarchical model necessitates the use of single-event PE results. As shown in the previous section, to infer a hierarchical model, it is necessary to

compute its evidence in the data from individual events, as a function of population-level quantities. To obtain the values of this evidence it is necessary to evaluate integrals of the form:

$$\langle f \rangle_i = \int d\theta p(\theta|d_i, F) \frac{f(\theta)}{p(\theta|F)}. \tag{2.55}$$

In most scenarios, the hierarchical model being inferred will only constrain a subset of CBC observables, in which case, the posterior $p(\theta|d_i, F)$ would need to be marginalized over the other parameters. These marginalization integrals as well as the integrals of Eq. (2.55) are often high dimensional due to which they are often implemented using a Monte Carlo sum over posterior samples:

$$\langle f \rangle_i \approx \frac{1}{N_{samples}} \sum_{\theta_j \sim p(\theta|d_i,F)} \frac{f(\theta_j)}{p(\theta_j|F)} \tag{2.56}$$

This reduces the computational cost of hierarchical inference since Bayesian PE for individual events already procures the posterior samples as part of the catalogs. However, in the context of deterministic hierarchical models such as the EoS, wherein $f(\theta)$ contains delta functions, the sum over posterior samples is not implementable. In such a scenario, a viable alternative is to approximate the marginalized posterior as a numerical function of the parameters sensitive to the hierarchical model from their samples, by means of kernel density estimation. The estimated density can be integrated numerically by means of techniques such as the trapezoidal rule:

$$\langle f \rangle_i = \int d\theta \hat{p}_{KDE,i}(\theta) \frac{f(\theta)}{p(\theta|F)}, \tag{2.57}$$

where $\hat{p}_{KDE,i}$ is the kernel density estimate of the posterior corresponding to the ith event, obtained from its samples $\{\theta_j \sim p(\theta|d_i, F)\}$.

2.6.3 Estimating the detection fraction and sensitive volumes using ranked simulated signals

In addition to individual event evidences, hierarchical inference and rate estimation necessitates, in some form, the computation detector sensitivity to signals from a population of CBCs occurring in the universe. In particular, for estimating both the sensitive hyper-volumes to a fixed population and the number of detectable events as a function of population-level quantities, it is necessary to compute integrals of the form:

$$\langle f \rangle_{det} = \int d\theta p(g = 1|\theta, F) f(\theta). \tag{2.58}$$

Given that direct evaluation $p(g = 1|\theta, F)$ involves an integral over all possible time-series data, it needs to be estimated through alternate means. For example, one can use a ranked set of simulated signals that survive the detection threshold to estimate $p(g = 1|\theta, F)$. This is achieved by first using Bayes theorem to obtain:

$$p(g = 1|\theta, F) = \frac{p(\theta|g = 1, draw, F)p(g = 1|draw, F)}{p(\theta|draw, F)}, \tag{2.59}$$

where $p(\theta|draw, F)$ is the fiducial population distribution used to draw the parameters of the simulated signals, and $p(g = 1|draw, F)$ is the fraction of simulated events that are recovered above the detection threshold. The parameters of the simulated signals that are found above the threshold therefore serve as samples drawn from $p(\theta|g = 1, draw, F)$. Under these considerations, for non-deterministic hierarchical models such as the population level distribution of CBC parameters, the integral of Eq. (2.58) can be approximated as a Monte Carlo sum over the parameters of simulated signals detected above threshold:

$$\langle f \rangle_{det} \approx \frac{1}{M_{draw}} \sum_{\theta_i \sim \mathrm{det}} \frac{f(\theta_i)}{p(\theta_i|draw)}. \tag{2.60}$$

While this implementation suffices for population inference and rate estimation, as before, for deterministic hierarchical models, the Monte Carlo sum over discrete samples is not implementable. For computing the integral over detection probability in such scenarios, $p(\theta|g=1, draw, F)$ can be approximated as a numerical function of θ by again using kernel density estimation on the parameters of detected simulations. $\langle f \rangle_{det}$ can then be estimated using numerical integration techniques such as the trapezoidal rule on:

$$\langle f \rangle_{det} \approx \int d\theta \hat{p}_{KDE,det}(\theta) \frac{f(\theta)}{p(\theta|draw)}, \tag{2.61}$$

where $\hat{p}_{KDE,det}$ is the kernel density estimate of $p(\theta|g=1, F, draw)$, obtained from its samples $\{\theta_j \sim det\}$.

2.6.4 Marginalizing over Monte Carlo uncertainties

When estimating integrals using Monte Carlo sums over samples, one must account for the effective number of independent samples that contribute to the integral to ensure that the sums are converged. Large Monte Carlo uncertainties in these integrals can cause biases in the inferred rates and hyper-parameters (Essick & Farr, 2022; Farr, 2019). It is therefore necessary to also compute the effective number of independent samples (n_{eff}) whenever a Monte Carlo sum is implemented and penalize hyper-parameter values that correspond to low values of n_{eff}.

To compute n_{eff} for Monte Carlo sums over posterior samples and the parameters of detectable simulations, the variance of the sum in question must first be evaluated such as:

$$\left\langle (f - \langle f \rangle)^2 \right\rangle_i = \frac{1}{N^2_{samples}} \sum_{\theta_j \sim p(\theta|d_i,F)} \left(\frac{f(\theta_j)}{p(\theta_j|F)} \right)^2 - \frac{\langle f \rangle^2_i}{N_{samples}}, \tag{2.62}$$

and

$$\left\langle (f - \langle f \rangle)^2 \right\rangle_{det} = \frac{1}{M_{draw}^2} \sum_{\theta_j \sim det} \left(\frac{f(\theta_j)}{p(\theta_j | draw, F)} \right)^2 - \frac{\langle f \rangle_{det}^2}{M_{draw}} \tag{2.63}$$

.

For reliably convergent Monte Carlo sums it is expected that $\left\langle (f - \langle f \rangle)^2 \right\rangle_i << \langle f \rangle_i^2$, and $\left\langle (f - \langle f \rangle)^2 \right\rangle_{det} << \langle f \rangle_{det}^2$, which both imply $n_{eff}^2 >> 1$. However, less stringent conditions can be derived by marginalizing the hyper-posterior over these Monte Carlo uncertainties.

According to the central limit theorem, for a large number of samples, realizations of Monte Carlo sums will follow a Normal distribution with mean $\langle f \rangle$ and variance $\left\langle (f - \langle f \rangle)^2 \right\rangle$. This distribution can be used to marginalize the hyper-posterior over Monte Carlo uncertainties (Farr, 2019). In order to find the local peak of the marginalized hyperposterior, it is necessary to impose conditions on $n_{eff} = \frac{\sqrt{\left\langle (f - \langle f \rangle)^2 \right\rangle}}{\langle f \rangle}$. These conditions are dependent on the nature of the hierarchical model being inferred and will be discussed in the corresponding chapters wherein they are used explicitly.

2.6.5 Sampling the hyper-posterior

Once the hyper-posterior is constructed, it can be used to compute credible intervals on the physically informative population-level quantities λ_F. The Bayesian credible interval I_α of confidence level $(1 - \alpha)\%$, on λ_F, is given by:

$$\int_{\lambda_F \in I_\alpha} d\lambda_F p(\lambda_F | \vec{d}) = (1 - \alpha) \tag{2.64}$$

where $\alpha \in [0, 1]$. These confidence intervals are not unique and can be computed using a variety of methods that use λ_F samples drawn from the hyper-posterior.

In the equal tails method, the samples are used to compute the empirical CDF of λ_F which is then used to compute $p \times 100\%$ quantiles q_p, with $p \in [0, 1]$. For example, given sorted samples

of $\{y\}$, one can find q_p such that $y_i < q_p$ for all $i < p \times N_{samples}$. The equal tail credible interval is then defined as $I_\alpha = [q_{\alpha/2.}, q_{1-\alpha/2}]$. By construction, the equal tail interval will include the median. While the equal tail intervals suffice for symmetric unimodial distributions, it is not a good representative of posterior support in the case of skewed and/or truncated distributions, for which, it will often exclude the mode/ the region of highest posterior support.

In such scenarios, the highest posterior density interval is a much better representation of posterior support. It is the smallest interval that contains $(1-\alpha) \times 100\%$ of the posterior mass and hence includes the mode by construction. It is constructed by first computing the number of samples that will be contained in the interval, namely $n_\alpha = (1-\alpha) \times N_{samples}$. Then, from the sorted set of samples, all possible interval widths that contain n_α samples are computed $\{w_\alpha^i = x_{i+n_\alpha} - x_i | i + n_\alpha < N_{samples}\}$. Then sample index i_0 corresponding to the shortest such width: $\min w_\alpha^i = w_\alpha^{i_0}$ is used to compute the highest posterior density interval: $I_\alpha = [x_{i_0}, x_{i_0+n_\alpha}]$.

Both these methods rely on posterior samples to numerically estimate credible intervals of the hyperparameters. Therefore, computing them would require one to first sample the hyper-posterior distribution of population-level quantities. For low-dimensional models and unimodal posteriors, MCMC implementations such as Metropolis-Hastings or Gibbs sampling may suffice. While multi-modal distributions necessitate the use of nested sampling (Skilling, 2006), higher dimensional models might require the use of Hamiltonian Monte Carlo (HMC) sampling which scales much more efficiently than MCMC and nested sampling with the dimensionality of the space of hyperparameters (Creutz, 1988; Neal, 1993, 2011; Homan & Gelman, 2014). To sample the hyperposteriors relevant to my thesis work, I mostly use the `emcee` package (Goodman & Weare, 2010; Foreman-Mackey et al., 2013) and the `pymc` package (Salvatier et al., 2016; Oriol et al., 2023) which implement MCMC and HMC sampling respectively.

2.7 APPLICATIONS THROUGHOUT THE REST OF THE THESIS

The frameworks derived in this chapter constitute the cornerstone upon which my thesis work relies to implement trigger classification and hyperparameter inference. The exact applications are summarized as follows.

In chapter 3, I develop novel implementations of the trigger classification and rate estimation frameworks discussed in Sec. 2.3.1. I summarize how it has been used on GWTC-3 data for categorizing GstLAL CBC triggers and estimating their astrophysical rate. I also developed a new and accurate low-latency classification algorithm which is currently being used by the LVK to compute multi-component P_{astro}'s for GstLAL candidates in real-time, with the resulting estimates comprising an integral part of the low-latency alerts corresponding to such candidates.

In chapters 4 and 5, building upon previous works, I construct a new data-driven population model for CBC parameters and use the significant-catalog population inference framework of Eq. (2.49) implemented through Eqs (2.56) and (2.60), to constrain correlations in the astrophysical BBH population distribution, and conduct astrophysics agnostic cosmological inference using GW data alone.

In chapters 6 and 7, I use the significant-catalog population inference framework of Eq. (2.53) in conjunction with the KDE-based posterior and selection function approximation schemes of Eqs. (2.57), and (2.61), to implement scalable inference of the NS EoS and alternative theories of gravity.

Part II

Categorizing gravitational wave sources

Chapter 3

Trigger counting and categorization

Man is fond of counting his troubles, but he does not count his joys. If he counted them up as he ought to, he would see that every lot has enough happiness provided for it.

Fyodor Dostoevsky

The counting and classification of GW candidates are crucial for constructing transient catalogs, determining the astrophysical implications of new CBC detections, and optimizing the search for multi-messenger counterparts to GW events. In addition to significant candidates, marginal triggers contribute non-negligibly to the astrophysical implications of CBC catalogs while also having a finite possibility of being multi-messenger events. Counting and categorizing such candidates rely not only on the ranking data provided by the search pipelines but also on prior expectations regarding the rate and population properties of various CBC classes. Uncertainties in current measurements of CBC rates therefore necessitate a self-consistent framework of Bayesian rate estimation which automatically determines the probability of astrophysical origin of the CBC triggers whose ranking data is used to obtain rate measurements (Farr et al., 2015; Kapadia et al., 2020).

Extracting the count of astrophysical events from a collection of GW candidates often strongly relies on the detection threshold used to truncate the dataset, particularly in the scenario where a stringent threshold value is chosen. In such situations. arbitrariness in the choice of the threshold value can migrate to the astrophysical implications of the estimated merger rates as well as to the classification probabilities that are determined by these rates. This is particularly relevant for rarely observed CBC systems such as BNSs and NSBHs, whose rates and classification probabilities can vary wildly depending on the inclusion and exclusion of one or two marginal events in the

analysis. Hence, to obtain reliable rate estimates for such CBC categories it is necessary to extend the counting analysis up to very low ranking statistic thresholds.

The methods proposed by Farr et al. (2015) and Kapadia et al. (2020) achieve this by modeling the occurrence of CBCs and noise transients as realizations of mixed inhomogeneous Poisson processes. By computing the posterior distribution of the detected count of triggers from their ranking data, and marginalizing over the uncertainty regarding each trigger's origin, they implement self-consistent rate estimation and trigger classification that is robust against variations of the low ranking statistic threshold used throughout the analyses. In this chapter, I discuss several improvements to these methods that I have contributed to developing, in the context of rate estimation, as well as offline and online trigger classification.

3.1 CATALOG CONSTRUCTION AND RATE ESTIMATION

At the end of an observing run, search data from various pipelines are aggregated to identify the CBC candidates among them and assess their system of origin. Triggers that are more than 50% likely to be of astrophysical origin are included in the cumulative gravitational wave transient catalogs. Once constructed, the catalogs are used for the inference of a wide range of physical models, which results in new discoveries in several fields of modern physics.

To calculate the probability of astrophysical origin, it is necessary to use thresholds that allow for higher FAR candidates than the ones that are almost certainly known to be astrophysical. In fact, these marginal candidates are the ones for which $\mathrm{P}_{\mathrm{astro}}$ plays a significant role in determining their inclusion into catalogs. However, given that $\mathrm{P}_{\mathrm{astro}}$ values rely on the expected number of counts above threshold, a low enough threshold must be chosen such that the estimated $\mathrm{P}_{\mathrm{astro}}$ values of these marginal candidates are invariant under a few orders of magnitude changes in the said threshold. This necessitates the use of the framework described in Sec. 2.3.1 to be implemented on a dataset of triggers the bulk of which are likely noise transients.

Similarly, for rarely observed CBCs such as BNSs and NSBHs, it is often necessary to estimate their rates using a broad, fixed population distribution to examine the astrophysical impact of new detections. Given the small number of existing detections, it is often necessary to account for the contributions of marginal candidates to the estimated merger rate and hence to apply the rate estimation framework derived in Sec. 2.3.1 on the extended dataset the bulk of which is dominated by noise.

To obtain accurate estimates of trigger classification and merger rates, it is desirable to use all of the information available at the end of an observing run including trigger data from previous observing runs. Prior to combining these fixed-population analyses across observations, it is necessary to first investigate a single run and estimate its sensitivity to CBC signals, since different observing runs have significantly different sensitivities.

3.1.1 Analyzing a single observing run and combining multiple ones

Given ranking data available at the end of an observing run, it is possible to construct the multi-component counts posterior of Eq. (2.33). For simplicity, I will use the change in notation: $N_{F_\alpha,det} \to \Lambda^r_{1\alpha}$, and $N_{B,det} \to \Lambda^r_0$, where r stands for the observing run in question. Furthermore, for a very low threshold, i.e. in scenarios where the number of background triggers is several times higher than that of astrophysical signals, it can be shown that the counts posterior is an extremely peaked function of Λ_0 about the total number of triggers N_r (Kapadia et al., 2020). In such a scenario, the posterior distribution of the foreground counts marginalized over Λ_0 takes the following form:

$$p(\vec{\Lambda}^r_1|\vec{x}) \propto p(\vec{\Lambda}^r_1) e^{-\sum_\alpha \Lambda^r_{1\alpha}} \prod_i^{N^r} \left(1 + K_\alpha(x_i) \sum_\beta \Lambda_{1,\beta}\right), \tag{3.1}$$

where,

$$K_\alpha(x_i) = \frac{p(x_i|F_\alpha)}{p(x_i|B)} \tag{3.2}$$

is the category-specific Bayesfactor for the ith trigger as a function of its ranking statistic and other associated trigger data.

Since the counts estimated from this posterior determine both source classification and rate estimation, it is desirable to aggregate data from past observing runs and obtain well-constrained count distributions. In order to do so, one must account for the change in sensitivity of the detectors between the observing runs. However, one must be careful in implementing the sensitivity calculation described in Eq. (2.35) by means of Eq. (2.60), since the thresholds chosen in this analysis go deep into the noise-dominated region of the dataset. In particular, it is not straightforward to map a given trigger to an injection and count that trigger as a recovered simulated signal (Farr et al., 2015; Kapadia et al., 2020).

Luckily, for a given observing run, it is possible to estimate the sensitive hyper-volumes in a manner that is self-consistent with the low ranking-statistic thresholds used in the analysis by using the counts posterior. For a trigger (γ_i) found within a short time window of an injection, its contribution to the overall count of triggers can be estimated in the following way:

$$\delta\Lambda_{1\alpha}^{r,\gamma_j} = \frac{N^r \left\langle \Lambda_{1\alpha}^r \right\rangle - \sum_\beta cov(\Lambda_{1\alpha}^r, \Lambda_{1\beta}^r) K_\beta(x_{\gamma_j})}{N^r + \sum_\beta K_\beta(x_{\gamma_j}) \left\langle \Lambda_{1\beta}^r \right\rangle} \tag{3.3}$$

where $cov(\Lambda_{1\alpha}^r, \Lambda_{1\beta}^r)$ and $\left\langle \Lambda_{1\beta}^r \right\rangle$ represent covariance and means of the counts with respect to the posterior distribution of Eq. (3.1). Using this count increment for all found injections in place of 1, the implementation of Eq. (2.35) through Eq. (2.60) yields the following $\langle VT \rangle$ estimate:

$$\langle VT \rangle_\alpha^r = \frac{\sum_{\gamma_j \in found} \delta\Lambda_{1\alpha}^{r,\gamma_j} \frac{p(\theta_{\gamma_j}|F_\alpha)}{p(\theta_{\gamma_j}|draw)}}{\sum_{\gamma_i \in all} \frac{p(\theta_{\gamma_i}|F_\alpha)}{p(\theta_{\gamma_i}|draw)}} \langle VT \rangle_\alpha^{inj} \tag{3.4}$$

where, $p(\theta|F_\alpha)$ is the fixed population-model for the α'th category of CBCs used throughout the analysis, $p(\theta|draw)$ is a fiducial population distribution used to draw the parameters of the simulated signals/injections, and $\langle VT \rangle_\alpha^{inj}$ is the total spacetime volume throughout which the injections

were distributed. See Eqs. (8) and (9) of Tiwari (2018) for details regarding how $\langle VT \rangle_{\alpha}^{inj}$ is calculated from $p(\theta|draw)$. Once the sensitive hyper-volume corresponding to each observing run is computed using the corresponding counts posterior, they can be used to combine count posteriors across observing runs in the following way.

While the detector sensitivity and the expected counts above the threshold corresponding to each category vary from one observing run to the other, the astrophysical merger rate is independent of any detector configuration. In other words, the ratio of the counts to the sensitive hyper-volume is expected to be the same for all observing runs, such as:

$$\mathcal{R}_{\alpha} = \frac{\Lambda_{1\alpha}^{r}}{\langle VT \rangle_{\alpha}^{r}} \quad \text{for all } r, \tag{3.5}$$

where $\mathcal{R}_{\alpha}$ is the astrophysical merger rate of the α'th category of CBCs. The astrophysical merger rate is further equal to the ratio of the combined counts $\Lambda_{1\alpha}$ and the sum of VTs over observing runs ($\sum_s \langle VT \rangle_{\alpha}^{r}$). Under these considerations, the posterior distribution of the combined counts can be expressed in the following way:

$$p(\vec{\Lambda}_1|\{x\}) = p(\vec{\Lambda}_1) e^{-\sum_{\alpha} \Lambda_{1\alpha}} \prod_{r} \prod_{i=1}^{N^r} \left(1 + \sum_{\alpha} k_{\alpha}^{r}(x_i) \Lambda_{1\alpha} \right). \tag{3.6}$$

where,

$$k_{\alpha}^{r}(x_i) = \frac{p(x_i|F_{\alpha})}{p(x_i|B)} \times \frac{\langle VT \rangle_{\alpha}^{r}}{\sum_s \langle VT \rangle_{\alpha}^{s}} \times \frac{1}{\sum_s N^s}, \tag{3.7}$$

is the reduced Bayes factor of the i'th trigger in the r'th observing run. Once this combined counts posterior is constructed it can be used to implement trigger classification and rate estimation for catalogs such as GWTC-3 and GWTC-2.1 (Abbott et al., 2023b, 2021d).

3.1.2 Bayes factors for `GstLAL` triggers

Even though the combined counts posterior can in principle be used on trigger data from any search pipeline, throughout the rest of this section, I will only focus on `GstLAL` triggers. For GstLAL, the ranking statistics is the log of the likelihood ratio itself which makes Bayes factor computation relatively straightforward. As mentioned in Sec. 2.6.1, and derived in appendix A.1, the category-specific Bayes factors for gstlal triggers take the following form:

$$K_\alpha(x_i) \propto \frac{p(\mathcal{L}_i|F)}{p(\mathcal{L}_i|B)} P(t_i, \rho_i|F_\alpha) \propto A e^{\mathcal{L}_i} P(t_i, \rho_i|F_\alpha). \tag{3.8}$$

Here, while A can be calculated straightforwardly from `GstLAL`'s background distribution, the category-specific template weights $P(t_i, \rho|F_\alpha)$ are harder to obtain. Upto GWTC-1, these template weights were obtained by injecting simulated signals from $p(\theta|F_\alpha)$ and histogramming the recovered templates. However, the sheer size of the O3 template bank which comprised more than a million templates, made such an approach inaccurate since there were not enough injections available per template.

To get around this problem, I used the semi-analytic formalism developed by Fong (2018) to compute similar quantities, as a proxy for these template weights. In the approach developed by Fong (2018), the template weights can be decomposed in the following way:

$$P(t_k, \rho_k|F_\alpha) = \sum_j P(\text{candidate in } t_k|\text{signal in } t_j, \rho) p(\theta_j, \rho|F_\alpha) V_j, \tag{3.9}$$

where $P(\text{candidate in } t_k|\text{signal in } t_j, \rho)$ is the probability of obtaining a candidate matched with template t_k given a true signal exactly matched with template t_j at SNR ρ, θ'_j are the CBC parameters corresponding to t_j, $p(\theta', \rho|F_\alpha)$ is obtainable from $p(\theta|F_\alpha)$, and V_j the volume element in parameter space associated with the template t_j. For the template bank used in O3, V_j can be computed by means of a Voronoi decomposition of the bank (Voronoi, 1908; Fortune, 1997; Fong,

2018).

To estimate $P(\text{candidate in } t_k|\text{signal in } t_j, \rho)$, one needs to model the probability distribution $p_N(n)$ of the random noise n that causes a true signal in t_j to be recovered as a candidate in t_k. Assuming the recovered signal to be $\rho_{obs} t_k = \rho t_j + n$, the probability of template migration takes the following form:

$$P(\text{candidate in } t_k|\text{signal in } t_j, \rho) = \int d\rho_{obs} p_N(\rho_{obs} t_k - \rho t_j) \rho_{obs}^{M-1}, \tag{3.10}$$

where M is the dimensionality of the parameter space corresponding to the template bank. For a bank spanned by masses and the z components of the spins, we have $M = 5$, where the extra dimension corresponds to the signal amplitude/SNR. For further details of this computation see Fong (2018).

The template weights are usually constructed from a fixed straw person representative of the population distribution of CBC parameters such as uniform in log masses and uniform in z-component spins, and an approximate noise distribution model such as the Gaussian distribution. With BH and NS binary components distinguishable by the NS maximum mass which is taken to be $3M_\odot$ and a maximum spin magnitude of NS components which is taken to be 0.5, the template weights corresponding to each CBC category can be pre-computed for every template, as a function of SNR, and stored.

At this point, I note that the calculation of template-weights summarized here, and detailed in Fong (2018) employs several approximations that might fall short in upcoming observing runs as detector sensitivity and the number of triggers continue to grow. However, it is an improvement over the injection-based histograms used up till GWTC-1. In particular, due to a lack of injections in certain regions of the BBH parameter space, the injection-based approach led to a factor of 2 underestimation of the BBH counts in O3, while the semianalytic method yields correct results (figure 3.1).

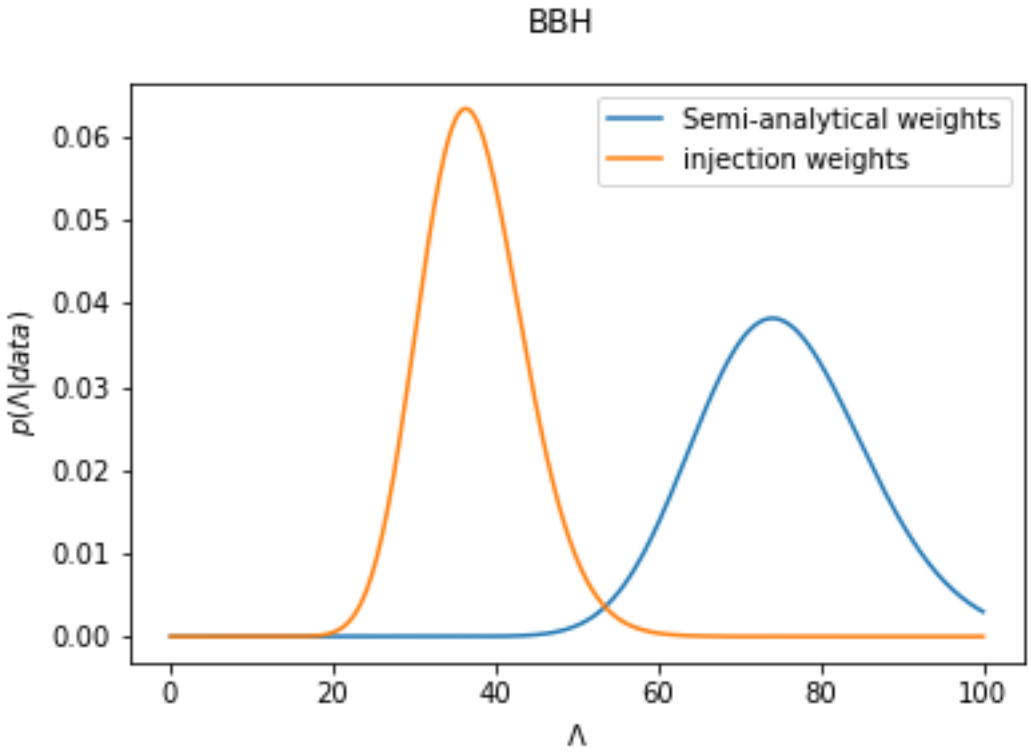


Figure 3.1: BBH counts posteriors with semianalytic vs injection based template weights from preliminary O3 data. Given that there are at least 69 BBHs with FAR less than 1 per year, it is clear that the injection-based weights are leading to even low FAR candidates being classified as terrestrial.

As fractional uncertainties continue to decrease with the aggregation of new data, more realistic calculations for the template weights than the ones presented in this section need to be implemented. In Sec. 3.2.2.3, I discuss a modified version of the template weights that we have used for trigger classification during the fourth observing run (O4) of the LVK. Further generalizations and improvements are part of an ongoing investigation, the plans for which are summarized in chapter 9.

3.1.3 Probability of astrophysical origin

Once the combined counts posterior is constructed and its moments evaluated using numerical integration techniques such as Gaussian quadrature, the probability of astrophysical origin for each

trigger can be computed using the following expression:

$$P_{i,r}(F_\alpha|\{x\}) = \frac{\langle\Lambda_{1\alpha}\rangle\, k^r_\alpha(x_i)}{1+\sum_\beta \langle\Lambda_{1\beta}\rangle\, k^r_\beta(x_i)}, \tag{3.11}$$

where $P_{i,r}(F_\alpha|\{x\})$ is the probability that the ith trigger in the rth observing run was caused by a CBC of category α, and $\langle\Lambda_{1\alpha}\rangle$ is the mean of $\Lambda_{1\alpha}$ computed using the posterior distribution of Eq. (3.6). This expression is obtained when the integral in Eq. (2.34) is evaluated using the counts posterior of Eq. (3.6). The combined probability of astrophysical origin for the i'th trigger can be obtained by summing over individual source categories: $P_{astro,i} = \sum_\alpha P_i(F_\alpha|\{x\})$

By combining data from O1 through O3, I computed the category-spcific probability of astrophysical origin of all `GstLAL` triggers which were published as part of Abbott et al. (2023b) and Abbott et al. (2021d). To compute the template weights I used a uniform in log population distribution for component masses and a uniform distribution for component spins with the maximum mass and dimensionless spin magnitudes of NSs taken to be $3M_\odot$ and 0.05 respectively. See figure 3.2 for a visual summary of the computed $\mathrm{P_{astro}}$ values.

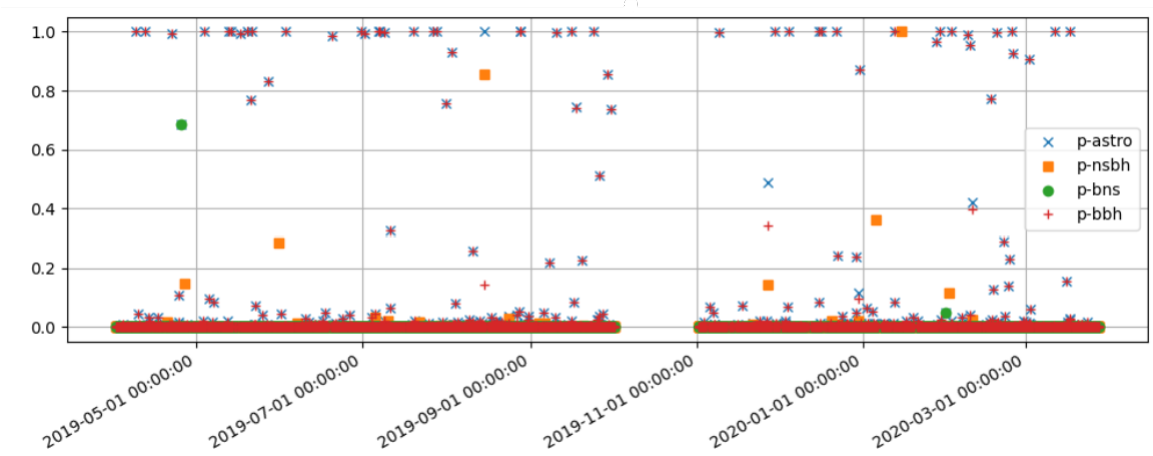


Figure 3.2: Category specific probability of astrophysical origin for all O3 `GstLAL` triggers found above a FAR threshold of 1 per hour vs the date and time of the trigger. The points labeled "p-astro" represent the sum of "p-bns", "p-bbh", and "p-nsbh". The $\mathrm{P_{astro}}$ values of the triggers that are above 0.5 are published as part of Abbott et al. (2021d) and Abbott et al. (2023b).

It is interesting to note that the $\mathrm{P_{astro}}$ values of the strongly ($P_{astro} \sim 1$) and marginally ($P_{astro} \in (0.5, 1)$) significant candidates are immune to the choice of the FAR threshold used throughout the analysis (see figure 3.3). This is indicative of the fact that the analysis framework can extract significant and marginal candidates correctly from deep into the noise-dominated region of the dataset and is stable against variations in the amount of background data incorporated in the analysis.

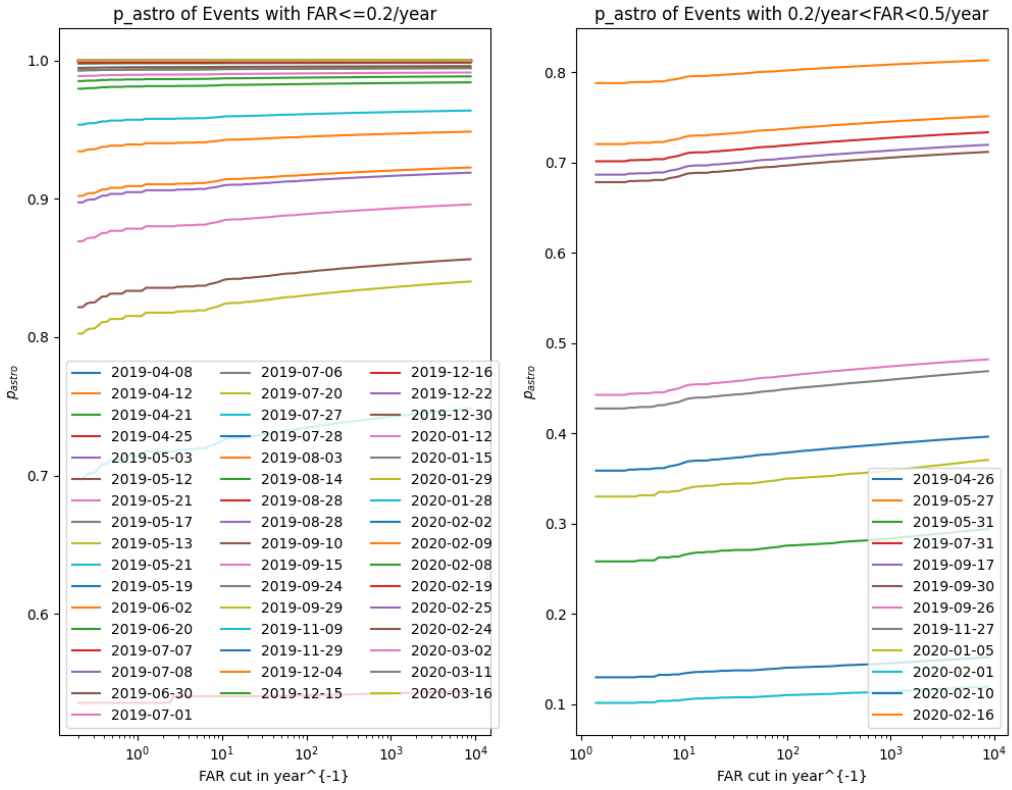


Figure 3.3: Variation of $\mathrm{P_{astro}}$ of significant and marginal candidates with the FAR threshold used in the analysis.

The list of $\mathrm{P_{astro}}$ values computed in this analysis serves as a guide for including events in the catalog and for providing a population-informed assessment of how likely we believe these

triggers to be of astrophysical origin. A $\mathrm{P_{astro}}$ threshold of 0.5 is used by the LVK collaboration for deciding whether or not to conduct Bayesian PE studies on a particular data segment since PE is computationally costly and it is undesirable to waste computational resources on GW candidates that are less than 50% likely to be astrophysical. Downstream studies that conduct hierarchical inference of physical models using individual event data from the catalog can also use $\mathrm{P_{astro}}$ to decide whether or not the inclusion of a particular event is relevant to the model in question.

3.1.4 Rate estimation

As mentioned before, the fixed population counts posterior can be used to compute the astrophysical merger rate of CBCs from a dataset containing both significant and marginal CBC candidates. It is often necessary to do so for rarely observed CBC categories such as BNS and NSBH, due to the following reasons. While estimating the rate of a Poisson distribution from a sparse dataset, the inclusion or exclusion of a single event can significantly change the inferred credible intervals. Therefore, to avoid making astrophysical statements that rely strongly on arbitrarily chosen FAR thresholds, it is necessary to include triggers up to high FARs and marginalize over the uncertainties regarding their origin (Farr et al., 2015; Kapadia et al., 2020).

The posterior distribution of the astrophysical merger rate of CBCs belonging to category α can be obtained from the combined counts posterior of Eq. (3.6) and the combined sensitive hypervolume of all observing runs:

$$p(\mathcal{R}_\alpha|\{x\}) = \int SdSp(\Lambda_{1\alpha} = \mathcal{R}S|\{x\})p(S|\left\langle VT\right\rangle_\alpha), \tag{3.12}$$

where

$$p(S|\left\langle VT\right\rangle_\alpha) = \frac{1}{\sqrt{2\pi}S\sigma}e^{-\frac{1}{2}\left(\frac{\ln S - \ln\left\langle VT\right\rangle_\alpha}{\sigma}\right)^2} \tag{3.13}$$

is a distribution quantifying Monte Carlo and systematic uncertainties in VT estimation,

$$\langle VT\rangle_\alpha = \frac{\sum_{\gamma_j \in found} \delta\Lambda_{1\alpha}^{\gamma_j} \frac{p(\theta_{\gamma_j}|\bar{F}_\alpha)}{p(\theta_{\gamma_j}|draw)}}{\sum_{\gamma_i \in all} \frac{p(\theta_{\gamma_i}|\bar{F}_\alpha)}{p(\theta_{\gamma_i}|draw)}} \langle VT\rangle_\alpha^{inj} \tag{3.14}$$

is the total sensitive hypervolume of all the observing runs with $\delta\Lambda_{1\alpha}^{\gamma_j}$ computed from the combined

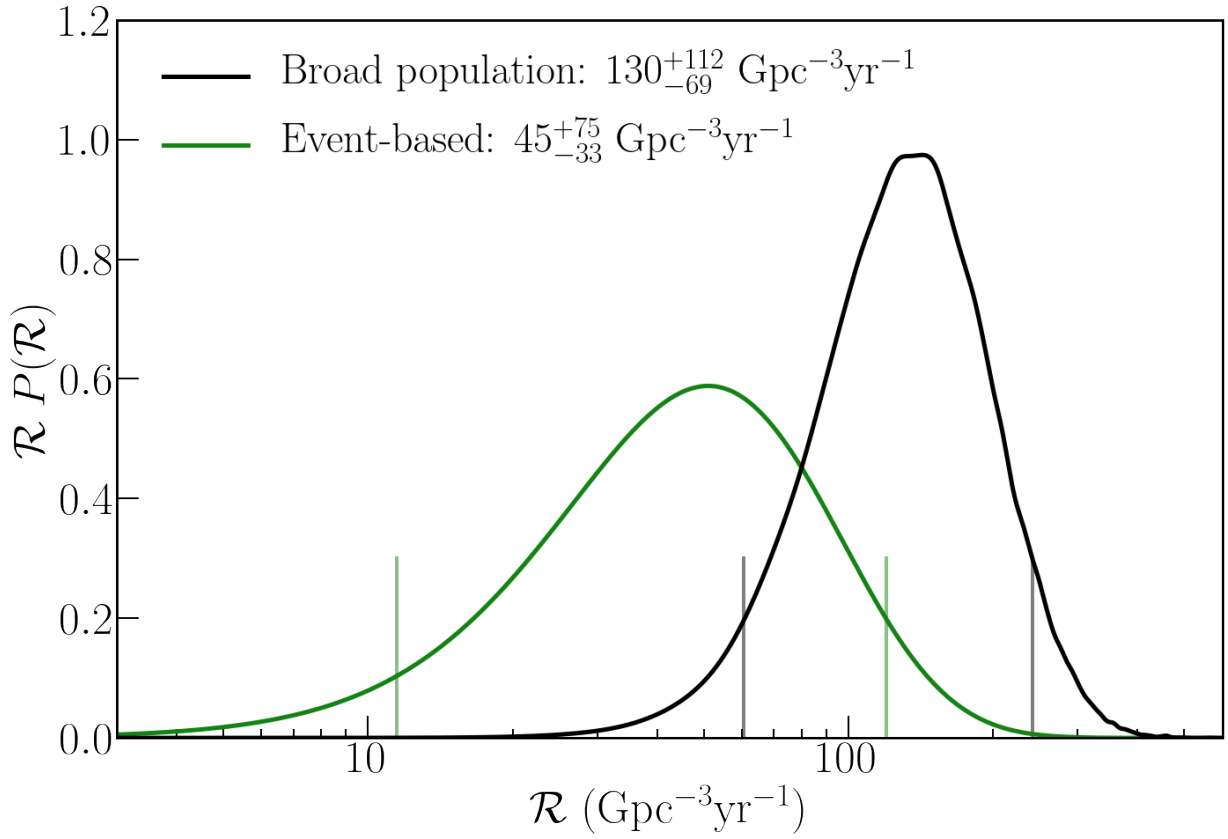


Figure 3.4: The NSBH merger rates obtained after the discovery of GW200105 and GW200115. The curve labeled "Broad population" was obtained using the method described here while the "Event-based" estimate was obtained using the approach of Kim et al. (2003) which excludes the contributions of marginal events to the merger rate. While I computed the broad population estimate, the event-based rate computation and figure generation credit goes to Charles Kimball.

counts posterior of Eq. (3.6) and $p(\Lambda_{1\alpha}|\{x\})$ is obtained by marginalizing the posterior of Eq. (3.6) overall counts except $\Lambda_{1\alpha}$. Note that I have used $\bar{F}_\alpha$ instead of F_α to emphasize the fact that the population distribution to which the VTs are being re-weighted need not be the same as the one used to compute the template-weights.

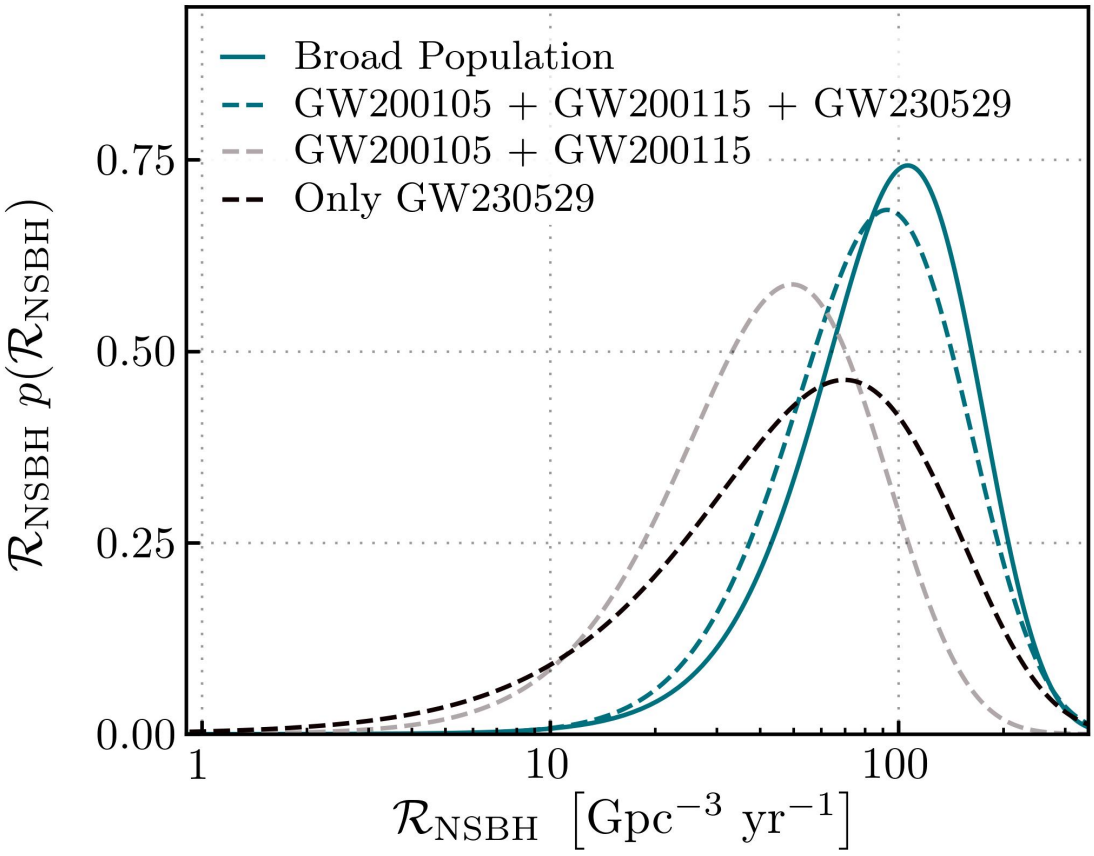


Figure 3.5: The rate of NSBHs including GW230529. The "Broad population" rate was computed using the method presented here and is published as part of Abac et al. (2024). While I computed the broad population estimate, the event-based rate computation was done by Jenifer Sanchez, and Shanika Galaudage generated the figure itself.

Using this framework I computed the merger rate of NSBHs after the discovery of GW200105 and GW200115 which were published in Abbott et al. (2021e). Figure 3.4 compares our rate estimate with the one obtained using the approach of Kim et al. (2003) which only includes GW200105 and GW200115 in the analysis and excludes all marginal candidates. While the two estimates are broadly consistent, our method predicts a higher merger rate of NSBHs.

Here I note that our rate estimate is much more stable than the event-based approach in the following sense. Upon the discovery of GW230529 which is likely an NSBH system (Abac et al., 2024), the updated event-based rate estimate increased quite a bit, even though still broadly con-

sistent with the earlier estimates. On the other hand, the updated rate estimate I obtained using our method is almost the same as the previous estimate only better constrained. The updated event-based rate agrees almost completely with the updated estimate of our method (figure 3.5). In other words, our method was able to better pin down the true rate at the end of O3 only and hence is stable against the inclusion of newly discovered events.

3.1.5 Marginalizing over population uncertainties

In light of our uncertainty regarding the population distribution of CBC parameters and the astrophysical formation channels of CBCs, it is often necessary to check the robustness of the fixed population rate-estimates against variations in the chosen distribution function of CBC parameters. In fact it is possible to use our current measurement of the CBC population distribution and marginalize our rate estimates over its uncertainties.

Population inference using significant CBCs yields posterior samples of hyper-parameters λ that characterize some distribution function $p(\theta|\lambda)$ of CBC parameters. It is possible to obtain estimates of VT corresponding to each hyper-parameters sample, such as:

$$\langle VT\rangle_{\alpha}(\lambda_{\alpha}) = \frac{\sum_{\gamma_j \in found} \delta\Lambda_{1\alpha}^{\gamma_j} \frac{p(\theta_{\gamma_j}|\lambda_{\alpha})}{p(\theta_{\gamma_j}|draw)}}{\sum_{\gamma_i \in all} \frac{p(\theta_{\gamma_i}|\lambda_{\alpha})}{p(\theta_{\gamma_i}|draw)}} \langle VT\rangle_{\alpha}^{inj} . \tag{3.15}$$

Using these VT's, one can obtain a rates-posterior that is marginalized over measurement uncertainties in the population hyper-parameters:

$$p(\mathcal{R}_{\alpha}|\{x\},\vec{d}) = \int d\lambda_{\alpha} p(\lambda_{\alpha}|\vec{d}) \int SdSp(\Lambda_{1\alpha} = \mathcal{R}S|\{x\}) p(S|\langle VT\rangle_{\alpha}(\lambda_{\alpha})), \tag{3.16}$$

where $p(\lambda_{\alpha}|\vec{d})$ is the posterior distribution of the hyper-parameters given data $\vec{d}$ from a collection of high significance events. After obtaining the samples of the hyper-parameters, this integral can

be implemented by means of a Monte Carlo sum:

$$p(\mathcal{R}_\alpha|\{x\},\vec{d}) \propto \sum_{ij} S_{ij} p(\Lambda_{1\alpha} = \mathcal{R}S|\{x\}) \tag{3.17}$$

where S_{ij} is the i'th sample drawn from $p(S|\langle VT\rangle_\alpha(\lambda_{\alpha,j}))$ and $\lambda_{\alpha,j}$ is the jth hyper-pparameter sample drawn from $p(\lambda_\alpha|\vec{d})$.

Using this method, I computed the BBH, BNS and NSBH rates using data from O1 through O3. I used the inferred hyper-parameters corresponding to the Powerlaw+Peak population model, the binned Gaussian process model, and the uniform model for estimating the merger rates of BBHs, NSBHs and BNSs respectively (Abbott et al., 2023d). Note that the hyperparameter inference was carried out using significant events only. Figure 3.6 represents the BBH, BNS, and NSBH rates I found using the GstLAL candidates from O1 through O3 with the corresponding numbers published in Abbott et al. (2023d).

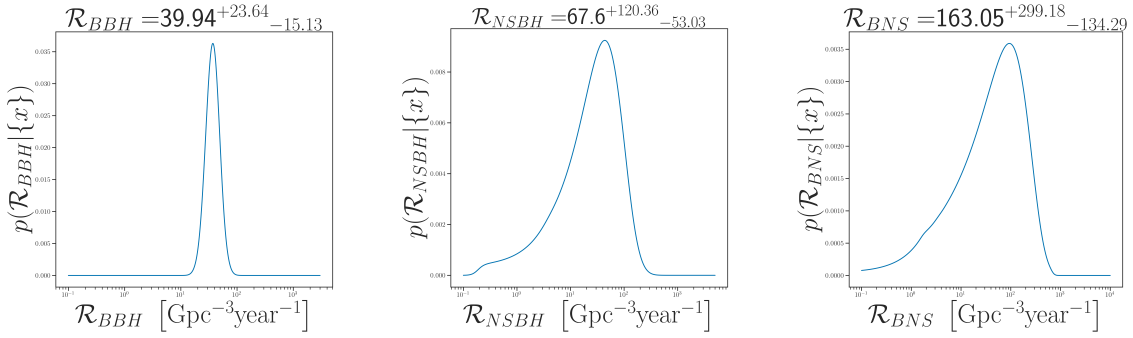


Figure 3.6: Rate estimates marginalized over population uncertainties using GWTC-3 gstlal triggers.

I note that the most general form of rate estimation using noisy datasets should be implemented using Eq. (2.42) wherein the population hyper-parameters are simultaneously inferred using the same dataset, i.e. one that contains both significant and marginal events. However, for the thresholds used in these analyses, Eq. (2.42) would require PE runs for thousands of triggers which was

computationally infeasible at the time these studies were carried out.

3.1.6 Summary

In this section, I delineated some novel implementations of fixed population trigger classification and rate estimation using noisy datasets, that I developed as part of my thesis work. The results of these investigations are crucial in constructing GW transient catalogs and for gauging the astrophysical implications of new CBC detections. I have communicated my findings for these investigations through the following LVK collaboration-wide papers: Abbott et al. (2023d,b, 2021d,e); Abac et al. (2024). I will continue these investigations for GWTC-4 and GWTC-5 to classify LVK's newly detected candidates and update the category-specific rate estimates. I further plan to improve the template-weight calculation presented here so as to implement more accurate trigger classification and rate estimation. These plans are summarized in Chapter 9.

Another highly crucial application of this analysis framework is the classification of GW candidates in real-time so as to optimize the multimessenger follow-up campaigns that are triggered by candidate alerts. However, due to the absence of a complete dataset of triggers for the ongoing observing run, the counts posterior cannot be constructed. Using the expected counts from obtained from the analyses of past observing runs will lead to underestimation of $\mathrm{P}_{\mathrm{astro}}$ values. In the next section, I describe my development of a new, self-consistent methodology for real-time trigger classification that is immune to this problem, and is currently being used by the LVK to classify GstLAL triggers amidst O4.

3.2 OPTIMIZING THE SEARCH FOR MULTI-MESSENGER COUNTERPARTS

We develop a robust and self-consistent framework to extract and classify gravitational wave candidates from noisy data, for the purpose of assisting in real-time multi-messenger follow-ups during LIGO-Virgo-KAGRA's fourth observing run (O4). Our formalism implements several improve-

ments to the low latency calculation of the probability of astrophysical origin ($\mathrm{P_{astro}}$), so as to correctly account for various factors such as the sensitivity change between observing runs, and the deviation of the recovered template waveform from the true gravitational wave signal that can strongly bias said calculation. We demonstrate the high accuracy with which our new formalism recovers and classifies gravitational wave triggers, by analyzing replay data from previous observing runs injected with simulated sources of different categories. We show that these improvements enable the correct identification of the majority of simulated sources, many of which would have otherwise been misclassified. We carry out the aforementioned analysis by implementing our formalism through the GstLAL search pipeline even though it can be used in conjunction with potentially any matched filtering pipeline. Armed with robust and self-consistent $\mathrm{P_{astro}}$ values, the GstLAL pipeline can be expected to provide accurate source classification information for assisting in multi-messenger follow-up observations to gravitational wave alerts sent out during O4.

3.2.1 Motivation

The observation of gravitational waves (GWs), a burst of gamma rays and other electromagnetic (EM) transients in X-ray through radio, from the binary neutron star (BNS) merger GW170817, marked a significant breakthrough in multi-messenger astronomy (Abbott et al., 2017c; Goldstein et al., 2017; Savchenko et al., 2017; Abbott et al., 2017b; Coulter et al., 2017; Abbott et al., 2017d). Interpretation of the complementary information carried by photons and GWs from a single BNS system has provided new insights into several mysteries of fundamental physics, beyond the scope of what either messenger could have conveyed individually. Some examples include explorations of the properties of matter at extreme densities (Radice et al., 2018), tests of general relativity in the strong gravity regime (Abbott et al., 2019e, 2017b), the measurement of cosmological parameters (Abbott et al., 2017a), and investigations on the production of heavy elements in the universe (Kasen et al., 2017).

To facilitate the prompt discovery of multi-messenger counterparts to GW observations, the LIGO, Virgo and KAGRA (LVK) (J. Aasi et al., 2015; Acernese et al., 2014; Akutsu et al., 2021) detector network sends out public alerts about GW candidates, within minutes of their detection (Abbott et al., 2018b). These alerts comprise information regarding source properties and candidate significance which can assist in optimizing the triggered follow-up campaigns. For example, the optical counterpart to GW170817 was detected while performing a targeted search that made use of the precise sky localization obtained through the analysis of multi-detector GW data (Coulter et al., 2017; Abbott et al., 2017d).

For GWs originating from compact binary coalescences (CBCs), information regarding the nature of the merging compact objects is also considered highly important for conducting real-time multi-messenger searches. The main reason for this consideration is as follows. Among the three kinds of CBC sources expected to be observable by the LVK, BNSs and neutron star–black hole binaries (NSBHs) are far more likely to be associated with multi-messenger counterparts than binary black holes (BBHs) (Metzger, 2019; Perna et al., 2019). Hence, estimates of the probability that a CBC candidate has originated from either a BNS or an NSBH, that are sent out as part of GW candidate alerts (Abbott et al., 2018b), can assist astronomers in making informed decisions on whether to follow-up on said alerts.

This so-called probability of astrophysical origin ($\mathrm{P_{astro}}$) can be computed from GW data, given knowledge of the rate at which BBH, BNS and NSBH systems merge in the observable universe and the population-level distributions of their system parameters (Kapadia et al., 2020; Abbott et al., 2019b). Informed by the relative rates and population properties of various CBC sources, estimates of $\mathrm{P_{astro}}$ can be expected to reveal a larger number of significant candidates in densely populated regions of the CBC parameter space, in complement to estimates of false alarm rate (FAR). Hence, in addition to facilitating source classification, the $\mathrm{P_{astro}}$ calculation also offers a population-informed measure of significance for GW candidates.

Several frameworks for estimating $\mathrm{P}_{\mathrm{astro}}$ self-consistently with the rates and population properties of CBCs from GW data have been studied in existing literature (Farr et al., 2015; Kapadia et al., 2020; Gaebel et al., 2019; Roulet et al., 2020). However, they all involve constructing the posterior probability distributions of said rates given data associated with all the GW candidates found above some FAR threshold during an observing run. Since the mentioned dataset can only become available in postprocessing, such analyses cannot be used straightforwardly for computing $\mathrm{P}_{\mathrm{astro}}$ in real-time, without resorting to various schemes for approximating the rate of detectable CBCs.

In this paper, we describe a self-consistent methodology for computing $\mathrm{P}_{\mathrm{astro}}$ in low-latency with the purpose of assisting in real-time multimessenger follow-up observations during LVK's fourth observing run (O4). Our framework comprises several improvements to the low-latency $\mathrm{P}_{\mathrm{astro}}$ calculation that was implemented in LVK's third observing run (O3). We implement our formalism through the GstLAL search pipeline's real-time (*online*) analysis of O4 data, even though our improved rate approximation scheme is applicable to other pipelines as well.

We validate our formalism through GstLAL's participation in a rigorous Mock Data Challenge (MDC) that comprised a replay of O3 data with simulated sources injected into the data stream (Chaudhary et al., 2023; Ewing et al., 2024). We show that our formalism yields highly accurate category specific $\mathrm{P}_{\mathrm{astro}}$ for both real and simulated events in the MDC. We also demonstrate how our improved rate-approximation prevents the underestimation of $\mathrm{P}_{\mathrm{astro}}$ values of real and simulated events that would have resulted from the use of O3's inaccurate rate-approximation scheme.

This paper is organized as follows. In section 3.2.2 we describe our formalism for computing low-latency $\mathrm{P}_{\mathrm{astro}}$ values in detail. In section 3.2.3, we summarize the performance of our methods in the aforementioned MDC and demonstrate the accuracy of our category-specific $\mathrm{P}_{\mathrm{astro}}$ values. In section 3.2.4, we conclude with a summary of our methods, their validation, and future prospects

regarding further improvements to our low-latency computation of $\mathrm{P_{astro}}$.

3.2.2 Methods

In this section, we describe our formalism for calculating accurate $\mathrm{P_{astro}}$ values in low latency amidst an ongoing observing run. We first summarize the offline calculation that can be implemented given the full set of GW triggers found during an observing run. We then derive several approximation schemes for accurate $\mathrm{P_{astro}}$ calculation in real-time, i.e. in the absence of the full list of triggers, which can only become available in post-processing.

3.2.2.1 Probability of astrophysical origin

GW triggers originating from different sources (including terrestrial ones) can be interpreted as realizations of independent Poisson processes (Farr et al., 2015; Kapadia et al., 2020). Hence, given the rate of events expected from each source category and the GW data associated with a particular trigger, one can compute the probability of its origin being any one of said categories. For example, the probability of the k^{th} trigger originating from category α is given by:

$$P_\alpha(d_k|\vec{R}) = \frac{P(d_k|H_\alpha)R_\alpha}{\sum_\beta P(d_k|H_\beta)R_\beta} \tag{3.18}$$

where, H_α is the hypothesis that a particular trigger has originated from the source category α, R_α the expected rate of *detectable* triggers per unit time from source $\alpha \in \{\mathrm{BBH}, \mathrm{BNS}, \mathrm{NSBH}, \mathrm{Terrestrial}\}$, and d_k the data associated with the k^{th} trigger[1]. Note that the analysis of only *detectable* triggers implies the existence of a threshold value for some detection statistic such as FAR that is used to rank events by the GW search pipelines.

In an offline analysis, wherein the data corresponding to the full list of N triggers found during

[1]Note that we express everything in terms of rates while Kapadia et al. (2020) used counts which relate to our rates as: $\Lambda_a = R_a \times T_{obs}$

an observing run is available, the probabilities in Eq. (3.18) can be inferred self-consistently with the rates from the same dataset (Farr et al., 2015; Kapadia et al., 2020). This is achieved by first constructing the posterior distribution of the rates using the trigger data from a particular observing run (say r), which looks like:

$$p(\vec{R}_r|\vec{d}_{N_r}) \propto p(\vec{R}_r)e^{-\sum_\alpha R_{\alpha,r}T_{obs,r}} \prod_i^{N_r} \sum_\alpha P(d_i|H_\alpha)R_{\alpha,r}T_{obs,r} \tag{3.19}$$

, where $\vec{d}_{N_r}$ is a set comprising data from all observed triggers, N_r the total number of observed triggers, $T_{obs,r}$ the total observation time and $p(\vec{R}_r)$ an uninformative prior on the rates. Note that we are using the capitalized P for probability and the regular p for probability distribution/density, a notational choice made consistently throughout the rest of the paper. Once the rate posterior is constructed, it is possible to marginalize the probabilities in Eq. (3.18) over the uncertainty in rate estimation:

$$P_\alpha(d_k|\vec{d}_{N_r}) = \int d\vec{R} P_\alpha(d_k|\vec{R}_r)p(\vec{R}_r|\vec{d}_{N_r}) \tag{3.20}$$

. Upon carrying out said marginalization, the probability of a GW trigger originating from a particular astrophysical source category $a \in \{\text{BBH}, \text{BNS}, \text{NSBH}\}$ can be expressed in terms of the expectation values of the rates w.r.t. the posterior in Eq. (3.19), as in:

$$P_a(d_k|\vec{d}_{N_r}) = \frac{\frac{P(d|H_a)}{P(d|H_0)}\frac{\langle R_a\rangle_{N_r/k}}{\langle R_0\rangle_{N_r/k}}}{1+\sum_b \frac{P(d|H_b)}{P(d|H_0)}\frac{\langle R_b\rangle_{N_r/k}}{\langle R_0\rangle_{N_r/k}}} \tag{3.21}$$

where, $\langle R_\alpha\rangle_{N/k} = \int d\vec{R} p(\vec{R}|\vec{d}_{N/k})R_\alpha$ is the mentioned expectation value, $\vec{d}_{N/k}$ the dataset corresponding to all N triggers except the k^{th} one, and the subscript 0 representative of the noise/terrestrial source category. Note that the sum over b in the denominator is over astrophysical categories only. Throughout the rest of this paper, Latin subscripts to rates and probabilities (such as a, b, c) are chosen to represent an astrophysical source category while a Greek subscript (such α, β) can

denote astrophysical as well as terrestrial categories. In terms of the quantities defined in Eq. 3.21, the combined probability of astrophysical origin can be written as $\mathrm{P}_{\mathrm{astro}} = \sum_a P_a$. For a derivation of Eq. (3.21) from Eq. (3.20) see Kapadia et al. (2020).

This framework has been used to compute offline $\mathrm{P}_{\mathrm{astro}}$ values that were used in the generation of GW transient catalogs for the first three observing runs of LVK (Abbott et al., 2019b, 2021c,d, 2023b). In the next subsection, we describe our rate approximation scheme that is needed to implement this framework accurately in low latency amidst an ongoing observing run.

3.2.2.2 Computing P_a in low latency: rate approximation

As mentioned before, in order to evaluate Eq. (3.21), the rate posterior needs to be constructed from the full list of triggers found during an observing run. However, for the purpose of assisting multi-messenger follow-up observations in real-time, P_a has to be computed amidst an on-ongoing observing run i.e. in the absence of the complete list of triggers.

In such a scenario, our only option is to approximate the expected rates of detectable GW sources from their corresponding estimates yielded by past observing runs, while accounting for the change in sensitivity of the detectors in between the previous and ongoing runs. In particular, we only need approximate the ratio $\frac{\langle R_{a,r} \rangle}{\langle R_{0,r} \rangle}$ of the foreground and background rates since it is the only term involving them that appears in Eq. (3.18). To construct such an approximation scheme, we can exploit the following facts:

- The astrophysical rate $\mathcal{R}$ of BBHs, BNSs, and NSBHs, per unit comoving volume per unit time, do not change in between observing runs.
- The rate of terrestrial triggers above some FAR threshold (FAR_0) can be expected to be equal to said threshold: $R_0 = FAR_0$.

Hence, estimates of $\mathcal{R}_a$ for $a \in \{\mathrm{BBH}, \mathrm{BNS}, \mathrm{NSBH}\}$ from previous observing runs can be used to approximate $R_{a,r}$ for an ongoing observing run r given knowledge of the hypervolume (Tiwari,

2018) to which the detector is sensitive during said run:

$$\langle R_{a,r} \rangle = \langle \mathcal{R}_a \rangle \frac{\langle VT \rangle_{a,r}}{T_{\mathrm{obs,r}}} \tag{3.22}$$

where, $\langle VT \rangle_{a,r}$ is the sensitive hypervolume during observing run r corresponding to the astrophysical source category a, $T_{obs,r}$ the corresponding live observation time, and $\langle \mathcal{R}_a \rangle$ an estimate of the astrophysical rate yielded through the combined analysis of data from previous observing runs. However, if the ongoing observing run has a different time-volume sensitivity than the previous ones, $\langle VT \rangle_{a,r}$ needs to be scaled appropriately from its values during past observing runs to account for said sensitivity change. While the sensitive hypervolume depends on various factors such as the source population model (Tiwari, 2018) (which itself remains uncertain to some degree (Abbott et al., 2023d)), as a first-order step, simplistic scaling factors can be constructed by using the BNS horizon distance of the detector at various observing runs (including the ongoing one for which we use projections based on simulations (Abbott et al., 2018b)). For example, if $d_{BNS,s}$ is the horizon distance of the detector network during observing run s at an SNR threshold of 8 and $T_{obs,s}$ the live observation time, then $\langle VT \rangle_{a,r}$ can be approximated in the following way:

$$\langle VT \rangle_{a,r} = \langle VT \rangle_{a,old} \times \frac{d^3_{\mathrm{BNS,r}} T_{\mathrm{obs,r}}}{\sum_{s \in old} d^3_{\mathrm{BNS,s}} T_{\mathrm{obs,s}}} \tag{3.23}$$

where old represents the combination of all past observing runs. The live observation times of the past runs $(T_{obs,s})$ can be approximated from the total number of terrestrial triggers found above threshold $(N_{0,old})$, their corresponding rate $(R_{0,old} = FAR_{0,old})$, and the individual run times $(T_{run,s})$ in the following way:

$$T_{obs,s} = \frac{T_{\mathrm{run,s}}}{\sum_{s \in old} T_{\mathrm{run,s}}} \times \frac{N_{0,old}}{R_{0,old}} \tag{3.24}$$

Putting Eqs. (3.22,3.23,3.24) together, we are able to construct an approximation for $\frac{\langle R_{a,r}\rangle}{\langle R_{0,r}\rangle}$ from estimates of various quantities available at the beginning of the ongoing observing run r, that takes the following form:

$$\frac{\langle R_{a,r}\rangle}{\langle R_{0,r}\rangle} = \langle \mathcal{R}_a \rangle \times \langle VT \rangle_{a,old} \times \underbrace{\left(\frac{d^3_{\mathrm{BNS,r}} \sum_{s \in old} T_{\mathrm{run,s}}}{\sum_{s \in old} d^3_{\mathrm{BNS,s}} T_{\mathrm{run,s}}} \times \frac{FAR_{0,old}}{N_{0,old} FAR_{0,r}} \right)}_{S_{r/old}} \tag{3.25}$$

Notice here that $T_{obs,r}$ conveniently cancels in our self-consistent approximation scheme. The scaling factor on the RHS of Eq. (3.25) is an improvement over the online $\mathrm{P_{astro}}$ calculations implemented upto O3 which assumed it to be 1 and hence did not account for the sensitivity change between O2 and O3, leading to potential underestimation of $\mathrm{P_{astro}}$ values. We demonstrate this through the analysis of O3 replay data in sec. 3.2.3.1 . We note that the rate approximation scheme described here is general in the sense that it can be used in conjunction with any low-latency matched filtering pipeline.

3.2.2.3 Computing P_a in low latency: category specific likelihood ratio

With the self-consistent rate approximation scheme at our disposal, we can proceed to assemble a methodology for computing the remaining ingredients required for evaluating Eq. (3.21), namely the marginalized likelihood ratios (also known as Bayes factors in standard literature):

$$K_a(d) = \frac{P(d|H_a)}{P(d|H_0)} \tag{3.26}$$

The Bayes factors quantify how much the data supports H_a over H_0. In order to compute these quantities, one must model the population of CBCs in each astrophysical category using some fiducial function characterizing the distribution of measurable CBC parameters. Measurements of said parameters from detector strain data obtained by means of matched filtering can then be

used to classify a potential GW candidate into various astrophysical categories given the different population models corresponding to each category.

Matched filtering search pipelines usually discretize the space of signal parameters to reduce computational cost, which results in a bank of template waveforms (Allen et al., 2012). A CBC signal is expected to *ring up* templates in the bank with some SNR (ρ). Multi-detector coincidences are formed when identical templates are triggered at different observatories within the propagation time of GWs between said observatories. The triggered templates are usually ranked in significance with some statistic (say x) that is often computed from additional quantities assigned to GW triggers by the search pipeline.

Among the triggers that pass a chosen detection threshold x_{th} in x, either the highest x or highest ρ template within a certain time-window is chosen to represent a potential GW candidate that might have occurred in said window. The data corresponding to a particular GW candidate (d_k) relevant for Bayes factor calculation thus comprises three quantities assigned to it by the search pipeline namely the representative template (t_k), the corresponding ranking statistic (x_k) and SNR ρ_k.

To compute the Bayes factor accurately from the trigger data, one needs to account for the possibility of random noise fluctuations causing a true signal originating from source category a to be recovered in a template whose parameters correspond to a different source category b. This is particularly true if the different source categories map onto neighboring regions of the template bank separated by some hard boundary.

Hence, a true signal corresponding to template t_j and SNR ρ, can have a finite probability ($P(t_k|t_j, \rho)$) of being found in a different template t_k, depending on how similar said templates are with each other. In terms of this probability and a category-specific distribution of template parameters, it is possible to evaluate the likelihood with which any template t_k in the bank might be triggered by a CBC signal originating from source category a:

$$P(t_k, \rho|H_a) = \sum_j P(t_k|t_j, \rho)p(\vec{\theta}(t_j)|H_a)p(\rho|H_a)V_j \tag{3.27}$$

where, $\vec{\theta}(t_j)$ is the CBC parameters corresponding to the template t_j, V_j the volume in parameter space associated with $\vec{\theta}(t_j)$ and $p(\rho|H_a)$ the prior distribution of SNRs corresponding to sources in category a (Fong, 2018).

To compute $P(t_k|t_j, \rho)$, we make the assumption that migration between true and recovered templates is solely due to Gaussian fluctuations (Fong, 2018). Furthermore, since the purpose of these *template-weights* ($P(t_k, \rho_k|H_a)$) is to account for the miss-classification of GW sources due to the template migration caused by noise, it is safe to assume, for a first approximation, that $\rho_k = \rho_j$. Under these considerations, the probability of template migration takes the following form:

$$P(t_k|t_j, \rho) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\rho^2\left[1-2t_k\cdot t_j+(t_j\cdot t_k)^2\right]} \tag{3.28}$$

where $t_k \cdot t_j$ is the noise averaged inner product of the normalized template waveforms. For a derivation of Eq. (3.28) see A.2.

The volume element V_j corresponding to each template needed for evaluating Eq. (3.27) can often be pre-computed from the template bank. For an arbitrary bank, it is possible to approximate the volume element V_j corresponding to a template t_j as the volume of the Voronoi cell (Voronoi, 1908; Fortune, 1997) that contains the point $\vec{\theta}(t_j)$ (Fong, 2018). However the method of Voronoi decomposition is computationally expensive and it is preferable to compute V_j as part of bank generation itself, if possible, as in the case of banks populated using methods like (Hanna et al., 2023). Since different search pipelines use different template banks and ranking statistics, the remainder of the Bayes factor calculation will also vary from pipeline to pipeline. We demonstrate how such a calculation can be implemented within the GstLAL search pipeline (Messick et al.,

2017; Cannon et al., 2021; Sachdev et al., 2019) while noting that our formalism can also be applied to other pipelines by simply replacing the GstLAL-specific part of the Bayes factor calculation that is described below.

The O4 template bank used by the GstLAL search pipeline places templates in the $\log_{10}(m_1) \times \log_{10}(m_2) \times \chi_{\text{eff}}$ space by splitting up the parameter space via a binary tree approach such that the volumes of the nearby templates are similar. Here, χ_{eff} is $\frac{m_1 \times s_{1,z} + m_2 \times s_{2,z}}{m_1 + m_2}$ where m_i and $s_{i,z}$ are the masses and z-component spins of the i^{th} component of the binary. The volumes surrounding templates are computed using the minimal match of the templates allowed in the template bank. For details regarding the volume calculation for the template bank used in O4 by the GstLAL search pipeline see Sakon et al. (2023).

Putting all of these pieces together yields a robust estimate of $P(t_k, \rho_k|H_a)$, one that accounts for the mismatch between true and recovered waveforms due to detector noise. The variation of these quantities as a function of SNR and template parameters are summarized in A.3. We find said variation to be consistent with our expectations regarding miss-classification among various categories at different SNRs.

We note that GstLAL's online analyses upto O3 did account for this migration between true and recovered templates accurately and classified sources into astrophysical categories either based solely on the recovered template parameters or based on simulation campaigns that were not only expensive to generate but also sparse compared to the full template bank. We demonstrate the accuracy of the source classification resulting from the use of our template weights in sec 3.2.3.2.

With the template weights accounting for miss-classification among astrophysical categories, the Bayes factors can be computed from said template weights and the probability distributions of the search pipeline's ranking statistic conditional on the signal and noise hypotheses (H_1 and H_0) (Kapadia et al., 2020). For the GstLAL search pipeline, its ranking statistic is itself the log of the likelihood ratio: $e^{x_k} \propto P(d_k|H_1)/P(d_k|H_0)$ (Cannon et al., 2015; Messick et al., 2017; Hanna

et al., 2020; Tsukada et al., 2023). Hence, in terms of GstLAL's ranking statistic (x_k), the Bayes factors can be written in the following way:

$$K(x_k, \rho_k, t_k) = A(FAR_0)e^{x_k} \times \frac{P(t_k, \rho_k|H_a)}{\left(\frac{1}{3}\right)\sum_b P(t_k, \rho_k|H_b)} \tag{3.29}$$

$$A(FAR_0) = \left[\int_{x(FAR_0)}^{\infty} dx e^x p(x|H_0)\right]^{-1} \tag{3.30}$$

where $x(FAR_0)$ is the log-likelihood ratio corresponding to the FAR threshold used throughout the $\mathrm{P}_{\mathrm{astro}}$ calculation and $p(x|H_0)$ is the normalized probability distribution of the ranking statistic conditional on the noise hypothesis. For a derivation of Eq. (3.29), see A.1.

With the Bayes factors calculable on the fly from trigger data and the rates pre-computed using the scaling relation (3.25), robust and accurate $\mathrm{P}_{\mathrm{astro}}$ values can be evaluated in low-latency through Eq. (3.21) for assisting in multi-messenger follow-up observations to GW triggers in real-time. The code developed to implement this analysis is publicly available in the python package GWSCI-PASTRO[2]

3.2.3 Illustrative results: mock data challenge

We have tested our formalism for calculating $\mathrm{P}_{\mathrm{astro}}$ through GstLAL's participation in a rigorous MDC (Chaudhary et al., 2023; Ewing et al., 2024). The MDC involved a replay of strain data from the detectors at Hanford, Livingston and Virgo spanning 40 days of O3, from 5 Jan. 2020 15:59:42 to 14 Feb. 2020 15:59:42, along with simulated signals injected into the data stream. Several low-latency search pipelines (including GstLAL) were used to analyze this mock dataset with our new $\mathrm{P}_{\mathrm{astro}}$ calculation implemented as part of GstLAL's analysis. For further details regarding the MDC see Chaudhary et al. (2023); Ewing et al. (2024).

[2]https://pypi.org/project/gwsci-pastro/

To validate our formalism for calculating $\mathrm{P}_{\mathrm{astro}}$, we first demonstrate the improvements resulting from our rate-approximation scheme in the context of real GW events replayed in the MDC. Afterward, we use the $\mathrm{P}_{\mathrm{astro}}$ values of simulated events computed using our formalism to provide a rigorous validation for the developed rate approximation and source classification schemes.

3.2.3.1 GW events replayed in the MDC

As mentioned before, the real-time $\mathrm{P}_{\mathrm{astro}}$ calculations necessitate the use of rate approximation schemes to compensate for the absence of the full list of triggers found during an observing run, which can only be available in post-processing. In that sense, our online $\mathrm{P}_{\mathrm{astro}}$ values are approximations to the ones computed self-consistently with the rates in an offline analysis carried out at the end of the observing runs for generating GW transient catalogs. Therefore, as a validity test of our rate-approximation scheme, $\mathrm{P}_{\mathrm{astro}}$ values computed using our formalism for O3 GW candidates replayed in a MDC can be compared with results from the O3 offline analysis.

For this MDC, we can thus pretend that O3 is the ongoing observing run while O1 and O2 are the previous ones. To compute rate estimates for calculating $\mathrm{P}_{\mathrm{astro}}$ values in the MDC, we therefore scale the sensitive hyper-volume of O1 and O2 obtained from offline analyses, using the projections of the O3 BNS ranges available at the beginning of O3 (Abbott et al., 2018b). We then use the estimates of the astrophysical rates available from offline analyses of O1 and O2 in conjunction with the scaled hypervolumes to find $\langle R_{a,O3} \rangle$ / $\langle R_{0,O3} \rangle$.

The accuracy of our rate approximation can be demonstrated by comparing the aforementioned scaled rate estimates to the unapproximated ones obtained from an offline analysis of public LVK data corresponding to O3 triggers found by GstLAL (LIGO Scientific Collaboration And Virgo Collaboration And KAGRA Collaboration, 2021). The comparison is summarized in table 3.1. It can be seen that the approximated scaling relation yields BBH rates that are remarkably accurate given how close they are to the offline results. On the other hand, the unscaled O1O2 rates are

smaller by a factor of 2.

The scaled estimates of BNS and NSBH are further away from the offline O3 estimates as compared to BBH, which is to be expected due to the following reasons. The number of confident BNS and NSBH detections in O1O2 are 1 and 0 respectively (Abbott et al., 2019b). Similarly, the corresponding numbers in O3 are 1 and 2 (Abbott et al., 2021c, 2023b). Hence, both sets of rate estimates, namely the ones scaled from O1O2 and the ones yielded by the O3 offline analysis, are subject to much larger Poisson uncertainties for BNS and NSBH than for BBH. The BNS and NSBH rate estimates for O4 (listed in table 3.2), which are scaled from the O1O2O3 rates and VTs, can be expected to be more accurate given the larger number of BNS and NSBH detections in O1O2O3 as compared to O1O2.

Category	Unscaled	Scaled	Offline
BBH	3.03×10^{-3}	7.67×10^{-3}	7.80×10^{-3}
BNS	4.23×10^{-4}	1.09×10^{-3}	1.7×10^{-4}
NSBH	3.04×10^{-4}	7.87×10^{-4}	5.9×10^{-4}

Table 3.1: Comparison of $\langle R_{a,O3} \rangle$ / $\langle R_{0,O3} \rangle$ for the three cases (unscaled, scaled and O3 offline) described above.

Furthermore, table 3.1 is indicative of the fact that using the unscaled rates and hypervolumes of O1 and O2 has led to the underestimation of $\mathrm{P_{astro}}$ values in the O3 online analysis. To verify this claim, we compare the $\mathrm{P_{astro}}$ of real GW events replayed in the MDC computed using the 3 different rate estimates listed in table 3.1. We choose the fiducial population model required for computing the template-weights (and hence $\mathrm{P_{astro}}$) to be a Salpeter function (Salpeter, 1955) in primary mass and uniform in all other parameters:

$$p(m_1, m_2, \chi_{1z}, \chi_{2z}|H_a) \propto m_1^{-2.35} \frac{\Theta(m_1 - m_2)}{m_1 - m_{min}} \begin{cases} 1 & (m_1, m_2, \chi_{1z}, \chi_{2z}) \in S_a \\ 0 & o.w. \end{cases} \tag{3.31}$$

. Here Θ is the Heaviside step function, m_{min} is the minimum mass of templates in the bank and

S_a are hyper-cubes in the space of the parameters $(m_1, m_2, \chi_{1z}, \chi_{2z})$ whose edges correspond to category specific upper and lower boundaries. The boundaries in parameter space separating BNS from NSBH and NSBH from BBH are determined by the maximum NS mass and spin which are taken to be $3\ M_\odot$ and 0.5 respectively. Masses and spins of BHs are allowed to go up to $300\ M_\odot$ and 0.99 respectively.

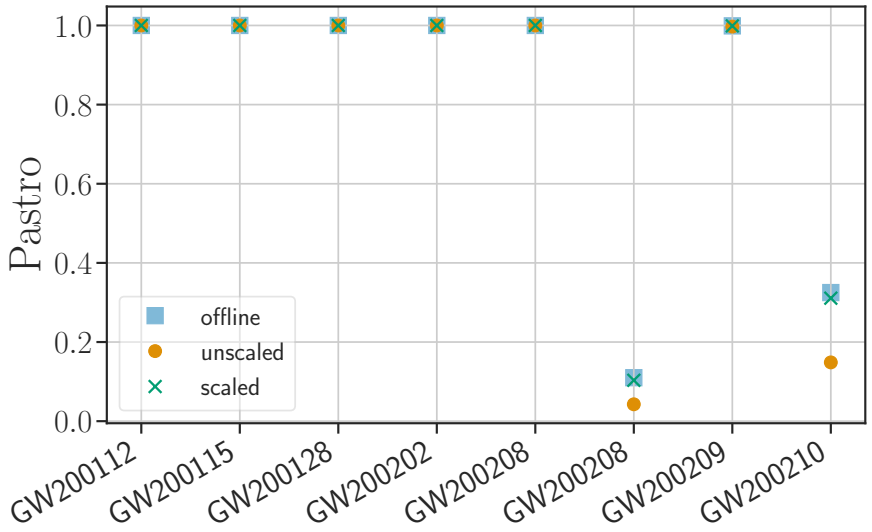


Figure 3.7: Comparison of $\mathrm{P_{astro}}$ values computed using scaled, unscaled, and offline rate estimates. The x-axis marks the actual dates on which the replayed GW events were discovered in O3.

With $P(t_k, \rho_k|H_a)$ pre-computed using Eq. (3.27) for the population model in Eq. (3.31), $\mathrm{P_{astro}}$ values for the real GW events replayed in the MDC can be computed for validating our rate-approximation schemes. For a visual representation of how the $\mathrm{P_{astro}}$ values computed using the aforementioned template weights and the 3 different rates listed on table 3.1 compare against each other see Fig. 3.7.

It can be seen that while both the scaled and unscaled rate estimates yield accurate $\mathrm{P_{astro}}$ values for the highly significant candidates, the scaled rate estimates perform significantly better for the

marginal events than the unscaled ones. In particular, it can be seen that the low-latency $\mathrm{P_{astro}}$ of the NSBH candidate GW200210_092254 would be underestimated by a factor of 2 if not for our accurate scaling approximation.

Given this performance in the MDC, we conclude that our formalism can be expected to yield $\mathrm{P_{astro}}$ values of real GW events in GstLAL's online analysis of O4 data, that approximate the results of a corresponding offline analysis with higher accuracy, as compared to what was implemented in O3. In sec. 3.2.3.2, We provide a more rigorous demonstration of the necessity of our rate scaling approximations through analysis of the set of simulated sources in the MDC.

The approximate rates to be used by GstLAL to calculate online $\mathrm{P_{astro}}$ values during O4 are listed in table 3.1. They are largely consistent with the estimates yielded by Monte Carlo simulations carried out for assessing LVK's observing capabilities in O4 (Abbott et al., 2018b). The rates in table 3.2 shall be used in conjunction with the template-weights corresponding to the Salpeter population-model in Eq. (3.31) by GstLAL in O4 for computing robust and accurate $\mathrm{P_{astro}}$ values in its online analysis of O4.

Category	$\langle \mathcal{R}_a \rangle$	$\langle VT_{a,old} \rangle$	$\langle R_{a,O4} \rangle$ / $\langle R_{0,O4} \rangle$
BBH	$40\ Gpc^{-3}yr^{-1}$	$1.6\ Gpc^{3}yr$	0.018
BNS	$163\ Gpc^{-3}yr^{-1}$	$0.014\ Gpc^{3}yr$	6.7×10^{-4}
NSBH	$68.0\ Gpc^{-3}yr^{-1}$	$0.048\ Gpc^{3}yr$	9.1×10^{-4}

Table 3.2: The estimates of $\langle R_{a,O4} \rangle$ / $\langle R_{0,O4} \rangle$ to be used by GstLAL in the O4 online analysis. The columns $\langle \mathcal{R}_a \rangle$ and $\langle VT_{a,old} \rangle$ list estimates of the astrophysical rates and the sensitive hyper-volume obtained from the offline analysis of O1O2O3 (appendix C3 of Abbott et al. (2023d)).

We note that the distribution in Eq. (3.31) is a straw-person representative of the true CBC mass and spin population which remains uncertain to this date (Abbott et al., 2023d). In particular, the population-level distributions of masses and spins of both BNS and NSBH systems are subject to significant Poisson uncertainties given the small number of detections made to date. If new BNSs and NSBHs discovered in the first half of O4 lead to more precise measurements of their population properties and if the new measurements are found to deviate significantly from our

straw-person population model, the template weights corresponding to some best estimate of the newly measured population can be re-computed straightforwardly leading to even more accurate source classification in the second half of O4.

However, improvements in our knowledge of the true CBC population can lead to better classification if and only if the deviation between true and recovered signal parameters due to noise fluctuations is correctly accounted for within the formalism of calculating $\mathrm{P}_{\mathrm{astro}}$. We thus want to test for the accuracy with which our formalism can correctly account for the probability of miss-classification among astrophysical source categories given knowledge of true CBC population distribution.

To perform such a validation study, we calculate the category-specific $\mathrm{P}_{\mathrm{astro}}$ values (P_a) for the simulated events in the MDC that comprise a known fiducial population of CBCs. We summarize the results obtained from said study in the next subsection.

3.2.3.2 Simulated sources in the MDC

The category-specific $\mathrm{P}_{\mathrm{astro}}$ values of the simulated sources that are injected into the replay data as part of the MDC, can be used to validate our source classification and rate-approximation schemes. The template weights as well as the rate of injected events per comoving volume, both of which are needed for calculating P_a, will depend on the population-level distributions of CBC parameters that were used to generate the mentioned injections.

The masses and spins characterizing the simulated signals in the MDC were drawn from fiducial population-level distributions (one corresponding to each source category), the functional forms of which were chosen to be truncated power-laws in masses and uniform in spins (Chaudhary

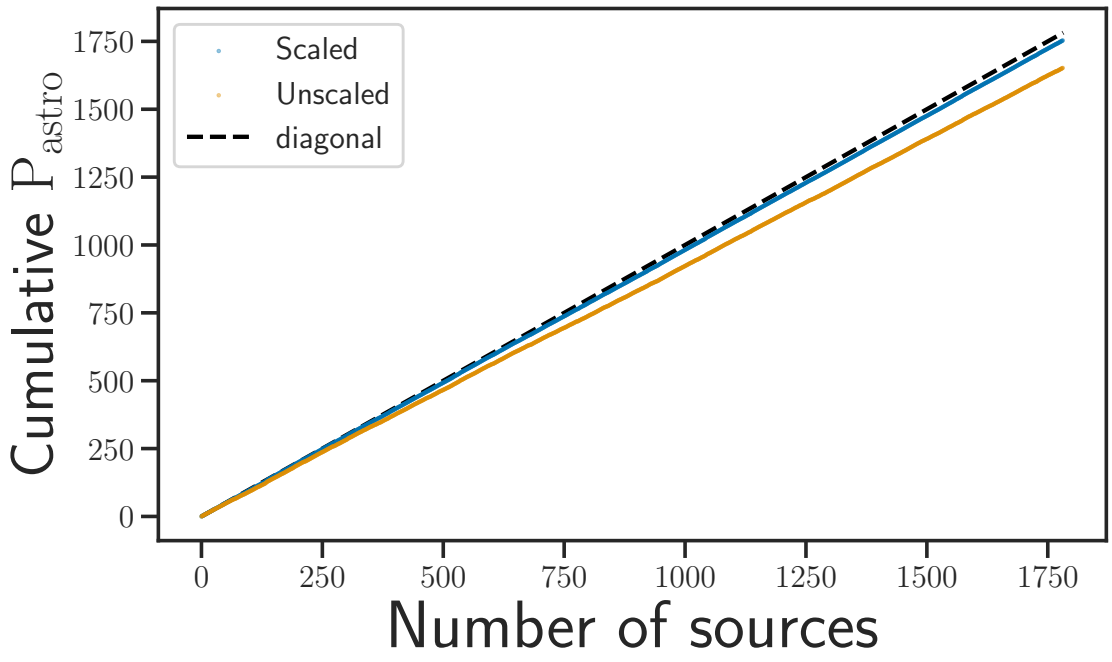


Figure 3.8: Recovery of simulated sources. The y-axis represents the cumulative sum of $\mathrm{P}_{\mathrm{astro}}$ values of all the triggers found above threshold by the search pipeline. The blue line corresponds to the scaled rate estimates in Eq. (3.33) while the orange curve corresponds to un-scaled rate estimates with $S_{new/old} = 1$

et al., 2023), as in:

$$p(m_1, m_2, \chi_{1z}, \chi_{2z}|H_a) \propto m_1^{-\beta_{1,a}} m_2^{\beta_{2,a}} \Theta(m_1 - m_2) \begin{cases} 1 & (m_1, m_2, \chi_{1z}, \chi_{2z}) \in S_a \\ 0 & o.w. \end{cases} \tag{3.32}$$

. Here, S_a are hyper-cubes in the space of the parameters $(m_1, m_2, \chi_{1z}, \chi_{2z})$ whose edges correspond to category-specific upper and lower boundaries. The redshift distribution of these injections is taken to be uniform in co-moving volume upto a category-specific maximum ($z_{max,a}$) (Chaudhary et al., 2023). The values of the power-law indices and the maximum redshifts corresponding to each category are listed in table 3.3. As before, the boundaries in parameter space separating BNS from NSBH and NSBH from BBH are determined by the maximum NS mass and spin. For

the population of simulated sources used in the MDC, these values were taken to be $2.05\ M_\odot$ and 0.4 respectively. Masses and spins of BH components were allowed to go upto $100\ M_\odot$ and 0.99 respectively for BBHs and upto $60\ M_\odot$ and 0.99 for NSBHs while no NSs lighter than $1M_\odot$ were drawn.

Category	β_1	β_2	z_{max}
BBH	2.35	1	1.9
BNS	0	0	0.15
NSBH	-1	0	0.25

Table 3.3: The powerlaw indices characterizing the mass-distribution of simulated sources corresponding to each category.

Given these population distributions, the template weights corresponding to each of them can be computed using Eqs. (3.27). The variation of these template weights with SNR, masses, and spins are discussed in A.3.

To compute $\mathrm{P}_{\mathrm{astro}}$ for the simulated sources, we also need the rate of injections which is significantly higher than the astrophysical rates listed in table 3.1. The rate of detectable injections can be estimated from the total number of injections drawn and the maximum redshift corresponding to each category, as in:

$$\frac{\langle R_{r,a}\rangle}{\langle R_0\rangle}|_{inj} = \frac{N_{inj,a}}{V(z_{max,a})T_{inj}}\,\langle VT_{old,a}\rangle\,S_{r/old} \tag{3.33}$$

, where $N_{inj,a}$ is the total number of injections drawn from category a, $S_{r/old}$ the scale factor derived in Eq. (3.22) and $V(z)$ is the co-moving volume upto redshift z. Armed with the rates in Eq. (3.33) and the template weights corresponding to the population distributions in Eq. (3.32), we can compute $\mathrm{P}_{\mathrm{astro}}$ for all the simulated sources in the MDC that are found by the search pipeline. We can then compare the cumulative sum of $\mathrm{P}_{\mathrm{astro}}$ values with the number of found injections for validating our rate-approximation and source classification schemes.

To demonstrate the improvements resulting from our rate-approximation scheme, we compute

two sets of $\mathrm{P_{astro}}$ values, one using rate estimates obtained from the correct values of $S_{r/old}$ and the other using rates that correspond to $S_{r/old} = 1$. We then compare the cumulative sum of $\mathrm{P_{astro}}$ values $(\sum_i \sum_a P_a(d_i))$ corresponding to both sets of rate estimates against the total number of found injections. The results of said comparison are summarized in Fig. 3.8.

In the case of perfect signal recovery, the plotted curve would lie exactly atop the diagonal. Fig. 3.8 thus validates that our formalism yields $\mathrm{P_{astro}}$ values that correspond to a near-perfect recovery of all simulated sources with very few injections lost to terrestrial. It also demonstrates how the rate-approximation part of our formalism prevents the underestimation of $\mathrm{P_{astro}}$ since the $\mathrm{P_{astro}}$ values generated using un-scaled rate estimates are significantly below the diagonal than the scaled ones.

To now demonstrate the accuracy with which our template weights can classify triggers into various astrophysical categories we compute the fraction of simulated sources from each category a that are classified into category b. The value of this fraction can be obtained from the cumulative sum of category-specific $\mathrm{P_{astro}}$ values of the simulated events in the following way:

$$f(b|a) = \frac{\sum_{i \in a} P_b(d_i)}{N_a} \tag{3.34}$$

where $i \in a$ represents the the trigger corresponding to the ith simulated event from source category a and N_a is the total number of simulated events from a that are recovered by the search pipeline with $FAR \geq FAR_0$. To compute the fraction of simulated sources lost to the terrestrial category, we use $f(\mathrm{Terrestrial}|a) = \left[\sum_i (1 - \sum_b P_b(d_i))\right] / N_a$.For the simulated astrophysical sources and the noise triggers in the MDC, this set of fractions obtained from our $\mathrm{P_{astro}}$ estimates are represented in Fig. 3.9.

We can see in Fig. 3.9 that the majority of simulated sources are correctly classified with a very small fraction of miss-classifications. For the BNS injections, we find that P_{BNS}, P_{NSBH} and $P_{\mathrm{Terrestrial}}$ accounts for 73%, 25% and 2% of them respectively with a negligible fraction lost

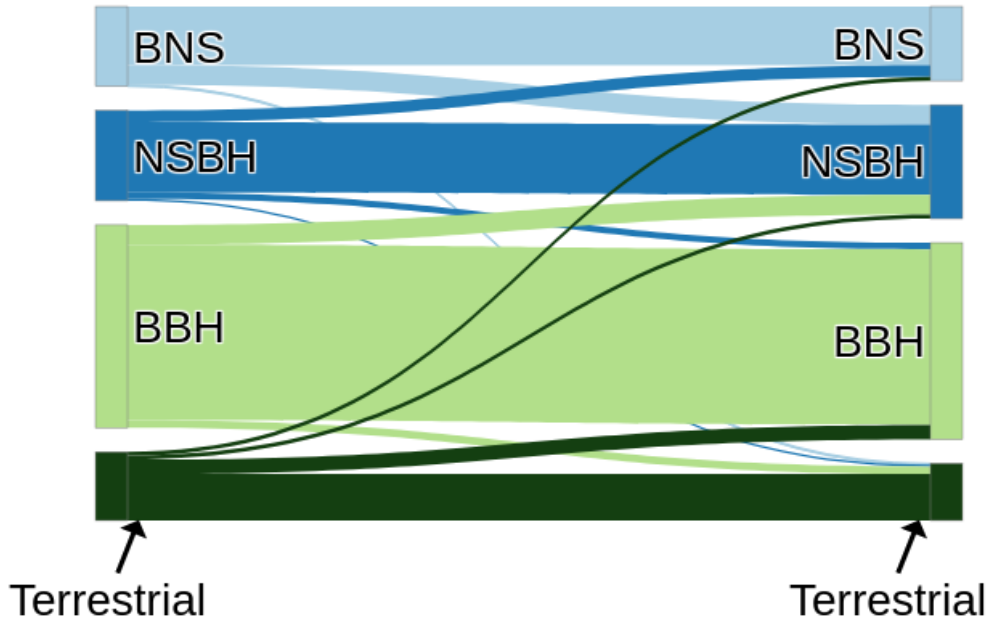


Figure 3.9: Classification of simulated sources. In this Sankey diagram, the size of each link is representative of $f(b|a)$ with a on the left and b on the right.

to BBH. Similarly, for the NSBH injections, we find that 77%, 13% and 8% are classified as NSBH, BNS, and BBH respectively with only 2% lost to terrestrial. 86% BBHs are correctly classified with 10% and 4% injections recovered as NSBH and terrestrial. Note that very few of the potentially EM bright sources (BNS and NSBH) are lost to BBH or terrestrial.

To summarize, we conclude that Figs 3.9 and 3.8 are implicative of the following. In a simulated universe with known populations of BBH BNS and NSBH systems that merge at some known astrophysical rate per comoving volume per time, our formalism can be used to compute $\mathrm{P}_{\mathrm{astro}}$ in real-time that enable highly accurate source classification in the presence of noisy data. Furthermore, we conclude that the rate approximation scheme developed as part of our formalism is a necessity since a significant fraction of simulated sources has been shown to be misclassified

as terrestrial without it.

3.2.4 Conclusion and future prospects

We have developed a self-consistent analysis framework capable of accurately classifying GW candidates in real-time, for the purpose of assisting multimessenger follow-up observations during O4. We have implemented our formalism through the GstLAL search pipeline while noting that our rate approximation scheme is applicable to other pipelines as well. We have demonstrated the accuracy of our classification scheme and the improvements resulting from our rate approximation scheme, both through the analysis of real and simulated GW data as part of an MDC. With the robust and accurate $\mathrm{P}_{\mathrm{astro}}$ values yielded by our framework, the GstLAL search pipeline can be expected to better assist in multimessenger follow-up campaigns to the GW alerts sent out during O4, as compared to O3.

While our formalism is an improvement over existing frameworks for calculating $\mathrm{P}_{\mathrm{astro}}$, further developments leading to even better source classification can potentially be implemented in the near future. For example, even though our template weights facilitate the correct classification of most GW candidates, they are still based on point estimates of CBC parameters computed by the search pipelines and hence susceptible to potential biases. On the other hand, Bayesian parameter estimation (PE) has been shown to yield estimates of these quantities that are significantly more accurate than the point estimates computed by search pipelines. Given the recent developments in the field of low-latency PE in real-time (Pankow et al., 2015; Rose et al., 2022), it is possible to replace our template-weights-based classification scheme with one that makes use of more accurate mass and spin measurements yielded by said PE. The related developments and the associated validation studies are part of an ongoing exploration.

To summarize, we have implemented several improvements in the low-latency computation of category-specific $\mathrm{P}_{\mathrm{astro}}$ values by the GstLAL search pipeline for assisting in the search for coun-

terparts to GW candidates in O4. Given the performance of our analysis framework in an MDC comprising real and simulated GW data, we conclude that these improvements will lead to the correct identification and classification of several GW candidates in O4 that would have otherwise been marked as terrestrial or miss-classified into an incorrect astrophysical category. While our classification scheme is susceptible to uncertainties in the population model, these uncertainties are expected to decrease as more and more events are found, leading to further improvements in classification in the later parts of O4 and beyond.

The entirety of this section is communicated in the form of a pre-print (Ray et al., 2023d) and will be submitted to a journal shortly.

3.3 UPDATING LOW-LATENCY TRIGGER CLASSIFICATION USING RAPID PARAMETER ESTIMATION

The real-time trigger classification framework described in the previous section depends on the CBC parameters of the best-matched template corresponding to a trigger, which are unreliable point estimates. While the template weights model the noise-induced template migration reasonably well, further improvements in trigger classification accuracy can be achieved by incorporating information from Bayesian PE which provides the most reliable estimates of CBC parameters and their measurement uncertainties. However, the high latency of traditional PE pipelines (Smith et al., 2020; Veitch et al., 2015) makes such updates to real-time $\mathrm{P}_{\mathrm{astro}}$ values practically useless.

Recently, Rose et al. (2022) and Pankow et al. (2015) have proposed a rapid Bayesian PE framework that relies on parallelized likelihood computation on grid points of intrinsic CBC parameters, efficient MCMC marginalization of extrinsic parameters and adaptive mesh-refinement to implement accurate PE within a matter of minutes. The `rapid-PE`-based CBC mass measurement is therefore a natural candidate for updating the search-based online $\mathrm{P}_{\mathrm{astro}}$ values to higher classification accuracy.

In Sec. 2.4.1, I described a method for updating pipeline $\mathrm{P}_{\mathrm{astro}}$ values in the most general sense possible, by self-consistently marginalizing over measurement uncertainties in the CBC parameters and their population distribution. However, due to the unavailability of the complete list of events to be found at the end of an observing run, such an analysis cannot be implemented in real-time trigger classification. Nevertheless, by combining the fixed straw-person population model and rate approximation scheme described in the previous section with `rapid-PE` based mass-measurements, it is possible to implement a $\mathrm{P}_{\mathrm{astro}}$ updating scheme that marginalizes over uncertainties in the measurements of CBC parameters to achieve one of the most accurate fixed population $\mathrm{P}_{\mathrm{astro}}$ computations possible.

In particular, we only intend to update the multi-component classification into astrophysical categories by incorporating accurate PE-based mass estimates. The probability of terrestrial origin is something we don't intend to contaminate with any PE-based information since the latter relies on the stationary Gaussian noise assumption and is hence unable to account for the existence of glitches/noies transients. To implement such a $\mathrm{P}_{\mathrm{astro}}$ computation, we can start by re-arranging Eq. (3.21) to obtain:

$$P_a(d_k|\vec{d}_{N_r}) = (1 - P_{\mathrm{Terr}})\frac{p(d|H_a)\left\langle R_a\right\rangle_{N_r/k}}{\sum_b p(d|H_b)\left\langle R_b\right\rangle_{N_r/k}}, \tag{3.35}$$

where

$$P_{\mathrm{Terr}} = 1 - \sum_a P_a(d_k|\vec{d}_{N_r}) = \frac{p(d|H_0)\left\langle R_0\right\rangle_{N_r/k}}{p(d|H_0)\left\langle R_0\right\rangle_{N_r/k} + \sum_b p(d|H_b)\left\langle R_b\right\rangle_{N_r/k}} \tag{3.36}$$

is the terrestrial probability returned by the search pipeline.

To improve classification accuracy in a manner that does not affect the terrestrial probability, it is desirable to replace $p(d|H_a)$ in Eq. (3.35) with PE-based evidences while keeping P_{Terr} unchanged from its search pipeline values. The PE-based evidence can be obtained in the following

way:

$$p(d|H_a) \equiv Z_a = \int p(d|\theta, F)p(\theta|H_a), \tag{3.37}$$

where $p(d|\theta, F)$ is the PE likelihood under the Gaussian noise assumption and $p(\theta|H_a)$ is the fixed population model used for classification throughout this chapter. Using the PE-based evidences and the rate expression derived in the previous section, we get the following expression:

$$P_a^{PE}(d_i) = (1 - P_{\rm Terr}^{pipe})\frac{\langle \mathcal{R}_a \rangle \times \langle VT \rangle_{a,old} \times Z_a}{\sum_b \langle \mathcal{R}_b \rangle \times \langle VT \rangle_{b,old} \times Z_b}, \tag{3.38}$$

where the scaling factor S_r/old has canceled out, and I have added the superscripts PE and $pipe$ to distinguish between PE and pipeline-based quantities. Improving real-time source classification estimates using this expression requires fast evidence computation schemes which will depend on the rapid PE implementation used.

3.3.1 Evidence computation using gridded likelihoods

In the rapid PE framework of Rose et al. (2022), the PE likelihood marginalized over extrinsic parameters is evaluated at various grid points of intrinsic parameters with grid resolution near the highest likelihood regions of the parameter space refined adaptively after each iteration. The marginalized likelihood as a function of CBC masses is then obtained by interpolating over grid points using a Gaussian interpolant:

$$p(d|m_1, m_2) = \sum_i^{N_{grid}} \mathcal{L}_i(d) \exp\left\{-\frac{1}{2\sigma^2_{\tau_1,i}}[\tau_1(m_1, m_2) - \tau_{1,i}]^2 - \frac{1}{2\sigma^2_{\tau_2,i}}[\tau_2(m_1, m_2) - \tau_{2,i}]^2\right\} \tag{3.39}$$

where τ_1, τ_2 are re-parameterizations of the component masses, $\tau_{1,i}, \tau_{2,i}$ are the coordinates of the ith grid point and $\mathcal{L}_i(d)$ is their marginalized likelihood. Naively, one would attempt to evalu-

ate the evidence integral as a Monte-Carlo sum over m_1, m_2 samples drawn from $p(m_1, m_2|\mathcal{H}_\alpha)$. However, if the Gaussian functions are narrow, that would lead to highly inefficient integration. Instead, one could draw samples from the Gaussians at each grid point(after appropriately normalizing them), for implementing the Monte Carlo integral. The evidence then becomes:

$$Z_a = \frac{1}{N_s} \sum_{i}^{N_{grid}} \sum_{j}^{N_s} 2\pi\sigma_{\tau_1,i}, \sigma_{\tau_2,i}\mathcal{L}_i(x)p(m_1(\tau_{1,j}, \tau_{2,j}), m_2(\tau_1^{ij}, \tau_2^{ij})|\mathcal{H}_\alpha)J^{ij}, \tag{3.40}$$

where

$$(\tau_1^{ij}, \tau_2^{ij}) \sim \frac{1}{2\pi\sigma_{\tau_1,i}, \sigma_{\tau_2,i}} \exp\left\{-\frac{1}{2\sigma^2_{\tau_1,i}}[\tau_1(m_1, m_2) - \tau_{1,i}]^2 - \frac{1}{2\sigma^2_{\tau_2,i}}[\tau_2(m_1, m_2) - \tau_{2,i}]^2\right\} \tag{3.41}$$

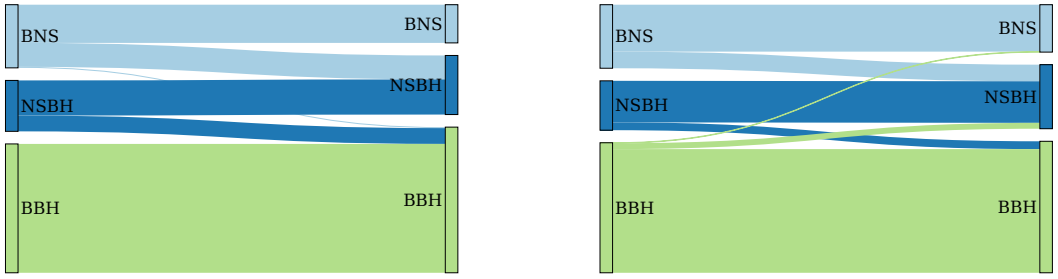
are samples drawn from the Gaussians at the ith grid point, and

$$J^{ij} = \left|\frac{\partial(m_1, m_2)}{\partial(\tau_1, \tau_2)}\right|_{\tau_1^{ij}, \tau_2^{ij}} \tag{3.42}$$

is the Jacobian for the transformation $(\tau_1, \tau_2) \rightarrow (m_1, m_2)$. Using this evidence approximation scheme, it is possible to compute $\mathrm{P}_{\mathrm{astro}}$ values for the simulated events of the MDC described in the previous section and compare those with the search pipeline-based estimates.

This formalism of `rapid-PE` based source classification is implemented as part of the `rapid-PE-RIFT` pipeline[3] and its development was led by Caitlin Rose, myself, and Vinaya Valsan, in collaboration with several others. The performance of our pipeline on MDC data demonstrated considerable improvement in classification accuracy over the search pipelines. Figure 3.10 shows astrophysical-only Sankey diagrams for `GstLAL` and `rapid-PE` which were made from $\mathrm{P}_{\mathrm{astro}}$ values computed using the same counts for both pipelines. The improvement in classification is

[3]The actual implementation of the evidence integral is slightly different. Instead of summing over the values of the integrand at the samples, those values are compared with uniform draws and the integral is evaluated as the acceptance ratio times the range of the integrand values.



(a) GstLAL (b) Rapid-PE

Figure 3.10: Comparison GstLAL and rapid-PE P_{astro} values for one week of MDC data. The same rates are used for both the figures, which were calculated by manually counting found injections. Note that the left plot is not representative of GstLAL's full classification pipeline which was represented in figure 3.9 of the previous section. This is instead a demonstration of how much rapid-PE improves classification in a particular situation where identical counts are used.

clearly visible, with quantitative estimates available in Table 3.4.

True Source	Percentage		
	BNS	**NSBH**	**BBH**
BNS	58.4	39.4	2.1
NSBH	0.0	69.3	30.7
BBH	0.0	0.0	100.0

(a) GstLAL

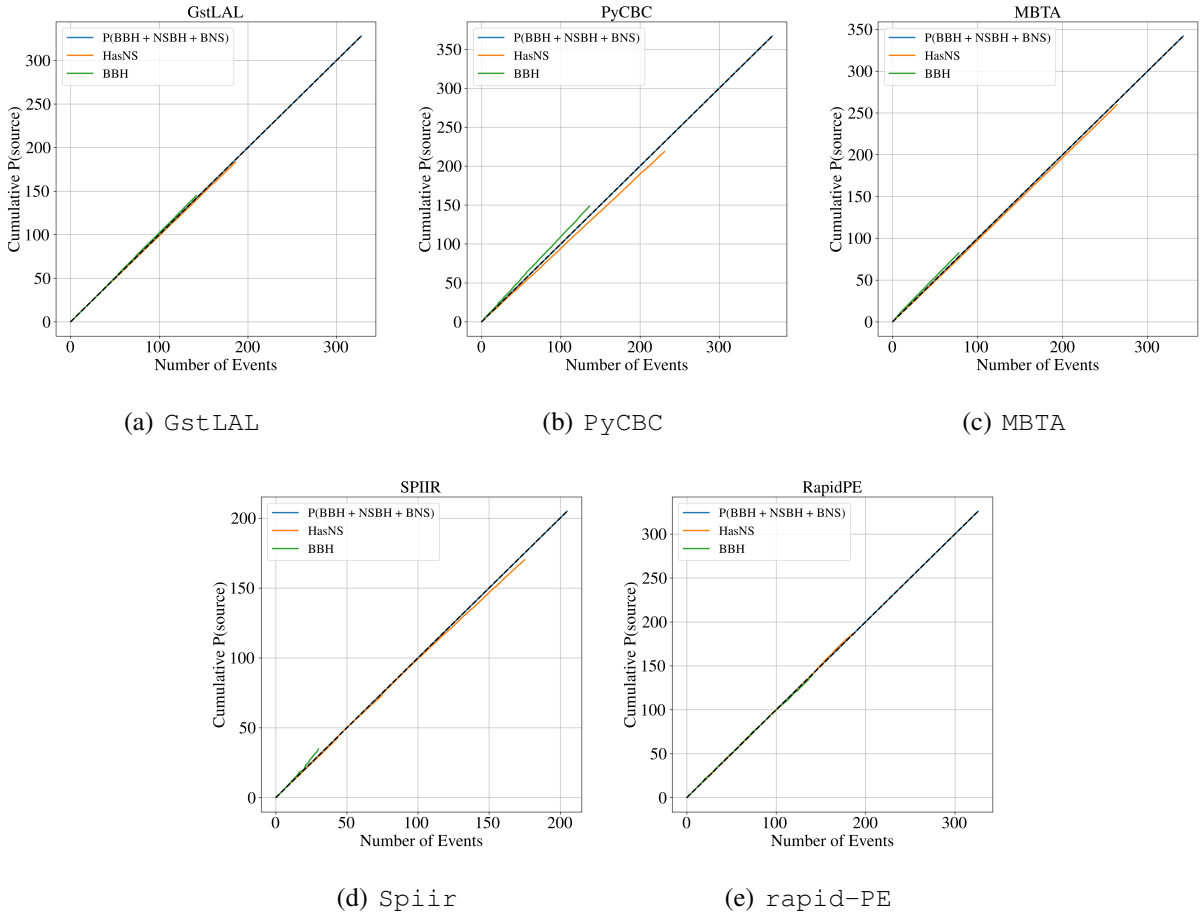
True Source	Percentage		
	BNS	**NSBH**	**BBH**
BNS	73.9	26.1	0.0
NSBH	0.8	83.2	16.0
BBH	0.7	4.8	94.5

(b) rapid-PE

Table 3.4: Quantitative Comparison GstLAL and rapid-PE P_{astro} values. The percentage of simulated sources from one category being recovered by other categories.

In addition to judging the performance of P_{astro} values based on classification fractions, another useful metric for assessing classification accuracy is the probability of correctly classifying an NS containing CBC. In other words, one can attempt to count how many NS containing CBCs have a high P_{BNS} or P_{NSBH} as compared to high P_{BBH} since astronomers decide to follow up GW candidates with both high P_{BNS} or high P_{NSBH}. The performance of classifiers in this metric can be assessed by plotting the probability $P(HasNS) = P_{BNS} + P_{NSBH}$ cumulated over the

simulated triggers found during the MDC vs the true number of simulated signals that had an NS containing component. Figure 3.11 demonstrates that the `rapid-PE` based $\mathrm{P_{astro}}$ values correspond to the almost near-perfect recovery of NS containing CBCs, greater than all of the four low-latency search pipelines.



(a) `GstLAL` (b) `PyCBC` (c) `MBTA`

(d) `Spiir` (e) `rapid-PE`

Figure 3.11: Comparison of $P(HasNS) = P_{BNS} + P_{NSBH}$ for `rapid-PE` $\mathrm{P_{astro}}$ values with that of other search pipelines. The y-axes represent cumulative probabilities while the x-axes represent the count of true signals, with the diagonal line representing perfect classification.

This development of the `rapid-PE` based $\mathrm{P_{astro}}$ computation and its performance studies will be communicated in the near future (Rose et al., In prep.). However, the algorithm itself is already being used by the LVK to update search pipeline $\mathrm{P_{astro}}$ values for GW triggers found above the PE

FAR threshold of 1 in 7 months, with the updated values sent out as part of secondary low-latency alerts and GCN circulars.

One principle caveat of this framework, akin to the search pipelines, is that a source frame population model is being applied to estimates of detector frame masses for implementing mass-based classification. While in standard PE, measurements of luminosity distance and a fixed cosmological model are used to convert detector frame quantities into source frame, it is not straightforward to reconstruct the distance posteriors from `rapid-PE`'s likelihoods, which are marginalized over distance at each intrinsic grid point. However, given that we are mostly interested in the classification of low-mass events and that CBCs near the NS-BH boundary will only be observable up to negligible redshifts, we do not expect this limitation of be much of a hindrance, as demonstrated by the MDC results. Nevertheless, it is still beneficial to implement distance reconstruction and conversion to source frame which is part of an ongoing investigation. On the other hand, distance measurement is straightforward in other low-latency PE pipelines, which can be used to validate `rapid-PE`'s $\mathrm{P_{astro}}$ values, albeit with a higher latency.

3.3.2 Evidence computation using Posterior samples

The PE-based source classification framework discussed so far can be easily extended to the other PE pipelines such as `bilby` (Ashton et al., 2019; Romero-Shaw et al., 2020) that often rely on techniques like focused reduced order quadrature (Morisaki & Raymond, 2020) to obtain posterior samples of component masses and luminosity distances with a latency of around 5-10 minutes for low mass and slowly spinning sources.

Given posterior samples of detector frame masses and luminosity distances, one can use Bayes theorem on Eq. (3.37) to evaluate the evidence integral as a Monte Carlo sum, such as:

$$Z_\alpha = \frac{Z_{PE}}{N_{samp}} \sum_{i\sim\text{posterior}}^{N_{samp}} \left[\frac{p(\frac{m_{1d,i}}{1+z_i}, \frac{m_{2d,i}}{1+z_i}, z_i|H_a)}{p(m_{1d,i}m_{2d,i}, d_{L,i}|F) \times \frac{dd_L}{dz}\frac{1}{(1+z_i)^2}} \right]_{z_i=f_z(d_{Li},\Omega)} . \tag{3.43}$$

Here $m_{1d,i}, m_{2d,i}, d_{L,i}$ are the posterior samples of detector frame masses and luminosity distances yielded by the PE pipeline, $p(m_{1d,i}m_{2d,i}, d_{L,i}|F)$ the prior used by PE to sample the likelihood, Z_{PE} the Bayesian evidence evaluated by PE, and Ω represents a fixed cosmological model that is used to convert detector frame quantities into source frame.

An important difference between this method and the `rapid-PE` based one is that here, we can correctly apply the fixed population model on source frame quantities. However, the PE-pipelines that can provide the required posterior samples are slower than `rapid-PE`, particularly for higher mass events, for which the PE will take much longer. Hence this method cannot be straightforwardly implemented on triggers corresponding to all low-latency alerts in its current stage and needs further development prior to its sanctioned use on live data.

3.3.3 Summary

In this section, I summarize my contributions to the development of PE-based classifiers that can upgrade the search pipeline's $\mathrm{P_{astro}}$ values to much higher accuracies. Accurate categorization of GW triggers is crucial in optimizing the search for multi-messenger counterparts to CBCs. The PE-based classifiers that I have co-developed significantly increase the chances of correctly alerting astronomers about potential multi-messenger candidates and reduce the risk of wasted efforts on BBH candidates substantially. These algorithms will undoubtedly play a central role in the detection of multimessenger counterparts to GW sources in the ongoing and upcoming observing runs of the LVK and hence facilitate, in a way, the major discoveries that these detections will entail.

The documentation of this method and the results of the MDC investigation will be communicated shortly (Rose et al., In prep.).

Part III

Astrophysics and Cosmology with gravitational waves

Chapter 4

Astrophysical inference from a population of compact binary mergers

When you change the model and make it more flexible or make different assumptions, you get a different answer about how black holes formed in the universe.

Sylvia Biscoveanu

4.1 INTRODUCTION AND MOTIVATION

The formation and evolution of compact object binaries that merge within a Hubble time via GW emission, particularly stellar-mass binary black holes (BBHs), is poorly understood to date (e.g., Mapelli, 2020b). The primary goal of GW-based population inference is to constrain the uncertain parameters, and initial conditions of the physical mechanisms that underlie various proposed formation channels by modeling how they characterize and correlate the observable population properties of the resulting CBC systems. *However, the astrophysical implications of GW-based population inference are sensitive to modeling choices* and hence require validation from model-independent characterizations of the CBC population that rely more on the data than on a priori assumptions to construct and constrain the functional form of the population distribution. In addition to verifying the findings of strongly model-dependent population inference, data-driven methods can extract previously un-modelled features in the population, which if found, might be indicative of new and exotic physics governing CBC formation.

Several non-parametric population inference methods have been proposed in literature that are independent of the astrophysical predictions of known formation scenarios, each with its own mer-

its and demerits. Compared to parametric approaches, these flexible ones are characterized by a significantly higher number of free parameters which often leads to exaggerated computational cost. This problem is worsened by the curse of dimensionality, which implies a rapidly growing computational cost with the number of CBC parameters whose populational trends are being modeled. In other words, when searching for population-level correlations between CBC parameters, highly flexible models can potentially become intractable to implement. For most non-parametric population models, it is prohibitive to search for higher dimensional correlations without loosing flexibility by means of strong assumptions on the nature of the correlations or the marginal distributions (e.g. Callister & Farr, 2024).

However, correlations in the CBC population can exclusively elucidate several outstanding questions in modern astrophysics and therefore need to be explored beyond the limitations of parametric population modeling. The development of a scalable yet highly data-driven population inference method for constraining various kinds of correlations in the compact binary population has been a major component of my thesis work, which I shall delineate throughout the rest of this chapter. I will further discuss the new astrophysical results that I have found by using the newly developed population models on GWTC-3 data.

4.2 THE BINNED GAUSSIAN PROCESS POPULATION MODEL

One of the simplest yet highly flexible population models that one can think of is a multi-dimensional histogram. Such a model can be constructed by binning up the space of CBC parameters whose population distribution we are interested in, with the merger rate density in each bin being a free parameter to be constrained by the data. The hierarchical likelihood of Eq. (2.49) can be constructed using this binned model and a suitably chosen prior on the rate densities can then be used to construct their posterior distribution.

In order to infer the shape of the joint population distribution of CBC parameters from sparse

data sets while being mostly independent of the binning choices, the histogram is regularized using a Gaussian process (GP) prior on the rate densities (Mohite, 2022; Foreman-Mackey et al., 2014; Mandel et al., 2017). This is achieved by choosing a Multi-variate normal distribution prior on $\vec{n}$, whose covariance matrix is a function of the distance between the bins in parameter space. The correlation length and amplitudes of the GP are themselves inferred from the data using suitable priors which in turn leads to stability against variations of binning choices after a certain level of resolution is achieved.

The primary merit of this binned Gaussian process model is its scalability to higher dimensions. As I discuss further in appendix B.7. the computational cost of this analysis scales additively with an increase in the number of dimensions, making it suitable for non-parametric explorations of the high dimensional correlations in the CBC population. Furthermore, as discussed in appendix B.1, the integrals over event-specific posterior samples and detectable simulations required to construct the hierarchical likelihood of Eq. (2.49) are pre-computable for the BGP model, leading to rapid inference even on central processing units, for as many as $O(2000)$ bins.

After realizing this potential of BGP models, I immediately started generalizing the two-dimensional models of Mohite (2022), into higher dimensional ones such as those of the joint mass-redshift and mass-spin distributions of CBCs. Analyzing GWTC-3 data with these models led to new astrophysical results with far-reaching implications, particularly for the analysis of mass-spin distributions. In the next couple of sections, I will summarize these investigations.

4.3 DATA-DRIVEN POPULATION INFERENCE FOR BINARY BLACK HOLES: MASS AND REDSHIFT DISTRIBUTIONS

The observation of gravitational waves from multiple compact binary coalescences by the LIGO-Virgo-KAGRA detector networks has enabled us to infer the underlying distribution of compact binaries across a wide range of masses, spins, and redshifts. In light of the new features found in

the mass spectrum of binary black holes and the uncertainty regarding binary formation models, non-parametric population inference has become increasingly popular. In this work, we develop a data-driven clustering framework that can identify features in the component mass distribution of compact binaries simultaneously with those in the corresponding redshift distribution, from gravitational wave data in the presence of significant measurement uncertainties, while making very few assumptions on the functional form of these distributions. Our generalized model is capable of inferring correlations among various population properties such as the redshift evolution of the shape of the mass distribution itself, in contrast to most existing non-parametric inference schemes. We test our model on simulated data and demonstrate the accuracy with which it can re-construct the underlying distributions of component masses and redshifts. We also re-analyze public LIGO-Virgo-KAGRA data from events in GWTC-3 using our model and compare our results with those from some alternative parametric and non-parametric population inference approaches. Finally, we investigate the potential presence of correlations between mass and redshift in the population of binary black holes in GWTC-3 (those observed by the LIGO-Virgo-KAGRA detector network in their first 3 observing runs), without making any assumptions about the specific nature of these correlations.

4.3.1 Motivation

The first direct observation of gravitational waves (GWs) from the binary black hole (BBH) merger GW150914 (Abbott et al., 2016a) by the Laser Interferometric Gravitational Wave Observatory (LIGO, J. Aasi et al. (2015)), has opened up a new window to the universe. Since then, the LIGO-Virgo-KAGRA (LVK, J. Aasi et al. (2015); Acernese et al. (2014); Akutsu et al. (2021)) detector network has observed about 70 BBHs with a false alarm rate (FAR) of less than 1 per year (Abbott et al., 2023b). Studying the ensemble of BBHs comprised by these detections has facilitated several important investigations such as the exploration of stellar evolution and binary

formation mechanisms (Abbott et al., 2023d), measurement of cosmological parameters (Abbott et al., 2023a), and tests of General Relativity in the strong gravity regime (Abbott et al., 2021h).

In particular, studying the population-level distributions of CBC masses, spins, and redshifts through the third Gravitational-Wave Transient Catalog (GWTC-3, Abbott et al. (2023b)) has allowed us to empirically probe several models that describe the astrophysical processes responsible for compact binary formation. For example, the existence of a steep fall-off in the BBH mass spectrum beyond $50M_{\odot}$ (Abbott et al., 2023d; Fishbach & Holz, 2017; Edelman et al., 2021) is indicative of the pair-instability process limiting the mass of stellar cores (Fowler & Hoyle, 1964; Barkat et al., 1967; Heger & Woosley, 2002; Heger et al., 2003; Woosley & Heger, 2015; Belczynski et al., 2016; Woosley, 2017; Marchant et al., 2019; Renzo et al., 2020). The mass-range near which this truncation happens has itself been shown to be informative on nuclear reaction rates in massive stars (Farmer et al., 2020). Similarly, the observed peak in the BBH mass-spectrum near the $30M_{\odot} - 40M_{\odot}$ range (Abbott et al., 2023d, 2021g, 2019a) has been thought to result from the pile up due to pulsational pair-instability supernovae (50truncreason7 Talbot & Thrane (2018); Woosley (2017)), with the location of the peak expected to be insensitive to stellar metalicity and hence redshift (Farmer et al., 2019). Additional sub-structure found in the BBH mass-spectrum, in the form of peaks and dips atop a smoothed power-law (Edelman et al., 2022a; Tiwari & Fairhurst, 2021; Tiwari, 2022; Edelman et al., 2023), has enabled us to constrain the relative contributions of different formation channels to the BBH population of the universe. On the other hand, studying the population-level distribution of BBH redshifts through GWTC-3 has led to the discovery that the BBH merger rate increases with redshift (Abbott et al., 2023d; Fishbach et al., 2021; van Son et al., 2022a; Karathanasis et al., 2023; van Son et al., 2022c; Fishbach et al., 2018), shedding light on the metalicity evolution and the star-formation history of our Universe. Hence, inferring the population properties of CBCs using GW measurements of their system parameters has been highly impactful to our understanding of several astrophysical processes that take place in the

universe.

LVK's fourth observing run (O4), can be expected to triple the number of observed BBHs by the end of its first half alone (Abbott et al., 2018b). Hence, analyzing the cumulative catalog of BBH observations post-O4 can enable us to precisely constrain several features in the BBH mass spectrum and redshift distribution. *Parametric* models that assume a specific functional form of the population-level distribution being inferred (Abbott et al., 2023d, 2021g; Talbot & Thrane, 2018; Wysocki et al., 2019; Edelman et al., 2021; Tiwari, 2022; Farah et al., 2022), can thus be used to constrain the parameters characterizing these functions with unforeseen precision given an O4-sized dataset, potentially leading to confident empirical validation of the astrophysical assumptions behind such models.

However, a parametric model is inherently restricted in the sense that it is unable to infer the existence of previously un-modeled features in the underlying population that are beyond its assumptions regarding the functional form of the population distribution. Furthermore, particular features in the functional form assumed by parametric population models can often manifest themselves as a posteriori even when the data strongly disfavors their existence (Callister & Farr, 2024; Callister et al., 2022), potentially leading to model-induced *false alarms* in the conclusions drawn from parametric population inference. Hence, given the uncertainties regarding the true population distribution and the motivation to search for new physics beyond the assumptions built into existing parametric models, data-driven population inference with minimal suppositions regarding the underlying population is highly important for O4 and beyond.

Several model-independent inference frameworks studied in existing literature, are able to extract features in the CBC population distribution from GW data, without assuming a priori the precise nature and location of these features. Some examples include population modelling based on autoregressive processes (Callister & Farr, 2024), splines (Edelman et al., 2022a, 2023), Gaussian mixture models (Tiwari, 2021; Tiwari & Fairhurst, 2021; Tiwari, 2022), adaptive kernel density

estimation (aKDE, Sadiq et al. (2022)), maximum population likelihood (Payne & Thrane, 2023), Dirichlet processes (Rinaldi & Del Pozzo, 2021) and binned Gaussian processes (GPs, Foreman-Mackey et al. (2014); Mandel et al. (2017); Abbott et al. (2023d); Mohite (2022)). While these methods have their individual pros and cons, apart from the aKDE-based one, they all implement a restrictive inference that does not allow for generic *correlations* between the CBC mass and redshift distributions. While most of the other methods can infer on the redshift evolution of the combined merger rate either individually or simultaneously with the mass population, they are all based on the simplifying assumption that the shape of the mass spectrum itself does not evolve with redshift. This artifact, when built into a population analysis framework, can prevent it from exploring several astrophysical phenomena, that in fact predict correlations between the shape of the BBH mass spectrum and the distribution of BBH redshifts (Fishbach et al., 2021; van Son et al., 2022a). On the other hand, the existing implementation of the aKDE method has been used to search for mass-redshift correlations only in the detectable population of CBCs, i.e. without accounting for selection biases (Sadiq et al., 2022). Hence, constraints inferred by Sadiq et al. (2022) cannot be used directly to probe the aforestated astrophysical phenomena without first converting them into constraints on the astrophysical population (as opposed to the detectable one) by means of appropriately constructed selection functions (Mandel et al., 2019; Vitale et al., 2020; Farr, 2019). Hence, previous studies on the existence of correlations in the astrophysical BBH population (such as the parametrized approaches of Fishbach et al. (2021); Callister et al. (2021); Biscoveanu et al. (2022) or the mixture-model based approaches of Abbott et al. (2023d); Godfrey et al. (2023); Wang et al. (2022); Li et al. (2023)), have mostly been carried out while making strong assumptions regarding the functional form of said correlations, rendering the inferred constraints susceptible to the previously mentioned limitations of parametric population modeling.

In addition, empirically probing the cosmological evolution of several astrophysical processes requires correlated population inference by means of data-driven frameworks that are free of the

limitations of parametric population modeling. For example, the evolution of initial conditions of zero-age main sequence stars with cosmic time (Kudritzki & Puls, 2000; Belczynski et al., 2010; Brott et al., 2011; Fryer et al., 2012; Dominik et al., 2015; Safarzadeh & Farr, 2019; Neijssel et al., 2019; Kinugawa et al., 2020; Farrell et al., 2020; Vink et al., 2021), the preference of dynamical BBH formation environments towards different BH mass ranges at different redshifts (Rodriguez & Loeb, 2018; El-Badry et al., 2018; Santoliquido et al., 2020; Romero-Shaw et al., 2021; Weatherford et al., 2021), the dependence of delay time between isolated BBH formation and merger on the corresponding BH masses (Li et al., 2018; Samsing, 2018; Mapelli et al., 2019), and the variation of the relative contributions of different BBH formation channels with redshift (Rodriguez & Loeb, 2018; Rodriguez et al., 2019; Santoliquido et al., 2020; Yang et al., 2020; Zevin et al., 2021), all predict a BBH mass spectrum whose shape correlates strongly with the corresponding redshift distribution. Many of these predictions are obtained through numerical simulations with the predicted population distribution lacking an obvious well-defined functional form as required for parametric modeling. Hence, in order to fully explore the aforestated astrophysical phenomena that govern BBH formation and evolution, non-parametric analysis schemes capable of inferring the correlations between the BBH mass and redshift populations from GW observations through O4 and beyond are of high importance and significance.

In this work, we develop a model-independent inference framework based on binned Gaussian processes that can infer the population-level distributions of CBC masses and redshifts from GW measurements of these quantities for a sample of BBHs. We allow for correlation between the mass and redshift distributions while also appropriately accounting for selection biases. We demonstrate the accuracy with which our method can constrain the underlying population by simulating astrophysically motivated fiducial populations of BBHs and realistic measurement uncertainties. We also re-analyze public LVK data for events in GWTC-3 and constrain the BBH mass and redshift distribution without *any* prior assumptions on the shape of these distributions. By particularizing to

an uncorrelated model, we show that our method yields constraints that are fully consistent with the fiducial underlying population for simulated sources and those yielded by uncorrelated parametric models for GWTC-3, even though our results show *hints* of new features beyond the standard POWER LAW + PEAK model (Talbot & Thrane, 2018). We then use our generalized model to constrain, for the first time, the correlations between the BBH mass and redshift distributions from GWTC-3 in a non-parametric manner.

We note that our work is *proof of concept* in the sense that we obtain the first non-parametric constraints on the correlations between the population-level distributions of BBH masses and redshifts. Our model thus enables us to probe new physics beyond the scope of existing non-parametric models which either restrict to an uncorrelated mass-redshift inference(Callister & Farr, 2024; Edelman et al., 2022a, 2023; Tiwari, 2021; Tiwari & Fairhurst, 2021; Tiwari, 2022) or attempt to infer those correlations in the detectable population without accounting for selection effects(Sadiq et al., 2022). Hence, straightforward generalizations of our method to simultaneously infer the spin population and other astrophysically significant ensemble properties of CBCs are left as future explorations. Similarly, studying the full CBC mass spectrum using our model through the inclusion of low-mass events in the analyzed datasets is an interesting venture beyond the scope of this paper and is also left for future work.

This paper is organized as follows. In section 4.3.2 we describe the construction of our binned population model, the Gaussian process prior, and the hierarchical inference framework used in conjunction to constrain the CBC population from multiple GW observations. In section 4.3.3 we summarize the results obtained by applying our method to real as well as simulated data and discuss their implications. In section 4.3.4 we conclude by summarizing the implications of our method in the context of O4 and beyond while highlighting potential generalizations which are left as future projects.

4.3.2 Methods

In this section, we develop our data-driven clustering algorithm within the framework of Bayesian hierarchical inference to constrain the population-level distributions of CBC parameters from GW observations. We make very few underlying assumptions regarding the functional form of the GW source population. We achieve this by first constructing a binned Gaussian process model over the aforestated population distribution.

4.3.2.1 Binned Model

In order to cluster across the 3-dimensional parameter space of binary component masses and redshift (m_1, m_2, z), we first divide the space into N_b bins. We then assume that the merger rate density per unit co-moving volume per log-component masses per unit source frame time is constant within a particular bin. Hence the choice of N_b determines the resolution with which our model can distinguish features in the CBC mass-redshift spectrum. For the γth bin, the rate density is thus defined as,

$$n^{\gamma} = \frac{dN^{\gamma}}{d\ln m_1 d\ln m_2 dV_C dt_R} \tag{4.1}$$

where N^{γ} is the number of events with masses and redshift lying in the γth bin, V_c is the comoving volume, and t_R is the source frame time. If we fix the lower and upper edges of the γth bin, denoted by $(l^{\gamma}_{m_1}, l^{\gamma}_{m_2}, l^{\gamma}_{z})$ and $(u^{\gamma}_{m_1}, u^{\gamma}_{m_2}, u^{\gamma}_{z})$ respectively, then the differential fraction of events with masses m_1, m_2 and at redshift z can be written in terms of rate densities as,

$$\begin{aligned}\frac{dN}{dm_1 dm_2 dz}(\vec{n}) \;=\; \sum_{\gamma}\Bigg(& n^{\gamma}\frac{1}{m_1 m_2}\frac{dV_c}{dz}\frac{T_{\rm obs}}{1+z}\Theta(m_1 - l^{\gamma}_{m_1})\Theta(u^{\gamma}_{m_1} - m_1)\Theta(m_2 - l^{\gamma}_{m_2}) \\ & \Theta(u^{\gamma}_{m_2} - m_2)\;\Theta(z - l^{\gamma}_{z})\Theta(u^{\gamma}_{z} - z)\Bigg)\end{aligned} \tag{4.2}$$

where Θ is the Heaviside step function and $T_{\rm obs}$ is the total observation time as measured in the detector frame which relates to the source frame time elapsed during observation $T_R = T_{\rm obs}/(1+z)$. We note that the model in Eq. (4.2) is general in the sense that it can even infer the existence of correlations between the mass and redshift distributions, in contrast to several studies in existing literature (Edelman et al., 2023; Callister & Farr, 2024) that only allow the total merger rate to vary with redshift. Furthermore, binning up a higher dimensional parameter space can allow for the straightforward generalization of our model to infer the distribution of other GW parameters such as spins. However, we note that while the higher dimensional models are straightforward to formulate, their implementation is susceptible to scalability issues, on which we elaborate further in Sec. 4.3.4.

To summarize, for the binned population model in Eq. (4.2) or any higher dimensional generalization thereof, constraining the rate densities from GW data amounts to inferring the functional form of the population-level distributions of CBC parameters up to the resolution limit imposed by our choice of binning.

4.3.2.2 Hierarchical Inference

To constrain the rate densities that characterize our model from multiple GW observations, we employ the framework of Bayesian hierarchical inference (Thrane & Talbot, 2019). Bayesian inference of GW data yields measurements of CBC parameters for each event in the form of posterior samples that can be re-weighted to the population model of interest. Multiple observations can be treated as independent realizations of an inhomogeneous Poisson process and hence combined hierarchically to yield the likelihood of population-level quantities given the combined dataset (Mandel et al., 2019; Loredo, 2004; Wysocki et al., 2019). In the context of our binned model, the joint likelihood of rate densities $\vec{n}$ given data $\vec{d}$ from a collection of $N_{\rm obs}$ observations takes the following

form,

$$p(\vec{d}|\vec{n}) = e^{-N_{\rm det}(\vec{n})} \prod_{i}^{N_{\rm obs}} \left\langle \frac{\frac{dN}{dm_1 dm_2 dz}(\vec{n})}{p_{\rm PE}(m_1, m_2, z)} \right\rangle_{\rm samples,i} \tag{4.3}$$

where $\langle\cdot\rangle_{\rm samples,i}$ denote an average over posterior samples of $(m_1, m2, z)$ obtained from the i'th observation and $N_{\rm det}$ is the number of CBCs expected to be *detectable* as a function of the rate-densities. For details regarding the convergence of the Monte Carlo integrals implemented by the afore-mentioned average over posterior samples see Sec. B.1. The implicit assumption in Eq. (4.3) is that every observation in the analyzed set is astrophysical which requires the imposition of a stringent detection threshold when selecting the list of candidate events to be used in population inference. The existence of this threshold introduces a Malmquist bias in the inferred population since arbitrary draws of (m_1, m_2, z) from Eq. (4.2) are not equally likely to be detectable. Hence integrating the right-hand side of Eq. (4.2) over (m_1, m_2, z) without accounting for selection effects yields a biased estimate of $N_{\rm det}$.

To account for selection biases, we compute $N_{\rm det}$ by simulating a large fiducial population of CBC signals and injecting them into detector-noise realizations. The parameters of simulated events that pass the detection criteria can be re-weighted to our binned population model in order to yield an unbiased estimate of $N_{\rm det}$ (Mandel et al., 2019; Vitale et al., 2020; Farr, 2019) and thus we write,

$$N_{\rm det}(\vec{n}) = \frac{K_{\rm det}}{K_{\rm draw}} \left\langle \frac{\frac{dN}{dm_1 dm_2 dz}(\vec{n})}{p_{\rm draw}(m_1, m_2, z)} \right\rangle_{\rm samples,det} \tag{4.4}$$

where $\langle\cdot\rangle_{\rm samples,det}$ denotes an average over detectable samples of (m_1, m_2, z) and $p_{\rm draw}$ is the fiducial population from which the simulations were generated. The numbers $K_{\rm draw}$ and $K_{\rm det}$ denote the total number of simulations generated and the number of simulated events that pass the detection threshold respectively. For details regarding the uncertainties in empirically estimating $N_{\rm det}$ from simulations, ways of marginalizing over them, and the corresponding accuracy requirements (Farr, 2019) see appendix B.2.

The average over samples in Eqs. (4.3) and (4.4) corresponding to each bin is proportional to the rate density of a given bin with the *constants* of proportionality pre-computable given the relevant samples and a choice of binning. The likelihood in Eq. (4.3) can then be used to infer the rate densities provided a suitable prior has been imposed on them. For details regarding the calculation of the likelihood and the aforementioned constants see appendix B.1

4.3.2.3 Gaussian Process Prior

We choose the prior on logarithmic rate densities to be a stationary Gaussian process so as to regularize and smoothen the inferred population distribution in the case of sparse datasets (Foreman-Mackey et al., 2014; Mandel et al., 2017). We represent this prior in the following way,

$$\ln \vec{n} \sim \mathcal{N}(\vec{\mu}, \Sigma) \tag{4.5}$$

where μ and Σ are the mean and covariance matrix of the Gaussian process. For the covariance matrix, we use an exponential quadratic function,

$$\Sigma_{\gamma\gamma'}(\sigma, \lambda) = \sigma^2 \exp\left(-\frac{(c^\gamma - c^{\gamma'})^2}{2\lambda^2}\right) \tag{4.6}$$

where the σ controls the amplitude of the covariances, λ determines the length scale over which the bins are correlated and c^γ is the $(\log m_1, \log m_2, z)$ co-ordinate of the γth bin center. The set of quantities $(\vec{\mu}, \sigma, \lambda)$ are treated as *hyper-parameters* of the model and are inferred simultaneously with the rate densities.

4.3.2.4 *Hamiltonian Monte Carlo sampling*

The Gaussian process prior and the likelihood together yield the joint posterior distribution of the rate densities and the hyper-parameters which takes the following form,

$$p(\vec{n}, \vec{\mu}, \sigma, \lambda|\vec{d}) \propto p(\vec{\mu}, \sigma, \lambda)p(\vec{n}|\vec{\mu}, \sigma, \lambda)p(\vec{d}|\vec{n}) \tag{4.7}$$

where $p(\vec{n}|\vec{\mu}, \sigma, \lambda)$ is the Gaussian process prior on the rate densities and $p(\vec{\mu}, \sigma, \lambda)$ are priors on the hyper-parameters (chosen to be broad normal, halfnormal and lognormal distributions in μ, σ and λ respectively). The constant of proportionality in Eq. (4.7) is the so-called Bayesian evidence of our binned model in the data and is independent of the rate densities and hyper-parameters. The stochastic sampling of the posterior density using Monte Carlo techniques can thus be carried out without computing the evidence, with the posterior samples of the rate densities being sufficient for constructing Bayesian credible intervals of the CBC population distribution.

However, for high-resolution inference, the number of quantities being sampled simultaneously becomes large leading to an increase in computational cost. In the context of our binned model, the quantities being simultaneously inferred $(\vec{n}, \vec{\mu}, \sigma, \lambda)$ span a $D = 2N_b + 2$ dimensional space. Algorithms such as random walk Metropolis or Gibbs sampling, which scale poorly with the dimensionality of the space being sampled (Neal, 1993; Homan & Gelman, 2014), can potentially render high-resolution population inference computationally prohibitive. For this reason, we sample the posterior in Eq. (4.7) using Hamiltonian Monte Carlo (HMC) (Neal, 2011; Homan & Gelman, 2014), which invokes a computational complexity of $O(D^{5/4})$ per independent sample and that is significantly more tractable than the $O(D^2)$ complexity of random walk Metropolis (Creutz, 1988) or Gibbs sampling (Homan & Gelman, 2014).

We perform HMC sampling by means of the No-U-Turn Sampler (Homan & Gelman, 2014) that improves upon standard HMC by efficiently auto-tuning the step-size, as implemented in the

`PyMC` software library (Salvatier et al., 2016). Once obtained, the stochastic samples of $\vec{n}$ can be used to re-construct credible intervals of the differential merger rate density as a function of (m_1, m_2, z), that is expected to contain the underlying population distribution of these CBC parameters with certain posterior probability.

4.3.2.5 Uncorrelated Inference for Small Datasets

As mentioned before, the generalized population model in Eq. (4.2) allows for and is able to infer correlations between the mass and redshift distributions of CBCs. However, simultaneous inference of a large number of quantities from small datasets such as GWTC-3 (or any sub-population thereof) can be expected to yield uninformative constraints on the marginal distributions. Furthermore, constraints yielded by the fully correlated inference cannot be compared to existing parametric and non-parametric population studies on GWTC-3, all of which restrict to population models comprising uncorrelated mass and redshift distributions. Hence for analyzing GWTC-3 size datasets, it is often preferable to restrict our generalized inference by prohibiting correlations between mass and redshift distributions. This is achievable by slightly re-defining the binned population model,

$$\frac{dN}{dm_1 dm_2 dz}(\vec{n}_m, \vec{n}_z) = \left(\sum_\alpha n_z^\alpha \frac{dV_c}{dz} \frac{T_{\rm obs}}{1+z} \Theta(z - l_z^\alpha)\Theta(u_z^\alpha - z) \right) \times \left(\sum_\beta \frac{n_m^\beta}{m_1 m_2} \Theta(m_1 - l_{m_1}^\beta)\Theta(u_{m_1}^\beta - m_1)\Theta(m_2 - l_{m_2}^\beta)\Theta(u_{m_2}^\beta - m_2) \right) \tag{4.8}$$

where $n_z^\alpha n_m^\beta$ holds the same meaning as the rate density defined in Eq. (4.1). With this correlation-free factoring of the rate densities, the GP prior can also be factored into two inde-

pendent GPs,

$$\ln \vec{n}_m \sim \mathcal{N}(\vec{\mu}_m, \Sigma_m) \tag{4.9}$$

$$\ln \vec{n}_z \sim \mathcal{N}(\vec{\mu}_z, \Sigma_z) \tag{4.10}$$

where both Σ_m and Σ_z are exponentially quadratic while being conditional on the different hyper-parameters: (σ_z, λ_z) and (σ_m, λ_m) respectively. This drastically reduces the number of independent quantities being simultaneously inferred, specifically from $N_b = (N_m - 1)^2(N_z - 1)$ to $N_b = (N_m - 1)^2 + N_z - 1$, where N_m and N_z are the numbers of bin edges along the mass and redshift axes respectively. The smaller number of quantities can be constrained informatively from a GWTC-3 size dataset with the results eligible for straightforward comparison to existing studies that all carry out uncorrelated parametric and non-parametric population modeling. (We note that the $m_1 \geq m_2$ convention that is often adapted in CBC population inference literature, when implemented in the context of our model, leads to $(N_m - 1)^2$ being replaced by $N_m(N_m - 1)/2$)

4.3.2.6 Choice of Binning

Prior to analyzing GW data with our models, we must choose the location and width of the bins along each parameter dimension whose population level distribution we intend to infer. To select from various possible binning choices one must weigh resolution against computational cost. A higher number of bins over the same region of parameter space can potentially lead to the identification of new features in the underlying population with increased resolution, while simultaneously increasing the computational cost of the ensuing analyses. In the context of our framework, drawing from a GP prior incurs a computational cost that scales with the total number of bins cubed. For the generalized correlated inference, since the number of bin edges along each parameter dimension contributes multiplicatively to the total number of bins, increasing the number

of bins by several factors can lead to intractability given the current central process unit (CPU) based implementation of our algorithms. Furthermore, a model with a larger number of bins, when used to analyze the same dataset, will converge to the underlying distribution for costlier sampler settings such as a larger number of samples and walkers, due to an increase in the number of quantities being simultaneously inferred from the same dataset. A fully scalable implementation of the described inference framework is part of ongoing development. Once achieved, it will enable a systematic study of the effects of bin choices on the inferred distributions which is currently limited by the computational cost of the existing implementations.

On the other hand, increasing the number of bins beyond the level of resolution required to identify *all* existing features in the underlying population, is not expected to further alter the shape of the inferred distribution significantly. This is because the GP hyper-parameters such as correlation lengths and amplitudes are simultaneously inferred with the rate densities, from the data itself. For example, once all the features in the underlying distribution along a particular parameter dimension are extracted, doubling the number of bins along said dimension leaves the inferred posterior of the corresponding length scale unchanged. This is indicative of the GP correlating roughly twice as many bins with each other, along that dimension. This results in the consistency of inferred constraints on the underlying population, among the two sets of binning choices. We have verified this for the version of BGP implemented in Abbott et al. (2023d) that infers only the mass distributions of CBCs and is hence computationally much cheaper than the three-dimensional models discussed in this work. We have analyzed simplistic populations of simulated CBCs using the two-dimensional model for two different sets of binning choices, one set having twice as many bins as the other. Both choices of binning yield constraints that are fully consistent with each other and also with the underlying true distribution within uncertainties. These results are discussed in appendix B.3.

Given these considerations we propose the following method of choosing the number of bins.

Initial results can be obtained with a preliminary binning choice that satisfies several conditions. For the analyses of GWTC-3 presented in this work, we have chosen bins such that the number of bins in every region of parameter space is at least higher than the number of features in the population that are *expected* in that region given the findings of existing population studies on the same dataset. For example, these expectations can be based on the number of features built into the functional forms of known parametric models that were found to best fit the GWTC-3 dataset (Abbott et al., 2023d). They can also be based on the findings of existing non-parametric models that were used to carry out uncorrelated population inference on the same dataset (Edelman et al., 2023; Callister & Farr, 2024). In our study of GWTC-3, we have taken into account the findings of both parametric and non-parametric population studies to choose our initial set of bins.

Once these initial results are obtained, one can keep on doubling the number of bins along a particular parameter dimension and reanalyzing the same dataset, until the length scale posteriors and the constraints on the rate densities stabilize. However, in the context of the three-dimensional models and their current implementation, even the first iteration of bin refinement can lead to computational intractability. For our proof-of-concept study, we thus focus on the results obtained from the initial binning choice to demonstrate the applicability of our method in producing non-parametric constraints on the mass-redshift correlations in the astrophysical BBH population, the first of their kind. We also focus on validating these constraints with simulation studies while leaving the higher resolution iterations by means of a scalable implementation of our framework as part of an upcoming study.

In this section, we summarize the results obtained upon analyzing real and simulated GW data with our non-parametric model. First, we re-analyze GWTC-3 with the uncorrelated model and compare the resulting constraints with those yielded by astrophysically motivated parametric models. We then validate those results for GWTC-3 by running the uncorrelated model on simulated sources drawn from an uncorrelated fiducial population. We also re-analyze GWTC-3 with the

generalized (fully correlated) model and infer non-parametric constraints on the redshift evolution of the shape of the BBH mass spectrum, for the first time. Finally, we test our generalized model on two different sets of simulated sources, one comprising an uncorrelated fiducial population of BBHs while the other correlated, so as to demonstrate the accuracy of our method in inferring the existence and nature of these correlations between the underlying distribution of BBH masses and redshifts.

4.3.3 Results

4.3.3.1 Uncorrelated Inference: GWTC-3

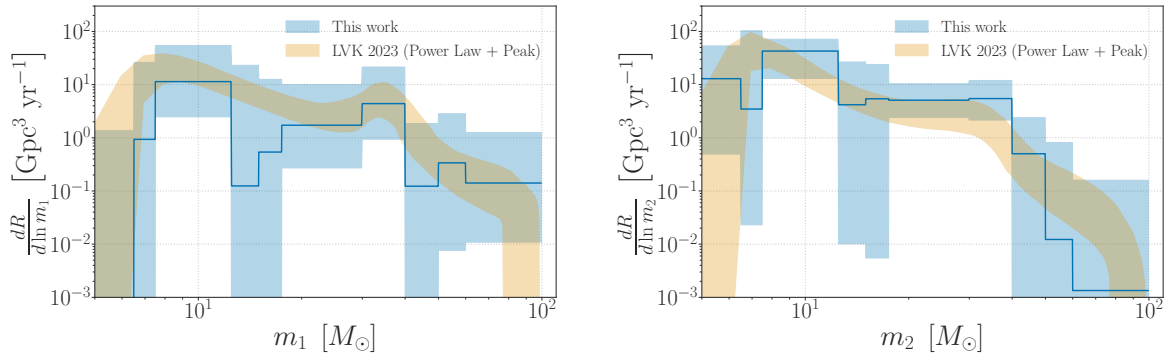


Figure 4.1: Constraints on the CBC mass-spectrum from GWTC-3 BBHs using the uncorrelated model. The constraints yielded by our non-parametric model corroborate the findings based on parametric population modeling by Abbott et al. (2023d) despite making minimal assumptions on the functional form of the underlying distribution. While there are signs of additional substructure in the primary mass-spectrum, for example, near the $15M_\odot$ and $55M_\odot$ bins, the large error bars on the rate density make any attempt at distinguishing such features from artifacts of Poisson uncertainty inconclusive.

To simultaneously infer the mass-spectrum and redshift distribution of BBHs in a model-independent manner, we re-analyze public LVK data (Abbott et al., 2023c) comprising all the BBH events that were observed through GWTC-3 with a FAR of less than 1 per year. Following previous works, we exclude the *outlier* event GW190814 given the uncertainty regarding its

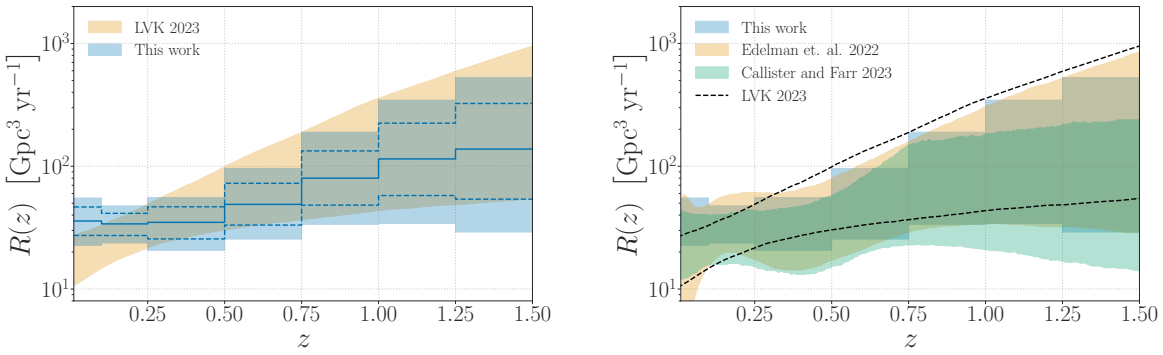


Figure 4.2: Constraints on the redshift evolution of the BBH merger rate from GWTC-3 using the uncorrelated model. *Left:* Comparison of our 90% (shaded) and 68% (dashed) credible intervals with the 90% interval obtained from parametric modeling (Abbott et al., 2023d). *Right* Comparison of our 90% credible intervals with that of other non-parametric studies such as those based on splines (Edelman et al., 2023, 2022b) and autoregressive processes (Callister & Farr, 2024, 2023)

system of origin, which leaves us with a set of 69 high confidence BBH observations (Abbott et al., 2021g, 2020e, 2023d; Essick et al., 2022). For each of these events, we use (m_1, m_2, z) PE samples calculated directly from the publicly released (m_1, m_2, d_L) samples by the LVK to compute Eq. (4.3). Specifically, following (Abbott et al., 2023d), we convert luminosity distance (d_L) samples to redshift by assuming a particular cosmological model, which we choose to be Planck 2015 (Ade et al., 2016). Since we do not infer the population properties of BBH spins, our inference essentially amounts to fixing the spin population to the default spin distributions used as PE priors. We note that previous implementations of the binned GP model (Abbott et al., 2023d) and some semi-parametric models (Edelman et al., 2022a) that focused on exploring features only in the mass-population of CBCs, used the same approach to dealing with spins. To summarize, we use the exact same dataset of PE samples used by (Abbott et al., 2023d) so as to facilitate an apples-to-apples comparison (The LIGO Scientific Collaboration et al., 2021, 2023). For further details of the single-event PE analyses see (Abbott et al., 2023d).

As shown in Fig. 4.1, we find that our uncorrelated model yields constraints on the BBH mass population that are fully consistent with the parametric inference carried out in (Abbott et al.,

2023d) using the POWERLAW+PEAK model, up to measurement uncertainties. It is able to identify both features in the primary mass spectrum in the form of peaks in the merger rate density at $10M_\odot$ and $35M_\odot$, which are also found by (Abbott et al., 2023d) using the same dataset. We also find hints of additional features in the form of a dip near $15M_\odot$ and a bump near $65M_\odot$. The measurement uncertainty of the inferred population in the bins corresponding to these new features allows for their interpretation as artifacts of Poisson noise. However, given the number of events expected to be observable in O4, the existence of these features can be verified empirically using our non-parametric analysis, unlike parametric models such as POWERLAW+PEAK, which are limited by their assumptions on the functional form of the mass population.

In addition to the mass spectrum, we simultaneously infer the redshift evolution of the BBH merger rate using our uncorrelated model. As shown in Fig. 4.2, our constraints are fully consistent with those obtained using a power-law in $(1 + z)$ model by (Abbott et al., 2023d) from the same dataset, again up to measurement uncertainties. We are able to recover a merger rate that increases with redshift without making any strong assumption regarding the functional form of the evolution. We note that our constraints rule out an unevolving merger rate with lesser confidence than the parametric model of Abbott et al. (2023d), which is to be expected given the non-parametric nature of our inference. In contrast to the parametric model, our 90% credible intervals are consistent with a nearly unevolving redshift distribution, similar to the findings of other model-independent explorations such as those based on splines (Edelman et al., 2023) and autoregressive processes (Callister & Farr, 2024). On the other hand, our 68% intervals (dashed lines) are fully inconsistent with a non-increasing merger rate. Hence, we are able to corroborate that the discovery of an increasing merger rate with redshift is not an artifact of the assumptions built into the parametric model used in that discovery since we are able to recover the same result, albeit with lesser confidence. We expect our model to yield more definitive conclusions regarding the redshift evolution of the merger rate given an O4-sized dataset, as evident from the results obtained from

the simulated catalogs in Sec. 4.3.3.1.

Validation of the Uncorrelated Inference

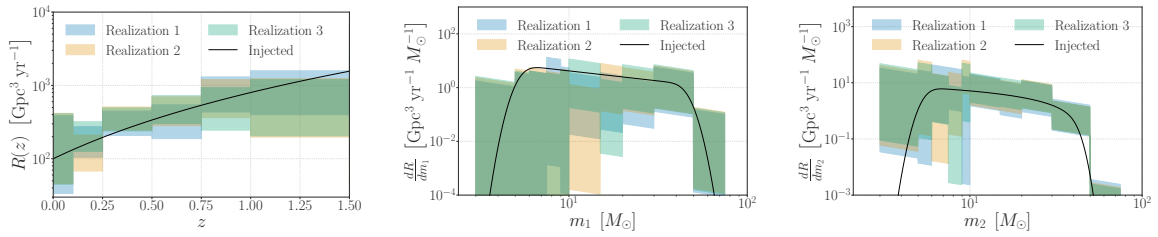


Figure 4.3: Constraints on the underlying population of simulated events using the uncorrelated model. We find that our uncorrelated model is able to recover the true underlying distributions of redshifts and masses up to measurement uncertainty and hence conclude that the results obtained from real-data displayed Figs. Figs. 4.1 and 4.2 are representative of the underlying BBH population and not of any artifacts built into our model construction

We validate our results obtained from real data using the uncorrelated model by testing our model on mock datasets comprised of simulated sources drawn from a known fiducial population. We choose the underlying distribution of masses to be a truncated power-law for both component masses. For the redshift evolution of the merger rate, we choose the underlying distribution to be a power-law in $(1+z)$. The intrinsic rate-densities for the aforementioned fiducial population thus take the following form:

$$\frac{dR}{dm_1 dm_2}(z) \propto R_0 (1+z)^{\kappa} m_1^{-\alpha} \frac{\Theta(m_1 - m_2)}{m_1 - m_{min}} \tag{4.11}$$

The hyper-parameter values characterizing our true population are listed in Table 4.1. After drawing the true values of masses and redshifts for our simulated events from the population Eq. (4.11), we generate the corresponding *observed* values using realistic estimates of measurement uncertainties following the methodology described in (Farah et al., 2023; Fishbach et al., 2020, 2018). We use the advanced LIGO design sensitivity noise curve (Abbott et al., 2018b) to simulate the signal-to-noise ratio (SNR) of mock events as a function of masses and redshifts,

Hyper-Parameter	True value
R_0	$100\ \mathrm{Gpc}^{-3}\mathrm{yr}^{-1}$
α	0.75
β	0.0
m_{min}	$4.5M_{\odot}$
m_{max}	$55M_{\odot}$
κ	3.0

Table 4.1: True values for the hyper-parameters characterizing the underlying population in Eq. (4.11).

in the case where all mock events are assumed to be optimally located and oriented with respect to the detector. We account for the distribution of source orientations and sky positions using a multiplicative factor that encodes information about the antenna response of the detector corresponding to randomly oriented sources, which itself has been shown to follow a well-modeled distribution for a single detector (Finn & Chernoff, 1993). We then interpolate the SNR over a grid of masses and redshifts to generate the PE samples for the simulated events that satisfy a given detection threshold, using the mock PE likelihood described in (Farah et al., 2023). We use the same SNR interpolation and detection threshold for generating a different set of detectable simulations required for estimating N_{det} through Eq. (4.4).

We draw *three* different realizations of our uncorrelated mock catalog, each spanning a one-month observation period with a duty cycle of 0.5, and analyze them individually using the uncorrelated model. With our choice of R_0, a single realization is found to comprise a mock catalog of 147 events. The resulting inference, summarized in Fig. 4.3, shows that our model is able to place constraints on the underlying mass population as well as the redshift evolution of the merger rate accurately, up to measurement uncertainty. This validates the results displayed in Figs. Figs. 4.1 and 4.2 obtained from analyzing real data using the uncorrelated model as being representative of the underlying BBH population as opposed to artifacts of the binned population model itself.

We note that, unlike the case of GWTC-3, the 90% credible intervals on the redshift distribution

of simulated events successfully rule out an unevoling merger rate. This is to be expected given that the simulated catalog has roughly twice as many events as GWTC-3 thereby enabling our model to extract much narrower constraints. Hence, we conclude that our uncorrelated model will be able to confidently constrain the nature of this redshift evolution of the BBH merger rate from an O4-sized catalog.

4.3.3.2 Correlated Inference: GWTC-3

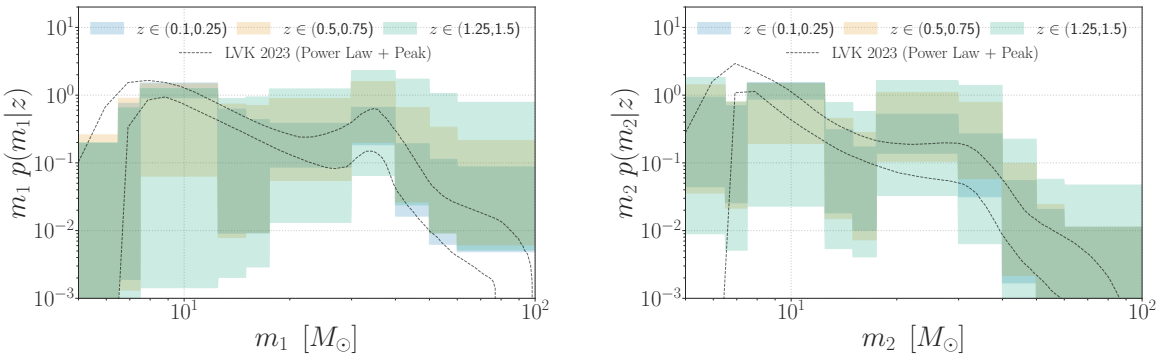


Figure 4.4: Non-parametric constraints on the redshift evolution of the shape of the BBH mass-spectrum: the first of their kind. We find that the shape of the BBH mass distribution across different redshifts agrees with each other up to measurement uncertainty. However, there are regions in the BBH component mass ranges wherein, the credible intervals of $p(m_1|z)$ have finite support for some redshift intervals, but not others. Hence we conclude that it is premature to rule out an evolving BBH mass distribution given current observations.

To search for correlations between the mass and redshift distributions of BBHs, we re-analyze the GWTC-3 dataset as in Sec. 4.3.3.1, but this time with the generalized population model in Eq. (4.2). We constrain the population level distributions of BBH component masses conditional on redshift, which we obtain from the merger rate density in the following way,

$$p(m_{1,2}|z) \quad = \quad \frac{1}{R(z)} \frac{dR}{dm_{1,2}}(z) \tag{4.12}$$

We note that uncorrelated population inference frameworks (Callister & Farr, 2024; Edelman et al., 2023) that assume $\frac{dR}{dm_1 dm_2}(z) \propto R_0 f(z) p(m_1, m_2)$ will always recover a distribution $p(m_{1,2}|z)$ that is independent of z regardless of how much the data favors otherwise. On the other hand, our generalized model in Eq. (4.2) has no restrictions built into its underlying assumptions and hence is capable of inferring a potentially evolving $p(m_1|z)$ from the data. We display the inferred credible intervals on $p(m_{1,2}|z)$ at three different redshift bins in Fig. 4.4

We find that the shape of the BBH mass-spectrum at different redshifts are fully consistent with each other up to measurement uncertainty, and hence conclude that there is no evidence of redshift evolution in GWTC-3. However, there are regions wherein the mass spectrum at lower redshifts has support while the ones at higher redshifts do not. This is indicative of the fact that we cannot rule out the existence of mass-redshift correlations either, given current observations. On the other hand, given the expectation that the width of the inferred credible intervals scales inversely with the square root of the number of observations, it might be possible to confidently validate or rule out the existence of mass-redshift correlations with our model from an O4-size dataset.

Validation of the Correlated Inference

We validate our correlated analysis of real data using two different mock datasets, one comprising an uncorrelated mass-redshift population of BBHs and the other a correlated one, so as to demonstrate the accuracy with which our model can infer the nature and existence of mass-redshift correlations in the underlying BBH population. For the first test, we analyze the exact same realizations of the uncorrelated mock observations described in Sec. 4.3.3.1, only this time using the fully correlated model in Eq. (4.2). As in the case of real data, we constrain the BBH mass distribution conditional on redshift, so as to demonstrate the accuracy with which our generalized model can recover the underlying densities corresponding to these distribution.

The resulting inference, summarized in Fig. 4.5, demonstrates that our generalized model re-

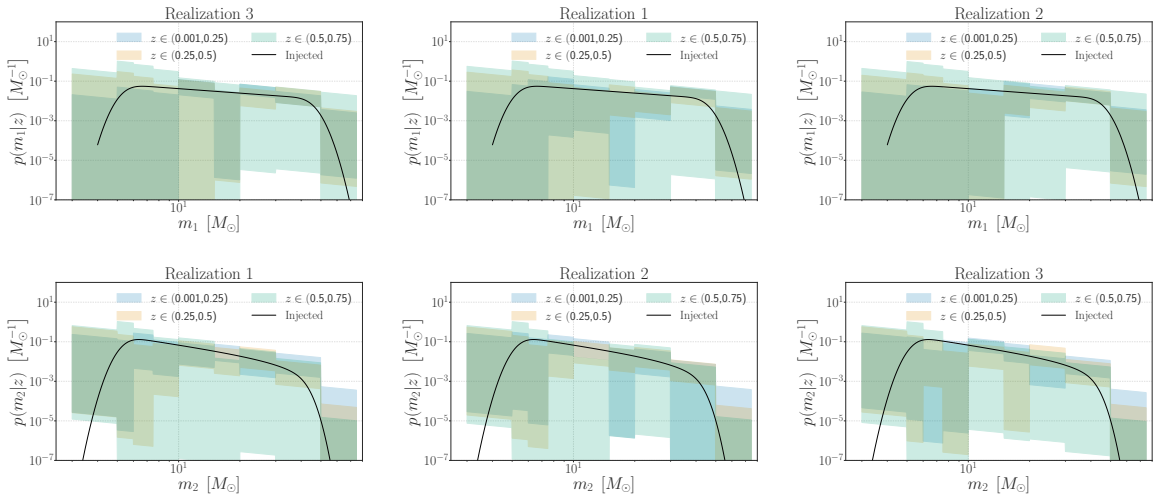


Figure 4.5: Constraints on the mass-population of simulated events conditional on redshifts for the mock dataset comprising an uncorrelated fiducial population. We find that our correlated model is able to recover an unevolving shape for the underlying distributions of component masses up to measurement uncertainty and hence conclude that constraints on mass-redshift correlations from real data displayed Fig. 4.4 are representative of the true BBH population and not of any artifacts built into our model construction.

covers the shapes of the BBH mass distribution at different redshifts to be fully consistent with the underlying population. In particular, it can be seen that, for all three realizations, the true mass distribution has significant support from the inferred constraints at all redshifts. This validates our model's ability to recover an unevolving mass spectrum from the data without assuming a priori, whether or not such evolution may exist. Hence, any potential evidence for (or against) the existence of mass-redshift correlations found using our model from future (post-O4) observations of CBCs can be interpreted as representative of the underlying CBC population and not as artifacts of the model itself since the latter would have otherwise manifested in this mock data challenge.

For our second test, we generate a different set of simulated sources that comprise a fiducial population whose mass distribution evolves in shape with redshift. We choose the distribution over masses to be a power-law in primary mass modulated by a Gaussian peak (Talbot & Thrane, 2018), with the fraction of events in the Gaussian component varying with redshift. The merger

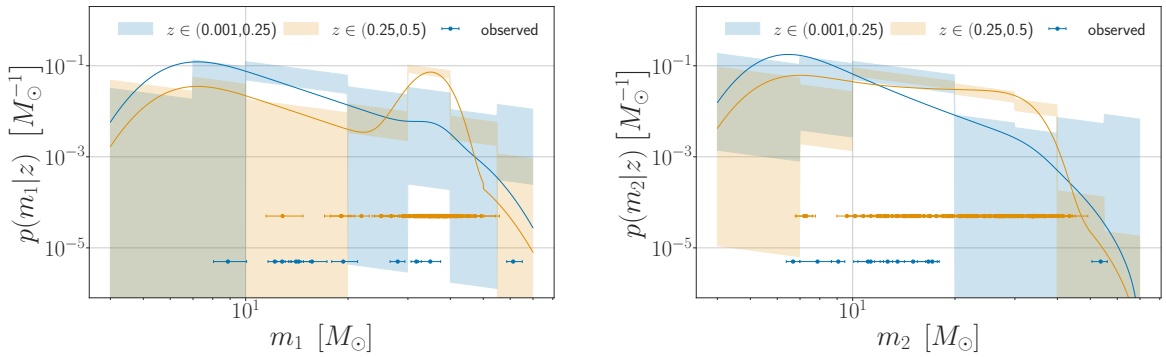


Figure 4.6: Constraints on the mass-population of simulated events conditional on redshifts, for the mock dataset comprising a correlated fiducial population. We find that our correlated model is able to accurately recover the true underlying distributions of redshifts and masses up to measurement uncertainty. The existence of disjoint grey and orange regions shows our generalized model can clearly distinguish between the different shapes of the BBH mass spectrum at different redshifts. Hence, we conclude that the results obtained from real-data displayed Fig. 4.4 are representative of the underlying BBH population and not of any artifacts built into our model-construction

rate density corresponding to the aforementioned underlying population thus takes the following form,

$$\begin{aligned}\frac{dR}{dm_1 dm_2}(z) \quad &\propto \quad R_0(1+z)^{\kappa} m_2^{\beta} \frac{\Theta(m_1 - m_2)}{m_1^{1+\beta} - m_{\rm min}^{1-\beta}} \left\{ \frac{m_1^{-\alpha}(1-\alpha)}{m_{\rm max}^{1-\alpha} - m_{\rm min}^{1-\alpha}} (1 - \lambda(z)) \right. \\ &+ \quad \left. \lambda(z) \mathcal{N}_T(m_1, \mu, \sigma, m_{\rm min}, m_{\rm max}) \right\} \end{aligned} \tag{4.13}$$

where $\mathcal{N}_T(m_1, \mu, \sigma, m_{\rm min}, m_{\rm max})$ is a truncated Gaussian distribution and $\lambda(z)$ the fraction of events in the Gaussian component which is chosen to be a piecewise function of redshift, as in $\lambda(z) = \lambda_0 \Theta(z_0 - z) + \lambda_1 \Theta(z - z_0)$. Here, $\lambda_0, \lambda_1, z_0$ are the additional hyper-parameters needed to describe a POWERLAW+PEAK mass distribution whose peak fraction evolves with redshift. The chosen fiducial values of all the hyper-parameters are listed in Table 4.2.

The generation of the mock posterior samples as well as the separate set of detectable simu-

Hyper-Parameter	True value
$m_{\rm max}$	$60M_{\odot}$
$m_{\rm min}$	$6.5M_{\odot}$
α	2.5
β	0
μ	$35M_{\odot}$
σ	$4M_{\odot}$
λ_0	0.001
λ_1	0.1
z_0	0.3
R_0	30 $\mathrm{Gpc}^3\ \mathrm{yr}^{-1}$

Table 4.2: True values for the hyper-parameters characterizing the underlying population in Eq. (4.13).

lations for estimating $N_{\rm det}$ are carried out using the exact same methodology and noise curve as the uncorrelated mock dataset, which is described in Sec. 4.3.3.1. However, instead of generating different realizations each spanning a small time window, we simulate a much longer observation time yielding a single realization of 507 mock observations. This choice of simulated run time was made so that the resulting number of mock events above the detection threshold becomes comparable to the upper bound on the expected number of observed BBHs post-O4. This in turn makes our constraints on mass-redshift correlations inferred from the described mock dataset an optimistic forecast for the post-O4 analysis of real data.

Upon analyzing the aforestated mock dataset, we find that our generalized model can correctly recover an underlying BBH mass spectrum that evolves in shape with redshift. As can be seen in Fig. 4.6, our generalized model is able to confidently identify the existence of correlations between the population-level distributions of BBH masses and redshifts, given enough observations. Furthermore, the credible intervals on the BBH mass distribution can be seen to be evolving in complete agreement with the true redshift evolution of the underlying curve. Hence, we conclude that given an O4-sized dataset, our non-parametric inference framework can potentially lead to confident empirical validation of several astrophysical BBH formation models that make informa-

tive predictions on the existence of mass-redshift correlations.

4.3.4 Conclusion and Future Prospects

We have developed a robust and non-parametric hierarchical inference framework based on binned Gaussian Processes that can constrain the population-level distributions of CBC masses and redshifts from GW data, while allowing for and being able to infer, the existence of correlations between the shapes of these distributions. We have demonstrated that our generalized population model has enabled the first non-parametric investigation of the correlations between the underlying distributions of BBH masses and redshifts.

To facilitate comparison with previous works, we have shown that our model can be restricted into one in which the shape of the mass distribution is independent of the redshift evolution of the merger rate. Using the restricted model, we have shown that our method yields measurements of the BBH mass and redshift distributions that are fully consistent with the results of parametric modeling despite being unassuming of the functional form of these distributions. In addition, using our generalized correlated model, we have inferred for the first time, a non-parametric constraint on the redshift evolution of the shape of the BBH mass distribution. We found that even though the credible intervals of the BBH mass distribution at different redshifts are broadly consistent with one another, the large error bars at high redshifts and the very small regions of tension between such intervals make it premature to rule out mass-redshift correlations given current observations.

We have validated the results from the uncorrelated inference by analyzing a fiducial uncorrelated population of simulated BBHs and realistic measurement uncertainty. In addition, using the same set of simulated sources, we have shown that correlated inference is capable of correctly recovering an uncorrelated population without assuming anything regarding the existence and shape of these correlations. Lastly, using a different set of simulated BBHs that comprise a correlated population, we have demonstrated that our generalized model is capable of confidently recovering

the correct redshift evolution of the shape of the underlying mass distribution given an O4-sized dataset.

Even though we restrict our work to a mass and redshift population inference, using only BBH observations, we note that such restrictions are straightforward to remove through simple generalizations of our robust and self-consistent inference framework. For example, the ability to simultaneously infer the population-level distribution of BBH spins and its underlying correlations with the corresponding mass and redshift populations can be incorporated within our framework by simply binning up the higher dimensional space of BBH parameters, now spanned by masses, redshifts, and spins.

However, as mentioned before, repeated draws from the GP algorithm used in this work would incur a computational cost that scales with the total number of bins cubed (N_b^3). Hence adding even a single dimension to the space of quantities being binned up can potentially lead to a drastic increase in the computational cost of the resulting inference, rendering the current CPU based implementation intractable. While several workarounds to this issue such as parallelized computing based on Graphics Process Units, sparse (Quiñonero-Candela & Rasmussen, 2005) and scalable (Gardner et al., 2018) GPs, etc., can potentially be implemented within our framework, the associated development is beyond the scope of this paper and are hence left as upcoming explorations.

On the other hand, including low-mass events from GWTC-3 in our analysis, thereby fitting across the entire CBC population, is a far simpler endeavor than say including spins, amounting to little to no increase in computational cost. However, to facilitate an apples-to-apples comparison between our uncorrelated mass-redshift inference and the corresponding analyses implemented in previous works such as (Abbott et al., 2023d; Edelman et al., 2023), we have chosen our dataset to be exactly identical to the one used by these studies and hence excluded from it the aforestated collection of low-mass events.

Furthermore, the sparsity of such events compared to BBHs, both in general, as well as at most of the higher redshift bins is indicative of the fact that such events would have contributed rather uninformatively to the search for population-level correlations between CBC masses and redshifts. While it is indeed of interest and astrophysical significance to study exactly how much the inclusion of these events affects our findings, the corresponding implications and explorations are beyond the scope of this proof-of-concept study and are hence left as follow-up investigations.

To summarize, we have developed a data-driven clustering framework to facilitate robust and self-consistent measurement of the underlying distributions of CBC masses and redshifts and their correlations, from GW observations, for the purpose of exploring new physics beyond the limitations of existing parametric and non-parametric population inference schemes. Armed with our model-independent hierarchical inference algorithm, future studies can hope to empirically probe several highly significant astrophysical phenomena that take place in our universe in a maximally model-independent manner.

4.4 THE BINARY BLACK HOLE MASS AND SPIN DISTRIBUTIONS: EVIDENCE OF MULTIPLE SUBPOPULATIONS

Astrophysically motivated population models for binary black hole observables are often insufficient to capture the imprints of multiple formation channels. This is mainly due to the strongly parametrized nature of such investigations. Using a non-parametric model for the joint population-level distributions of binary black hole component masses and effective inspiral spins, we find hints of multiple subpopulations in the third gravitational wave transient catalog. The higher (more positive) spin subpopulation is found to have a mass-spectrum without any feature at the $30 - 40M_{\odot}$, which is consistent with the predictions of isolated stellar binary evolution, simulations for which place the pile up due to pulsational pair-instability supernovae near $50M_{\odot}$ or higher. The other sub-population with effective spins closer to zero shows a feature at $30 - 40M_{\odot}$ and is consis-

tent with binary black holes formed dynamically in globular clusters, which are expected to peak around $30M_{\odot}$. We also compute merger rates for these two subpopulations and find that they are consistent with the theoretical predictions of the corresponding formation channels. We validate our results by checking their robustness against variations of several model configurations and by analyzing large simulated catalogs with the same model.

4.4.1 Introduction

The formation of compact object binaries that merge within a Hubble time via gravitational wave (GW) emission, particularly stellar-mass binary black holes (BBHs), is poorly understood yet critically important to several fields of modern astrophysics (Mandel & Farmer, 2022; Mapelli, 2020a,b). The proposed formation scenarios for such BBH systems can be broadly classified into three categories: isolated evolution of massive stellar binaries that undergo orbital hardening, via either common-envelope, stable mass transfer, or chemical mixing (e.g., Postnov & Yungelson, 2006; Paczynski, 1976; van den Heuvel et al., 2017; Marchant et al., 2016; Mandel & de Mink, 2016), dynamical assembly assisted by either a tertiary companion, multiple exchanges in dense clusters, or gas-assisted migration (e.g., Wen, 2003; Antonini & Perets, 2012; Benacquista & Downing, 2013; Bartos et al., 2017), and hierarchical mergers in star clusters and AGN disks (Gerosa & Fishbach, 2021). These formation channels are usually characterized by several unknown parameters and initial conditions which can uniquely shape and correlate the observable features in an astrophysical BBH population, and hence be constrained from a catalog of GW observations.

At the end of its third observing run, the LIGO-Virgo-KAGRA detector network (LVK, J. Aasi et al., 2015; Acernese et al., 2014; Akutsu et al., 2021) unveiled a catalog of GW observations (GWTC-3, Abbott et al., 2023b) around 70 of which were confidently[1] identified to be BBH signals. Studying the ensemble properties of these BBHs has offered a plethora of new insights

[1] With a false alarm rate of less than 1 per year

into the underlying astrophysics of their progenitor systems and environments. With the LVK's ongoing fourth observing run promising a large number of new detections, the search for the imprints of BBH formation mechanisms on the observable population properties of BBHs will continue to enable exciting new discoveries. Traditional approaches towards conducting such investigations have relied on the theoretical predictions of one or more formation models to construct a fiducial form of the distribution function of BBH parameters. The hyperparameters characterizing this distribution function encode information on the astrophysical processes underlying the corresponding formation scenarios and are inferred from a collection of GW observations (e.g. Fishbach & Holz, 2017; Edelman et al., 2021; Kimball et al., 2021).

These *parametric* population models, while useful in extracting information on a restrictive set of astrophysical assumptions regarding BBH formation, are not flexible enough to capture all of the physics that underlies BBH formation and are also susceptible to model-induced biases. For example, BBHs formed in isolation are expected to be distributed as a Power-law in component masses extending up to a maximum range followed by a steep fall-off known as the pair-instability gap and a small bump right before the truncation arising from the pile-up due to pulsational pair-instability supernovae (50truncreason7, Fishbach & Holz, 2017; Woosley, 2017; Talbot & Thrane, 2018). However, when this POWERLAW+PEAK model for the BBH mass distribution was constrained using GWTC-3 data (Abbott et al., 2023d), several concerns regarding our understanding of massive stellar interiors came to light. Firstly, the location of the 50truncreason7 feature was found to be within the $30 - 40M_{\odot}$ mass range, leading to tension with the predictions of stellar evolution simulations which place it above $50M_{\odot}$ (see eg, Hendriks et al., 2023, and references therein). Secondly, flexible alternatives to the POWERLAW+PEAK model have found hints of additional features in the BBH mass spectrum (Edelman et al., 2023; Toubiana et al., 2023; Abbott et al., 2023d; Callister & Farr, 2024; Tiwari, 2021), which might be indicative of additional physics at play, beyond the restrictions built into the underlying assumptions of POWERLAW+PEAK.

A plausible explanation for both of these issues could be the existence of multiple sub-populations of BBHs in GWTC-3 data, each corresponding to different formation channels. The peak at the $30-40M_{\odot}$ range, for example, can be explained by a sub-population of dynamically formed BBHs in globular clusters, which are expected to peak around the same mass range (Antonini et al., 2023; Wong et al., 2021). Given that the distribution of the effective inspiral spin parameters for dynamically formed BBHs in globular clusters is expected to be more symmetric about zero than that of BBHs formed in isolation (Mapelli et al., 2022; Chattopadhyay et al., 2023; Rodriguez et al., 2018, 2022), observational evidence for these two sub-populations coexisting together can manifest in the form of population-level correlations between BBH masses and spins, and hence be searched for in current and future GW catalogs (Baibhav et al., 2023). However, several other astrophysical processes such as mass-ratio reversal in isolated binaries (Broekgaarden et al., 2022) can introduce mass-spin correlations in the BBH population. Hence, the use of flexible population models that can account for a wide variety of known, as well as previously unmodelled astrophysical scenarios, is crucial to these kinds of investigations.

In this letter, using a non-parametric model for the joint population-level distribution of BBH component masses and effective inspiral spins based on binned Gaussian processes (BGP, Ray et al., 2023c; Mohite, 2022; Mandel et al., 2017), we find hints of multiple subpopulations of BBHs in GWTC-3 data, one consistent with the predictions of isolated BBH formation, and the other dynamical. In particular, we find two distinct shapes in the BBH mass spectrum, one of which has a powerlaw-like fall-off with no feature in the $30-40M_{\odot}$ region and is associated with higher (more positive) values of effective inspiral spins. The other one is found to have a bump at the $30-40M_{\odot}$ and is associated with smaller values of effective inspiral spins. Furthermore, BBHs with at least one component in the $30-40M_{\odot}$ range are found to have a more symmetric distribution of effective inspiral spins about zero, as compared to that of the complementary case which is characterized by a spin distribution skewed towards more positive values. We also constrain the total merger rate

of BBHs within different mass and effective spin ranges and find that they are consistent with the theoretical predictions of the corresponding formation models.

Given the size of contemporary datasets, we restrict the dimensionality of our population model to account for the distributions of only the best-measured among BBH observables that are relevant to the astrophysics we aim to extract. In particular, we only model the joint distribution of source frame component masses and effective inspiral spins using BGPs, and fix that of other BBH observables to be of much simpler and less flexible functional forms. While this restriction amounts to ignoring higher dimensional correlations in the BBH population that can in principle bias our conclusions, we validate our analysis using large simulated catalogs and demonstrate that the ignored correlations cannot manifest into the trends we find in the joint mass and effective spin distributions. Even though a higher dimensional BGP on masses, redshifts, and all of the spin parameters (component spin magnitudes and tilts) would be the most generalized and robust approach to this investigation, we expect such a model to yield uninformative constraints given GWTC-3-sized datasets. However, as catalogs continue to grow, we aim to implement higher dimensional BGP models for obtaining a more generic and unbiased characterization of the various sub-populations of BBHs.

In a previous study, Godfrey et al. (2023) had also looked at mass-spin correlations in the BBH population using a mixture of semi-parametric models and had come to conclusions that are broadly consistent with ours except for a few subtle distinctions. They identify a sub-population of binaries with a mass distribution that demonstrates a strong peak near $10M_{\odot}$ followed by a sharp fall-off, and a spin distribution that slightly prefers aligned components over isotropic ones. They identify a second sub-population that demonstrates a peak in the 30-40$M_{\odot}$ range and has a spin distribution consistent with that of the other subpopulation. In addition to component spins, they infer an effective spin distribution that is completely consistent between the two subpopulations they identify. On the other hand, our study identifies two subpopulations that correspond to signifi-

cantly different effective spin distributions. The astrophysical implications of Godfrey et al. (2023) are therefore slightly different from ours even though both are broadly in agreement regarding the existence of a subpopulation in the mass-spectrum that falls off steadily beyond $15M_{\odot}$. These differences might be attributed to the differences in our modeling choices which are discussed further in Sec. 4.4.2.

This letter is organized as follows. In Sec. 4.4.2 we describe in detail our flexible population model and the associated inference framework. In Sec. 4.4.3, we summarize our results for GWTC-3 and describe the data used to obtain it. In Sec. 4.4.4, we validate our model by analyzing large simulated catalogs of five different BBH populations and demonstrate that our results for real data are robust against known sources of systematic biases. In Sec. 4.4.5, we discuss the astrophysical implications of our findings. Lastly, In Sec. 4.4.6, we conclude with a summary of our model, findings, and future investigations.

4.4.2 Population model and hierarchical inference

To search for multiple sub-populations in the BBH mass spectrum that potentially correspond to different distributions of effective inspiral spin, we construct a flexible model for the population properties of BBH observables, agnostic of astrophysical predictions. Specifically, we model the joint distribution of BBH component masses (m_1, m_2), and effective inspiral spins $(\chi_{\rm eff})$ as a piecewise binned function, such as:

$$\frac{dN}{dm_1 dm_2 dz d\chi_{\rm eff}}(\vec{n}, \kappa) = \sum_{\gamma} \frac{n^{\gamma}}{m_1 m_2} \frac{dV}{dz} T_{\rm obs} (1+z)^{\kappa-1} \times \begin{cases} 1 & \text{if } (m_1, m_2, \chi_{\rm eff}) \in \gamma^{th} \text{ bin} \\ 0 & \text{otherwise} \end{cases} \tag{4.14}$$

where, the LHS is the number of CBCs per component masses, effective spin and redshift, n^{γ} is the merger rate density per comoving volume, source-frame time, log component mass, and

effective spin in the γ^{th} bin, and κ is the parameter that controls the redshift evolution of the overall merger rate. Simultaneously inferring the merger rate density in each bin from multiple uncertain measurements amounts to learning the shape of the population distribution from GW data up to the resolution limit imposed by our choice of binning.

While Eq. (4.14) is flexible enough to capture any correlation between the mass and effective spin populations, it is unable to inform on the existence of mass-redshift and spin-redshift correlations. Even though contemporary datasets are mostly uninformative on the existence of mass-redshift correlations (Abbott et al., 2023d), particularly in the context of binned models (Ray et al., 2023c), the studies of Heinzel et al. (2023) and Biscoveanu et al. (2022) have revealed strong evidence for spin-redshift correlations in the same. Biscoveanu et al. (2022) has further shown that certain population models designed to search only for a mass-spin correlation can falsely infer their existence from a simulated catalog with spin-redshift correlations. We show is Sec. 4.4.4 that our results are robust against such model-induced systematics by testing our model on large simulated catalogs characterized by various correlated populations. See Rinaldi et al. (2023); Karathanasis et al. (2023) for more investigations on mass-redshift correlations using alternative models.

To infer the merger rate density in each bin from multiple uncertain measurements of BBH masses and effective spins we implement the framework of Bayesian hierarchical inference (Thrane & Talbot, 2019; Mandel et al., 2019; Loredo, 2004; Wysocki et al., 2019). By modeling the occurrence of BBHs as an inhomogeneous Poisson process, we construct the likelihood function of the rate densities from a collection of single event measurements and use a simulated population of detectable BBHs to account for Malmquist biases in the inferred distributions. In the context of our population model, the likelihood function takes the following form:

$$\log p(\vec{d}|\vec{n}) = -\sum_{\gamma} n^{\gamma} \left\langle VT \right\rangle^{\gamma} + \sum_{i} \log \left(\sum_{\gamma} n^{\gamma} w^{\gamma}(d_i) \right) \tag{4.15}$$

where $w^{\gamma}(d_i)$ is the posterior support of the γ^{th} bin from the i^{th} event and $\langle VT \rangle^{\gamma}$ is the sensitive

time-volume within which BBHs in the γ^{th} bin are expected to be detectable (Ray et al., 2023c).

To regularize the inferred shape of the population distribution over sparse regions of parameter space we choose the prior on logarithmic rate densities to be a Gaussian process (GP) with an exponential quadratic kernel. The means, correlation lengths, and covariance amplitudes characterizing the GP are themselves modeled using normal, log-normal, and half-normal priors respectively (Ray et al., 2023c). Using Hamiltonian Monte Carlo (HMC) sampling implemented through the No U-Turn Sampler (Neal, 2011; Homan & Gelman, 2014), we infer the GP hyper-parameters along with the rate densities characterizing our population model from their joint posterior distribution. The code developed to implement this analysis is publicly available as the python package `gppop`[2], which in turn relies on the `Pymc` package (Oriol et al., 2023) for conducting the HMC sampling.

Previous studies by Farr (2019); Essick & Farr (2022) have demonstrated that the convergence of the Monte Carlo sums used to compute the posterior weights and detectable time-volumes of Eq. (4.15) needs to be ensured in order to avoid biases in the corresponding hierarchical inference. Following the implementation of their convergence criteria in the context of the binned model, as derived by Ray et al. (2023c), we verify that all of our inferred hyper-parameter samples support regions of parameter space that have enough effectively independent Monte Carlo samples, both for event-specific posteriors and detectable injections to avoid biases resulting from improperly converged estimates of posterior weights and detectable time-volumes.

Even though we do not fit for the redshift evolution parameter κ and instead fix it to be $\kappa = 2.9$, we show in appendix B.5 that varying κ across its measured confidence interval leaves our results unchanged. The reason for this choice is as follows. For fixed κ, the posterior weights and sensitive time-volumes of Eq. (4.15) are pre-computable leading to computationally cheap inference of the joint population posterior distribution. Inferring κ amounts to re-computing the weights and time-volumes for every step of the sampling which is intractable for our current Central Processing

[2]https://github.com/AnaryaRay1/gppop/tree/spin-dev

Unit (CPU) based implementation. While parallelizing these computations on Graphics Processing Units (GPUs) offers a potential solution to this problem, the associated code developments are part of an ongoing study and beyond the scope of this work given the insensitivity of our conclusions to the value of κ. We however note that our current implementation is much more scalable on CPUs than the earlier ones of Ray et al. (2023c). See appendix B.7 for more details.

We further note that while previous attempts (Godfrey et al., 2023; Li et al., 2023) at obtaining data-driven constraints on mass-spin correlations in the BBH population have accounted for more spin parameters such as component spin magnitudes and orientations, most of these studies have imposed several restrictions on the nature of mass-spin correlations, in contrast to our flexible model. In particular, they have used mixture models wherein, each mixture component is assumed to have an uncorrelated mass-spin distribution. Therefore, within their framework, mass-spin correlations can only manifest in the form of multiple sub-populations coexisting together. Hence, mass-spin correlations arising from the astrophysical processes within a single formation channel (such as mass-ratio reversal during stable mass transfer introducing a spin-mass-ratio anti-correlation during field formation Broekgaarden et al., 2022), can potentially bias the astrophysical conclusions of such models. On the other hand, we impose no similar restrictions on our joint mass-effective-spin distribution and reconstruct its functional form directly from the data, up to the resolution limit imposed by our choice of binning.

An additional source of systematics in our population model involves the potential sensitivity of the inferred shapes of the population distribution to binning choices. Similar to the example presented in Ray et al. (2023c) we show in appendix B.6 that our results are stable against multiple binning choices by varying the mass and $\chi_{\rm eff}$ bin locations and demonstrating that the trends we infer are robust against this variation.

4.4.3 Results

We re-analyze public LVK data (Abbott et al., 2023c) comprising all the BBH events that were observed through GWTC-3 with a FAR of less than 1 per year. Following previous works, we exclude the *outlier* event GW190814 given the uncertainty regarding its system of origin, which leaves us with a set of 69 high confidence BBH observations (Abbott et al., 2021g, 2020e, 2023d; Essick et al., 2022). For each of these events, we use $(m_1, m_2, \chi_{\rm eff})$ samples publicly released by the LVK (Abbott et al., 2023c) to compute the posterior weights in Eq. (4.15). Specifically, following (Abbott et al., 2023d), we convert detector frame mass and luminosity distance samples to source frame by assuming a particular cosmological model, which we choose to be Planck 2015 (Ade et al., 2016). For further details of the single-event PE analyses see (Abbott et al., 2023d). We also use LVK's publicly released set of detectable injections (Abbott et al., 2023c) to compute the detectable time-volumes of Eq. (4.15).

As seen in Figure 4.7, we find trends indicative of at least two subpopulations in the BBH mass-spectrum that differ by 90% in shape for the $30-40M_\odot$ bin and are each conditional on different ranges of effective inspiral spin. In particular, we find that for $0.1 < \chi_{\rm eff} < 0.2$, the primary and secondary mass distributions resemble a power-law like fall off with no feature in the $30-40M_\odot$ range. On the other hand, for $0.05 < \chi_{\rm eff} < 0.1$, we find that the primary and secondary mass-distributions both show an excess of BBHs in the $30-40M_\odot$ range. Similarly, constraints on the conditional spin distribution depict that binaries with one or both components in the $30-40M_\odot$ mass range have a more symmetric effective spin distribution about zero as compared to other binaries which are found to have a skewed distribution of $\chi_{\rm eff}$ that prefers more positive values. We also compute the fraction of events with $\chi_{\rm eff} < 0$ to be $21^{+43}_{-18}\%$ for binaries with at least one component in the the $30-40M_\odot$ mass range and $10^{+22}_{-7}\%$ for all other binaries.

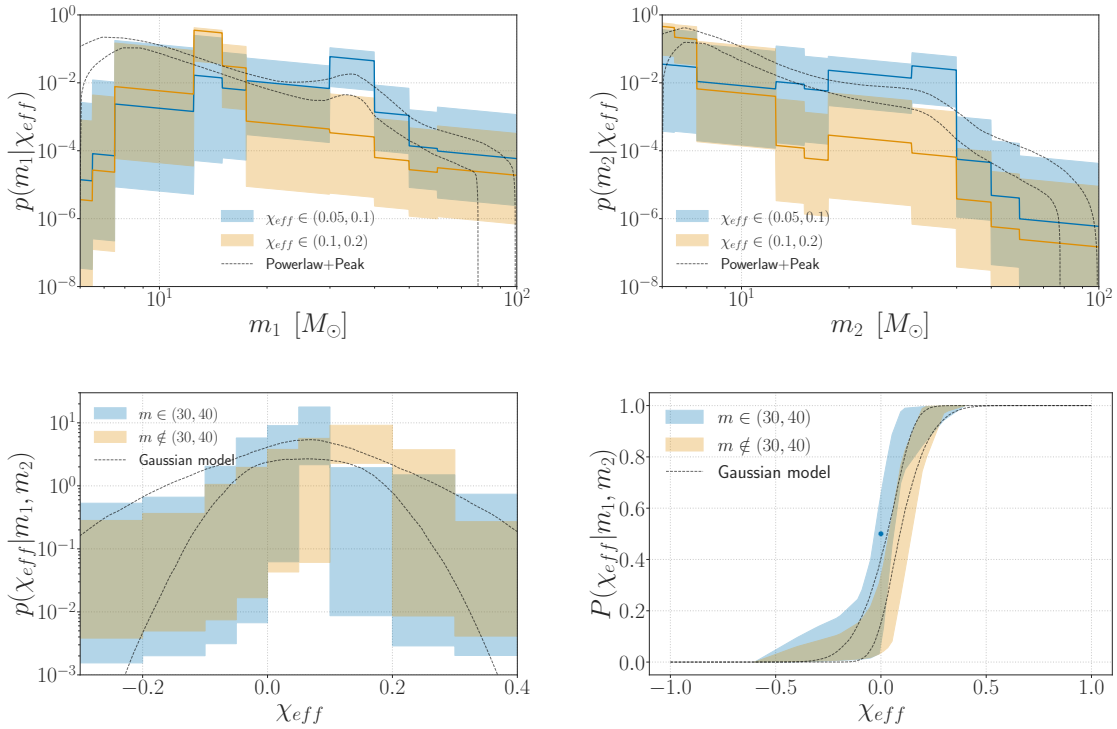


Figure 4.7: The parametric inferences of Abbott et al. (2023d) that used the Powerlaw+Peak and Gaussian effective spin models are overplotted for comparison. The top panels show the primary (*left*) and secondary (*right*) mass distributions conditioned on two different ranges of effective inspiral spins, namely $0.05 < \chi_{\rm eff} < 0.1$, and $0.1 < \chi_{\rm eff} < 0.2$. The bottom panels show the density (*left*) and cumulative densitiy (*right*) functions corresponding to the distribution of effective inspiral spins, conditioned on two different mass ranges, namely $\{m_1 \in (30M_\odot, 40M_\odot)$ or $m_2 \in (30M_\odot, 40M_\odot)\}$, and its complement. The blue dot on the lower right figure is the point (0,0.5).

4.4.4 Validation Study

To validate the robustness of our results obtained from real data we test our model on large simulated catalogs. These datasets are generated by drawing BBH parameters from a known population distribution and injecting the corresponding signals into detector noise realizations. We use Bayesian parameter estimation on the injected signals using the `BILBY` package (Ashton et al., 2019; Romero-Shaw et al., 2020), its `DYNESTY`-based nested sampler (Speagle, 2020), and the aligned spin waveform approximant known as IMPRphenomD (Husa et al., 2016; Khan et al.,

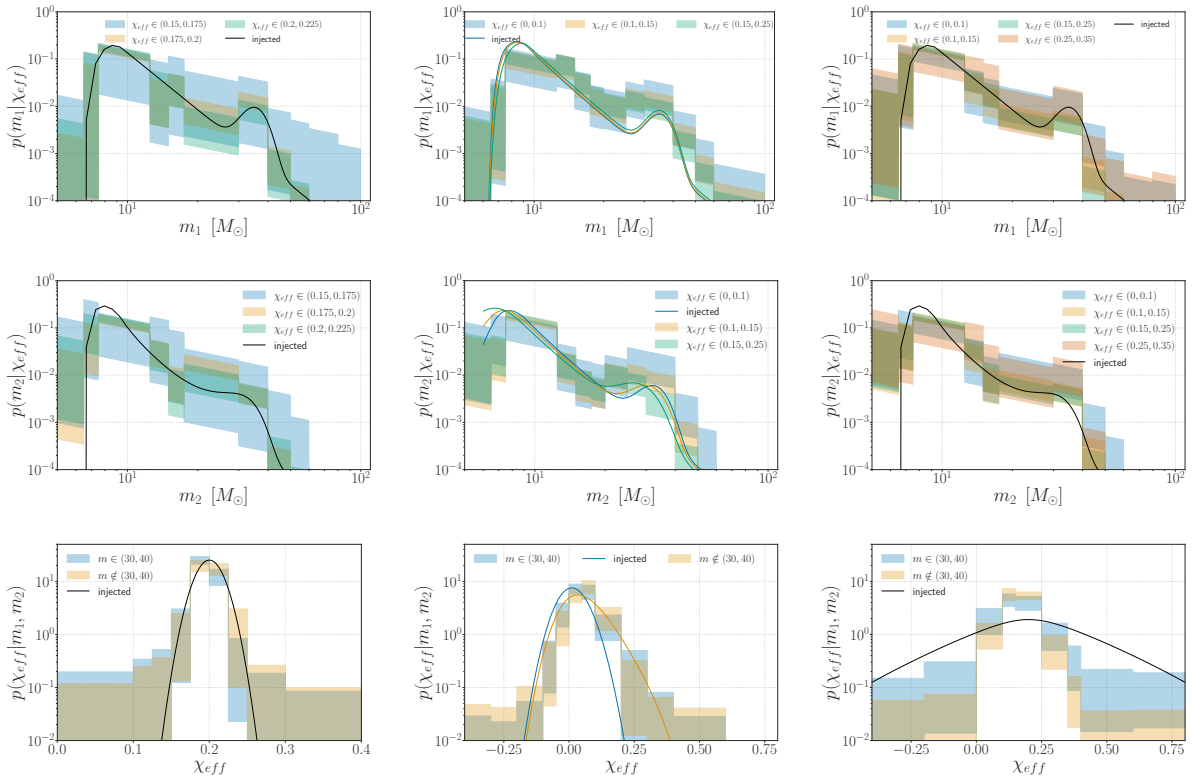


Figure 4.8: The uncorrelated population is shown on the *left*, the one with χ_{eff}-q correlations in the middle and the one with spin-redshift correlations on the right. Note that for the right panels, the reconstructed χ_{eff} distributions are not expected to match the injected population since our model assumes the joint $\chi_{eff} - z$ distribution to be separable, i.e., $p(\chi_{\rm eff}, z) = p(\chi_{\rm eff})p(z)$.

2016). We use LVK's noise power spectral densities that were measured during the first three months of its third observing run (Collaboration et al., 2022) and draw 276 events for each simulated population which is roughly 4 times the size of the GWTC-3 BBH catalog. We use sensitivity injections found during the third observing run as publicly released by the LVK (Abbott et al., 2023c) to compute the detectable time-volumes required for analyzing these simulated populations.

4.4.4.1 *Testing for false positives*

We first simulate three BBH populations each of which is characterized by a POWERLAW + PEAK model for the marginal primary mass-distribution (Talbot & Thrane, 2018), a mass-ratio dependent pairing function (Fishbach & Holz, 2020), a power-law in one plus redshift for the evolution of the merger rate (Fishbach et al., 2018), and a Gaussian distribution for effective inspiral spins. In the first population, we impose no intrinsic correlations between masses, spins, and redshifts. In the second one, we introduce an anti-correlation between mass-ratio and effective inspiral spins by making the mean of the $\chi_{\rm eff}$ Gaussian a linear function of mass-ratio with a negative slope (Callister et al., 2021). In the third population we allow the $\chi_{\rm eff}$ distribution to broaden with redshift, similar to the models of Biscoveanu et al. (2022).

All three population distributions can be expressed in the following form:

$$\begin{aligned}\frac{dN}{dm_1 dq dz d\chi_{\rm eff}} &= N_0 \left(\lambda \frac{(1-\alpha) m_1^{-\alpha}}{m_{\rm max}^{1-\alpha} - m_{\rm min}^{1-\alpha}} + (1-\lambda) \frac{e^{-\frac{1}{2}\left(\frac{m_1-\mu_m}{\sigma_m}\right)^2}}{\sqrt{2\pi}\sigma_m} \right) \\ &\quad \frac{(1+\beta) q^\beta}{1 - \left(\frac{m_1}{m_{\rm min}}\right)^{\beta+1}} \frac{dV}{dz} (1+z)^{\kappa-1} \frac{e^{-\frac{1}{2}\left(\frac{\chi_{\rm eff}-\mu_\chi+\delta\mu_\chi(q-0.5)}{\sigma_\chi+\delta\sigma_\chi(z-0.001)}\right)^2}}{\sqrt{2\pi}(\sigma_\chi + \delta\sigma_\chi(z-0.001))} \end{aligned} \tag{4.16}$$

where $q = \frac{m_2}{m_1}$ is the mass-ratio and N_0 the total number of BBHs that have occured through out the observation time. The hyper-parameters $(m_{\rm min}, m_{\rm max}, \lambda, \alpha, \beta, \mu_m, \sigma_m, \kappa, \mu_\chi, \delta\mu_\chi, \sigma_\chi, \delta\sigma_\chi)$ corresponding to the three simulated populations are listed in Table 4.3.

The reasons for choosing these three populations to conduct our validation study are as follows. Using the un-correlated mock dataset, we demonstrate that the kind of mass-spin correlations we are seeing in real data are not spuriously manifesting from any artifacts built into our model, nor from the correlations between the measurement uncertainties of the BBH observables of interest (left panels of Fig. 4.8). Using the simulated population with intrinsic q-$\chi_{\rm eff}$ correlation, we demonstrate that such a correlation cannot manifest into multiple subpopulations in the marginal

m_1-$\chi_{\rm eff}$ plane (middle panels of Fig. 4.8). Finally, using the population with intrinsic spin-redshift correlations we validate that our results for GWTC-3 are not indicative of spin-redshift correlations but instead of new astrophysics underlying BBH formation (right panels of Fig. 4.8).

We note that for the simulated dataset with intrinsic spin-redshift correlation, the reconstructed $\chi_{\rm eff}$ distributions are not expected to match the injected population since our model assumes the joint $\chi_{\rm eff}$-z distribution to be separable, i.e., $p(\chi_{\rm eff}, z) = p(\chi_{\rm eff})p(z)$. The point of this validation study is to instead show that the ignored spin-redshift correlations cannot manifest into the kind of mass-spin correlations we are seeing, in the context of our binned model. On the other hand, for the simulations with intrinsic spin-mass ratio correlation, we choose a higher signal to noise ratio (SNR) threshold than the other two to make sure that the poor measurability of both mass ratio and effective spin at low SNR does not bias the inferred shapes.

Parameter	Uncorrelated	Spin-mass-ratio	Spin-redshift
$m_{\rm min}/M_\odot$	6.0	6.0	6.0
$m_{\rm max}/M_\odot$	70.0	70.0	70.0
λ	0.04	0.04	0.04
α	4.0	4.0	4.0
β	1.5	1.5	1.5
$\mu_m/M_\odot$	35.0	35.0	35.0
$\sigma_m/M_\odot$	4.0	4.0	4.0
κ	3.0	3.0	3.0
μ_χ	0.2	0.2	0.2
$\delta\mu_\chi$	0.0	-0.5	0.0
σ_χ	$10^{-1.8}$	$10^{-1.8}$	$10^{-0.85}$
$\delta\sigma_\chi$	0.0	0.0	$10^{-0.88}$

Table 4.3: True values for the hyper-parameters characterizing the underlying population in Eq. (4.16) for the three simulated catalogs. The parameters controlling the correlations are chosen to have best-fit values from Abbott et al. (2023d)

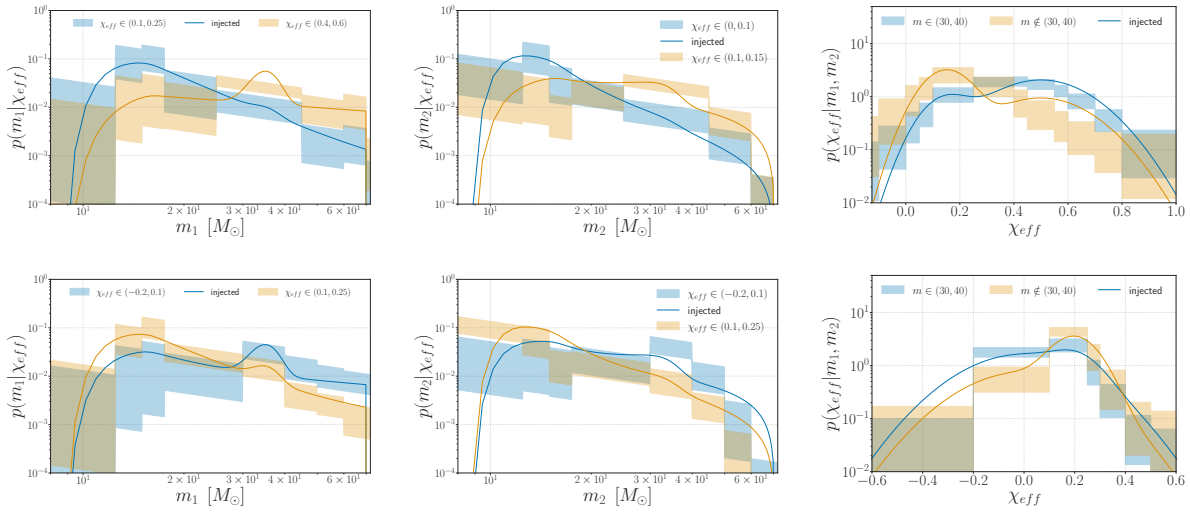


Figure 4.9: The top and bottom pannel represrents the peak+aligned and peak+isotropic populations respectively from table 4.4

4.4.4.2 Recovering different kinds of mass-spin correlations

In the previous section, we had shown that various forms of χ_{eff} correlations which have already been found in GWTC-3, cannot manifest into the trends we have found in the same dataset. In this section we simulate two additional populations with intrinsic mass-spin correlations, similar to the trends we have found in GWTC-3, and show that our model can accurately recover injected conditional distributions of mass given effective spin and vice-versa. For the primary mass distribution, we choose a mixture of two subpopulations, one of which is a power-law without any peak and the other a Powerlaw+Peak with a very shallow spectral index and a large peak fraction. Each of these subpopulations is associated with a Gaussian χ_{eff} distribution whose mean and standard deviations are different from that of the other subpopulations. The mass-ratio distribution and redshift evolution of the merger rate are both chosen to be the same as that of the mock populations in the previous section.

These two population distributions can be expressed in the following form:

$$\frac{dN}{dm_1 dq dz d\chi_{\rm eff}} = \sum_{i=1}^{2} f_i N_{0,i} \left(\lambda_i \frac{(1-\alpha) m_1^{-\alpha_i}}{m_{\rm max}^{1-\alpha_i} - m_{\rm min}^{1-\alpha_1}} + (1-\lambda_i) \frac{e^{-\frac{1}{2}\left(\frac{m_1-\mu_{m,i}}{\sigma_{m,i}}\right)^2}}{\sqrt{2\pi}\sigma_{m,i}} \right) \frac{(1+\beta) q^{\beta}}{1-\left(\frac{m_1}{m_{\rm min}}\right)^{\beta+1}} \frac{dV}{dz} (1+z)^{\kappa-1} \frac{e^{-\frac{1}{2}\left(\frac{\chi_{\rm eff}-\mu_{\chi,i}}{\sigma_{\chi,i}}\right)^2}}{\sqrt{2\pi}(\sigma_{\chi,i})} \tag{4.17}$$

where the hyper-parameters corresponding to the two simulated populations are listed in Table 4.4

We set the mean of the effective-spin Gaussian associated with the non-zero peak mass distribution to be 0.5 in one of the simulated datasets (peak+aligned) and 0 in the other (peak+isotropic, table 4.4). The subpopulation without the peak is chosen to have an effective spin distribution with mean 0.2 for both cases. We analyze both mock catalogs with our model and show that it can accurately reconstruct the injected distribution (Figure 4.9). In other words, for GWTC-3, had the subpopulation with the $30-40M_{\odot}$ feature been associated with more positively aligned spin distributions, our model would have recovered it correctly.

4.4.5 Astrophysical Implications

We have found evidence for a subpopulation of BBH mergers in the mass range $30-40\ M_{\odot}$ with preference for a symmetric $\chi_{\rm eff}$ population distribution. As explored by different studies, the $30-40\ M_{\odot}$ feature seen in the GWTC-3 BBH inferred mass spectrum with models such as POWERLAW + PEAK, cannot be explained by the PISN/50truncreason7 process as this feature is expected to be above the $50M_{\odot}$ mass range (Hendriks et al., 2023; Woosley, 2017; Thrane & Talbot, 2019; Golomb et al., 2023). In the rest of this section we consider different proposed formation channels which could be responsible for this subpopulation.

Isolated binary evolution formation channels generally rely on Roche-overflow mass transfer or close initial binary separations to produce BBHs that are close enough to merge within a Hubble

Parameter	Peak+aligned	Peak+isotropic
$m_{\rm min}/M_\odot$	8.0	8.0
$m_{\rm max}/M_\odot$	70.0	70.0
λ_1	0	0
λ_2	0.4	0.4
α_1	3	3
α_2	0.5	0.5
β	1.5	1.5
$\mu_m/M_\odot$	35.0	35.0
$\sigma_m/M_\odot$	3.0	3.0
κ	3.0	3.0
$\mu_{\chi,1}$	0.2	0.2
$\mu_{\chi,2}$	0.5	0
$\sigma_{\chi,1}$	$10^{-1.1}$	$10^{-1.1}$
$\sigma_{\chi,2}$	$10^{-0.7}$	$10^{-0.8}$
$f1$	0.6	0.6
$f2$	0.4	0.4

Table 4.4: True values for the hyper-parameters characterizing the underlying population in Eq. (4.17) for the three simulated catalogs. The parameters controlling the correlations are chosen to have best-fit values from Abbott et al. (2023d)

time. In the case of Roche overflow, the mass transfer can proceed stably in all cases potentially leading to mass ratio reversals (e.g. Gallegos-Garcia et al., 2021; Broekgaarden et al., 2022; van Son et al., 2022c; Bavera et al., 2023) or unstably leading to common envelope evolution (e.g. Mapelli, 2020a; Zevin et al., 2021; Wong et al., 2021; Belczynski et al., 2022; van Son et al., 2022b). In the case of close initial binary separations, the stellar components are expected to rotate with periods synchronized to the orbital period such that the stars evolve in a chemically homogeneous fashion and thus do not experience significant radius increases over their lifetimes (e.g. Mandel & de Mink, 2016; Marchant et al., 2016; de Mink & Mandel, 2016). While common envelope evolution is primarily expected to produce BBHs with masses $< 30M_\odot$, both stable mass transfer and chemically homogeneous evolution could produce BBHS in this mass range. However, due to the coevolution of the stars in the binary, our results suggest that most of the merger rate contribution from either of these channels are likely not significantly contributing to

the $30-40\ M_{\odot}$ merger rate due to the preference for a symmetric $\chi_{\rm eff}$ population distribution (e.g, Bavera et al., 2020; Briel et al., 2023).

Dynamically formed BBHs can originate from a wide variety of astrophysical environments from triple star systems (e.g., Antonini et al., 2017; Martinez et al., 2022) to young stellar clusters (e.g., Di Carlo et al., 2019, 2020), globular clusters (e.g., Rodriguez et al., 2018; Mapelli et al., 2022), and nuclear star clusters (e.g. Antonini & Perets, 2012; Petrovich & Antonini, 2017), as well as the disks of active galactic nuclei (AGN) (e.g. Bartos et al., 2017; Stone et al., 2017; McKernan et al., 2020). While each of these environments are capable of producing masses near $30-40\ M_{\odot}$, only globular clusters are predicted to produce BBHs with a mass distributions that peaks near $30-40\ M_{\odot}$ and a $\chi_{\rm eff}$ distribution that is symmetric about 0. In the case of nuclear star clusters the mass peaks toward lower values, while in the case of young stellar clusters the mass peaks toward higher values (Cheng et al., 2023; Mapelli et al., 2022). In the case of BBH formation in AGN disks, the mass and spin distributions are heavily influenced by the existence (or lack) of migration traps which predominantly produce low-mass-ratio mergers (e.g., Bellovary et al., 2016; McKernan et al., 2020).

We can isolate the $30-40\ M_{\odot}$ subpopulation and compute the measured local merger rate $R_0^{\rm peak}$ (at $z=0$) under various assumptions. First, we compute the contribution by marginalizing over the remaining mass bins and across the entire $\chi_{\rm eff}$ range. Second, we can compute this contribution $R_0^{{\rm low}\chi}$ over the range $0.05<\chi_{\rm eff}<0.1$ only and also provide an estimate for the local merger rate $R_0^{\rm other}$ in the complementary mass and $\chi_{\rm eff}$ bins, a proxy, for the merger rate due to field binaries and other formation channels which we neglect. In the first case we find, $R_0^{peak}=2.04^{+0.79}_{-0.73}\ {\rm Gpc^{-3}yr^{-1}}$ with the merger rate of other binaries to be $R_0^{\rm other}=22.4^{+10.2}_{-9.4}\ {\rm Gpc^{-3}yr^{-1}}$. Similarly, for the second more restrictive case we consider, we find, $R_0^{{\rm low}\chi}=1.09^{+0.81}_{-0.95}\ {\rm Gpc^{-3}yr^{-1}}$, while for the remaining binaries we find, $R_0^{\rm other}=23.27^{+10.51}_{-9.50}$. We measure a total merger rate of $R_0=24.4^{+10.7}_{-9.2}\ {\rm Gpc^{-3}yr^{-1}}$ for reference. From the estimated local merger rates we estimate the

percentage of binaries contributing to the $30-40\ M_\odot$ feature to be in the range of $1-8\%$.

Under the assumption that the $30-40\ M_\odot$ and $0.05 < \chi_{\rm eff} < 0.1$ regions contain dominant contributions from the dynamical channel, our results indicate that we find merger rates in the range $R_0^{\rm dyn} = 0.1-2.8\ {\rm Gpc}^{-3}{\rm yr}^{-1}$ for these subpopulations. We also find merger rates in the range $R_0^{\rm other} = 13-34\ {\rm Gpc}^{-3}{\rm yr}^{-1}$ for the remaining contribution due to other formation channels. If we assume that this subpopulation is due entirely to mergers occuring in GCs, we expect to have merger rates in the range $1-4\ {\rm Gpc}^{-3}{\rm yr}^{-1}$ and hence find consistency with theoretical models (Mapelli et al., 2022; Rodriguez et al., 2018). However, we find a preference towards the lower and more pessimistic end of the rate predictions. We note that our measured rates are also broadly consistent with other dynamical formation environments such as young star clusters and nuclear star clusters. However, the mass feature being at $30-40\ M_\odot$, makes GCs most consistent with our findings (Mapelli et al., 2022).

4.4.6 Discussion and Conclusion

In this work, for the first time, we have characterized the joint distribution of BBH component masses and effective inspiral spins using a highly data-driven BGP-based population model. Using GWTC-3 data, we have found hints of multiple sub-populations in the BBH mass spectrum that are associated with different ranges of effective inspiral spins. The high (more positive) spin sub-population is found to have a powerlaw-like shape with no feature at the $30-40M_\odot$ range, which is consistent with the isolated binary formation channel and with the predictions of stellar evolution models that place the 50truncreason7 feature at $50M_\odot$ or higher (Hendriks et al., 2023). The low (closer to zero) spin sub-population demonstrates a feature at the $30-40M_\odot$ range and is consistent with the expectations for BBHs formed dynamically in globular clusters, which are thought to peak around $30M_\odot$. These implications are corroborated by the conditional effective spin distributions that we have inferred. We have found that BBHs in the $30-40M_\odot$ range have an effecetive spin

distribution more symmetric about zero as compared to that of other BBHs which correspond to a more positively skewed effective spin distribution. We have also computed the combined merger rates in the $30 - 40M_{\odot}$ and $0.05 < \chi_{\rm eff} < 0.1$ ranges and have found that they are consistent with the theoretical predictions for dynamically formed BBHs in globular clusters.

Our conclusions align broadly with the findings of Godfrey et al. (2023) who employ flexible mixture models to search for BBH subpopulations in GWTC-3 data, except a few subtle distinctions. We both identify two subpopulations in the mass spectrum one of which peaks near $10M_{\odot}$ and falls off steeply afterward, and the other one with a feature near the $30 - 40M_{\odot}$ range. However, Godfrey et al. (2023) find that the spin distributions of these two subpopulations are mostly consistent with each other, particularly the effective-spin distributions which they find to be almost completely overlapping. On the other hand, we find that the subpopulation that is comprised mostly of BBHs with at least one component in the $30 - 40M_{\odot}$ range has a significantly different effective spin distribution from the complementary subpopulation of binaries and hence arrive at slightly different astrophysical interpretations.

We have verified that our results are not susceptible to the inflexibilities that remain in our highly data-driven model. We have shown that our findings from GWTC-3 are robust against variations of binning choices as well as that of the redshift evolution parameter κ. Using large simulated catalogs corresponding to three different populations of BBHs, we have shown that the mass-spin correlations we are seeing cannot manifest spuriously due to artifacts built into our model assumptions, nor from any other kinds of correlations that are known to exist in the astrophysical BBH population and yet have been ignored in our simplified model. Additionally, using two simulated populations of BBHs with intrinsic mass-spin correlations, we have shown that if the subpopulation contributing dominantly to the $30 - 40M_{\odot}$ range had instead been associated with a more positively skewed effective spin distribution, our model would have accurately identified that trend.

At the end of LVK's ongoing fourth observing run, we expect our model to yield tighter constraints on the nature and existence of subpopulations in the BBH mass-spin distributions. As the size of GW catalogs continue to grow, we aim to generalize our model by constructing higher-dimensional BGPs that will self-consistently inform on mass-spin, spin-redshift and mass-redshift correlations in the BBH population and thereby constrain the underlying astrophysics of several BBH formation channels, beyond the limitations of parametric population modeling.

4.5 SUMMARY

In this chapter, I summarized my contributions to the development of a scalable, data-driven population model and how I have used it to constrain the correlations in the astrophysical BBH population given data from growing GW catalogs. In addition to obtaining the first data-driven constraint on BBH mass-redshift correlations I have found, for the first time, hints of a dynamical origin of the $35M_{\odot}$ peak in the BBH mass-spectrum. The mass-redshift results were communicated through Ray et al. (2023c), while the mass-spin results have been communicated in the form of the pre-print Ray et al. (2024) which is currently submitted to a journal. Furthermore, I am involved in implementing these analyses on GWTC-4 data, the exciting results from which will be communicated in the near future.

Chapter 5

Cosmological inference from a catalog of compact binary mergers

Cosmologists are often in error but seldom in doubt.

Lev Landau

5.1 INTRODUCTION

Key questions persist in our understanding of cosmological expansion, such as the mysterious disparity between early and late universe measurements of the local expansion rate (Hubble parameter) and the unconstrained characteristics of the elusive substance (dark energy) driving the accelerating expansion. GW sources such as CBCs are standard sirens in the sense that they provide a direct measurement of their luminosity distance (e.g., Schutz, 1986). Possible measurements of the redshifts of GW standard sirens will facilitate precise and independent measurements of cosmological parameters such as the Hubble constant and the dark energy equation of state, given the large number and potentially high redshifts of CBCs that are expected to be detectable in the future by current (2G) and next-generation (3G) detectors. However, *redshift measurements of GW sources* that rely on the cross-correlation of their measured sky positions with large galaxy catalogs (e.g. Schutz, 1986; Del Pozzo, 2012) or the detection of their EM counterparts (e.g., Abbott et al., 2017a) *are limited by the reach of EM telescopes* which will quickly be exceeded even by the 2G GW detectors. Hence alternative approaches to GW-based cosmological inference are necessary to fully exploit the potential of 2G and 3G GW detectors in shedding light on the Hubble

tension and measuring the dark energy equation of state.

A cosmological model that determines the scaling of detector frame (redshifted) CBC masses with their measured luminosity distances can be constrained simultaneously with the CBC mass spectrum by tracing features in the latter across different redshift bins (Farr, 2019; Ezquiaga & Holz, 2022). This *spectral-siren* method for cosmological inference is independent of EM observations and hence a prime candidate for only GW-based cosmological probes using 2G and 3G detectors. However, the spectral siren method is susceptible to the limitations of strongly model-dependent population analyses discussed in the previous chapter. *Missed or falsely identified features in the inferred mass spectrum can underinform or bias the associated cosmological inference.* The flexible and unbiased population inference offered by the model-independent BGP method discussed in chapter 4 is hence a natural candidate for doing unbiased spectral-siren cosmology.

5.2 DATA-DRIVEN SPECTRAL SIREN METHODS

Gravitational wave standard sirens typically require electromagnetic (EM) data to obtain redshift information to constrain cosmology. Difficult to find EM counterparts for bright sirens and galaxy survey systematics for dark sirens make cosmological constraints with spectral sirens, a gravitational wave data-only approach, extremely appealing. In this work, we use the GWTC-3 BBH detections as spectral sirens to constrain the BBH population and the underlying cosmological expansion with a flexible model for the black hole mass spectrum. We use a binned Gaussian process to model the BBH mass distribution in the source frame without any astrophysical assumptions on the shape and or inclusion (or lack of) features that drive the cosmological constraints as the redshifted detector frame masses become consistent with the underlying astrophysical mass distribution features. For GWTC-3 we find a measurement on the Hubble constant of $H_0 = 73.0^{+13.3}_{-7.7}$ km s^{-1} Mpc^{-1} at 68% C.L. when combined with that obtained from the bright

standard siren analysis with GW170817 and its associated host galaxy NGC 4993. We find an improved estimate for the Hubble constant of around a factor of 1.4 times better than the GW170817 measurement alone. We validate our nonparametric spectral siren approach with simulations and benchmark its scalability and constraining performance when compared with parametric methods. $H_0 = 73.0^{+13.3}_{-7.7}\ \mathrm{km\ s^{-1}\ Mpc^{-1}}$

With the recent release of the third Gravitational-wave Transient Catalog (Abbott et al., 2023b, GWTC-3) from the LIGO Scientific, Virgo, and KAGRA Collaborations (LVK, J. Aasi et al., 2015; Acernese et al., 2014; Akutsu et al., 2021; Abbott et al., 2018b), the number of confident gravitational wave (GW) detections from compact binary mergers is around 70, most being from binary black hole (BBH) mergers. The increasing size of GW catalogs has enabled the study of the BBH population (Abbott et al., 2021g, 2023d), the cosmic expansion history (Abbott et al., 2021a,b) as well as allowing for extensive tests of general relativity (GR) within the strong field regime (Abbott et al., 2023d, 2021h, 2023e). With half of the LVK's fourth observing run (O4) having concluded, the number of additional BBH merger detections has already doubled, and by its conclusion, we expect to have $\mathcal{O}(300)$ additional BBH mergers and $\mathcal{O}(10)$ mergers that contain at least one neutron star with $\mathcal{O}(1)$ being a multimessenger event.

Gravitational waves have become a promising avenue to study the cosmic expansion history of the universe. Measurements of the Hubble constant (H_0) with local probes such as type 1A supernovae standard candles (Riess et al., 2022) are in tension with early universe constraints from the cosmic microwave background (CMB) (Planck Collaboration, 2018). These two state-of-the-art measurements provide an independent way of determining the Hubble constant at the percent level, however, there is currently a larger than 5-sigma tension between these constraints, implying either new physics or unaccounted sources of systematics (Riess et al., 2022; Valentino et al., 2021).

To potentially resolve this tension, one ideally needs a third independent cosmological probe.

Gravitational wave sources are so-called standard sirens (Schutz, 1986; Holz & Hughes, 2005), that is, they provide an absolute measurement for the luminosity distance to the source without the need for the cosmic distance ladder as a calibrating step. To do cosmology with GW sources, we require an independent measurement of the source redshift.

For bright sirens, such as binary neutron star mergers we can attempt to detect the associated electromagnetic counterpart (EM) e.g., a kilonovae, and measure its redshift directly (Abbott et al., 2017c,a). For dark sirens, such as binary black hole mergers or bright sirens without a detectable (or missed) EM counterpart, we can determine the redshift statistically using galaxy surveys as prior information on the potential host galaxies for these sources (Schutz, 1986; Del Pozzo, 2012; Nair et al., 2018; Chen et al., 2018; Fishbach et al., 2019; Gray et al., 2019; Soares-Santos et al., 2019; Abbott et al., 2021a; Palmese et al., 2020; Mukherjee et al., 2021a; Diaz & Mukherjee, 2022; Ghosh et al., 2023b; Mukherjee et al., 2021b; Gray et al., 2023, 2022). The growing catalog of BBH mergers is thus critical for cosmological studies since multimessenger events have proven difficult to find.

However, even without accessible electromagnetic information, either as EM counterparts for direct H_0 measurements or complete enough galaxy surveys, one can still make a statistical measurement of redshift using the features of the population distribution of compact binary mergers. Since we measure redshifted detector-frame masses, $m^{\rm det} = m(1+z)$, one can model the expected source frame mass distribution to estimate the redshift z. So called spectral sirens (Ezquiaga, 2021; Ezquiaga & Holz, 2022) therefore allow for a direct measurement of the cosmic expansion history with gravitational wave data alone.

The mass spectrum of LIGO–Virgo–KAGRA events introduces at least five independent mass features (Abbott et al., 2023d): the upper and lower edges of the pair-instability supernova (PISN) gap (Woosley et al., 2002; Heger & Woosley, 2002; Heger et al., 2003; Woosley, 2017, 2019), the upper and lower edges of the neutron star–black hole gap (Ye & Fishbach, 2022), and the

minimum neutron star mass (Suwa et al., 2018). However, the location of these features is still uncertain, and some BBH formation channels might form events with masses in the gaps, e.g., hierarchical mergers (Gerosa & Fishbach, 2021) leading to a more complex mass spectrum. By using the full mass distribution, degeneracies between mass evolution and cosmological evolution can be broken. This self-calibrating spectral siren method has the potential to provide precision constraints of both cosmology and the potential evolution of the mass distribution. For recent works on cosmological inference that relies on the astrophysical mass distribution of CBCs see Mastrogiovanni et al. (2021, 2023, 2024); Gray et al. (2023)

However, these studies (Farr, 2019; Ezquiaga & Holz, 2022; Chen et al., 2024) have assumed a priori some knowledge about the underlying astrophysical processes to forward model the CBC population with simple parametric population models. For example, modeling expected features in the mass spectrum which are then used to measure the cosmological expansion. However, Pierra et al. (2023) has shown that incorrect assumptions regarding the shape of the mass-spectrum and its redshift evolution can lead to significant (3σ) biases in the inferred cosmological parameters. For more details regarding the redshift evolution of the mass spectrum and its impact on the associated cosmological inference, see Mukherjee (2022); Karathanasis et al. (2023).

Therefore, given the significant uncertainty regarding known BBH formation models and the new features found in the BBH mass spectrum (Abbott et al., 2023d), there is a need for a flexible and non-parametric approach to spectral siren cosmology. Examples of non-parametric population models include: autoregressive processes (Callister & Farr, 2024), splines (Edelman et al., 2022a, 2023), Gaussian mixture models (Tiwari, 2021; Tiwari & Fairhurst, 2021; Tiwari, 2022), adaptive kernel density estimation Sadiq et al. (2022), maximum population likelihood (Payne & Thrane, 2023), Dirichlet processes (Rinaldi & Del Pozzo, 2021) and binned Gaussian processes (GPs, Ray et al. (2023c); Mohite (2022); Abbott et al. (2023d)).

In this work, we employ, for the first time, a flexible population model on the BBH popula-

tion distribution and perform hierarchical Bayesian inference on the GWTC-3 BBH detections to measure the cosmic expansion history using the spectral siren methodology. We model the BBH population using the binned gaussian process (BGP) approach (Ray et al., 2023c; Mohite, 2022; Abbott et al., 2023d) to place constraints on the Hubble constant without assuming a particular shape on the mass distribution of BBH mergers. For a discussion regarding how our method compares with a parallel investigation on non-parametric spectral sirens (Farah et al., In prep.), see Sec. 5.2.3.

This paper is organized as follows. In Section 5.2.1, we summarize the hierarchical Bayesian framework describing the BGP spectral siren model used in our analysis. In Section 5.2.1.4, we present the main results of this paper by analyzing the GWTC-3 BBH observations with our framework. In Section 5.2.2, we test our methodology with a simulated population that mimicks the GWTC-3 population and current GW catalog sizes. Finally, in Section 5.2.3 we provide a summary of our work and discuss future directions.

5.2.1 Methods

In this section, we describe the hierarchical inference framework used to simultaneously infer the cosmological parameters with the shape of the CBC mass spectrum.

5.2.1.1 Flexible population model

Following Ray et al. (2023c); Mohite (2022), we construct our flexible population model in the source frame as a piece-wise binned function in masses and redshift, such that,

$$\frac{dN}{dm_1 dm_2 dz}(m_1, m_2, z|n^\gamma, n^\alpha) = \frac{n_m^\gamma n_z^\alpha}{m_1 m_2}\frac{dV}{dz}T_{\rm obs}(1+z)^{-1}, \tag{5.1}$$

where $n_m^\gamma n_z^\alpha$ is the merger rate density of CBCs per log component mass per co-moving volume per unit time in the γth mass bin and the αth redshift bin. While the model in Equation 5.1 is more

flexible than the one used in Mohite (2022) which assumes no evolution of the merger rate, it is less so than the correlated model of Ray et al. (2023c) where the joint distribution of masses and redshifts were modelled using a single piece-wise binned function.

Keeping in mind that the primary objective of measuring the Hubble parameter independent of the uncertainties regarding CBC formation channels, and the large measurement uncertainties in the mass-redshift correlations reported by Ray et al. (2023c), we expect the un-correlated mass-redshift model given by Equation 5.1 to cause no significant biases in our results given the size of current datasets. However, as catalogs continue to grow, we hope to relax this restriction on the existence of mass-redshift correlations in future studies for which we will use the generalized correlated model of Ray et al. (2023c).

To fit a population model such as Equation 5.1 to GW observations of CBC systems, it is necessary to convert detector frame observables such as redshifted masses and luminosity distances into their corresponding source-frame counterparts using a fixed cosmological model. In the next subsection, following the spectral siren method of Farr (2019); Ezquiaga & Holz (2022), we describe how to vary and infer the cosmological parameters simultaneously with the population distribution in the context of our flexible model.

5.2.1.2 Simultaneous cosmological inference

Given a cosmological model characterized by parameters Ω, it is possible to express the population model described in the previous section as a function of detector frame CBC observables instead,

$$\begin{aligned}&\frac{dN}{dm_1 dm_2 dz}(m_1^d, m_2^d, D_L|n_m^\gamma, n_z^\alpha, \Omega)\\&= \frac{n_m^\gamma n_z^\alpha}{m_1^d m_2^d}\frac{dV_c}{dz}\left(D_L|\Omega\right)T_{\rm obs}\left(1+z(D_L|\Omega)\right),\end{aligned} \tag{5.2}$$

where m_1^d, m_2^d and D_L are the observed detector frame masses and luminosity distances of CBCs and $z(D_L|\Omega)$ is the redshift of the source as a function of luminosity distance given a set of cosmological parameters. As usual, dV_c/dz is the differential uniform-in-comoving volume element, $T_{\rm obs}$ is the total observation time, and the extra factor of $1/(1+z)$ converts source-frame time to detector-frame time.

By modeling the occurrence of CBCs as realizations of an inhomogeneous Poisson process, it is possible to infer hyper-parameters characterizing the spectral siren model of Equation 5.2 from the posterior samples of masses and luminosity distances of each observed CBC using Bayesian hierarchical inference (Mandel et al., 2019; Loredo, 2004; Vitale et al., 2020). Within this framework, the likelihood of population and cosmological hyper-parameters given GW data (d) from all events in an observed catalog can be constructed as,

$$\begin{aligned} &p(\vec{d}|\vec{n}_m, \vec{n}_z, \Omega) \\ &= e^{-N_{\rm det}(\vec{n}_m, \vec{n}_z, \Omega)} \prod_i^{N_{\rm obs}} \left\langle \frac{\frac{dN}{dm_1 dm_2 dz}(\vec{n}_m, \vec{n_z}, \Omega)}{p_{\rm PE}(m_1, m_2, z)} \right\rangle_{{\rm samples},i}, \end{aligned} \tag{5.3}$$

where $\langle\cdot\rangle_{{\rm samples},i}$ represents a Monte Carlo sum over posterior samples from the ith event reweighted by the prior used in its parameter estimation and $N_{\rm det}$ is the expected number of detections as a function of hyper-parameters.

Since the criteria for detection takes the form of a threshold imposed on some statistic such as false alarm rate, the estimation of $N_{\rm det}$ has to account for Malmquist biases resulting from the imposition of the detection criteria (Mandel et al., 2019; Loredo, 2004). This is implemented by simulating a fiducial population of CBCs and injecting them into detector noise realizations Essick & Farr (2022); Farr (2019). The set of simulated sources that are detected above thresholds can be re-weighted to the population model of Equation 5.2 to estimate $N_{\rm det}(\vec{n}_m, \vec{n}_z, \Omega)$ in the following

way:

$$N_{\rm det}(\vec{n}_m, \vec{n}_z, \Omega) \\ = \frac{K_{\rm det}}{K_{\rm draw}} \left\langle \frac{\frac{dN}{dm_1 dm_2 dz}(\vec{n}_m, \vec{n}_z, \Omega)}{p_{\rm draw}(m_1, m_2, z)} \right\rangle_{\rm samples,det} \tag{5.4}$$

where similarly, $\langle \cdot \rangle_{\rm samples,det}$ represents a Monte Carlo sum over the parameters of detectable events re-weighted by the population used to generate the mock simulations. For a discussion regarding the convergence of Monte Carlo sums used to evaluate various terms in the likelihood see Appendix B.2

The likelihood defined in Equation 5.3 can then be used to inform the shape of the CBC mass spectrum, the redshift evolution of the merger rate and the cosmological parameters self-consistently given a catalog of GW events.

5.2.1.3 Gaussian process prior and HMC sampling

To infer the shape of the mass distribution using the likelihood of Eq. (5.3), we draw the rate densities in each mass bin from a Gaussian process prior (Ray et al., 2023c; Mohite, 2022) so that the posterior on the rate densities is given by,

$$p(\vec{n}_m, \vec{n}_z, \vec{\mu}, \vec{\sigma}, \vec{l} | \vec{d}, \Omega) \propto p(\vec{d} | \vec{n}_m, \vec{n}_z, \Omega) \\ \times p(\vec{\mu}, \vec{\sigma}, \vec{l}) p(\vec{n}_m | \vec{\mu}_m, \sigma_m, l_m) \\ \times p(\vec{n}_z | \vec{\mu}_z, \sigma_z, l_z) \tag{5.5}$$

where $p(\vec{n}|\vec{\mu}, \sigma, \lambda)$ is the GP prior and $\vec{\mu}$ is the mean function of the GP while l, σ are parameters that control the correlation length and amplitudes of the GP's covariance matrix.

The prior on the GP hyperparameters is defined by $p(\vec{\mu}, \sigma, l)$. Following previous implemen-

tations of the binned Gaussian process population analysis, we chose the priors on the GP hyper-parameters $(\vec{\mu}, l, \sigma)$ to be Normal, Log Normal, and Half Normal respectively while modeling the covariance matrix as an exponential quadratic function.

Along with the rate-densities, we simultaneously draw the cosmological parameters uniform priors and use Equation 5.3 to sample the posterior distribution,

$$p(\vec{n}_m, \vec{n}_z, \Omega|\vec{d}) \propto p(\vec{n}_m, \vec{n}_z, \vec{\mu}, \vec{\sigma}, \vec{l}|\vec{d}, \Omega)p(\vec{\Omega}) \tag{5.6}$$

The posterior samples of $\vec{\Omega}$ represent measurements of the cosmological expansion that have been marginalized over the uncertainties regarding the shape of the CBC mass spectrum without making any astrophysical assumptions about any features present in the mass spectrum. Simultaneously, the samples of $\vec{n}$ can then be used to reconstruct the shape of the CBC mass distribution in a data-driven way, while being free of any biases that might result from the uncertainties in the measurements of cosmological parameters.

5.2.1.4 Results for GWTC-3

In this section, we present our BGP spectral siren constraints using GWTC-3 BBH mergers. Following the LVK GWTC-3 population analysis choices Abbott et al. (2023d), we make use only of the BBH events that pass a 1 per year IFAR threshold. That is, we analyze a total of 69 confidently detected BBH mergers and exclude GW190814 as a population outlier. For the events incorporated in our analysis, the parameter estimation samples of detector frame masses and luminosity distances are used to compute the posterior weights of Equation 5.3. We also use the publicly available LVK's GWTC-3 sensitivity estimate injection campaigns to compute Equation 5.4 (The LIGO Scientific Collaboration et al., 2021, 2023).

We infer the mass rate densities as well as the redshift evolution of the BBH merger rate jointly with the cosmological parameters H_0, Ω_m. In Figure 5.1, we show the inferred marginal posterior

distributions on both the primary mass m_1 and secondary mass m_2 along with the redshift evolution of the combined merger rate. We also show the corresponding LVK marginal posterior distributions on the masses for the POWERLAW+PEAK model for reference Talbot & Thrane (2018); Abbott et al. (2023d). Our results show broad consistency with the LVK POWERLAW+PEAK posterior. Our results are also consistent with the BGP results presented in Ray et al. (2023c) which assume a fixed cosmological model.

In Figure 5.2, we show the corner plot for the joint posterior distribution on the Hubble constant H_0, the matter density parameter Ω_m, as well as the GP kernel length scales l_m, l_z. We find a BGP spectral siren BBH measurement on the Hubble constant of $H_0 = 82.2^{+51.7}_{-33.7}$ km s^{-1} Mpc^{-1} while for the matter density parameter, we find $\Omega_m = 0.3^{+0.1}_{-0.1}$ both at 68% C.L. We measure length scales $l_m = 2.1^{+1.0}_{-1.1}$, $l_z = 2.1^{+1.0}_{-1.1}$ at 68% C.L.

In Figure 5.3, we show joint constraints for our BGP spectral siren measurement on the Hubble constant with the 69 BBHs from GWTC-3 only when combined with the bright siren H_0 measurement from GW170817 and its EM counterpart NGC 4993 (Abbott et al., 2017c,a). We find a joint H_0 measurement of $H_0 = 73.0^{+13.3}_{-7.7}$ km s^{-1} Mpc^{-1} at 68% C.L. which is an improvement with a factor of 1.5 times better relative to the GW180817 with NCG 4993 measurement alone ($H_0 = 71^{+23}_{-8}$ km s^{-1} Mpc^{-1} at 68% C.L.). In all of our reported H_0 measurements, we used a uniform prior in the range $[20, 140]$ km s^{-1} Mpc^{-1}.

5.2.2 Validation of Results

We validate our inference on simulated GW catalogs made from events drawn from a known fiducial population and cosmological model. We draw simulated BBH observations from the following models: for the primary mass, we draw from the POWERLAW + PEAK model (Talbot & Thrane, 2018), mass ratios are drawn from a powerlaw in q with slope β (Fishbach & Holz, 2020) and we allow for the merger rate to evolve with redshift (Fishbach et al., 2018). The fiducial hyperparam-

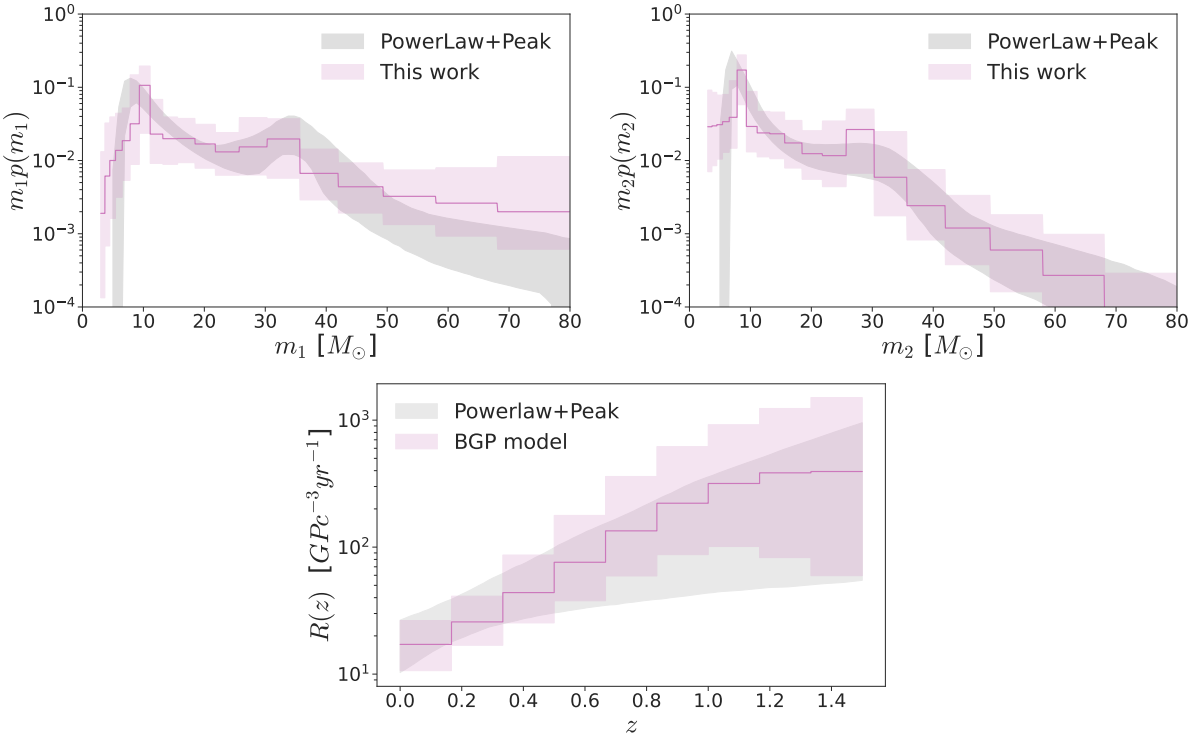


Figure 5.1: We show our results as the binned piecewise function plotted in purple and for comparison, we also show the corresponding results obtained by the LVK using the POWERLAW + PEAK model in Abbott et al. (2023d) as the gray band depicting the 95% posterior credible region.

eter values for the simulated population correspond to $m_{\rm min} = 5M_\odot$, $m_{\rm max} = 65M_\odot$, $\alpha = 3.14$, $\beta = 1.4$, $m_{\rm peak} = 35M_\odot$, $\sigma_{\rm peak} = 5M_\odot$, $f_{\rm peak} = 0.01$, and $\kappa = 3$. For explicit expressions, we refer the reader to Appendix 1 of X and to Figure 5.4 for the marginal posterior distributions over m_1 and m_2 that we use in our simulations. We use a spatially flat ΛCDM cosmological model with assumed Planck 2015 cosmological parameters, that is, with $H_0 = 67.8\ {\rm km\ s^{-1}\ Mpc^{-1}}$ and $\Omega_m = 0.308$ in our simulations.

The drawn masses and redshifts from the fiducial population are then assigned corresponding observed values calibrated to the expected measurement uncertainties for BBH mergers at advanced LIGO design sensitivity (Chen et al., 2018; Abbott et al., 2018b; Fishbach et al., 2018) and selected so that only a subset is observable, i.e., simulating the selection effects seen in GW

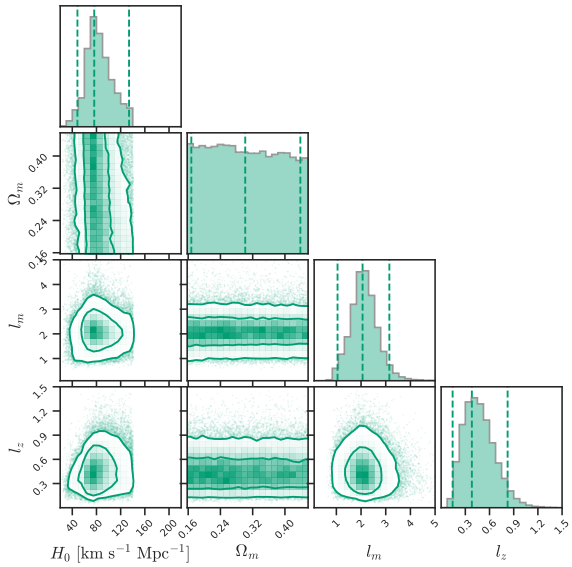


Figure 5.2: Corner plot showing the joint marginalized posterior distribution on the Hubble constant H_0, the matter density parameter Ω_m, and the inferred length scales l_m, l_z from our BGP spectral siren model and the 69 GWTC-3 BBH events.

observations. The details of this procedure are described in Farah et al. (2023); Fishbach et al. (2020, 2018). A corresponding set of simulated detectable injections is generated from a broad distribution to take into account the estimation of selection effects, a requirement for estimating $N_{\rm det}$ through Equation 5.4.

Our catalog is comprised of 144 observed BBH events drawn from the fiducial population model and underlying cosmological model. As with our analysis of the GWTC-3 BBH events, we employ the BGP spectral siren model to fit our simulated population with the model defined in Equation 5.2. In Figure 5.4 we show the inferred marginal posterior distributions on both the primary mass m_1 and secondary mass m_2 using our simulated data. We can see that the BGP model is sufficiently accurate at recovering the shape of the fiducial mass population with the 121 events that we consider. Although there is some discrepancy in the inferred shape of the primary mass population at high masses, we are still broadly consistent with the simulated population. We note

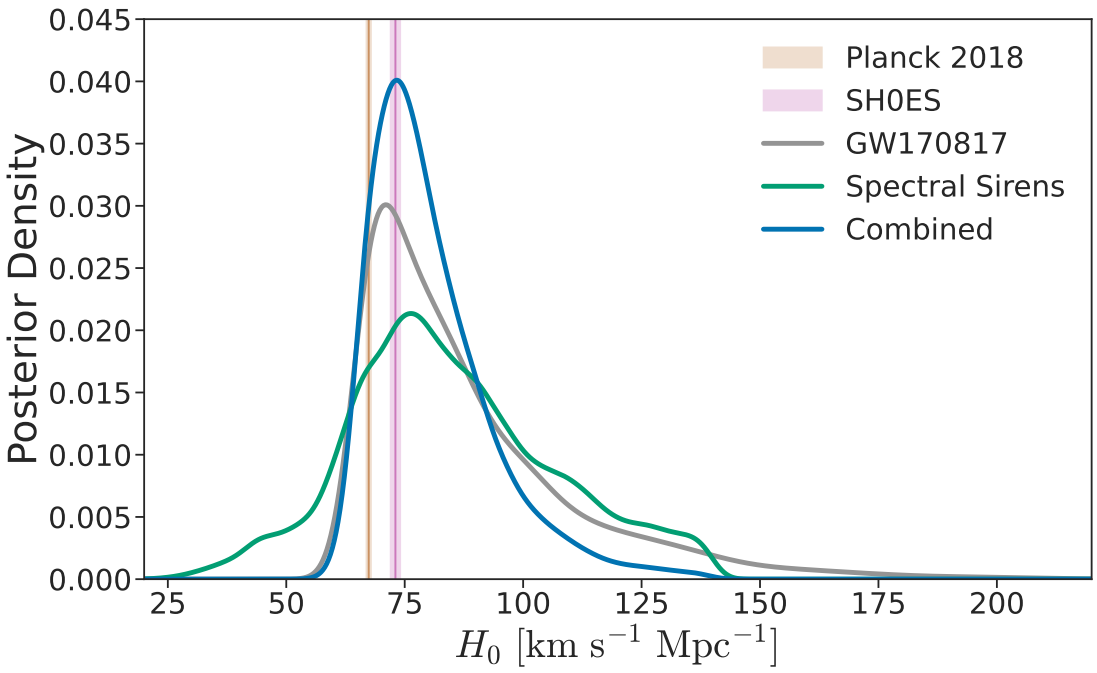


Figure 5.3: In green, we show the posterior from the 69 GWTC-3 BBH events we have considered in the spectral siren analysis using our BGP model. In gray, we show the GW170817 H_0 posterior obtained with its host galaxy NGC4993. In blue, we show the resulting combined GW170817 and GWTC-3 BBH spectral siren posterior to obtain a joint measurement on H_0. For reference, in the orange band we show measurements on H_0 from the CMB (Aghanim et al., 2020) and in the pink shaded band from standard candle type 1A SN measurements (Riess et al., 2019).

that this discrepancy is due to a small number of events out of the total in our catalog being at the higher end of the mass spectrum as well as due to the increased uncertainty in their corresponding single-event parameter estimation posterior distributions.

We infer the BBH mass distribution as well as the redshift evolution of the BBH merger rate jointly with the Hubble constant H_0. In Figure 5.4, we show the inferred marginal posterior distributions on both the primary mass m_1, secondary mass m_2, and redshift. We also show the fiducial populations for comparison. In figure 5.5, we show the posterior distributions of the cosmological parameters. The BGP spectral siren method can accurately reconstruct the shape of the injected population as well as the fiducial values of the cosmological parameters.

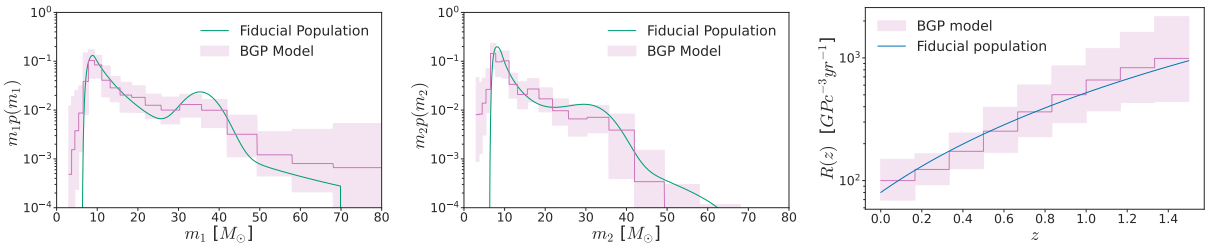


Figure 5.4: We show our results as the binned piecewise function plotted in purple. The simulated population is shown as the solid green lines.

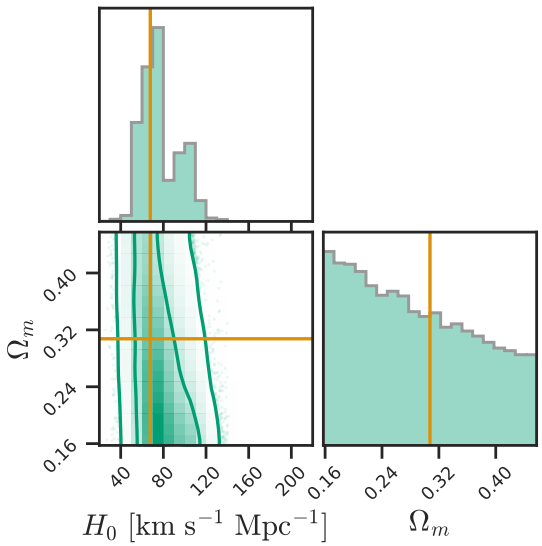


Figure 5.5: Corner plot showing the joint posterior distributions on the Hubble constant H_0, the matter density parameter Ω_m, and the inferred GP length scales for the simulated population described in Section 5.2.2. We plot the simulated values for both $H_0 = 67.8$ km s^{-1} Mpc^{-1}, $\Omega_m = 0.308$ for reference in gray.

5.2.3 Discussion and Conclusion

In this work, we have explored how the spectral siren framework can be used jointly with non-parametric approaches in the modeling of the BBH population. In particular, we used a binned Gaussian process to model the black hole mass distribution without assuming any particular shape inclusive (or exclusive) of features expected from the astrophysics of BBH formation channels. Our proof of principle demonstration made use of the GWTC-3 confident BBH detections to place the first nonparametric constraints on the Hubble constant. We have validated our approach with GWTC-3 like populations and have made use of current catalog sizes to demonstrate the constraining power of our nonparametric method relative to traditional parametric approaches to spectral siren cosmology.

In a parallel investigation, Farah et al. (In prep.) have also attempted to use Gaussian-process-based mass models on a simulated dataset for a proof of concept demonstration of astrophysics agnostic spectral siren cosmology. Even though their non-parametric population model is not restricted by bin resolutions, the mock data analysis presented in Farah et al. (In prep.) models only the primary mass distribution of CBCs with gaussian processes. They further assume that all BBHs are equal mass systems and use a parametric model for the redshift evolution of the merger rate. By simulating a large population of equal mass BBHs with a known functional form of the redshift evolution of the merger rate, they demonstrate that their method can implement accurate spectral siren cosmology in a simplistic Universe while being free of the necessity to bin up the space of BBH parameters. We note that while an unbinned population model in both component masses and redshifts could in principle be more flexible than ours, Farah et al. (In prep.) have not demonstrated whether their method can be scalably generalized and applied to more realistic datasets wherein BBH masses are expected to follow more complicated pairing functions (Abbott et al., 2023d) than a delta function in mass ratio centered around one. Furthermore, given the strong correlations between the measurement uncertainties of the Hubble parameter and the red-

shift evolution parameters (Abbott et al., 2021b), previously unmodelled features in the redshift evolution function can serve to equally bias the cosmological parameters. On the other hand, our model can reconstruct features in the joint distribution of component masses and the redshift evolution function completely driven by the data, up to the resolution limit imposed by our choice of binning, variations of which are found to yield consistent results.

As the size of GW catalogs grow, we expect to expand our nonparametric data-driven approach to take advantage of potential correlations between mass and redshift. In this work, we have assumed an uncorrelated mass-redshift distribution for simplicity. This choice can be relaxed as was shown in Ray et al. (2023c) where a single three dimensional BGP infers the shape of the joint mass-redshift distribution. However, as it was shown in Ray et al. (2023c), most of the structure is washed out due to the increased dimensionality of the problem and thus would require larger GW catalogs to measure such correlations. This is important, if say for examples, the BBH mass distribution evolves at roughly the same rate as the cosmic expansion history as argued by Ezquiaga & Holz (2022). We leave these simulated and systematic studies as future work.

Another caveat of our analysis is the fact that we only use BBH mergers to infer the cosmic expansion rate. Our methodology is CBC agnostic, that is, we fit for the merger rate density in mass bins. Including the whole population of CBC mergers (such as the inclusion of BNS and NSBH mergers) should make use of the full mass distribution of CBCs, a more powerful and event category agnostic spectral siren probe as explored by Ezquiaga & Holz (2022).

Our method can also be expanded to be model agnostic concerning the underlying functional of the background cosmological model. In our work, we have assumed a spatially flat ΛCDM cosmological model which is sufficient given the number of observations that we have analyzed and their associated single event parameter uncertainties. Fitting the luminosity distance and redshift relation, $D_L(z)$, with a GP may provide a way to test the underlying cosmological model with future networks of GW detectors.

Finally, nonparametric methods are data-driven by nature and therefore require large amounts of observations for meaningful inference. As such, the scalability of such methods is critical. However, for the next generation of GW detectors (XG/3G), the development of scalable and potentially approximate inference will likely be necessary to make use of the $\mathcal{O}(10^4 - 10^6)$ expected GW detections catalog sizes (Chen et al., 2024).

5.3 SUMMARY

In this chapter I delineated my contributions to the development of a data-driven spectral siren method based on the BGP population model of chapter 4, that can conduct cosmological inference with GW data alone, agnostic of both the distance ladder and the astrophysics of compact binary formation. Using this approach we obtained the first data-driven constraint on the cosmological parameters using GW data alone. With future catalogs (O5 and beyond), this method will be crucial in obtaining unbiased, sub-percent level measurements of the Hubble parameter and potentially other cosmological parameters. The development of this data-driven analysis, a first-of-its-kind approach towards cosmological inference, and the results yielded by it are communicated in the form of a pre-print: Magaña Hernandez & Ray (2024) and is currently submitted to a journal.

Part IV

Probing the nature of fundamental interactions using gravitational waves

Chapter 6

Inferring the neutron star equation of state from multiple observations

To achieve this density of a neutron star at home, just cram a herd of 50 million elephants into the volume of a thimble.

Neil deGrasse Tyson

6.1 INTRODUCTION

The NS EoS is a universal relationship between the density and pressure of degenerate matter at supranuclear densities which is known to determine the unique relationship between source frame mass and finite size properties (such as tidal deformability) of NSs (Özel & Freire, 2016; Lindblom, 1992). Measurements of the EoS are therefore fundamentally important for understanding the poorly constrained nature of low energy strong nuclear interactions and for facilitating cosmological measurements by 3G GW detectors in the absence of EM counterparts (Del Pozzo et al., 2017; Messenger & Read, 2012). Direct observation of NSs provides a natural laboratory for measuring the EoS, in complement to and beyond the reach of terrestrial experiments (e.g. Özel & Freire, 2016).

The observation of GWs from NS containing compact binary mergers by the LVK, X-rays from pulsars in isolated or binary systems by NASA's Neutron Star Interior Composition Explorer Mission (NICER, Watts et al., 2016) and European Space Agency's X-ray Multi-Mirror Mission (XMM Newton), and radio waves from binary pulsar systems[1] can each facilitate mea-

[1]The EoS uniquely determines the maximum mass of NSs. Mass measurements by radio observations of binary pulsars will constrain the lower-bound on the maximum mass.

surements of the EoS. The universality of the EoS can be leveraged to combine information from multiple observations, even across different channels (such as the GW and EM channels) to obtain joint measurements of the EoS whose precision increases with the number of observations analyzed. With the LVK expecting O(10-100) new BNS detections in the near future (Abbott et al., 2018b) and a growing list of pulsars observed in X-ray and radio, precise EoS measurements are imminent.

To extract information on the EoS from multiple astrophysical observations, several analysis frameworks have been developed based on Bayesian Hierarchical inference. Unfortunately, hierarchical EoS inference suffers from dimensional reduction which prevents the direct re-use of EoS agnostic inferences of the higher dimensional parameter space (Wysocki et al., 2020). Hence, such an analysis incurs a heavy computational cost that grows with the number of observations analyzed. This lack of scalability with dataset size can lead to computational infeasibility even for a GW-only dataset post LVK's upcoming observing runs.

Throughout the rest of this chapter, I will summarize my contributions to the development of a scalable EoS inference pipeline that surmounts the problem of dimensional reduction by re-using individual event measurements using kernel density estimation. In particular, I will use the hierarchical likelihood of Eq. (2.53) implemented through Eq. (2.57), to conduct rapid EoS inference from multiple BNS and NSBH observations.

6.2 SCALABLE INFERENCE FROM MULTIPLE BNS OBSERVATIONS

Bayesian hierarchical inference of phenomenological parameterized neutron star equations of state (EoS) from multiple gravitational wave observations of binary neutron star mergers is of fundamental importance in improving our understanding of neutron star structure, the general properties of matter at supra nuclear densities and the strong nuclear force. However, such an analysis is computationally costly as it is unable to re-use single-event EoS agnostic parameter estimation

runs that are carried out regardless for generating gravitational wave transient catalogs. With the number of events expected to be observable during the 4th observing run (O4) of LIGO/Virgo/KAGRA, this problem can only be expected to worsen. We develop a novel and robust algorithm for rapid and computationally cheap hierarchical inference of parameterized EoSs from gravitational wave data which re-uses single event EoS agnostic parameter estimation samples to significantly reduce computational cost. We efficiently include a priori knowledge of neutron star physics as Bayesian priors on the EoS parameters. The high speed and low computational cost of our method allow for efficient re-computation of EoS inference every time a new binary neutron star event is discovered or whenever new observations and theoretical discoveries change the prior on EoS parameters. We test our method on both real and simulated gravitational wave data to demonstrate its accuracy. We show that our computationally cheap method produces EoS constraints that are completely consistent with existing analysis for real data, the chosen fiducial EoS for simulated data. Armed with our fast analysis scheme, we also study the variability of EoS constraints with binary neutron star properties for sets of simulated events drawn in different signal-to-noise ratio and mass ranges.

6.2.1 Motivation

The interior of a neutron star (NS) is one of the most extreme environments known to exist in the observable universe. The hydrostatic equilibrium established by the NSs own gravity acting against the degeneracy pressure of neutrons in its core is described by the Tolman-Oppenheimer-Volkoff (TOV) equations (Oppenheimer & Volkoff, 1939). In order to solve the TOV equations, we need an equation of state (EoS) that relates the neutron degeneracy pressure inside the NSs core to its density at zero temperature. Due to the rapid cooling of NSs via neutrino emission and hence their fast β-equilibration, finite temperature and dissipative corrections to the EoS are negligible, allowing for NS matter to be modeled as a perfect fluid (Friedman & Stergioulas, 2013).

The EoS of cold matter at supra nuclear densities is thought to be universal and has a one-to-one correspondence with the mass radius relationship of NSs (Lindblom, 1992). Hence, knowledge of the cold matter nuclear EoS is of extreme importance in understanding not only the structure of NSs, but also the general properties of matter at such densities. However, limitations in our understanding of the strong nuclear force restrict our knowledge of the NS EoS to a number of approximate models. Quantitatively, the pressure density and the implied mass radius relationship for an EoS model $\mathcal{E}$ becomes:

$$p = p_{\mathcal{E}}(\rho) \iff r = r_{\mathcal{E}}(m) \tag{6.1}$$

where p, ρ are the pressure and rest mass density at some point in the NS core, and m, r are the mass and radius of the NS.

The tabulated set of equation of state models constitutes a discrete collection of models based on different theoretical descriptions of nuclear theories (for a recent list of tabulated EoS models, see (Özel & Freire, 2016)), allowing us to compare one model against the others. However, this does not give us the ability to fully survey the pressure-density parameter space of matter at extreme densities. Parameterized EoS models such as the spectral decomposition of the adiabatic index in powers of logarithmic pressure (Lindblom, 2010) and the piecewise polytropic parameterization (Read et al., 2009), are more flexible in the sense that a chosen range of the parameters correspond to a continuum in the pressure density space, which can be constrained empirically, given data. However, the data required for inferring the EoS are challenging to acquire from controlled experiments. The high density of NS cores, orders of magnitude higher than the nuclear saturation density (Özel & Freire, 2016), is impossible to re-create in a terrestrial laboratory. This has resulted in the NS EoS remaining largely unconstrained for a long time (Özel & Freire, 2016).

Observations of systems involving neutron stars often enable physicists to extract information regarding the NS EoS from measurements of the mass and/or radii of the NSs in such systems (Özel

& Freire, 2016). The one-to-one correspondence between the NS mass-radius relationship and the EoS pressure-density relationship translates simultaneous mass-radius measurement into a measurement of the NS EoS itself. Observations of electromagnetic (EM) signals from such astrophysical systems, such as X-ray pulsar observations (Bogdanov et al., 2019; Miller et al., 2019, 2021; Silva et al., 2021), thermonuclear bursts and quiescent low mass X-ray binaries (Özel et al., 2016) etc, have been used to simultaneously measure the mass and radii of NSs and hence the NS EoS. However, these analyses are model-dependent and can be subject to the systematics of the individual EM emission models (Özel & Freire, 2016). Mass measurements of NSs through pulsar timing experiments (Hu et al., 2020), have also been informative about the NS EoSs using the fact that an EoS model implies a particular maximum mass of NSs. However, such precise mass measurements are possible only for pulsars in binary systems which comprise only 10 percent of the galactic pulsar population (Manchester et al., 2005; ATNF, 2021). Thus, observations of EM signals from astrophysical NS systems have provided us with a natural laboratory for exploring the properties of cold nuclear matter at extreme densities, albeit in a rather limited notion. For a review of EoS measurements through electromagnetic signals from astrophysical NS systems see (Özel & Freire, 2016).

Observations of gravitational waves (GWs) from merging binary neutron stars (BNS)s at ground based GW observatories such as LIGO (J. Aasi et al., 2015) and Virgo (Acernese et al., 2014) has resulted in significant advances in the field of NS EoS measurement. The GW waveform, in addition to being sensitive to the component NS masses, carries an imprint of finite size effects in the evolution of BNS systems. Thus analyzing GW data from ground based observatories allows for the simultaneous measurements of NS mass and tidal deformability (akin to the radius), that are independent of EM emission model systematics. Hence GW based NS mass-radii measurements yield EoS measurements that are free of the limitations of EM signal based EoS measurements. The universality of cold nuclear matter EoS can be exploited to combine GW observations from

multiple BNS events together, or along with other astrophysical and terrestrial observations, to yield joint measurements of the NS EoS. GW data from the two merging BNSs observed to date, GW170817 (Abbott et al., 2017c) and GW190425 (Abbott et al., 2020b), analyzed together with EM counterpart observations of GW170817 (Goldstein et al., 2017; Savchenko et al., 2017; Coulter et al., 2017), other EM based astrophysical observations, terrestrial measurements and theoretical considerations have yielded joint EoS constraints that have greatly improved our understanding of dense nuclear matter (Dietrich et al., 2020). With future GW observations from more BNSs expected to further improve these measurements, a robust and efficient scheme EoS inference, jointly from the GW observations of multiple BNS mergers, needs to be developed and/or perfected before the next observing run, O4, of the LIGO/Virgo/KAGRA (Akutsu et al., 2021) (LVK) detector network begins.

Bayesian hierarchical inference from GW data (Thrane & Talbot, 2019) is a statistical framework that has been used to accurately measure the EoS from multiple GW observations. However, numerical implementations of such an analysis incur a heavy computational cost, one that grows with the number of events analyzed. Thus, such implementations are likely to be problematic given the number of events expected to be observable during O4, potentially requiring several weeks of computation time. Several techniques (Golomb & Talbot, 2022; Wysocki et al., 2020; Hernandez Vivanco et al., 2019; Hernandez Vivanco et al., 2020; Ghosh et al., 2021) have been developed to circumvent this problem that achieve their speed-up gains by re-using information from single event EoS agnostic parameter estimation (PE) runs that are carried out regardless for GW transient catalogs. Among them, the likelihood approximation algorithm GWXTREME (Ghosh, 2020) has been shown to perform accurate Bayesian model comparison on the tabulated set of known EoS models with a latency of a few minutes per event (Ghosh et al., 2021). However, being a model selection scheme, the current version of GWXTREME cannot compute empirical credible intervals on the EoS pressure-density space from the data using parameterized EoS models. It can only

comment on how some discrete lines on that space, one corresponding to each known EoS model, compare against each other.

In this work, we develop an algorithm based on GWXTREME to hierarchically infer parameterized EoSs from multiple GW observations. We incorporate our physical knowledge of NS physics such as the requirements of causality, thermal stability, and observational consistency of the NS maximum mass as Bayesian priors on the EoS parameters and compute their posterior distributions using our generalized version of GWXTREME, with a run time of 20 hours to a day for order 10 events. The posterior samples of EoS parameters then translate naturally into empirically measured confidence intervals in the EoS pressure-density plane. Since the GWXTREME method is based on an approximation scheme (Ghosh et al., 2021) rather than distribution of computation over large number of computational resources, we do not need GPU parallelization to achieve our results, making this method easier to implement on most machines. Our analysis needs no prior assumption about the fiducial population of BNSs.

Following previous works (Wysocki et al., 2020; Golomb & Talbot, 2022) we test our algorithm on a set of simulated events drawn from the galactic BNS population and demonstrate its high accuracy. In addition, we test our algorithm on data from the real events, GW170817 and GW190425, to show that the results produced by our algorithm are fully consistent with the existing un-approximated Bayesian PE results, despite being orders of magnitude faster. Furthermore, armed with this fast and cheap algorithm we study the variability of EoS constraints with component NS masses, the signal-to-noise ratio (SNR) of events analyzed and the number of events analyzed by drawing additional simulated events in narrow SNR bins and over a wide range of BNS total masses. We note that this work is proof of concept, where-in we show that GWXTREME's likelihood approximation scheme can be generalized to perform fast, cheap and accurate hierarchical inference of parameterized NS EoS models from multiple GW observations, making GWXTREME a strong candidate for hierarchical EoS inference in O4. We leave further improve-

ments to our algorithm for potentially more generalized EoS inference as part of a future work, while outlining the blueprints of such generalizations.

This paper is organized as follows. In Sec. 6.2.2, we describe our method of EoS inference in detail. First, we review the Bayesian hierarchical framework for EoS inference from GW data in Sec. 6.2.2.1 and the problematic computational cost of its implementation. Then, in Sec. 6.2.2.2 we describe our algorithm, its approximations and how it resolves the aforementioned problem. Next, Sec. 6.2.2.3 summarizes the EoS parameterization and the priors on the EoS parameters that we have chosen for studying and testing our algorithm. We wrap up the discussion of methods in Sec. 6.2.2.4 with a note on the compatibility of our method with various GW waveform models and the associated systematics. In Sec. 6.2.3 we describe the details of the data on which we test our algorithm, which include both real and simulated events. In Sec. 6.2.4, we display the results of our study and discuss their implications. In Sec. 6.2.5 we conclude with a summary of our method, its virtues and its potential for performing efficient hierarchical EoS inference in O4 given the results of our study. We also outline schemes for potential improvements to our algorithm which are left as upcoming explorations.

6.2.2 Methods

6.2.2.1 Bayesian Hierarchical Inference of Parameterized EoS

In this section we review the framework of Bayesian hierarchical inference using GW data in the context of EoS inference. For a review of Bayesian inference from GW data see (Thrane & Talbot, 2019). Time series data from a GW detector, $d(t)$ can be thought to comprise random noise $n(t)$, with a modeled distribution, and possibly a signal $h(t, \vec{\theta})$ characterized by the parameter θ as dictated by the assumed signal model:

$$d(t) = h(t, \vec{\theta}) + n(t) \tag{6.2}$$

The parameters $\vec{\theta}$ consist of both EoS sensitive observables (masses and tidal deformability) and other observables such as distance, sky position, etc., to which the EoS is not sensitive. The finite size effects of NSs manifest themselves in the data through the dependence of the waveform model on the component NS masses m_i and the tidal deformability of the NSs, λ_i:

$$\vec{\theta} = \{m_1, m_2, \lambda_1, \lambda_2, \vec{\theta}_{\rm ne}\} \tag{6.3}$$

$$\lambda_i = \frac{2}{3G} k_2(m_i) r_i^5 \tag{6.4}$$

where r_i, are radii of the component NSs, k_2 the tidal love number (Hinderer, 2008), and $\vec{\theta}_{\rm ne}$ are other non EoS sensitive parameters that characterize the GW waveform. Given a noise model which specifies the probability distribution of the random noise time series $n(t)$, one can compute the likelihood of the observed time-series data as a function of the parameters characterizing the waveform model: $\mathcal{L}(d|\vec{\theta})$. Since the EoS only constrains the EoS sensitive observables, we can marginalize the likelihood over the remaining parameters $\vec{\theta}_{\rm ne}$ using uninformative priors $p(\vec{\theta}_{\rm ne})$ and construct the ingredients of our EoS inference analysis from that marginalized likelihood:

$$\mathcal{L}(d|m_1, m_2, \lambda_1, \lambda_2) = \int \mathcal{L}(d|m_1, m_2, \lambda_1, \lambda_2, \vec{\theta}_{\rm ne}) p(\vec{\theta}_{\rm ne})\, d\vec{\theta}_{\rm ne} \tag{6.5}$$

An EoS model, $\mathcal{E}$ implies a deterministic relationship between m_i and λ_i, which imposes a delta-function prior on these quantities through Eq. (6.1):

$$p(\lambda_1, \lambda_2|m_1, m_2, \mathcal{E}) = \delta(\lambda_1 - \lambda_{\mathcal{E}}(m_1))\delta(\lambda_2 - \lambda_{\mathcal{E}}(m_2)) \tag{6.6}$$

where, the tidal parameter λ as a function of mass is obtained by substituting Eq. (6.1) into Eq. (6.4)

Using the likelihood of GW data given EoS sensitive BNS parameters and an EoS model that imposes a prior on those parameters, one can construct the Bayesian evidence of the EoS model

by marginalizing the likelihood over the prior:

$$\mathcal{Z}(\mathcal{E}|d) = \int \mathcal{L}(d|m_1, m_2, \lambda_1, \lambda_2) p(\lambda_1, \lambda_2|m_1, m_2, \mathcal{E}) p(m_1, m_2)\, dm_1\, dm_2\, d\lambda_1\, d\lambda_2 \tag{6.7}$$

where $p(m_1, m_2)$ is an uninformative prior on the component masses. The overall evidence of an EoS model given multiple observations $\{d\} = \{d_1, d_2, ...\}$ can be obtained by multiplying the individual evidences, yielding the joint evidence: $\mathcal{Z}(\mathcal{E}|\{d\}) = \prod_i \mathcal{Z}(\mathcal{E}|d_i)$. The ratio of the joint evidence for two different EoS models $\mathcal{E}_1$ and $\mathcal{E}_2$ yields the Bayes factor which can be compared against unity to perform model selection:

$$BF^{\mathcal{E}_1}_{\mathcal{E}_2}(\{d\}) = \frac{\mathcal{Z}(\mathcal{E}_1|\{d\})}{\mathcal{Z}(\mathcal{E}_2|\{d\})} \tag{6.8}$$

This framework was used to perform model comparison for the set of known tabulated EoSs for the two BNS events observed by LVK to date: GW170817 (Abbott et al., 2020c), and GW190425. However, model selection can only compare known EoS models with each other. While such an analysis sheds light on the physics of the NS EoS since each model has its own set of assumptions about said physics, empirical constraints on the pressure density space in the form of continuous credible intervals inferred from data are beyond its scope. Phenomenologically parameterized EoSs such as the spectral parameterization and the piecewise polytropic parameterization are free of this limitation.

Such an EoS model $\mathcal{E}$, characterized by the parameters $\vec{\gamma}$, would imply a pressure-density and the corresponding mass-tidal deformability relation:

$$p \;=\; p(\rho, \vec{\gamma}) \implies \lambda = \lambda(m, \vec{\gamma}) \tag{6.9}$$

The deterministic relation Eq. (6.9) would then impose a prior on λ_i conditional on m_i in the form of

$$p(\lambda_1', \lambda_2'|m_1, m_2, \vec{\gamma}) \quad = \quad \delta(\lambda_1' - \lambda(m_1, \vec{\gamma}))\delta(\lambda_2' - \lambda(m_2, \vec{\gamma})) \tag{6.10}$$

Replacing the EoS sensitive prior in Eq. (6.7) with Eq. (6.10) would allow for the reinterpretation of the left hand side of Eq. (6.7) as the marginalized hierarchical likelihood $\mathcal{L}_h$ of GW data d, given the EoS parameters $\vec{\gamma}$ (viewed as hyper-parameters):

$$\mathcal{L}_h(d|\vec{\gamma}) = \int \mathcal{L}(d|m_1, m_2, \lambda_1, \lambda_2)p(\lambda_1, \lambda_2|m_1, m_2, \vec{\gamma})p(m_1, m_2)\, dm_1\, dm_2\, d\lambda_1\, d\lambda_2 \tag{6.11}$$

The universality of the cold matter nuclear EoS then implies that the EoS hyper-parameters $\vec{\gamma}$ are also universal and will have the same value for all NS systems. This can be exploited to construct the joint "quasi" likelihood of GW data from multiple events, $\{d\}$ given EoS hyper-parameters by multiplying the individual event hierarchical likelihoods. One can then use Bayes Theorem to convert that joint likelihood into a posterior distribution of EoS hyper-parameters given GW data from multiple events, by imposing a prior on those hyper-parameters based on our knowledge of NS physics:

$$p(\vec{\gamma}|\{d\}) \propto p(\vec{\gamma}, I) \prod_{i=1}^{N} \mathcal{L}_h(d_i|\vec{\gamma}) \tag{6.12}$$

where $p(\vec{\gamma}, I)$ encodes our a priori knowledge of NS physics by vanishing at values of $\vec{\gamma}$ for which the EoS becomes unphysical (examples of unphysicallity include acausal sound speed, thermal instability etc.). The abstract symbol I represents our understanding of NS physics, to which our priors on the EoS hyper-parameters are conditional. This posterior distribution of the EoS hyper-parameters can be used to compute their Bayesian credible intervals which translate naturally into

a credible interval on the EoS pressure-density space.

To compute credible intervals on the EoS hyper-parameters, one can in principle use standard Bayesian inference engines (such as direct quadrature or MCMC) on Eq. (6.12) directly. However, the product of integrals in Eq. (6.12) is difficult to deal with due to the presence of the delta functions in the integrands. The delta function priors imposed by the EoS model prevents us from approximating each integral in the product as a Monte Carlo sum over posterior samples yielded by single event EoS agnostic PE runs, that are carried out for generating GW transient catalogs. Unable to re-use information from single event PE runs, the alternative we are left with is to essentially redo the PE of BNS waveform parameters with delta function priors imposed on a subset of those parameters. This boils down to simultaneous inference of EoS hyper-parameters and the event specific waveform parameters, for all events. Numerical implementations of such PE would then require a large number of costly GW template waveform evaluations per event. This would lead to a rapidly increasing computational cost of hierarchical EoS inference with the number of events analyzed, potentially requiring several weeks of computation time per analysis in O4.

It would be much more efficient if an algorithm can be developed to numerically compute the marginalized hierarchical likelihood as a fast evaluating function of its arguments, by somehow re-using information from single event EoS agnostic PE runs, which are carried out regardless, for GW transient catalogs. In the next section, we describe how to achieve this using the GWXTREME likelihood approximation scheme and the algorithm we developed based on it.

6.2.2.2 Likelihood Approximation Scheme

To evaluate the hierarchical likelihood of GW data given EoS parameters, we perform the integral in Eq. (6.11) numerically, after approximating the integrand in a way that drastically reduces computational cost and latency. We base our approximation on (Ghosh et al., 2021), where repa-

rameterization of the EoS sensitive observables to reduce the dimensionality of the integrand and KDEs to approximate the lower dimensional (marginalized) integrand were implemented, for rapid and computationally cheap computation of EoS evidences, i.e., Eq. (6.7). Since the integral in Eq. (6.11) is essentially the same as that in Eq. (6.7) with the only difference being the parameterization of the EoS sensitive prior, similar approximations as in (Ghosh et al., 2021) can be used for the fast and cheap evaluation of our hierarchical marginalized likelihood.

To reduce the dimensionality of the integral in Eq. (6.11) and evaluate it numerically, we first note that by Bayes theorem, the single event likelihood given EoS sensitive BNS parameters, multiplied by EoS uninformative priors on those parameters, is proportional to the posterior distribution of those parameters given GW data: $p(m_1, m_2, \lambda_1, \lambda_2|d) \propto \mathcal{L}(d|m_1, m_2, \lambda_1, \lambda_2, \vec{\theta}_{\rm ne})p(m_1, m_2, \lambda_1, \lambda_2)$. Note that this posterior has already been sampled during the single event EoS agnostic PE run for some choice of the uninformative priors: $p(m_1, m_2, \lambda_1, \lambda_2) = p_{\rm PE}(m_1, m_2, \lambda_1, \lambda_2)$. This can be used to rewrite the integral in Eq. (6.11) in terms of the single-event posterior distribution of BNS parameters $p(m_1, m_2, \lambda_1, \lambda_2|d)$:

$$\mathcal{L}_h(d|\vec{\gamma}) = \int \frac{p(m_1, m_2, \lambda_1, \lambda_2|d)}{p_{\rm PE}(m_1, m_2, \lambda_1, \lambda_2)} p(\lambda_1, \lambda_2|m_1, m_2, \vec{\gamma}) p(m_1, m_2)\, dm_1\, dm_2\, d\lambda_1\, d\lambda_2 \tag{6.13}$$

As we shall demonstrate shortly, the single-event posterior distribution of BNS parameters given GW data can be accurately approximated as a fast evaluable numerical function, from its stochastic samples which are already generated and written to disk during the creation of GW transient catalogs. Before implementing such an approximation, we first show that the dimensionality of the integral can be now be reduced by reparameterizing the EoS sensitive observables. We re-parameterize $(m_1, m_2, \lambda_1, \lambda_2)$ into the chirp mass, mass ratio, and tidal parameters, $\mathcal{M}(m_1, m_2), q(m_1, m_2), \tilde{\Lambda}(\lambda_1, \lambda_2, m_1, m_2), \delta\tilde{\Lambda}(\lambda_1, \lambda_2, m_1, m_2)$, which are defined as follows:

$$\mathcal{M} = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}} \tag{6.14}$$

$$q = m_2/m_1 \tag{6.15}$$

$$\tilde{\Lambda} = \frac{8}{13}\left[(1 + 7\eta - 31\eta^2)(\Lambda_1 + \Lambda_2) + \sqrt{1 - 4\eta}(1 + 9\eta - 11\eta^2)(\Lambda_1 - \Lambda_2)\right] \tag{6.16}$$

$$\begin{aligned}\delta\tilde{\Lambda} = \frac{1}{2}\Bigg[&\sqrt{1 - 4\eta}\left(1 - \frac{13272}{1319}\eta + \frac{8944}{1319}\eta^2\right)(\Lambda_1 + \Lambda_2) + \\ &\left(1 - \frac{15910}{1319}\eta + \frac{32850}{1319}\eta^2 + \frac{3380}{1319}\eta^3\right)(\Lambda_1 - \Lambda_2)\Bigg]\end{aligned} \tag{6.17}$$

where $\eta = m_1 m_2/(m_1 + m_2)^2$ is the symmetric mass ratio and $\Lambda_i = G\lambda_i[c^2/(Gm_i)]^5$ is the dimensionless tidal deformability. We note that the definitions Eq. (6.16) and Eq. (6.17) assume $m_1 > m_2$. Under this reparameterization, Eq. (6.13) becomes:

$$\mathcal{L}_h(d|\vec{\gamma}) \propto \int \frac{p(\mathcal{M}, q, \tilde{\Lambda}, \delta\tilde{\Lambda}|d)}{p_{\rm PE}(\mathcal{M}, q, \tilde{\Lambda}, \delta\tilde{\Lambda})} p(\tilde{\Lambda}, \delta\tilde{\Lambda}|\mathcal{M}, q, \vec{\gamma}) p(\mathcal{M}, q)\, d\mathcal{M}\, dq\, d\tilde{\Lambda}\, d\delta\tilde{\Lambda} \tag{6.18}$$

where the Jacobians associated with the variable change cancel out and the reparameterized EoS sensitive priors are

$$p(\tilde{\Lambda}', \delta\tilde{\Lambda}'|\mathcal{M}, q, \vec{\gamma}) = \delta(\tilde{\Lambda}' - \tilde{\Lambda}(\mathcal{M}, q, \vec{\gamma}))\delta(\delta\tilde{\Lambda}' - \delta\tilde{\Lambda}(\mathcal{M}, q, \vec{\gamma})) \tag{6.19}$$

To further simplify Eq. (6.18), we can choose the uninformative priors in the denominator of the integrand to be uniform in the reparameterized tidal deformabilities: $p_{\rm PE}(\mathcal{M}, q, \tilde{\Lambda}, \delta\tilde{\Lambda}) \propto p(\mathcal{M}, q)$. This choice is compatible with at least one standard GW waveform for BNSs: the TaylorF2 Waveform model (Bini et al., 2012). We elaborate more on Generalizations to other waveforms

and the associated systematics in Sec. 6.2.2.4. Under this choice of priors, Eq. (6.18) becomes

$$\mathcal{L}_h(d|\vec{\gamma}) \propto \int p(\mathcal{M}, q, \tilde{\Lambda}, \delta\tilde{\Lambda}|d) p(\tilde{\Lambda}, \delta\tilde{\Lambda}|\mathcal{M}, q, \vec{\gamma})\, d\mathcal{M}\, dq\, d\tilde{\Lambda}\, d\delta\tilde{\Lambda} \tag{6.20}$$

The reason for this reparameterization is as follows. The chirp mass $\mathcal{M}$ is known to be extremely well measured, with its posterior distribution being sharply peaked about a number equal to the mean of the samples of $\mathcal{M}$ which are available from single event PE runs. Among the tidal parameters, $\tilde{\Lambda}$, which enters the GW waveform model at the 5th post-Newtonian order, has the dominant contribution as compared to $\delta\tilde{\Lambda}$, which enters the waveform at the 6th post-Newtonian order. As a result, the posterior distribution of BNS parameters is largely independent of $\delta\tilde{\Lambda}$. Under these considerations, the posterior distribution of BNS parameters given GW data can be approximated as $p(\mathcal{M}, q, \tilde{\Lambda}, \delta\tilde{\Lambda}|d) \approx p(q, \tilde{\Lambda}|d)\delta(\mathcal{M} - \bar{\mathcal{M}})$, where $\bar{\mathcal{M}}$ is the mean of the chirp mass samples obtained from EoS agnostic single event PE runs. Substituting these into Eq. (6.18) allows us to use the delta functions and reduce the dimensionality of the integral to one, provided the marginalized posterior $p(q, \tilde{\Lambda}|d)$ can be evaluated, at least approximately, as a numerical function of its parameters.

$$\mathcal{L}_h(d|\vec{\gamma}) \propto \int p(q, \tilde{\Lambda}(\bar{\mathcal{M}}, q, \vec{\gamma})|d)\, dq \tag{6.21}$$

To evaluate $p(q, \tilde{\Lambda}|d)$ as a numerical function of $(q, \tilde{\lambda})$, we approximate it from EoS agnostic single event PE samples of its parameters via Gaussian kernel density estimation (KDE), customized to be immune to edge effects. Traditional Gaussian KDE fits a Gaussian around each posterior sample and approximates the density of those samples as the sum of those Gaussians (Silverman, 1986; Scott, 1992). The covariance matrix of each of the Gaussians is approximated from the sample covariance matrix of the posterior samples. While the KDE evaluation speed is very fast, it is susceptible to edge effects in its traditional form, becoming inaccurate for distributions with sharp edges. Single event PE runs implicitly assign the heavier NS as the primary mass m_1,

resulting in a sharp edge on the distribution of mass ratio at $q = 1$. To circumvent the incompatibility of this sharp edge with KDE, we add to the KDE probability distribution function (pdf) at each point, the value of the pdf at a point symmetric about the sharp edge, resulting in a bounded KDE, similar to (Ghosh et al., 2021). This results in our KDE being free of edge effects and compatible with the sharp edge at $q = 1$. The accuracy of this approximation is demonstrated in Fig. 6.1. With this approximation of the marginalized posterior, the hierarchical likelihood for the ith GW event becomes

$$\mathcal{L}_h(d_i|\vec{\gamma}) \propto \int_0^1 K_i(q, \tilde{\Lambda}(\bar{\mathcal{M}}_i, q, \vec{\gamma}))\, dq \tag{6.22}$$

where $K_i(q, \tilde{\Lambda})$ is the KDE approximation of $p(q, \tilde{\Lambda}|d_i)$ obtained from the EoS agnostic single event posterior samples $\{(q, \tilde{\Lambda})_i\}$. Since K_i is a fast evaluating function of its arguments, the definite integral with a finite range in Eq. (6.22) can be evaluated efficiently using numerical techniques such as the trapezoidal rule. Substituting Eq. (6.22) into Eq. (6.12) yields the approximate joint hierarchical posterior of EoS parameters given GW data from multiple BNS observations, the EoS parameterization model and the prior knowledge on NS matter physics, that is numerically evaluable almost instantaneously:

$$p(\vec{\gamma}|\{d\}, I) \propto p(\vec{\gamma}, \mathbf{I}) \prod_{i=1}^{N} \int_0^1 K_i(q, \tilde{\Lambda}(\bar{\mathcal{M}}_i, q, \vec{\gamma}))\, dq \tag{6.23}$$

The posterior distribution in Eq. (6.23) can be sampled stochastically to produce empirical constraints on the EoS pressure density relation. We use affine-invariant Markov Chain Monte Carlo (MCMC) ensemble sampling (Goodman & Weare, 2010), as implemented in the the package EMCEE, with CPU parallelization inbuilt (Foreman-Mackey et al., 2013), to sample the posterior in Eq. (6.23). We parallelize emcee on 50 CPU cores to sample the posterior in Eq. (6.23) within 20 hours to a day, for 10 BNS events. Using those posterior samples, for each value of the density ρ in Eq. (6.9), we evaluate N_{samples} number of pressures $\{p(\rho, \vec{\gamma}_j)\}$, one corresponding to every posterior

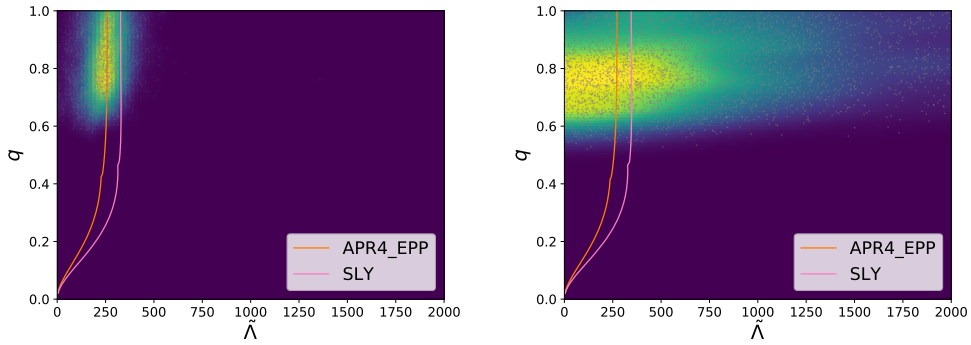
sample $\{\vec{\gamma}_j\}$ where, $j = 1, 2, ..., N_{\rm samples}$. We then compute the median, 5 and 95 percent quantiles from the set $\{p(\rho, \vec{\gamma}_j)\}$, which we plot against ρ and interpret them as empirical constraints/credible intervals on the NS EoS. The posterior predictive distribution of the dimensionless tidal parameter at $1.4\,M_\odot$ can also be produced as a different representation of how well the EoS is constrained. These are obtained by computing the histograms of the values of the dimensionless tidal parameter at $1.4\,M_\odot$ $\Lambda_i = \Lambda(m = 1.4\,M_\odot, \vec{\gamma}_i)$ corresponding to each posterior sample $\vec{\gamma}_i$ of the EoS parameters drawn form the joint posterior Eq. (6.23). We produce these constraints by performing our analysis on GW data from both real and simulated events and demonstrate the accuracy and efficiency of our method.

Even though we do not simultaneously infer BNS population parameters, we note that our algorithm is generalizable to do so while retaining its low computational cost. We discuss the blueprints of such a calculation in the conclusion section and leave it as part of an upcoming work. In the next section, we describe our choice of EoS parameterization and the priors imposed on the EoS parameters, which are based on existing knowledge of NS matter physics.

In the next section, we describe our choice of EoS parameterization and the priors imposed on the EoS parameters, which are based on existing knowledge of NS matter physics.

6.2.2.3 EoS parameterization and priors

Previous studies like (Lackey & Wade, 2015) have shown the effectiveness of using phenomenologically parameterized EoSs such as the piecewise polytrope, in measuring the EoS pressure-density relation empirically from GW data. Parameterizing the pressure density relation instead of the mass-tidal parameter relation and deriving the latter from the former makes the inclusion of a priori knowledge of NS physics into the analysis straightforward, in the form of Bayesian priors on the EoS parameters. However, the non-differentiability in the EoS at the joining point of two consecutive polytrope pieces leads to increases statistical errors in the EoS measurements inferred



(a) SNR=32 (b) SNR=11

Figure 6.1: KDE visualization for two simulated events, with APR4_EPP as the injected EoS, one at low (bottom) and the other at high (top) SNR. In both the figures we have plotted $K_i(q, \tilde{\Lambda})$ as a 2D density plots. We have overplotted the PE samples of $q, \tilde{\Lambda}$ as discrete gray points to demonstrate the KDE accurately approximates the posterior distribution from which the samples are drawn. The integral in Eq. (6.23) can be visualized as line integrals of these 2D density along the $\tilde{\Lambda}(q, \mathcal{M}, \vec{\gamma})$ line for some particular value of $\vec{\gamma}$. Two such lines, one corresponding to the injected EoS APR4_EPP and a different SLY are plotted as an example of the EoS lines along which the integrals are performed.

from GW data using the piecewise parameterizations, at those points (Carney et al., 2018). The spectral parameterization of the EoS which expands the adiabatic index in powers of logarithmic pressure, has been shown to be free of this deficiency (Carney et al., 2018). For this reason, we choose the four parameter spectral decomposition as our EoS parameterization for this study while noting that our analysis works for any parameterized model including the piecewise-polytrope, which we show in the appendix as a validation study. Under the spectral parameterization, the adiabatic index in terms of the pressure is

$$\ln \Gamma(p, \vec{\gamma}) = \sum_{k=1}^{4} \gamma_k \left(\ln \frac{p}{p_0} \right)^k \tag{6.24}$$

where $\Gamma(p)$, is the adiabatic index at pressure p and p_0 is the minimum pressure above which the representation is valid. Eq. (6.24) can be used to directly compute pressure as a function of energy

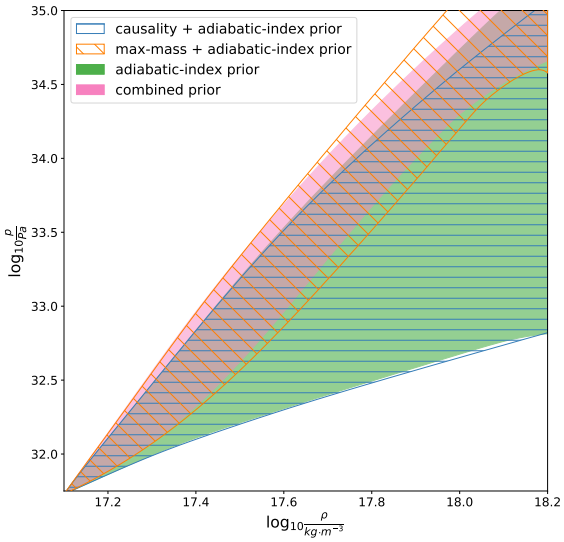


Figure 6.2: Visualization of the priors used on EoS parameters as bounds on the pressure density plane. We draw a large number of samples of the spectral parameters from the uniform distributions on the right hand side of Eq. (6.26). From those samples, we only keep the ones that satisfy the physicality conditions imposed by the various component of the prior. We then calculate the pressure density curves for each of those samples and plot credible intervals on the pressure density plane that contain 90 percent of those curves.

density e using the thermodynamic relation

$$\frac{de}{dp} = \frac{e+p}{p\Gamma(p,\vec{\gamma})} \tag{6.25}$$

which can then be used to find rest mass density as a function of pressure using $\rho(p,\vec{\gamma}) = [e(p,\vec{\gamma}) + p]\exp[-h(p,\vec{\gamma})]$ where $dh/dp = 1/[e(p,\vec{\gamma}) + p]$. The pressure density relation can then be used to solve the TOV equation to yeild $\tilde{\Lambda} = \tilde{\Lambda}(\mathcal{M}, q, \vec{\gamma})$. We use the LALSIMULATIONNEUTRONSTAREOS module of the LIGO Algorithm Library package (LALSUITE) (LIGO Scientific Collaboration, 2018), that implements TOV solving algorithms described in (Lindblom, 1992) and (Damour & Nagar, 2009), to solve the TOV equation, for each $\vec{\gamma}$.

As implemented in LALSIMULATIONNEUTRONSTAREOS, the minimum pressure p_0 is cho-

sen to be $5.37 \times 10^{34}\,\mathrm{dyn\,cm^{-2}}$. At pressures below this value, a different EoS, the SLY EoS model of (Douchin & Haensel, 2001), is used, which is stitched to the high density spectral EoS at $p = p_0$, as implemented in LALSIMULATIONNEUTRONSTAREOS, consistent with previous works like (Carney et al., 2018; Wysocki et al., 2020). The range of $\vec{\gamma}$ for which the EoS is physical and stable can be determined using existing knowledge of NS physics. Following (Carney et al., 2018) and (Wysocki et al., 2020) we demand that for an EoS characterized by a particular value of the parameters $\vec{\gamma}$ to be physical and observationally consistent, it must be thermally stable, causal and result in a maximum NS mass that is larger than the mass of the most massive NS observed with precise mass measurement to date. The thermal stability requirement, which demands that $de/dp > 0$, is satisfied by imposing $\Gamma(p, \vec{\gamma}) \in [0.6, 4.5]$ for all $p \in (5.37 \times 10^{32}, 1.19 \times 10^{38})\,\mathrm{dyn\,cm^{-2}}$, in addition to the uniform priors on the $\vec{\gamma}$: $\gamma_0 \in [0.2, 2]$, $\gamma_1 \in [-1.6, 1.7]$, $\gamma_2 \in [-0.6, 0.6]$, $\gamma_3 \in [-0.02, 0.02]$. The causality prior demands that the speed of sound in NS matter, $c_s = \sqrt{dp/de}$, up to and at the central pressure $p_{c,max}$ of the heaviest NS supported by the EoS, has to be less than the speed of light. We allow for a 10 percent buffer in the speed of sound inequality, to account for causal EoS models in the list of tabulated EoSs to be a fit by acausal spectral EoSs. Finally, we impose the maximum mass constraint by demanding that spectral parameters corresponding to physical EoSs must correspond to a maximum NS mass that is larger than that of the most massive NS observed, for which precise mass measurements are possible: $m_{\mathrm{max}}(\vec{\gamma}) > 1.97\,M_\odot$ (Antoniadis et al., 2013). All of these priors are consistent with previous work on hierarchical inference of parameterized EoSs from GW data, for example (Carney et al., 2018; Lackey & Wade, 2015; Wysocki et al., 2020). We note that the recent mass measurements of PSR J0952-0607 (Romani et al., 2022) renders this choice of $1.97\,M_\odot$ for the heaviest observed NS mass outdated. However, since we validate our results for the real events by comparing with previous analyses like (Abbott et al., 2018a, 2020b) that were carried out before this recent observation, we choose $1.97\,M_\odot$ for the heaviest observed NS mass so as to be consistent

with those works.

Combining all of these individual priors, our resultant prior on the EoS parameters is:

$$p(\vec{\gamma}, I) \propto \begin{cases} 1 & \text{where } 0.2 \leq \gamma_0 \leq 2,\ -1.6 \leq \gamma_1 \leq 1.7,\ -0.6 \leq \gamma_2 \leq 0.6,\ -0.02 \leq \gamma_3 \leq 0.02, \\ & \quad 0.6 \leq \Gamma(p, \vec{\gamma}) \leq 4.5,\ c_s(p_{c,max}, \vec{\gamma}) < 1.1c, \text{ and } m_{\text{max}}(\vec{\gamma}) > 1.97 M_\odot \\ 0 & \text{otherwise} \end{cases} \tag{6.26}$$

For a visualization of how these priors constrain the EoS pressure-density relation, we draw samples of the EoS parameters from the prior in Eq. (6.26) and compute pressure density credible intervals from those samples. We plot the intervals for 50000 combined prior samples in Fig. 6.2.

In the next section, we summarize the compatibility of our approximation, with various GW waveform models that might be used to perform the single-event PE.

6.2.2.4 Waveform Systematics

The compatibility of our approximation scheme with various GW waveform models manifests through the dependence of the choice of uninformative priors $p_{\text{PE}}(\mathcal{M}, q, \tilde{\Lambda}, \delta\tilde{\Lambda})$ used in the single-event PE runs on said waveform models. The equivalence of Eq. (6.18) and Eq. (6.20) is conditional on whether the $p_{\text{PE}}(\mathcal{M}, q, \tilde{\Lambda}, \delta\tilde{\Lambda})$ is uniform in $\tilde{\Lambda}$. However, such a prior typically has support in regions of the $\tilde{\Lambda}$-$\delta\tilde{\Lambda}$ space that corresponds to negative values of the tidal parameters λ_1, λ_2, which can cause certain waveform generators for certain families of GW waveform models to fail. PHENOMNRT (Hannam et al., 2014; Schmidt et al., 2015; Husa et al., 2016; Khan et al., 2016), is one such example. On the other hand, a uniform in Λ_1-Λ_2 prior, which is compatible to PHENOMNRT waveforms, will imply a $p_{\text{PE}}(\mathcal{M}, q, \tilde{\Lambda}, \delta\tilde{\Lambda})$ that is not uniform in $\tilde{\Lambda}$ and has vanishing support at $\tilde{\Lambda} = 0$, as can be seen in Fig. 6.3. This will tend to blow up the integrand in

Eq. (6.18) near $\tilde{\Lambda} = 0$ which might result in unacceptable numerical errors.

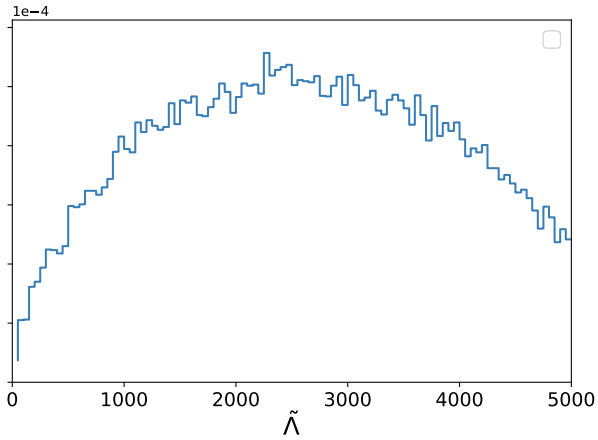


Figure 6.3: Visualization of uninformative priors on tidal parameters used in single event EoS agnostic PE runs which can be one or the other depending on the waveform model being used. The uniform in $\tilde{\Lambda}$ prior, compatible with TAYLORF2 and hence the current version of GWXTREME, would be a horizontal line in this plot. The uniform in positive Λ_1, Λ_2, which is compatible with the PHENOMPNRT waveform families leads to a $\tilde{\Lambda}$ prior distribution that is shown in this plot and evidently has vanishing support at $\tilde{\Lambda} = 0$. This can tend to blow up the integrand in Eq. (6.18) as its denominator $p_{\rm PE}(\mathcal{M}, q, \tilde{\Lambda}, \delta\tilde{\Lambda}$ will then be proportional to a quantity that vanishes a region inside the integration range.

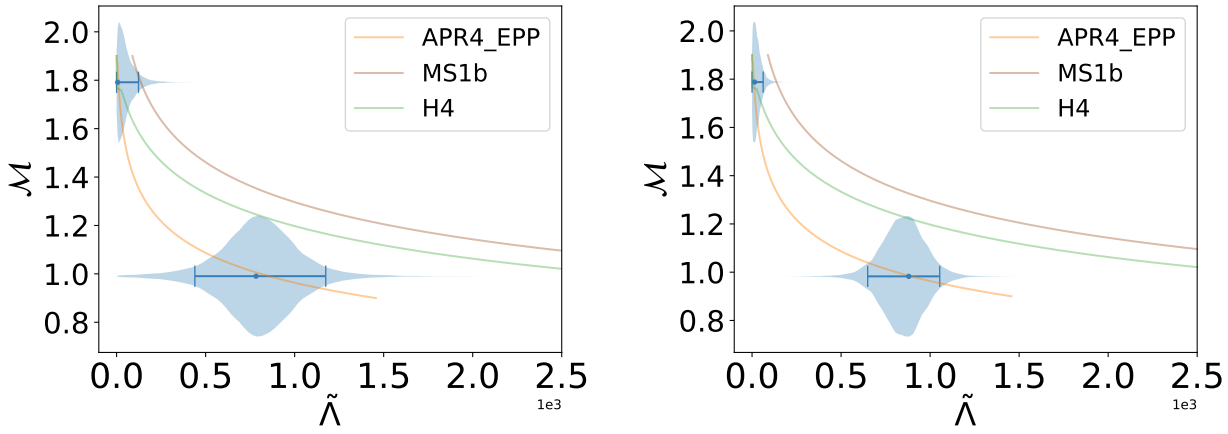
The TAYLORF2 waveform (Bini et al., 2012), is immune to this issue since it is evaluated originally as a function of $(\tilde{\Lambda}, \delta\tilde{\Lambda})$ and does not need a conversion to Λ_1-Λ_2 space. A uniform in $\tilde{\Lambda}$ prior for the single event PE runs is thus compatible with TAYLORF2. Hence we choose the TAYLORF2 waveform model to perform our single event EoS agnostic PE runs for both real and simulated events.

Following (Ghosh et al., 2021; Carney et al., 2018; Lackey & Wade, 2015), we truncate TAYLORF2's frequency domain waveform model at a stage of the binary's evolution, where the separation between the two NSs become comparable to the innermost stable circular orbit (ISCO), at which point the frequency of GW has the value $f_{\rm ISCO} = c^3/[6^{2/3}\pi G(m_1 + m_2)]$. While this choice of upper frequency cutoff can be unrealistic for EoSs that predict large NS radii, leading to merger happening before the separation of the NSs reaching ISCO, GW detectors are largely insensitive

to such high frequencies, rendering the unphysicallity of our frequency cutoff irrelevant. It has been shown in previous studies, such as (Wade et al., 2014), that varying the cutoff frequency has a negligible effect on the EoS Inference from GW data.

We note that our algorithm can be made compatible with other waveforms such as PHENOMNRT that require positive Λ_1 and Λ_2 by switching to a 3-dimensional KDE, as described in Sec. 6.2.5. We leave such a generalization as part of an upcoming work. In the next section, we describe in detail the simulation study we performed, the single event PE method used in those simulations and the real event data, on which we run our analysis to test its accuracy.

6.2.3 Data: Simulation Study and real events



(a) SNR $\in (23, 25)$ (b) SNR $\in (33, 35)$

Figure 6.4: Visualization of why high mass events are less informative on the EoS. We plot 90% highest posterior density intervals of the $\tilde{\Lambda}$ posteriors for a low mass and a high mass simulated event drawn from the narrow SNR bins, top: (23,25) and bottom: (33,35). We also display the shape of the posterior in the form of violin plots. The solid lines are the $\mathcal{M}$-$\tilde{\Lambda}$ curves at $q = 1$ for some of the known EoS models.

We run our algorithm on the single event PE data for real events GW170817 and GW190425 as released publicly by the LVC (LVC, 2018, 2020), both individually and jointly. We compare our results with the EoS constraints obtained from the full unapproximated PE run that infers the

event-specific BNS parameters simultaneously with the EoS hyper-parameters. We show that the results from our approximate technique are in very good agreement with that of the full PE results, while being orders of magnitude faster. To test the accuracy of our algorithm in recovering the true EoS from data we run our analysis on simulated events which were chosen as follows.

We randomly draw BNS masses from the Galactic BNS population as inferred in (Farrow et al., 2019). Astrometric parameters were chosen in such a way that the sky positions and orientations of the coalescences are isotropic and distributed uniformly in co-moving volume within a luminosity distance range of 30 Mpc to 200 Mpc. This is consistent with previous work: (Golomb & Talbot, 2022; Wysocki et al., 2020) except for the distance distribution which (Golomb & Talbot, 2022) chose to be uniform in distance instead. However, unlike the aforementioned works, we also set the dimensionless spin parameters to be zero for simplicity, since neutron stars are expected to spin down and loose rotational energy to magnetically driven plasma winds (Goldreich & Julian, 1969; Contopoulos et al., 1999; Spitkovsky, 2006). This choice is further justified by noting that galactic double neutron star systems have been observed to have very low spins (with dimensionless spin-parameter $\chi < 0.05$) (Zhu & Ashton, 2020; Zhu et al., 2018). We assign our fiducial/"true" EoS to be APR4_EPP (Akmal et al., 1998) and compute the tidal parameters using said EoS from the drawn masses. Then, using these parameters and the TAYLORF2 waveform model, we simulate a GW waveform and inject it into GW detector noise realizations corresponding to power spectral densities of the 4 LVK detectors at projected O4 sensitivity (Abbott et al., 2018b).

Among the events drawn, we choose the first $N_{\rm O4}$ with a signal-to-noise ratio greater than or equal to 8 in at least one detector and perform PE on them to infer the EoS agnostic posteriors of single event BNS parameters given the simulated GW data. Here the expected number of observable events in O4, $N_{\rm O4}$, is calculated using Poisson statistics, the number of events discovered till date and the projected O4 sensitivity, as derived in appendix C.2. With 2 confident BNSs observed till date and the projected O4 sensitivity estimated in (Abbott et al., 2018b), we find the upper

bound on $N_{\rm O4}$ to be 16 with 90% confidence. This effectively makes our EoS constraints obtained from analyzing the simulated events a forecast for how well we might be able to constrain the NS EoS from GW data alone, by the time O4 ends.

To study the variability of EoS constraints we also draw equal mass simulated events from narrow SNR bins, distributed uniformly over a broad range of chirp masses. For nearly identical SNRs, we expect the EoS constraints to be dominated by low-mass events which is due to the following reason. Many candidate EoSs all predict small $\tilde{\Lambda}$ close to 0 for higher mass BNS systems. However, they predict large and often vastly different $\tilde{\Lambda}$ for lighter systems. Thus for larger mass systems, many EoSs in addition to the true EoS are expected to have posterior support since all of them predict $\tilde{\Lambda}$ within a narrow range of 0. On the other hand, for smaller mass systems only the true EoS has support. As an example, see Fig. 6.4. This leads to larger mass system being much less informative about the EoS than low mass systems with the later dominating the EoS constraints. We verify this by running our algorithm on these sets of of simulated events drawn from narrow SNR bins which yield EoS constraints consistent with this expectation.

In the next section we describe the single event PE runs which were used to generate the posterior samples that serve as input to our analyses.

6.2.3.1 Single Event EoS-Agnostic Parameter Estimation runs

The single event EoS-agnostic PE runs stochastically sample the posterior distribution of parameters $\vec{\theta}$ that characterize the frequency-domain GW waveform model $h(f, \vec{\theta})$, given GW data $p(\vec{\theta}|d)$. By Bayes theorem, this posterior distribution can be expressed as being proportional to the likelihood of obtaining GW data, given said parameters and a noise model, multiplied by uninformative priors on those parameters: $p(\vec{\theta}|d) \propto \mathcal{L}(d|\vec{\theta})\ p(\vec{\theta})$. Under the assumption of stationary Gaussian noise (Finn, 1992; Cutler & Flanagan, 1994; Abbott et al., 2020a; Thrane & Talbot, 2019) , the

likelihood function is

$$\mathcal{L}(d|\vec{\theta}) \propto \exp[-(d-h(\vec{\theta})|d-h(\vec{\theta}))/2] \tag{6.27}$$

$$(a|b) = 4\mathrm{Re}\int_0^\infty \frac{a^*(f)b(f)}{S_n(f)} \tag{6.28}$$

where $S_n(f)$ is the noise power spectral density of detector and the frequency domain quantities are obtained by taking a Fourier transform of the time domain quantities in Eq. (6.2) and noise. The integral in Eq. (6.27) can be evaluated numerically by truncating the waveform model at $f = f_{\rm ISCO}$, as mentioned in Sec. 6.2.2.4. For the simulated events, we choose a lower frequncy cutoff at 20 Hz. One can now evaluate the posterior $p(\vec{\theta}|d) \propto \mathcal{L}(d|\vec{\theta})p(\vec{\theta})$ and sample it stochastically. For the real events GW170817 and GW190425, we re-use the PE samples released by the LIGO/Virgo Collaboration (LVC, 2018, 2020), that were generated using LALINFERENCE_NEST, of the LALSUITE software package (LIGO Scientific Collaboration, 2018), which implements nested sampling, to sample the posterior distribution.

However, such nested sampling runs involve potentially millions of likelihood evaluations (Veitch et al., 2015; Smith et al., 2020), which are computationally costly. Since each likelihood evaluation involves the computation of the entire frequency-domain GW waveform at the corresponding values of $\vec{\theta}$, such a PE run can take potentially weeks per event (Canizares et al., 2015; Smith et al., 2016). For the purpose of our simulation study, wherein we draw of order 20 events from the Galactic BNS population and of order 10 events each in different SNR and NS mass ranges, we need single event PE analysis techniques that are much more computationally efficient.

For accelerating the analyses, we employ the reduced order quadrature (ROQ) technique (Canizares et al., 2015; Smith et al., 2016). We constructed linear and quadratic ROQ basis vectors of TAYLORF2 waveform over the parameter space we consider, employing the procedure described in the previous works. To obtain highly compressed basis sets, we constructed tens of linear ROQ basis sets, each of which is constructed over a narrow chirp-mass range, as done in (Morisaki &

Raymond, 2020). The resultant speed-up gain is $\sim 10^3$ to $\sim 10^4$, reducing the run time to a few hours.

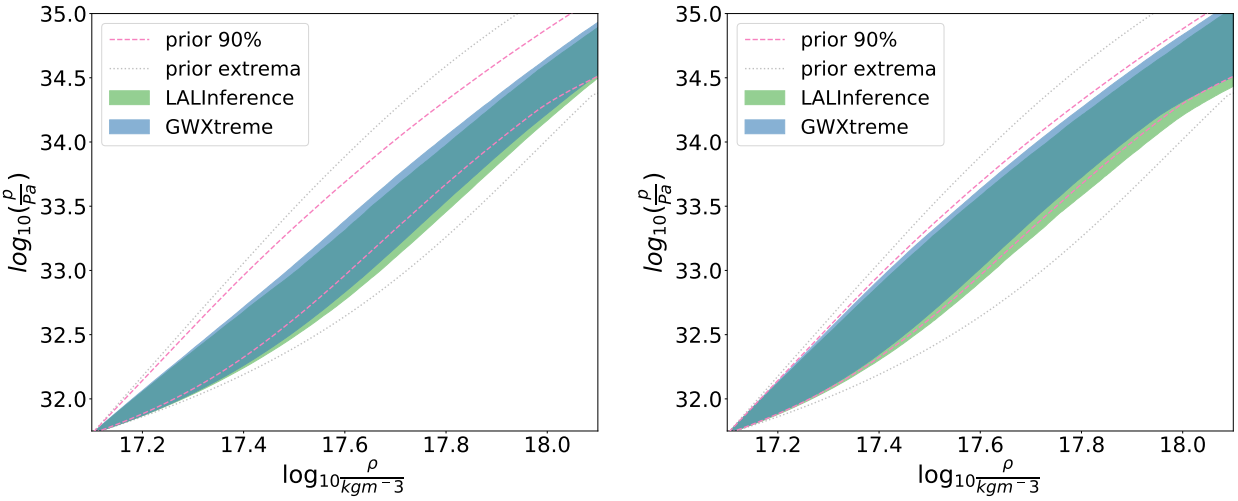
We use the BILBY (Ashton et al., 2019; Romero-Shaw et al., 2020) package that implements ROQ for single event EoS agnostic PE of simulated BNSs. We choose priors on the BNS parameters for the single event PE runs, that are consistent with previous work. We choose uniform priors in the mass ratio $q \sim U(0, 1)$, tidal parameters $\tilde{\Lambda} \sim U(0, 5000)$ and $\delta\tilde{\Lambda} \sim U(-5000, 5000)$, and the dimensionless component spin parameters $\chi_i \sim U(-0.05, 0.05)$. We choose a prior that is uniform in sky position, orientation, and polarization angle. For the real events, the priors used in EoS agnostic PE are listed in the public release by LVK (LVC, 2018, 2020) of the PE samples we use.

We also use BILBY's implementation of analytic marginalization over extrinsic parameters such as arrival time, coalescence phase and luminosity distance to further accelerate the EoS agnostic PE for the simulated events(Thrane & Talbot, 2019). We set uniform priors in time and phase and a powerlaw with index 2 prior on luminosity distance for carrying out the marginalizations. We then exploit BILBY's ability to re-construct posterior samples of the marginalized parameters (Romero-Shaw et al., 2020) in post-processing to obtain posterior samples of luminosity distance which is necessary for EoS inference due to the following reason.

The single event likelihood in Eq. (6.27) is implicitly a function of the detector frame chirp mass $\mathcal{M}_d$ which is related to the source frame chirp mass $\mathcal{M}$ by the redshift: $\mathcal{M}_d = \mathcal{M}(1 + z)$. Since GW observations alone cannot break the degeneracy between mass and redshift, BILBY samples the the likelihood in detector frame chirp mass. However, since the EoS is sensitive to the source frame masses, we need a way to break the mass-redshift degeneracy which is achievable by imposing a cosmological model. Given a cosmological model, posterior samples of luminosity distance (that are reconstructed by BILBY in post-processing) can be converted to samples of redshift. These can then be used to convert posterior samples of detector frame chirp-mass to source

frame. Following previous works such as (Golomb & Talbot, 2022), we use the Planck15 cosmology (Ade et al., 2016) for converting our chirp masses to detector frame before feeding them into our algorithm for EoS inference. We note that for the low luminosity distances we consider in this study (< 300 Mpc), Planck15 yields redshifts that are small, leading to source frame masses within at most 6-8% of the detector frame masses.

6.2.4 Results



(a) GW170817

(b) GW190425

Figure 6.5: Comparison of GWXTREME with LALINFERENCE_NEST for the real events GW170817 (left) and GW190425 (right). The posterior samples of the spectral parameters are used to compute pressure density curves and the shaded regions in the plots are equal tail confidence intervals that contain 90 percent of those curves. We also display the 90% prior intervals in dashed lines and the prior extrema in dotted lines computed using 50000 samples of $\vec{\gamma}$ drawn from the prior. It can be seen that GWXTREME results are consistent with the LALINFERENCE_NEST found by the LVK, results despite being orders of magnitude faster. The slight preference of LALINFERENCE_NEST towards softer EoSs can be attributed to the difference in waveform models as well as the priors on the tidal parameters that are used by the two algorithms.

In this section, we display the results we got by running GWXTREME-parameterized on real and synthetic data, which were obtained/generated by using the techniques and details summarized in the previous sections. The algorithm developed based on the method described in Sec. 6.2.2.2 is

publicly available and documented in the release of GWXTREME-0.3.1. We also reproduce each of these results with the piecewise-polytropic parameterization in the appendix to demonstrate the compatibility of our algorithm with any EoS parametrization.

6.2.4.1 Real Events

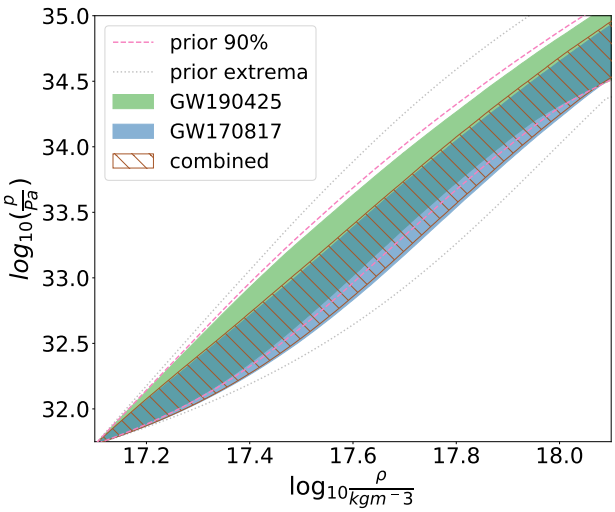
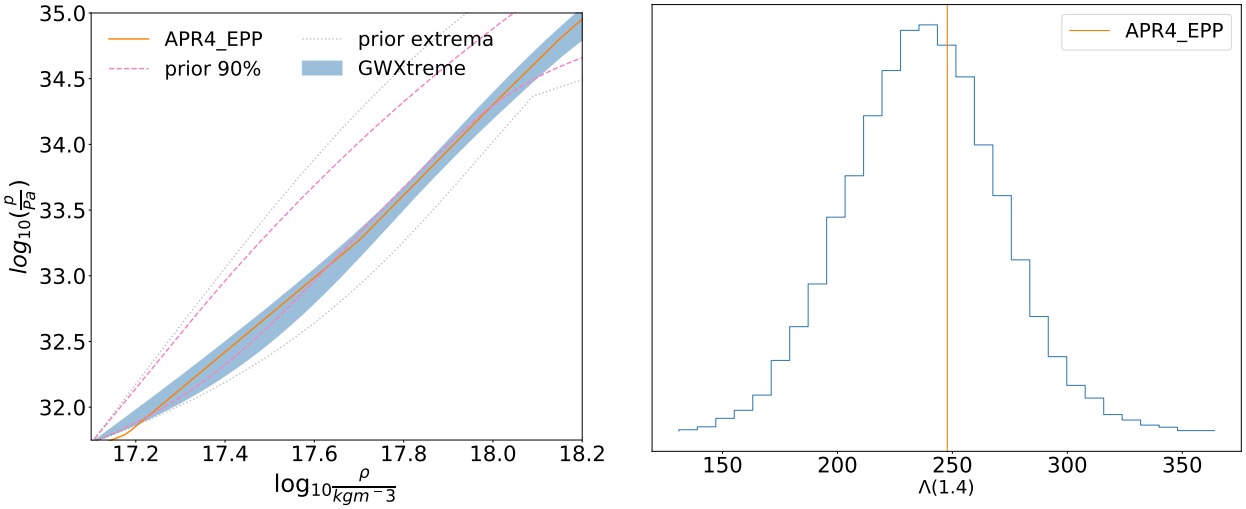


Figure 6.6: EoS constraints obtained by jointly analyzing GW170817 and GW190425 using GWXTREME. It can be seen that the joint EoS constraints are dominated by that of GW170817 which is to be expected due its larger SNR and smaller masses of component NSs than GW190425. However, the joint constraint can be seen to be very slightly narrower than that of GW170817 due to the contribution from GW190425.

We re-used the single event EoS agnostic posterior samples of masses and the tidal parameters given GW data from GW170817 and GW190425, as released by the LVC (LVC, 2018, 2020), that were generated using narrow spin priors and ran our analysis on them to produce EoS constraints in the form of credible intervals on the EoS pressure density plain. We compare our EoS constraints with those obtained by the joint un-approximated PE of EoS hyper-parameters and BNS parameters obtained by the LVK (Abbott et al., 2018a, 2020b) using LALINFERENCE_NEST module of the LALSUITE package. Even though different waveform models were used (TAYLORF2 for the EoS agnostic single event PE for GWXTREME's input and IMRPHENOMNRT-V2 for LAL-



(a) EoS Constraint

(b) $\Lambda(1.4M_{\odot})$ constraint

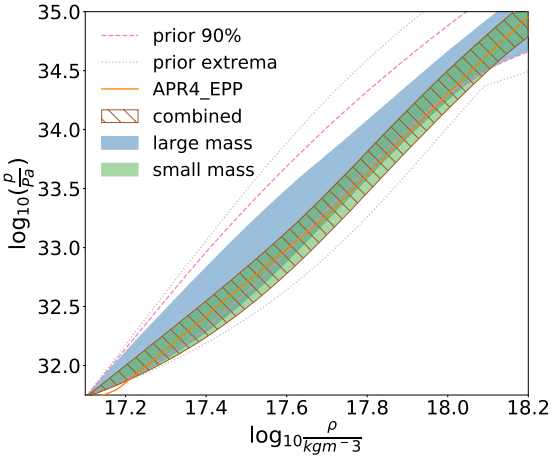
Figure 6.7: EoS constraints obtained by jointly analyzing 16 events drawn from the Galactic BNS population with APR4_EPP as the true EoS and injected into O4 sensitivity. The posterior samples of the spectral parameters generated by GWXTREME can be used to compute $p(\rho, \vec{\gamma})$ and $\Lambda(m, \vec{\gamma})$ curves. In plot (a), the shaded region marks the equal-tail confidence interval that contains the 90 percent of the $p(\rho, \vec{\gamma})$ curves. Plot (b) is obtained by histogramming $\Lambda(1.4\,M_{\odot}, \vec{\gamma}_i)$ corresponding to each posterior sample $\vec{\gamma}_i$. Both plots demonstrate that our computationally cheap and fast algorithm accurately measures the injected EoS.

INFERENCE_NEST), the EoS constraints are largely consistent, as can be seen in Fig. 6.5. The slightly higher preference of the LALINFERENCE_NEST constraints towards softer EoSs than GWXTREME's, can be attributed to the difference in the waveform models used in the two analyses. We note that the EoS agnostic posterior distribution of $\tilde{\Lambda}$ for the events computed using the two waveform models are themselves different. The $\tilde{\Lambda}$ posterior obtained using TAYLORF2 favors larger $\tilde{\Lambda}$ than PHENOMNRT for both GW170817 and GW190425 (see Fig. 11 of (Abbott et al., 2019d) for GW170817 and Fig. 14 of (LVC, 2020) for GW190415). This explains the preference of stiffer equations of state for TAYLORF2 and hence GWXTREME which takes the TAYLORF2-based EoS agnostic posterior samples of $\tilde{\Lambda}$ as input. From this we can conclude that the slight mismatch in EoS constraints between the two analysis is not due to an artifact of our algorithm but rather due to the difference in waveform models being used.

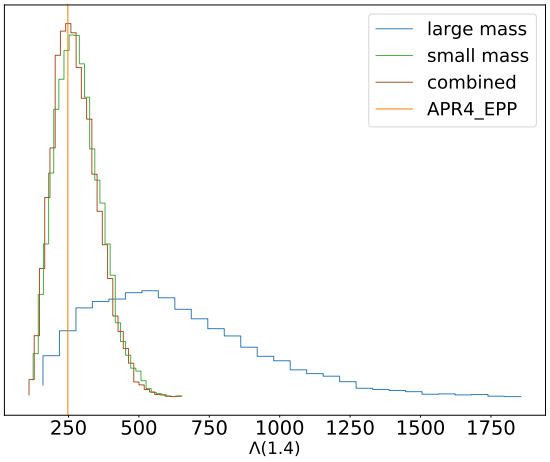
We also produced a joint constraint by hierarchically combining GW170817 and GW190425 data which we show in Fig. 6.6. Due to the large mass of its components and lower SNR as compared to GW170817, the GW190425 event does not contribute much to the joint EoS constraint even though the constraints change slightly from GW170817, which is barely discernible in Fig. 6.6.

6.2.4.2 Simulated Events

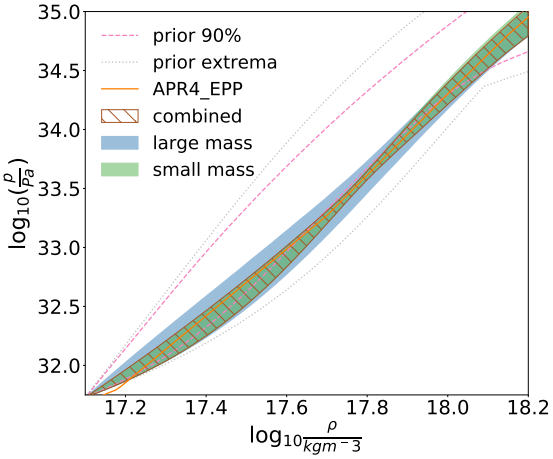
In this section we display the results we found upon analyzing the data from 16 simulated events whose masses are distributed according to the Galactic double neutron star population using our algorithm. The details of the simulation and associated modelling of measurement uncertainties are described in Sec. 6.2.3. Upon analyzing data generated from such events using our computationally cheap algorithm, we compute EoS constraints that are completely consistent with the chosen "true" EoS, as displayed in Fig. 6.7. We note that the increase in the width of the credible intervals near the high and decrease near the low density ends of the EoS pressure density relation are expected given our choice of the parameterized EoS, the population of simulated events and the injected EoS. The widening of the constraints at high densities occurs due to those densities being higher than the central density corresponding to the injected EoS, of the most massive component NS in our chosen set of events. On the other hand, the narrowing in the low density end is an artifact of our EoS model. As mentioned in Sec. 6.2.2.3 our high density spectral EoS is stitched to SLY at low densities. Due to this stitching with SLY at low densities, variations of the spectral parameters do not affect the pressure density relation at low densities, resulting in the credible interval obtained using the spectral parameters to converge into to a single line (corresponding to SLY) at low densities. We also display the posterior predictive distribution of $\Lambda(1.4\,M_{\odot})$ in Fig. 6.7. The uncertainties in the distribution of $\Lambda(1.4\,M_{\odot})$ can be thought of as estimates of how well GWXTREME can measure the EoS from GW observations and can be used for the comparison



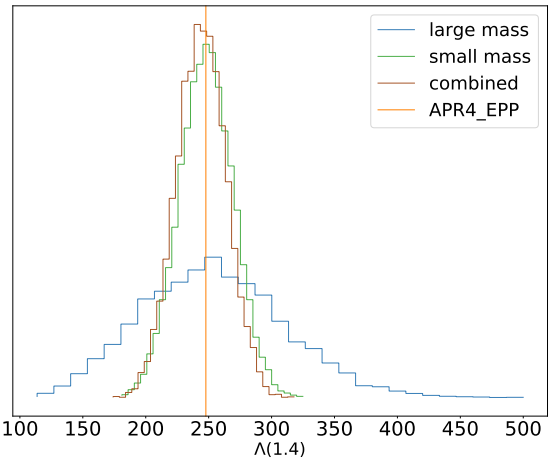
(a) SNR $\in (23, 25)$: EoS Constraint



(b) SNR $\in (23, 25)$: $\Lambda(1.4M_\odot)$ constraint Constraint



(c) SNR $\in (33, 35)$: EoS Constraint



(d) SNR $\in (33, 35)$: $\Lambda(1.4M_\odot)$ constraint Constraint

Figure 6.8: EoS constraints obtained using GWXTREME for events drawn in narrow SNR ranges. For each chosen SNR range, we draw events uniformly in chirp masses and then draw the extrinsic parameters so that when injected into O4 sensitivity, the simulated signals will have an SNR that belongs in said range of SNRs. We then group the highest mass and lowest mass events in each SNR range and analyze them using our algorithm both separately and jointly. Then using the posterior samples of the spectral parameters generated by our algorithm, we produce EoS constraints as both pressure-density credible intervals and quantile ranges of the tidal deformability at $1.4\,M_\odot$. It can be seen that EoS constraints are dominated by the lower mass events in each SNR bin and that they become narrower with increasing SNR.

of our method with competing ones.

6.2.4.3 Variability of EoS Constraints with SNR and Mass

To explore the variability of the EoS constraints with the SNR and chirp mass of events, we simulated events randomly in narrow SNR bins of $(23, 25)$ and $(33, 35)$, with the chirp masses for each bin chosen uniformly in the range $(0.8\,M_\odot, 1.8\,M_\odot)$. The joint EoS constraints along with the predictive distributions of the measured dimensionless tidal parameter $\Lambda(1.4\,M_\odot)$, obtained from analyzing these events using our algorithm is displayed in Fig. 6.8. For the $(23, 25)$ SNR range we see that the high mass events are much less informative than the low mass once and that the joint constraint is completely dominated by the lower mass events even in the high density regime. This is consistent with what we expect, as described in Sec. 6.2.3. We see similar trends in the $(33, 35)$ SNR range along with a couple of additional features. First, both high mass and low mass events produce narrower constraints than their $(23, 25)$ SNR counterparts. Second, in the high density regime, the joint constraint appears more informative than the one obtained from low mass events only due to contribution from the high mass events. This implies that in joint EoS inference from multiple events, a high mass event can result in a non-neglible information gain only if it is loud enough.

6.2.5 Conclusion and Future Prospects

We have developed an algorithm for fast and computationally cheap hierarchical inference of the NS EoS using observations of GWs from multiple BNSs that re-uses single event EoS agnostic PE results to achieve its latency and efficiency. We demonstrated the accuracy of our method by showing its results to be fully consistent with the existing EoS constraints for the events GW190425 and GW170817 that were computed using un-approximated and hence much costlier analyses. We also demonstrated the accuracy with which our method can constrain the true EoS by performing

a simulation study with realistic modeling of the measurement uncertainties in the EoS sensitive BNS parameters and a population of BNSs consistent with the latest studies of the Galactic double neutron star populations. We further studied the variability of EoS constraints with BNS chirp mass and SNRs by drawing simulated events in narrow SNR bins and over a broad range of BNS chirp masses. We found that EoS constraints are dominated by lower mass and higher SNR events, in agreement with our understanding of NS structure, which can be used for potentially truncating the list of events that need to be analyzed for joint EoS constraints without losing precision. While the variation with SNR is consistent with previous works, we have shown that the variation with chirp mass is also equally significant in selecting the events that will have the most dominant contribution to the EoS constraints.

Even though we do not simultaneously infer the BNS mass distributions which might have non-negligible corelations with the EoS constraints (Wysocki et al., 2020), we note that our method is generalizable to do so. Simultaneous inference of BNS mass distributions in our framework would require the uninformative prior $p(\mathcal{M}, q)$ in Eq. (6.18) to be replaced by a population prior conditional on the mass population model and its parameters that we are trying to infer: $p(\mathcal{M}, q|\vec{\gamma}_{\rm pop})$, where $\vec{\gamma}_{\rm pop}$ are the universal parameters characterizing the BNS mass distribution. This will induce an additional factor of $p(q, \bar{\mathcal{M}}|\vec{\gamma}_{\rm pop})/p(\vec{q}, \bar{\mathcal{M}})$ in the integrand of Eq. (6.20) and hence those of Eq. (6.23). Then, Including an additional prior on the mass population parameters $p(\vec{\gamma}_{\rm pop})$ to be multiplied with the prior on the EoS parameters $p(\vec{\gamma}|I)$ in Eq. (6.23) effectively makes its LHS the joint posterior of the universal EoS and population parameters given GW data from multiple events, which can be sampled stochastically to produce simultaneous mass-population and EoS constraints. Since the KDE and rest of the approximations along with the dimensionality of the integrals remain the same, this generalization will not affect the latency and computational cost of our algorithm. We leave such a generalization as an upcoming work.

We have shown that our algorithm is compatible with the TaylorF2 waveform model. We

note that switching to a 3-dimensional KDE in $(\Lambda_1, \Lambda_2, q)$ instead of our 2 dimensional one, has the potential of making our analysis compatible with other waveforms as well, such as PHENOMPNRT. As described in Secs. 6.2.2.4 and 6.2.2.2, our 2-dimensional KDE-based approximation necessitates the use of a uniform in $\tilde{\Lambda}$ prior in the single event EoS agnostic PE runs which prevents the use of existing PE results with the PHENOMPNRT waveforms. However the 3-dimensional KDE based generalization would work with the uniform in Λ_1-Λ_2 prior being used in the single event PE with PHENOMPNRT waveforms, enabling the use of these PE results. For such a prior, i.e., $p_{\rm PE}(\Lambda_1, \Lambda_2, \mathcal{M}, q) \propto p(\mathcal{M}, q)$, the 3-dimensional KDE approximation to the EoS agnostic posterior, $p(q, \Lambda_1, \Lambda_2|d_i) \approx K_i(q, \Lambda_1, \Lambda_2)$, would lead to the modification of Eq. (6.23) to $p(\vec{\gamma}|\{d\}, \mathcal{E}, I) \propto p(\vec{\gamma}, I) \prod_{i=1}^{N} \int_0^1 K_i(q, \Lambda_1(\bar{\mathcal{M}}_i, q, \vec{\gamma}), \Lambda_2(\bar{\mathcal{M}}_i, q, \vec{\gamma}))\, dq$. This posterior can be then be sampled stochastically using the same techniques outlined in Sec. 6.2.2.2 to produce EoS constraints. Similar calculations with higher dimensional KDE's have been shown to produce sensible results in non-parametric EoS inference studies like refs Landry & Essick (2019); Essick et al. (2020); Landry et al. (2020). In the context of our algorithm, such a calculation is a generalization we leave as part of future work.

To summarize, in this proof of concept work, we develop an algorithm for fast and efficient hierarchical Inference of parameterized NS EoS from multiple GW observations and demonstrate its accuracy. With this development, `GWXtreme` is now a strong candidate for performing fast and computationally cheap hierarchical EoS inference using both tabulated and parameterized EoS models from multiple GW observations in O4. We have noted that generalizations of our algorithm to increase accuracy and applicability while maintaining efficiency is straightforward and will be available soon in future releases of GWXTREME. We further note that our demonstrations serve as proof of concept for the applicability of bounded KDEs in increasing the efficiency of other GW based hierarchical inference problems. Similar problems where-in the parameterized physical model being inferred implies deterministic relationships between event-specific observ-

ables leading to delta function priors on them, can be efficiently handled with customized KDEs. Our algorithm can serve to guide such analyses which can re-use the basic concept of our framework while needing to modify only the model-imposed priors and set of observables sensitive to the model.

6.3 GENERALIZING TO NSBH OBSERVATIONS

The scalable inference method described in the previous section can be straightforwardly generalized to NSBH observations, by generalizing to a three-dimensional KDE. This is necessary due to the following reason. Inspiral-only waveforms such as Taylor-F2 are insufficient for analyzing NSBH signals for which the merger can happen within LIGO-Virgo's sensitive bands. Therefore, more sophisticated waveform models such as `Phenom-NRT`, which cannot handle a uniform in $\tilde{\Lambda}, \delta\tilde{\Lambda}$ PE priors.

As mentioned in the previous section, posterior samples obtained using a uniform in Λ_1, Λ_2 priors need a 3D-KDE based evidence approximation scheme:

$$p(\vec{\gamma}|\{d\}, I) \propto p(\vec{\gamma}, \mathbf{I}) \prod_{i=1}^{N} \int_0^1 K_i(q, \Lambda_1(\bar{\mathcal{M}}_i, q, \vec{\gamma}), \Lambda_2(\bar{\mathcal{M}}_i, q, \vec{\gamma}))dq \tag{6.29}$$

where $K_i(q, \Lambda_1, \Lambda_2)$ is a three-dimensional bounded KDE on the posterior samples of $(q, \Lambda_1, \Lambda_2)$ obtained from the ith event. Using this framework, I analyzed the data from GW230529 in an attempt to obtain EoS constraints, which are shown in figure 6.9, and communicated as part of Abac et al. (2024). The lack of EoS information from such an NSBH candidate is expected given the poor measurability of Λ_2 for NSBH mergers at low SNRs (Kumar et al., 2017). However, using a simulated source at a much higher SNR, I show in figure 6.10, that if such a source is indeed detected, our framework can be used to quickly extract EoS constraints from multiple such events.

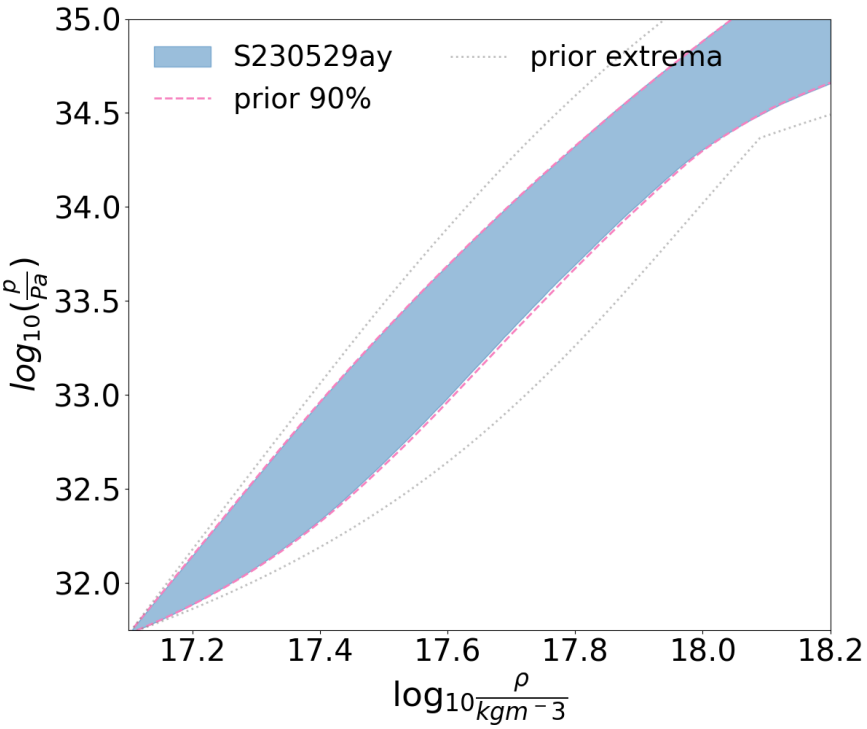


Figure 6.9: EoS constraints obtained from analyzing GW230529's data.

6.4 SUMMARY

In this chapter, I highlighted my contributions to the development of a scalable inference framework that can analyze multiple BNS and NSBH observations to constrain the nuclear EoS. This framework will be crucial in probing the properties of matter at extreme densities using GW observations as more and more detections are made by LVK's ongoing and upcoming observing runs. The results of the investigation delineated in Sec. 6.2, is published through Ray et al. (2023a).

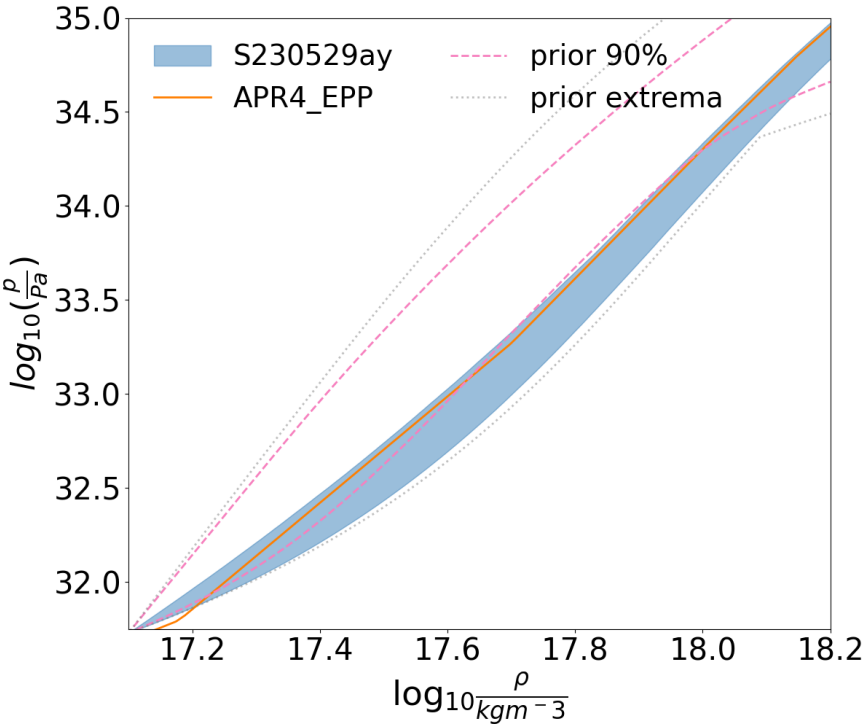


Figure 6.10: EoS constraints obtained from analyzing a GW230529-like simulation, with masses and tidal deformability determined by the APR4_EPP EoS, injected at a much closer distance (40Mpc).

Chapter 7

Searching for deviations from general relativity

When I was a first-term graduate student some 36 years ago, it was said that the field of general relativity is "a theorist's paradise and an experimentalist's purgatory".

C.M. Will

7.1 INTRODUCTION

The invariance of physical results under boosts and rotations, also known as Lorentz invariance, is a fundamental assumption of both the standard model of particle physics and GR. Despite their enormous success, these two theories that comprise our present description of nature, are widely believed to be the low-energy limit of some more fundamental description of nature whose predictions will emerge as the characteristic energies involved in experiments approach the Planck scale, 10^{19} GeV. Given the infeasibility of Planck energy experiments, searching for small deviations from the predictions of GR and the standard model in present day experiments can offer new insights on the underlying theory. The standard model extension (SME) (e.g. Colladay & Kostelecký, 1998) is an effective field theory framework for studying small deviations from the predictions of GR and the standard model in low-energy experiments, that can be caused by the violation of fundamental symmetries such as Lorentz invariance at the level of the action.

The gravity sector of the SME (Kostelecký, 2004; Bailey & Kostelecký, 2006, e.g.) that describes all Lorentz and CPT-violating theories of modified gravity can be fully constrained by measuring the propagation properties of GWs. Such imprints of gravitational Lorentz violation on GW propagation can be quantified into three possible effects: rotational anisotropy, birefringence, and modified dispersion. Given a large catalog of GW observations, all of these effects can be

constrained simultaneously, leading to robust and precise conclusions on several modified theories of gravity. However, previous studies have attempted to constrain these effects only individually and with unreliable modeling of measurement uncertainties. For example, studies such as Haegel et al. (2023) have used a mixture of Bayesian parameter estimation and χ^2 fitting to constrain these coefficients leading to ambiguity in the meaning of their confidence intervals. Furthermore, measuring GW properties such as propagation speed in different directions (Cornish et al., 2017; Liu et al., 2020) from multiple multi-detector GW observations is highly susceptible to selection biases given the stringent timing restrictions employed by GW search pipelines to identify multi-detector coincidences. Previous studies have ignored these selection effects completely.

Throughout the rest of this chapter, I will delineate my contributions to constraining Lorentz violating GW propagation using GWTC-3 data. Using the likelihood of Eq. (2.53) implemented through Eqs. (2.57), and (2.61), I developed a fully Bayesian framework of hierarchical inference for constraining SME coefficients in the non-birefringent and non-dispersive limit. Our method, for the first time, correctly accounted for selection biases in these kinds of studies. I further used this method to simultaneously contain all 9 coefficients of anisotropic GW propagation from GWTC-3 data.

7.2 MEASURING THE SPEED OF GRAVITATIONAL WAVES AND LORENTZ VIOLATION IN THE PRESENCE OF SELECTION EFFECTS

The speed of gravitational waves v_g can be measured with the time delay between gravitational-wave detectors. Our study provides a more precise measurement of v_g using gravitational-wave signals only, compared with previous studies. We select 52 gravitational-wave events that were detected with high confidence by at least two detectors in the first three observing runs (O1, O2, and O3) of Advanced LIGO and Advanced Virgo. We use Markov chain Monte Carlo and nested sampling to estimate the v_g posterior distribution for each of those events. We then combine their

posterior distributions to find the 90% credible interval of the combined v_g distribution for which we obtain $0.99^{+0.02}_{-0.02}c$ without the use of more accurate sky localization from the electromagnetic signal associated with GW170817. Restricting attention to the 50 binary black hole events generates the same result, while the use of the electromagnetic sky localization for GW170817 gives a tighter constraint of $0.99^{+0.01}_{-0.02}c$. The abundance of gravitational wave events allows us to apply hierarchical Bayesian inference on the posterior samples to simultaneously constrain all nine coefficients for Lorentz violation in the nondispersive, nonbirefringent limit of the gravitational sector of the Standard-Model Extension test framework. We compare the hierarchical Bayesian inference method with other methods of combining limits on Lorentz violation in the gravity sector that are found in the literature.

7.2.1 Motivation

The third observing run (O3) of Advanced LIGO (J. Aasi et al., 2015) and Advanced Virgo (Acernese et al., 2014) was the first complete run in which all three detectors were used (Abbott et al., 2021c, 2023b). In total, O3 adds 79 gravitational-wave (GW) candidates, more than seven times the 11 GW candidates from the first (O1) and second (O2) observing runs combined (Abbott et al., 2019b). With the availability of many more GW events, it becomes possible to measure the speed of gravitational waves v_g more precisely than previous works that used similar methods (Cornish et al., 2017; Liu et al., 2020). Furthermore, it allows a direct and comprehensive exploration of the isotropy of v_g for the first time.

General relativity predicts that the speed of GWs is the same as the vacuum speed of light c. The GWs detected by Advanced LIGO and Advanced Virgo can be used to make statistical inferences about v_g, thereby testing the theory of general relativity. The first measurement of v_g using the time delay between the GW detectors was performed by Cornish et al. (2017). By applying Bayesian inference, the 90% credible interval of v_g distribution was constrained to be

$(0.55c, 1.42c)$ (Cornish et al., 2017). Reference (Liu et al., 2020) further constrained the 90% credible interval to $(0.97c, 1.05c)$, by applying similar methods to 11 events from O1 and O2. With a total of 52 high-confidence multi-detector GW events accrued through the end of O3, we are able to perform a similar analysis using more events, more robustly testing the theory of general relativity. The method used here remains much less sensitive than the multimessenger astronomy approach used in Ref. (Abbott et al., 2017b), which placed the constraint $+7 \times 10^{-16} \leq \frac{v_g - c}{c} \leq -3 \times 10^{-15}$. Rather an improved precision, the present approach provides confirmation of the basic conclusion of $v_g = c$ via an alternative approach. More significantly, the many events available that come from different sky directions permit the exploration of the isotropy of v_g.

The large body of events now available, which arrive from a multitude of sky directions, allows for a complete exploration of the isotropy of v_g in the context of the Lorentz invariance test framework provided by the gravitational Standard-Model Extension (SME).[1] Reference (Liu et al., 2020) simultaneously constrained the first four of nine coefficients for Lorentz violation in the nondispersive, nonbirefringent limit of the gravity sector using four GW events from O1 and O2. Other recent works (Haegel et al., 2023; Niu et al., 2022; Gong et al., 2023) have sought the effects of birefringence and dispersion using the SME. Still others have sought the dependence of GW speed on the motion of the source (Ghosh et al., 2023a). In this paper, we use 24 of 52 high-significance multi-detector GW events to simultaneously constrain all nine coefficients in the nondispersive, nonbirefringent limit of the SME. While our constraints are much weaker than previous works such as Abbott et al. (2017d), which have constrained the coefficients for Lorentz violation in the gravity sector down to the order of 10^{-15} to 10^{-14} via multimessenger astronomy, these constraints were obtained using models with only one parameter each. Therefore, our work is the first to provide direct limits from GW observations on all nine coefficients simultaneously.

[1] For an annually updated review of observational and experimental results, see (Kostelecký & Russell, 2008). For early foundational work on the SME, see (Colladay & Kostelecký, 1998). For foundational gravity-sector work, see (Kostelecký, 2004; Bailey & Kostelecký, 2006).

The remainder of this paper is organized as follows. In Sec. 7.2.2, we discuss the methods used to extract v_g estimates for each event and present the results. Section 7.2.3 presents and compares a number of methods for extracting simultaneous limits on the nine coefficients for Lorentz violation before presenting our final estimate of these coefficients from the O1-O3 data.

7.2.2 Speed of Gravitational Waves

7.2.2.1 Bayesian Inference Methods

Here, we briefly describe our method for obtaining the speed of GWs. Interested readers are invited to refer to Liu et al. (2020) for full details.

When a GW passes through Earth, if two or more detectors detect the signal, we can use the relative locations of the detectors and the difference in detection times from those detectors to simultaneously estimate the sky location of the GW event and v_g. With only one detector, we cannot find any v_g information, as there is no difference in detection times in this case. Therefore, we select those events that are detected by at least two GW detectors.

Furthermore, we only consider those events whose median signal-to-noise ratios (SNR) are no smaller than 10.0, as reported in the GWTC-2 and GWTC-3 catalog papers (Abbott et al., 2021c, 2023b). In total, 41 O3 events meet our selection criteria and are listed in Tables 7.2 and 7.1. All O1 and O2 events meet these two selection criteria, so we include their posterior distributions used in Liu et al. (2020) in our analysis. Note that the SNR values used to select the O1 and O2 events (which is the same as what was used in Ref. (Liu et al., 2020)) correspond to the network SNR with which the events were found by the GstLAL search pipeline as reported in Ref. (Abbott et al., 2019b).

The standard parameter estimation using GW data from multiple detectors imposes the constraint that GWs travel at the speed of light (LIGO Scientific Collaboration, 2018). In this work, we remove this constraint such that v_g becomes a parameter to be estimated with all other signal

parameters. This causes wider distributions for certain parameter estimations. For example, the calculated sky area is often larger because a defined v_g aids sky localization.

Gravitational wave data d, can be decomposed into a pure GW signal $h(t)$ plus random noise $n(t)$,

$$d(t) = h(t) + n(t). \tag{7.1}$$

Within the framework of Bayesian inference, the posterior distribution of the parameters $\vec{\theta}$ characterizing a GW signal is computed from the likelihood of obtaining GW data given particular values of said parameters and the a priori knowledge of what we expect those values to be. The likelihood function is constructed by assuming the noise $n(t)$ to be stationary and Gaussian distributed. For details regarding the exact forms of the likelihood see Ref. (Liu et al., 2020). Once obtained, the joint posterior distribution of the signal parameters can be used to compute the marginalized posterior distribution of v_g as in:

$$p(v_g|d) = \int p(\vec{\theta}|d) d\vec{\theta'}, \tag{7.2}$$

where $\vec{\theta'}$ is the set of parameters in $\vec{\theta}$ except for v_g (Liu et al., 2020).

To carry out PE for each event that passes our selection criteria, we use public data (Abbott et al., 2021f, 2023c) from GWTC-1 through GWTC-3. We use `LALINFERENCE_MCMC` (LIGO Scientific Collaboration, 2018; Metropolis et al., 1953; Hastings, 1970; Veitch et al., 2015) which implements MCMC with Metropolis-Hastings algorithm and `LALINFERENCE_NEST` which implements nested sampling to run the Bayesian parameter estimation (LIGO Scientific Collaboration, 2018; Veitch & Vecchio, 2010; Skilling, 2006). For our purposes of extracting v_g distributions, these two algorithms generate comparable results. We use the publicly available power spectral densities and calibration envelopes from the LIGO Scientific, Virgo and KAGRA (LVK) collaborations in our analysis. In this paper, we use a uniform prior in v_g between $0.1c$ and $10c$. When

the v_g posterior rails against the prior, we increase the upper limit of the prior by another $10c$. The broadest prior we use is from $0.1c$ to $30c$, which we only use for one event, GW190929_012149. For parameters such as binary masses and spins, we use the same uniform and isotropic priors as those used by the LVK (Abbott et al., 2019b, 2021c, 2023b). We choose a distance prior that is proportional to luminosity distance squared, similar to Abbott et al. (2019b). We do not use the more complicated cosmological priors used by Abbott et al. (2021c, 2023b). For O1 and O2 events, we use the posterior samples from Liu et al. (2020), which used the IMRPhenomPv2 (Khan et al., 2016; Husa et al., 2016; Hannam et al., 2014) waveform for all events except for the binary neutron star (BNS) event GW170817 which was analyzed with the TaylorF2 waveform (Arun et al., 2009; Buonanno et al., 2009; Mikóczi et al., 2005; Vines et al., 2011; Bohé et al., 2013, 2015). For most O3 events, we use the phenom4 waveform (Khan et al., 2016; Husa et al., 2016), which is an aligned spin waveform model for black-hole binaries. We do not use the more sophisticated IMRPhenomPv2 model for these events since in the context of our study, we do not expect any siginificant change in v_g measurements to result from the additional intricacies of the more sophisticated model. We have verified this lack of change for a subset of these events and hence chosen to stick to the phenom4 model consistently for all O3 events except for GW190521. For the extremely high-mass BBH event GW190521, we use the NRSur7dq4 waveform (Varma et al., 2019) which is one of the waveform models used by Ref. (Abbott et al., 2020d) for inferring this event's source properties. We note that IMRPhenomPv2, phenom4, and NRSur7dq4 are all waveform models with inspiral, merger as well as ringdown.

We can achieve a more precise measurement of v_g by combing data from multiple GW events. By interpreting each observation as an independent experiment, we can multiply the marginalized likelihood as a function of v_g corresponding to each event and obtain the joint posterior distribution of v_g given data from multiple events. For a uniform prior on v_g, the joint posterior can be expressed as a product of individual event posteriors.

Suppose the GW detectors observe n independent GW events with data $d_1, d_2, ..., d_n$. For a uniform prior distribution of v_g, the combined posterior distribution of v_g is

$$p(v_g|d_1, d_2, ..., d_n) \propto p(v_g|d_1)p(v_g|d_2)\cdots p(v_g|d_n). \tag{7.3}$$

The single event posterior distributions $p(v_g|d_i)$ are obtained as a numerical function of v_g from its PE samples by means of Gaussian Kernel Density estimation (KDE) (Silverman, 1986; Scott, 1992). We use the package SCIPY's implementation of Gaussian KDE to obtained the posteriors (Virtanen et al., 2020). The joint posterior distribution is then obtained through Eq. (7.3).

Then, for individual and combined posteriors, we calculate Bayes factors K, via the Savage-Dickey density ratio

$$K = \frac{p(v_g = c \mid d_1, d_2, ...)}{p(v_g = c)}, \tag{7.4}$$

where $p(v_g = c \mid d_1, d_2, ...)$ is the posterior probability of $v_g = c$, and $p(v_g = c)$ is the prior probability of $v_g = c$ (Wagenmakers et al., 2010). Higher Bayes factors suggest stronger evidence for $v_g = c$.

7.2.2.2 Results

In Tables 7.2 and 7.1, we show the v_g estimates with 90% credible intervals, network SNRs, sky areas at 90% credible level, and Bayes factors for these selected 41 O3 events. Also shown are the analogous quantities obtained from their combined posteriors. Out of the 41 selected O3 events, 40 events are binary black hole (BBH) candidate events. GW200115_042309 is a neutron star-black hole (NSBH) event, with masses of $5.9^{+2.0}_{-2.5}\mathrm{M}_\odot$ and $1.44^{+0.85}_{-0.29}\mathrm{M}_\odot$ at 90% credible interval (Abbott et al., 2023b). Here, by combining the 41 selected O3 events, we constrain the 90% credible interval of v_g to be $0.99^{+0.02}_{-0.03}c$, with a Bayes factor of 205.9.

We combine the O3 results with the O1 and O2 results discussed in Ref. (Liu et al., 2020).

The eleven O1 and O2 events are run with LALINFERENCE_MCMC, which shows results that are consistent with LALINFERENCE_NEST used for O3a runs (Liu et al., 2020; Abbott et al., 2019b, 2021c). In Table 7.3, we show the v_g estimates with 90% credible intervals, network SNRs, sky areas at 90% credible level, and Bayes factors for the 11 O1 and O2 events and their combined posteriors. We use the same posterior samples as used by Liu et al. (2020), but Table 7.3 shows slightly different 90% credible intervals from those in Liu et al. (2020), because we use Gaussian KDE smoothing in this study to extract the credible intervals while Liu et al. (2020) directly used the posterior samples without KDE smoothing (Liu et al., 2020). These 11 events were detected by at least two detectors and had median GstLAL network SNR values greater than 10.0 (Abbott et al., 2019b). GWTC-2.1 (Abbott et al., 2021d) shows network SNR values for O1 and O2 events based on LALINFERENCE parameter estimations, but we choose GstLAL SNR values to be consistent with Liu et al. (2020) from which we obtain the v_g posterior samples. GW170817 is a BNS event that was also detected in the electromagnetic spectrum (Abbott et al., 2017c,d). The “fixed" label means that the result uses the sky localization from the electromagnetic detections, which is much more precise than the localization generated by GW detection pipelines.

Combining the 41 O3 events and 11 O1 and O2 events without fixing GW170817's sky localization at the detected EM signal, we obtain the 90% credible interval of v_g to be $0.99^{+0.02}_{-0.02}c$, with a Bayes factor of 291.9. With GW170817 sky localization fixed, v_g is $0.99^{+0.01}_{-0.02}c$, with a Bayes factor of 249.0. For a total of 49 BBH events, i.e. excluding GW170817, GW190924, and GW200115, v_g is $0.99^{+0.02}_{-0.02}c$, with a Bayes factor of 221.2. FIG. 7.1 shows the combined posterior of v_g.

O3a Event	v_g(c)	SNR	Ω(deg^2)	Bayes Factor
*GW190408_181802	$1.66^{+0.}_{-0.}$	15.:	1216	3.5
*GW190412	$1.49^{+0.47}_{-0.53}$	18.9	594	2.8
GW190421_213856	$1.15^{+0.}_{-0.}$	10.'	2837	11.9
*GW190503_185404	$0.55^{+0.26}_{-0.24}$	12.4	1237	0.5
*GW190512_180714	$1.42^{+0.}_{-0.}$	12.:	1637	4.9
*GW190513_205428	$1.26^{+1.51}_{-0.65}$	12.9	1075	7.6
GW190517_055101	$0.88^{+0.}_{-0.}$	10.'	2125	13.8
GW190519_153544	$2.04^{+1.75}_{-1.21}$	15.6	2070	3.7
GW190521	$1.82^{+5.}_{-1.}$	14.:	2279	3.4
GW190521_074359	$1.20^{+0.64}_{-0.44}$	25.8	2318	11.8
*GW190602_175927	$0.98^{+0.}_{-0.}$	12.:	1567	27.1
GW190630_185205	$6.07^{+3.52}_{-4.90}$	15.6	3983	0.5
*GW190701_203306	$0.85^{+0.}_{-0.}$	11.:	203	21.4
GW190706_222641	$6.67^{+2.91}_{-4.88}$	12.6	3726	0.3
GW190707_093326	$2.77^{+5.}_{-1.}$	13.:	5708	0.9
*GW190720_000836	$1.54^{+0.13}_{-0.37}$	11.0	388	1.3
*GW190727_060333	$3.42^{+2.}_{-2.}$	11.'	1521	0.6
*GW190728_064510	$0.97^{+0.82}_{-0.53}$	13.0	1873	14.9
*GW190814	$1.33^{+0.}_{-0.}$	24.'	334	4.1
GW190828_063405	$6.11^{+3.41}_{-4.90}$	16.2	2498	0.5
GW190828_065509	$1.64^{+4.}_{-0.}$	10.(	2917	4.0
*GW190915_235702	$0.40^{+0.71}_{-0.11}$	13.6	904	5.1
*GW190924_021846	$0.92^{+0.}_{-0.}$	11.:	918	21.1
GW190929_012149	$6.08^{+3.53}_{-4.90}$	10.1	5761	0.5

Table 7.2: The 90% credible intervals of v_g from individual O3a events posteriors. Median network SNR values are reported from GWTC-2 (Abbott et al., 2021c). The 90% credible regions of the sky localization (Ω) without fixing v_g at c are calculated from the individual posteriors (Abbott et al., 2021c). The Bayes factor K indicates how strong the posterior distributions support $v_g = c$. The asterisks (*) in front of GW event names represent the GW events chosen for obtaining constraints on all nine coefficients for Lorentz violation in Sec. 7.2.3.

O3b Event	v_g(c)	SNI	Ω(de	Bayes Factor
GW191109_010717	$1.80^{+1.20}_{-0.93}$	17.3	4033	2.4
GW191129_134029	$1.69^{+4.}_{-1.}$	13.	4891	4.2
GW191204_171526	$1.18^{+1.26}_{-0.99}$	17.5	3009	5.5
GW191215_223052	$1.44^{+2.}_{-0.}$	11.	3280	5.9
GW191216_213338	$1.31^{+0.59}_{-0.37}$	18.6	2076	11.0
GW191222_033537	$5.11^{+4.}_{-3.}$	12.	3206	0.4
GW191230_180458	$1.42^{+1.07}_{-0.91}$	10.4	2538	5.9
GW200115_042309	$3.02^{+5.}_{-2.}$	11.	3271	1.2
GW200128_022011	$5.57^{+3.79}_{-3.90}$	10.6	8988	0.3
*GW200129_065458	$0.99^{+0.}_{-0.}$	26.	149	53.7
*GW200202_154313	$0.69^{+0.20}_{-0.33}$	10.8	1551	0.7
*GW200208_130117	$1.39^{+0.}_{-0.}$	10.	706	5.6
GW200219_094415	$1.92^{+2.20}_{-1.59}$	10.7	3114	3.2
*GW200224_222234	$1.03^{+0.}_{-0.}$	20.	94	48.1
GW200225_060421	$1.24^{+0.66}_{-0.75}$	12.5	3583	8.0
*GW200311_115853	$0.96^{+0.}_{-0.}$	17.	102	32.1
*GW200316_215756	$3.76^{+5.26}_{-2.76}$	10.3	1881	0.8
O3 Combined (BBHs)	$0.99^{+0.}_{-0.}$			203.3
O3 Combined	$0.99^{+0.02}_{-0.03}$			205.9

Table 7.1: The 90% credible intervals of v_g from individual O3b events posteriors and combined posteriors using all O3 events. Median network SNR values are reported from GWTC-3 (Abbott et al., 2021c). The 90% credible regions of the sky localization (Ω) without fixing v_g at c are calculated from the individual posteriors (Abbott et al., 2023b). The Bayes factor K indicates how strong the posterior distributions support $v_g = c$. The asterisks (*) in front of GW event names represent the GW events chosen for obtaining constraints on all nine coefficients for Lorentz violation in Sec. 7.2.3.

Event	v_g(c)	SN	Ω(de	Bayes Factor
GW150914	$0.68^{+0.50}_{-0.29}$	24.4	2385	7.0
GW151012	$6.01^{+3.}_{-4.}$	10.(	6607	0.4
GW151226	$4.04^{+5.17}_{-3.28}$	13.1	6515	1.3
GW170104	$1.61^{+1.}_{-1.}$	13.(	5313	4.3
*GW170608	$1.14^{+0.22}_{-0.26}$	14.9	1269	16.9
*GW170729	$3.70^{+1.}_{-2.}$	10.:	1287	0.3
GW170809	$0.65^{+0.36}_{-0.34}$	12.4	2252	8.6
*GW170814	$1.00^{+0.}_{-0.}$	15.(	250	68.0
GW170817 (unfixed)	$1.01^{+0.04}_{-0.06}$	33.0	53	111.9
*GW170817(fixed)	$0.99^{+0.}_{-0.}$	33.(	0	228.5
*GW170818	$0.94^{+0.24}_{-0.38}$	11.3	168	20.5
GW170823	$3.90^{+4.}_{-3.}$	11.:	6412	0.9
Combined (All BBHs)	$0.99^{+0.02}_{-0.02}$			221.2
Combined (All, fixed)	$0.99^{+0.}_{-0.}$			249.0
Combined (All, unfixed)	$0.99^{+0.02}_{-0.02}$			291.9
Combined (O1/2, fixed)	$0.99^{+0.}_{-0.}$			249.0
Combined (O1/2, unfixed)	$1.01^{+0.04}_{-0.05}$			149.0

Table 7.3: The 90% credible intervals of v_g from individual O1 and O2 events posteriors. Combined posteriors are for all selected O2 and O3 events. Network SNR values are reported from the GstLAL search pipeline in GWTC-1 (Liu et al., 2020; Abbott et al., 2019b). The 90% credible regions of the sky localization (Ω) without fixing v_g at c are calculated from the individual posteriors (Liu et al., 2020; Abbott et al., 2019b). The Bayes factor indicates how strong the posterior distributions support $v_g = c$. The asterisks (*) in front of GW event names represent the GW events chosen for obtaining constraints on all nine coefficients for Lorentz violation in Sec. 7.2.3. The "fixed" and "unfixed" labels represent whether we fixed the sky localization of GW170817 at the source of its EM counterpart.

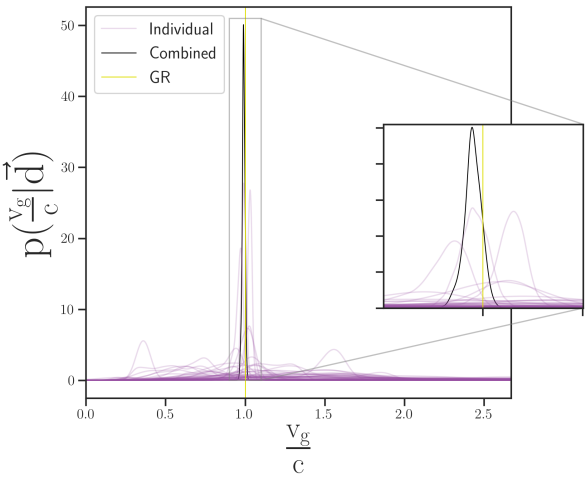


Figure 7.1: Posterior distributions of v_g inferred jointly from all 41 events. The black solid line represents the joint posterior, the thin purple lines represent individual event posteriors and GR ($v_g = c$) is marked by the yellow vertical line.

7.2.2.3 Discussion

In Liu et al. (2020), with 11 O1 and O2 events and GW170817's sky localization unfixed, the combined posterior distribution of v_g was measured to be $1.01^{+0.04}_{-0.05}c$, while here we measure v_g to be $0.99^{+0.02}_{-0.02}c$ with the 52 selected events. With GW170817's localization fixed, in Liu et al. (2020), the combined posterior distribution of v_g was measured to be $0.99^{+0.02}_{-0.02}c$ for 11 events, while here we find $0.99^{+0.02}_{-0.01}c$ for the 52 events. Given that $1c$ is the relativistic prediction of v_g, the combined posterior distributions of v_g measured using 52 selected events show no evidence for a violation of general relativity. All of these combined results have Bayes factors on the order of 10^2, providing strong evidence for $v_g = c$.

Here, our measured distribution of v_g is much narrower than that measured with 11 O1 and O2 events in Liu et al. (2020) using GW signals alone. This is reasonable, given the larger sample size of events included in this study. When we assume that the v_g distributions of individual events are independent and identically distributed, we expect that the measurement errors would decrease by $1/\sqrt{n}$. In our calculations, we find that the combined v_g distribution roughly follows such a pattern

as more events are added. For example, with 11 O1 and O2 events, the combined v_g posterior had an error bar of $0.09c$. With 52 events in total, the combined posterior had an error bar of $0.04c$, which follows $0.09c/\sqrt{52/11} \approx 0.04c$. With GW170817's sky localization unfixed, we find that Bayes factor more than doubles from the value of 149.0 obtained from 11 O1 and O2 events to the value of 291.9 obtained with all 52 events. This, in conjunction with the error bar being reduced by half, implies that our measurement with 52 GW events in total has provided approximately twice stronger evidence for $v_g = c$.

Interestingly, we find that the combined 90% v_g credible interval using the 41 O3 events is approximately the same as the 90% v_g credible interval obtained by only considering GW170817 with the fixed sky localization. GW170817 had an SNR of 33.0, while only four of the 41 O3 events had SNRs above 20.0, with the highest being 26.8 for GW200129_065458. The similarity between the v_g posterior of GW170817 alone and the 41 O3 events combined suggests that some combination of higher SNRs and better sky localization do help put tighter constraints on v_g. This shows that our decision to exclude events with SNRs lower than 10.0 should not have a high impact on the v_g estimates.

Looking to the future, additional two- and three-detector BBH events with SNRs typical of those above will lead to a slow improvement in v_g measurements as improvements proceed as $1/\sqrt{n}$. However, as GW detectors become more sensitive and the network of detectors expands, we expect more high-SNR, multi-detector GW events that would likely lead to a more rapid pace of progress in v_g estimations via the methods used here. Meanwhile, future multimessenger detections can provide more precise sky localizations, which will likely improve the error bars on the 90% v_g credible interval further.

7.2.3 Simultaneous SME Limits

7.2.3.1 Basics

In the non-birefringent, non-dispersive limit of the SME (mass dimension $d = 4$), using natural units and assuming that the nongravitational sectors, including the photon sector, are Lorentz invariant, the difference between the group velocities of gravity and light takes the form (Kostelecký & Mewes, 2016):

$$\Delta v = -\sum_{lm} Y_{lm}(\hat{n})\tfrac{1}{2}(-1)^{1+l}\bar{s}_{lm}, \tag{7.5}$$

where the Y_{lm}'s are the spherical harmonics with $l \leq 2$. Here the nine Lorentz-violating degrees of freedom are characterized by the spherical coefficients for Lorentz violation $\bar{s}_{lm}$, and $\hat{n}$ is the sky location of the source of the GWs. We can expand Eq.(7.5) over positive m to get its equivalent expression:

$$\Delta v = \sum_{l}(-1)^{l}\Big(\tfrac{1}{2}\bar{s}_{l0}Y_{l0} + \sum_{m>0}[\mathrm{Re}\ \bar{s}_{lm}\mathrm{Re}\ Y_{lm} - \mathrm{Im}\ \bar{s}_{lm}\mathrm{Im}\ Y_{lm}]\Big). \tag{7.6}$$

The SME is a broad and general test framework for testing Lorentz invariance. Unlike models that attempt to describe specific effects with a small number of parameters, test frameworks, because of their generality, have a large number of undetermined coefficients to be explored in experimental data. While a number of studies have proceeded under a simplified approach, sometimes referred to as a maximum reach analysis (Flowers et al., 2017), in which only one coefficient at a time is considered, it is also common to study a family of coefficients together in what is sometimes referred to as a coefficient separation approach (Flowers et al., 2017). In the context of the maximum reach approach, many coefficients can sometimes be constrained one at a time using a single measurement, while a number of measurements that is greater than or equal to the number of coefficients considered is typically required to simultaneously measure the entire family.

A number of approaches to simultaneously estimating multiple coefficients exist in the liter-

ature (H. Pihan-le Bars et al., 2019; Kostelecký & Tasson, 2015; Liu et al., 2020; Haegel et al., 2023). One approach involves directly fitting a single data stream to a model involving all of the coefficients in the family.[2] This approach is well-suited to experiments that take data as the lab is boosted and rotated.

In the context of astrophysical observations, each individual event provides a measurement of a linear combination of coefficients for Lorentz violation. A system of these inequalities must then be solved, or otherwise disentangled, for estimates of the coefficients for Lorentz violation. Several methods of addressing this issue exist in the literature. In this section, we will compare the implications of several of these approaches in the context of the speed of gravitational wave data, as well as introduce new methods based on hierarchical Bayesian inference. Our goal is to consolidate information about these methods and help illuminate their relative merits. We achieve that goal by performing a Mock Data Challenge (MDC), where-in we generate synthetic data corresponding to a chosen set of "true" values of the SME coefficients and test the efficacy of each method in recovering the true values from the synthetic data.

Given their performance in the MDC, along with other considerations, we choose one of these methods whose merits outweigh that of the others and use it to analyze the real speed of gravity data from a subset of the events analyzed in Sec. 7.2.2 to generate the final results of our SME analysis. Because well-localized events are most informative for the SME analysis, we choose the 24-event subset of those considered in Sec. 7.2.2 with 90% credible posterior sky areas under 2000 square degrees as obtained from our parameter estimation with v_g as a free parameter.

7.2.3.2 Linear Programming Method

A number of past studies that have performed maximum reach analysis using limits from astrophysical events have taken a linear programming approach. See, for example, Díaz et al. (2014);

[2]For a recent example of this approach in the gravity sector, see (H. Pihan-le Bars et al., 2019).

Lau & Seifert (2017); Kostelecký & Tasson (2015). The basic idea translated to the speed of gravitational waves problem proceeds as follows.

From a given event we have an upper and lower bound on Δv. If we suppose that we know an exact sky location, as is effectively the case for GW170817 when the electromagnetic signal's localization is used, then Eq. (7.6) can be understood as generating a pair of hyperplanes in $\overline{s}_{lm}$ space that are the boundaries of the parameter space excluded by the event. A subsequent event at a different sky location will generate a distinct pair of hyperplanes. Once a set of n events are collected at distinct sky locations, where n is greater than or equal to the dimensionality of the coefficient space, then a finite maximum and minimum allowed value for each coefficient can be identified via a linear programming scheme such as the simplex method.

In the applications of Díaz et al. (2014); Lau & Seifert (2017); Kostelecký & Tasson (2015), the sky localizations were sufficiently well known that analysis could proceed directly via the above prescription. In the current problem, for all events except GW170817, the sky localization is comparatively poorly known. This makes the slopes of the hyperplanes bounding the allowed region poorly known.

To address our uncertainty in sky positions, the linear programming scheme can be adapted as follows. The linear programming process can be applied with all possible hyperplanes generated by samples from our inference that fall within the 68% credible sky localization bands. The worst-case limits generated by the set of linear programming analysis can then be taken as bounds. As might be expected, this method generates very conservative bounds relative to the methods to follow. Testing this approach using four test events and a skymap resolution of $N_{\text{side}} = 64$, which corresponds to $12 \times 64^2 = 49152$ pixels on the celestial sphere (Zonca et al., 2019; Górski et al., 2005), we generate bounds that are about an order of magnitude greater than the 1σ credible intervals found via the application of the random draw method that we present in the next subsection. Hence we do not consider this approach further as a method of extracting SME limits from the

speed of gravitational waves data at this time.

7.2.3.3 Random Draw Method

In Ref. (Liu et al., 2020) the random draw method for extracting simultaneous limits on coefficients for Lorentz violation was first used. In that work, simultaneous limits were achieved for the set of four $l = 0$ and $l = 1$ $\overline{s}_{lm}$ coefficients using the 4 high-confidence, well-localized events available at the time. In this section, we review this method and discuss ways of extending it to cases in which the number of events exceeds the number of coefficients to be estimated.

The result of the inference discussed in Sec. 7.2.2.1 is a set of samples with each sample consisting of values for each of the parameters including the speed of GWs and the sky localization. Hence distributions for each of the sampled parameters are generated. If one randomly draws one sample associated with each event, one can then solve for the coefficients for Lorentz violation that are consistent with that set of samples using Eq. (7.6). The process of randomly drawing one sample from each event and solving for the coefficients can be iterated to build up a set of samples for the $\overline{s}_{lm}$ coefficients. In other words, a set of points in $\overline{s}_{lm}$ space is built up.

The process described above is straightforward when the number of events observed is equal to the number of coefficients for Lorentz violation to be estimated. Furthermore, in such a scenario, quantile ranges of $\overline{s}_{lm}$ computed from the set of samples of $\overline{s}_{lm}$, accurately represent the uncertainty in our measurement of the SME coefficients. This is because using one posterior sample of $(\Delta v, \theta, \phi)$ from each event and exactly calculating $\overline{s}_{lm}$ from them by solving a set of non-degenerate linear equations, is equivalent to computing and multiplying the posterior distributions of $\overline{s}_{lm}$ for each event and then drawing one sample from that joint posterior. However, in the case where the number of GW observations exceeds the number of SME coefficients, the linear equations become degenerate and hence no longer exactly solvable. While one can be tempted to cherry-pick the top 9 events with the highest SNRs and lowest sky areas from the set of observa-

tions and perform random draw on those, such an analysis will not be maximally informative given the data we have. We can do better using Bayesian hierarchical inference techniques which can combine information from a large number of events, producing much more informative bounds on the SME coefficients with accurate estimation of measurement uncertainties.

Before discussing our robust Bayesian methods we show how the random draw method can be extended to the case in which the number of observations exceeds the number of coefficients for Lorentz violation by means of Singular Value Decomposition (SVD). However, this extension of the random draw method is susceptible to the limitations of the approximation used to perform the SVD and hence cannot produce reliable uncertainty estimates for the measured Lorentz violation parameters. We elaborate more on this near the end of this section while informing the reader beforehand that this SVD-assisted random draw generalization is useful in the present context only as a consistency check and an optimization tool for the hierarchical Bayesian methods on which we rely for our final results.

For n_{SME} number of Lorentz Violation coefficients and N_E number of events with $n_{\mathrm{SME}} < N_E$, for each random draw, we need to solve the degenerate system of linear equations:

$$\boldsymbol{A}[\bar{\boldsymbol{s}}_{\boldsymbol{lm}}] \;=\; [\boldsymbol{\Delta v}]. \tag{7.7}$$

Here $\boldsymbol{A}$ is an $N_E \times n_{\mathrm{SME}}$ matrix in which each row corresponds to one of the N_E events under consideration. The entries in each of the n_{SME} columns moving across a given row consist of the coefficients of $\bar{\boldsymbol{s}}_{\boldsymbol{lm}}$ in Eq. (7.6), computed for a random sample of θ, ϕ drawn from the event corresponding to that row. The n_{SME} SME coefficients to be computed are organized into a column vector denoted $[\bar{\boldsymbol{s}}_{\boldsymbol{lm}}]$, while $[\boldsymbol{\Delta v}]$ denotes a column vector of the randomly drawn Δv corresponding to the samples used in constructing the rows of $\boldsymbol{A}$. Before factorizing the non-square matrix, we scale both sides of each line of Eq. (7.7) by the standard deviation of the Δv samples corresponding to that event. We define $[\boldsymbol{\sigma}_{\boldsymbol{\Delta v}}]$ to be a column vector in which each element corresponds

to the standard deviation of the Δv samples from that particular event, then we can write the scaled version of Eq. (7.7) as:

$$\boldsymbol{A'}[\bar{\boldsymbol{s}}_{\boldsymbol{lm}}] \;=\; [\boldsymbol{\Delta v'}], \tag{7.8}$$

where

$$\boldsymbol{A'_{ij}} \;=\; \frac{\boldsymbol{A}_{ij}}{[\boldsymbol{\sigma}_{\boldsymbol{\Delta v}}]_i} \tag{7.9}$$

$$[\boldsymbol{\Delta v'}]_i \;=\; \frac{[\boldsymbol{\Delta v}]_i}{[\boldsymbol{\sigma}_{\boldsymbol{\Delta v}}]_i}. \tag{7.10}$$

The SVD factorizes the non-square matrix $\boldsymbol{A'}$ into two orthogonal square matrices $\boldsymbol{U}$ and $\boldsymbol{V}$, that are $N_E \times N_E$ and $n_{\rm SME} \times n_{\rm SME}$ respectively, and a diagonal $N_E \times n_{\rm SME}$ matrix $[\Sigma]$ with non-negative entries:

$$\boldsymbol{A'} \;=\; \boldsymbol{U\Sigma V}^T, \tag{7.11}$$

where Σ has the form:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{S} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \tag{7.12}$$

with

$$\boldsymbol{S} = diagonal\{\sigma_1, ..., \sigma_{n_{\rm SME}}\}. \tag{7.13}$$

The non-negative values $\sigma_1 > \sigma_2 > ... > \sigma_{n_{\rm SME}}$ are known as singular values and are estimated along with $\boldsymbol{U}$ and $\boldsymbol{V}$ by a linear least squares algorithm(Golub & Reinsch, 1970). The scaling with the standard deviation of Δv essentially transforms a least square minimized SVD on $\boldsymbol{A'}$ into a Chi-square minimized SVD on $\boldsymbol{A}$. This allows us to properly account for the fact that some events in our list are less significant than others. Proceeding without this scaling biases the SVD. Once

computed, the singular values can be used to solve for $\bar{\boldsymbol{s}}_{\boldsymbol{lm}}$ in Eq. (7.7) :

$$[\bar{\boldsymbol{s}}_{\boldsymbol{lm}}]_i = \frac{1}{\sigma_i} \sum_{k=1}^{n_{\mathrm{SME}}} \boldsymbol{V}_{ik} [\boldsymbol{U}^T \boldsymbol{\Delta v'}]_k \tag{7.14}$$

for each draw. We can then estimate the densities of the SME parameters from all draws and produce constraints on them.

We note that despite being a computationally cheap method for computing constraints on the SME coefficients from multiple GW events, the SVD-assisted random draw method has certain inadequacies. There is ambiguity in the exact meaning and interpretation of the uncertainty estimates produced by this method. In the case where the number of events is larger than the number of SME coefficients, this implementation of the random draw method boils down to randomly choosing a posterior sample of $(\Delta v, \theta, \phi)$ from each event and doing a least chi-square fit for the SME parameters. This procedure is then repeated a large number of times, producing a least chi-square fit of the SME coefficients for each draw. However this is not equivalent to the multiplication of posterior probabilities of the SME coefficients, over all events, and drawing samples from that joint posterior. Thus the quantile ranges of the set of chi-square fitted SME coefficients do not hold the same meaning as Bayesian credible intervals. While the Bayesian intervals represent regions of the SME parameter space wherein their true values lie with a particular posterior probability given the data, the SVD-based random draw constraints can be expected to have a different meaning, the exact nature of which remains ambiguous.

Due to these considerations, we conclude that the weighted SVD-assisted random draw method produces constraints that are unreliable and are likely to be underestimates of the true uncertainties in the measurement of SME coefficients. We verify this claim by testing this method against its Bayesian counterparts in an MDC that we describe later in this work. The results of the MDC show that the samples of SME parameters produced by this method are concentrated in a narrow region around the true values of the parameters, which also coincide with the peaks of the poste-

rior distributions inferred by the Bayesian methods. Therein lies the merit of this method in the present context and its potential to serve as a rapid consistency check for the Bayesian methods. Furthermore, this method is extremely fast and computationally cheap and hence can be used to quickly find the narrow region in the parameter space inside which the peak of the posteriors lies. The stochastic MCMC sampling employed by our Bayesian methods is expected to converge much faster if the MCMC chains are initialized near the maxima of the posterior being sampled. Thus the SVD-assisted random draw method can be used to optimize the MCMC sampling used in our Bayesian methods with significant speed-up gains for narrowly peaked SME posteriors. Given the large number of events expected to be observed in O4 and the width of the Bayesian intervals we compute using our current set of events, the posterior distributions of the SME coefficients can be expected to be very narrow post O4, and hence lead to a drastic increase in the computational cost and latency of the Bayesian methods being applied to such a data set. This will likely make the optimization of the Bayesian methods as offered by the SVD-assisted random draw method a necessary tool in the near future.

7.2.3.4 Hierarchical Bayesian Inference

Since the SME coefficients are properties that are expected to be the same for all events, one can perform Bayesian Hierarchical Inference on them from the GW data of multiple events. To do so, we can construct the marginalized likelihood of GW data given a particular value of the SME coefficients, jointly from multiple events

$$L(\bar{s}_{lm}) = \prod_{i \in \{\text{events}\}} \int L(d_i|\Delta v', \theta, \phi)\Pi(\Delta v', \theta, \phi|\bar{s}_{lm}) d\Delta v' d\theta d\phi, \tag{7.15}$$

where the SME sensitive part of the prior imposes the relationship (7.6) on $\Delta v, \theta, \phi$ for a given value of the SME coefficients :

$$\Pi(\Delta v', \theta, \phi|\bar{s}_{lm}) = \delta(\Delta v' - \Delta v(\bar{s}_{lm}, \theta, \phi))\pi(\theta)\pi(\phi). \tag{7.16}$$

Here $\Delta v(\bar{s}_{lm}, \theta, \phi)$ is the right hand side of Eq. (7.6). Note that we have chosen to represent the deviation of the speed of gravity from the speed of light by the dummy variable $\Delta v'$ whenever a probabilistic quantity (such as likelihood, posterior, prior, or detection fraction) is expressed as a function of it, so as to distinguish it from the quantity $\Delta v(\bar{s}_{lm}, \theta, \phi)$. The presence of the delta function in Eq. (7.16) is due to the deterministic nature of the Eq. (7.6).

By Bayes' theorem, for a uniform prior on $\bar{s}_{lm}$, the likelihood $L(\bar{s}_{lm})$ is proportional to the posterior of these parameters given GW data. We can now sample this posterior using MCMC to produce joint SME constraints from multiple GW observations. However, this procedure involves a very large number of evaluations of the likelihoods $L(d_i|\Delta v', \theta, \phi)$ which is so computationally expensive that it's practically infeasible.

To get around this problem, one can again use Bayes' theorem to write the likelihood $L(d_i|\Delta v', \theta, \phi)$ as proportional to the ratio of the posterior $p(\Delta v', \theta, \phi|d_i)$ to the prior:

$$L(d_i|\Delta v', \theta, \phi) \quad \propto \quad \frac{p(\Delta v', \theta, \phi|d_i)}{\pi(\Delta v')\pi(\theta)\pi(\phi)} \tag{7.17}$$

Substituting this into Eq. (7.15) gives us:

$$L(\bar{s}_{lm}) \propto \prod_{i \in \{\text{events}\}} \int p(\Delta v', \theta, \phi|d_i)\delta(\Delta v' - \Delta v(\bar{s}_{lm}, \theta, \phi))d\Delta v' d\theta d\phi. \tag{7.18}$$

We can now use the samples drawn from the posterior $p(\Delta v', \theta, \phi|d_i)$ obtained using the parameter estimation run described above to evaluate the integral in Eq. (7.18). Note that we have ignored a

factor of $1/\pi(\Delta v')$ in Eq. (7.18) which is constant since we choose $\pi(\Delta v')$ to be uniform in our parameter estimation runs. However, the presence of the Dirac delta makes it slightly complicated to evaluate this integral directly as a sum over posterior samples. We describe shortly two approximation schemes that can be used to smooth out the discrete sum of Dirac deltas over posterior samples that would entail the evaluation of the integral in Eq. (7.18) and hence constrain the SME coefficients jointly from multiple GW observations. Before that, we first describe why Bayesian Inference of this form is subject to selection biases and how we account for them.

Bayesian Hierarchical Inference from a set of GW events selected based on a particular criterion introduces selection biases into the inferred posterior distribution of hyper-parameters (Mandel et al., 2019; Vitale et al., 2020). Since we are selecting events based on whether they were found with a signal-to-noise Ratio (SNR) greater than some threshold in at least 3 detectors, and since each detector has an antenna pattern that makes it more sensitive to certain sky directions than others at the time of detection(Payne et al., 2020), our analysis might be biased towards some values s_{lm} against others. Particularly, the fact that GW search pipelines such as GstLAL only report multi-detector coincidences based on whether or not the time-delays between the detectors being triggered are smaller than the light travel time between detectors plus a 5 millisecond window, has the potential to bias our results greatly (Messick et al., 2017). Furthermore, non-coincident events are down-ranked in significance by means of single's penalties (Messick et al., 2017), making events even less likely to be detectable for certain cases. Other pipelines such as PyCBC use similar methods for identifying multi-detector coincidences albeit with a different value for the timing error window (which is 2 milliseconds for PyCBC (Davies et al., 2020)). The existence of this restriction for coincidence formation in search pipelines implies that we are more likely to discover a multi-detector event if the speed of gravitational waves is greater than or equal to c, as compared to if it were lower than c. Thus, our speed of gravitational wave measurements may be biased towards measuring $\Delta v \geq 0$ against $\Delta v < 0$ along any particular sky position.

To account for this bias, we must normalize our hierarchical likelihood over the true rate of events as opposed to the detected rate, with the latter being different from the former, due to selection biases. The constant of normalization is the fraction of events that are detectable given a particular value of the hyper-parameters and the detection criteria:

$$L(\bar{s}_{lm}) \propto \frac{1}{\beta^N_{\rm det}(\bar{s}_{lm})} \times \prod_{i\in\{\rm events\}}^{N} \int p(\Delta v', \theta, \phi|d_i)\delta(\Delta v' - \Delta v(\bar{s}_{lm}, \theta, \phi))d\Delta v' d\theta d\phi, \tag{7.19}$$

where $\beta_{\rm det}(\bar{s}_{lm}) = \frac{R_{\rm det}(\bar{s}_{lm})}{R_{\rm true}}$, the fraction of detectable events is the ratio of the detectable rate of events to the true Rate of events (Farr, 2019). To calculate the fraction accurately we must simulate a large number of events whose parameters are drawn from broad enough distributions, inject them into the detector noise realizations, and see what fraction of them are recovered given our selection criteria. To do that we must first quantify our selection criteria in terms of the parameters that characterize the GW signal. Accurate modeling would require us to recalculate the search pipeline's ranking statistic of a simulated event while allowing for non-zero Δv and to find the corresponding False Alarm Rate(FAR) of that trigger from said ranking statistics. One can then apply a threshold on the combined FAR of the event to classify them as detectable or non-detectable. However, such a calculation would require a pipeline-specific analysis which is beyond the scope of this work. Instead, we use an approximated selection criteria: for the i-th event to be detectable, its recovered parameters must satisfy:

$$\begin{aligned} \text{det} \implies \{&\rho_H \geq \rho_{\rm th}, \rho_L \geq \rho_{\rm th}, \rho_V \geq \rho_{\rm th}, \rho_{\rm net} \geq \rho_{\rm net,th}, \\ &\Delta t_{HL}(\Delta v) \leq \Delta t_{HL}(0) + 5\,{\rm ms}, t_{HV}(\Delta v) \leq \\ &\Delta t_{HV}(0) + 5\,{\rm ms}, t_{VL}(\Delta v) \leq \Delta t_{VL}(0) + 5\,{\rm ms}\}, \end{aligned} \tag{7.20}$$

where ρ_A is the SNR in detector A, ρ_{net} is the network SNR, $\Delta t_{AB}(\Delta v)$ is the time-delay of signal arrival between detectors A and B as a function of Δv and ρ_{th} is the SNR threshold used

for selecting events. Even though we do not select events depending on which search pipeline found them, we use GstLAL's timing error window to quantify our selection criteria, instead of say PyCBC's, due to the following reason. Among the events that survive our three detector SNR thresholds, most are found by both GstLAL and PyCBC except for GW170818, GW190701, and GW190814 which are found only by GstLAL. Hence, it is sufficient to model the selection biases that might have appeared in this particular study based on GstLAL's value of the timing error window. This would not have been possible if there were events found by PyCBC and not GstLAL with SNR greater than 10 in three detectors during O3. In such a scenario, a more generalized treatment of selection biases would have been necessary, one that accounts for the difference in timing errors allowed by GstLAL and PyCBC.

Now that we have a quantifiable detection criterion, we can carry out our simulations. Once the simulated events are injected into detector noise realizations and classified as detectable or non-detectable depending on their recovered parameters, it is possible to compute the fraction of events detectable given a choice of CBC parameters:

$$f_{\rm det}(\Delta v', \theta, \phi, \vec{\gamma}) \quad \propto \quad \frac{p(\Delta v', \theta, \phi, \vec{\gamma}|det)}{\Pi_{\rm sim}(\Delta v', \theta, \phi, \vec{\gamma})}. \tag{7.21}$$

Here, $\vec{\gamma}$ are additional CBC parameters such as masses, spins, etc. that characterize the waveform, $p(\Delta v', \theta, \phi, \vec{\gamma}|det)$ is the probability of detection, which can be calculated from the set of simulated events that are detectable, and $\Pi_{\rm sim}$ is the prior from which the simulations are drawn, which has to be broad enough so that we have enough events in both the detectable and non-detectable parts of the parameter space. We can marginalize Eq. (7.21) over suitable priors to get:

$$\beta_{\rm det}(\bar{s}_{lm}) = \int f_{\rm det}(\Delta v', \theta, \phi, \vec{\gamma}) \Pi(\Delta v', \theta, \phi|\bar{s}_{lm}) \times \Pi(\vec{\gamma}) d\Delta v' d\theta d\phi d\vec{\gamma}. \tag{7.22}$$

If we choose $\Pi_{\rm sim}(\Delta v', \theta, \phi, \vec{\gamma}) = \pi(\Delta v')\pi(\theta)\pi(\phi)\Pi(\vec{\gamma})$, where $\pi(\Delta v'), \pi, (\theta), \pi(\phi)$ are the same

as the ones defined in Eqs. (7.16) and (7.17), then priors in the denominator and numerator of the integrand in (7.21) cancel out and we can define the marginalized fraction of detectable events (up to the factors that cancel out later):

$$f_{\rm det}^{\rm marg}(\Delta v', \theta, \phi) \quad \propto \quad \int p(\Delta v', \theta, \phi, \vec{\gamma}|det) d\vec{\gamma}. \tag{7.23}$$

As in the case of Eq. (7.18), we have ignored a factor of $1/\pi(\Delta v')$ in Eq. (7.23) for the same reason mentioned before. In terms of this marginalized fraction, $\beta_{\rm det}$ becomes:

$$\beta_{\rm det}(\bar{s}_{lm}) \propto \int f_{\rm det}^{\rm marg}(\Delta v', \theta, \phi) \delta(\Delta v' - \Delta v(\theta, \phi, \bar{s}_{lm})) d\Delta v' d\theta d\phi.$$

To estimate $p(\Delta v', \theta, \phi, \vec{\gamma}|det)$ and hence $f_{\rm det}^{\rm marg}(\Delta v', \theta, \phi)$ we simulate a large number of events whose parameters are drawn from a broad distribution. We then inject the corresponding signals into detector noise realizations and record their SNRs and arrival times. We then apply our selection criteria to find which of these simulated events are detectable given our criteria and estimate $p(\Delta v', \theta, \phi, \vec{\gamma}|det)$. The estimation schemes will depend on which of the two approximations referred to before are used to smooth out the delta function integral and are hence described in more detail in the corresponding subsections below.

The priors we use to draw the simulated events are truncated power-law in the primary mass and mass ratio, uniform in spin, sky position, orientation, co-moving volume, geocentric time, and speed of gravitational waves. Particularly, for each observing run, the mass distributions are chosen to be consistent with corresponding population analyses performed by the LVC such that the distributions used have are have support in regions of the mass space where the events being analyzed are found. For O2, we choose $p(m_1) \propto m_1^{-1.6}$, $m_1 \in (7.9M_\odot, 42M_\odot)$ and $p(q) \propto q^{6.7}$ where $q = \frac{m_2}{m_1}$ which is consistent with Ref. (Abbott et al., 2019a), and is identical to the mass distributions used for similar selection function computations (Payne et al., 2020). For O3 we

choose $p(m_1) \propto m_1^{-1.6}$, $m_1 \in (7M_\odot, 80M_\odot)$ and $p(q) \propto q^{6.7}$, which is broad enough for the O3 events as evident from Ref. (Abbott et al., 2023d). In the next two subsections, we describe the details of our smoothing approximations and the computation $\beta_{\rm det}$ in each approximation scheme.

Narrow Gaussian Method

The approach introduced here involves estimating the delta function in Eq. (7.18) as a narrow Gaussian distribution. For each sample with measured speed difference $\Delta v'$ and sky location θ and ϕ. We construct a Gaussian distribution for the random variable $\Delta v' - \Delta v(\bar{s}_{lm}, \theta, \phi)$ with mean zero and standard deviation σ. Thus, Eq. (7.19) becomes:

$$L(\bar{s}_{lm}) = \frac{1}{\beta^N_{\rm det}(\bar{s}_{lm})} \prod_{i\in\{\rm events\}}^{N} \int p(\Delta v', \theta, \phi|d_i) \times \mathcal{N}(\Delta v' - \Delta v(\bar{s}_{lm}, \theta, \phi)) d\Delta v' d\theta d\phi, \tag{7.24}$$

where $\mathcal{N}$ represents Gaussian distributions. Similarly, we can also apply the Narrow Gaussian approximation to the computation of $\beta_{\rm det}$ in Eq. (7.24):

$$\beta_{\rm det}(\bar{s}_{lm}) = \int f^{\rm marg}_{\rm det}(\Delta v', \theta, \phi) \mathcal{N}(\Delta v' - \Delta v(\bar{s}_{lm}, \theta, \phi)) \times d\Delta v' d\theta d\phi. \tag{7.25}$$

Since $\mathcal{N}$ is a smooth function of its arguments we can evaluate the two integrals in Eqns. (7.24) and (7.25) as a Monte Carlo sum over samples drawn from $p(\Delta v', \theta, \phi|d_i)$ and $f^{\rm marg}_{\rm det}(\Delta v', \theta, \phi)$ respectively. Since we already have posterior samples drawn from $p(\Delta v', \theta, \phi|d_i)$ for each event during the v_g inference described in Sec. 7.2.2, and since the samples drawn from $f^{\rm marg}_{\rm det}(\Delta v', \theta, \phi)$ are the parameters of simulated events that survive our selection criteria, we can compute the log-likelihood of $\bar{s}_{lm}$:

$$\ln L(\bar{s}_{lm}) = \sum_{i\in\{\rm events\}} \ln \frac{\sum_{\{j\}} \mathcal{N}(\Delta v'_j - \Delta v(\bar{s}_{lm}, \theta_j, \phi_j))}{\sum_k \mathcal{N}(\Delta v'_k - \Delta v(\bar{s}_{lm}, \theta_k, \phi_k)}, \tag{7.26}$$

where the sum in the numerator is over posterior samples corresponding to the ith event while the

one in the denominator is over detectable samples. After choosing a width σ for our Gaussian $\mathcal{N}$ appropriately, we can thus use Eq. (7.26) for fast evaluation of the log-likelihood $\ln L(\bar{s}_{lm})$ as a numerical function of the SME coefficients. Hence we can use MCMC algorithms to draw samples from $\ln L(\bar{s}_{lm})$ and interpret the quantile ranges of said samples as Bayesian credible intervals of the SME coefficients given GW data.

To determine the appropriate width of our Gaussian distribution σ, we consider the effect of varying its size. Because the Gaussian distribution is an estimation of the delta distribution, theoretically, as the size of σ decreases, the approximation should be more accurate. However, because we sample the log-likelihood with an MCMC algorithm, we encounter numerical difficulties when the σ is too small. Thus our choice of σ has to be tuned in accordance with how the MCMC is implemented numerically.

In the MCMC process, the walkers only make use of local information at each step. Thus, it is possible for walkers to be trapped inside of islands of high likelihood. This is what happens when σ is set too small. Since most samples have high likelihood around zero, walkers can explore freely the region near zero. However, at more peripheral locations in the parameter space, the peaks are usually scattered. Thus, when σ is too small, these peripheral samples form isolated islands of high likelihood. In this case, the walkers will not be able to explore these isolated islands, resulting in false small constraints. On the other hand, as σ gets larger, our approximation becomes less accurate and distributions are artificially broadened. Therefore, we aim to find a σ such that it is big enough for the walkers to explore the sample space fully and small enough such that it gives us useful results.

One way to determine the appropriate value of σ is by applying both the random draw method and the Narrow Gaussian method on the same set of data and comparing the results. We divide our list of events into subsets of nine events and apply both methods to each subset using various sizes of σ. The solid lines in Fig. 7.2 represent the average uncertainty of the resultant $\bar{s}_{lm}$'s against

σ for the nine O3 events from this paper with the smallest sky areas. The average uncertainty is calculated by taking the average of the absolute value of the upper and lower one-standard-deviation value for each $\bar{s}_{lm}$. For the same sets of events, the random draw method produces uncertainties on the order of 10^0-10^1, which corresponds to dashed lines in Fig. 7.2. To select a suitable σ, we use superimposed plots such as Fig. 7.2 to select a σ such that each uncertainty produced by the Narrow Gaussian method is marginally larger than its counterpart from the random draw method. In the example shown, $\sigma = 0.0005$ is a good choice because every solid line lies marginally above the corresponding dashed line in the same color, which means that the choice σ does not artificially tighten the constraints. We perform such analysis for every subset of events and produce a "good σ". We pick the largest of such "good σ's" as the final choice. Even though this final σ produces wider constraints for each subset of events as compared to the random draw method, because the Narrow Gaussian method incorporates information from more than 9 events, ideally we could still produce tighter constraints than the random draw method.

Another procedure for determining σ is to use information directly from the distribution of $\Delta v'$ samples for each event. One specific procedure is to sort the list of $\Delta v'$ samples and compute the average difference between adjacent values. Choosing σ as half this value produces results that align well with the prior method for the specific distributions tested. That is, the value of σ for a given event is half of the average of the differences between adjacent $\Delta v'$ samples for that event. This procedure is used in the MDC shown in Fig. 7.3, and has the advantage of not needing to construct plots to determine σ. However, there still is subjectivity in choosing what fraction of the average to use. Moreover, the distribution and quantity of samples will impact the value of σ. Both considerations will affect the final credible intervals for $\bar{s}_{lm}$.

This method performs much better than the SVD-assisted random draw in the MDC performed in Sec. 7.2.3.6. However, we note that both processes for choosing σ involve a significant amount of user-controlled fine-tuning and can potentially lead to an over/under-estimate of measurement

uncertainties of the SME coefficients. Due to these reasons, we do not choose this method for our final results. We instead choose a different smoothing approximation to the Bayesian method by means of Kernel Density Estimation (KDE) which can be shown to produce either equally or more accurate results, while requiring almost no user-controlled fine-tuning.

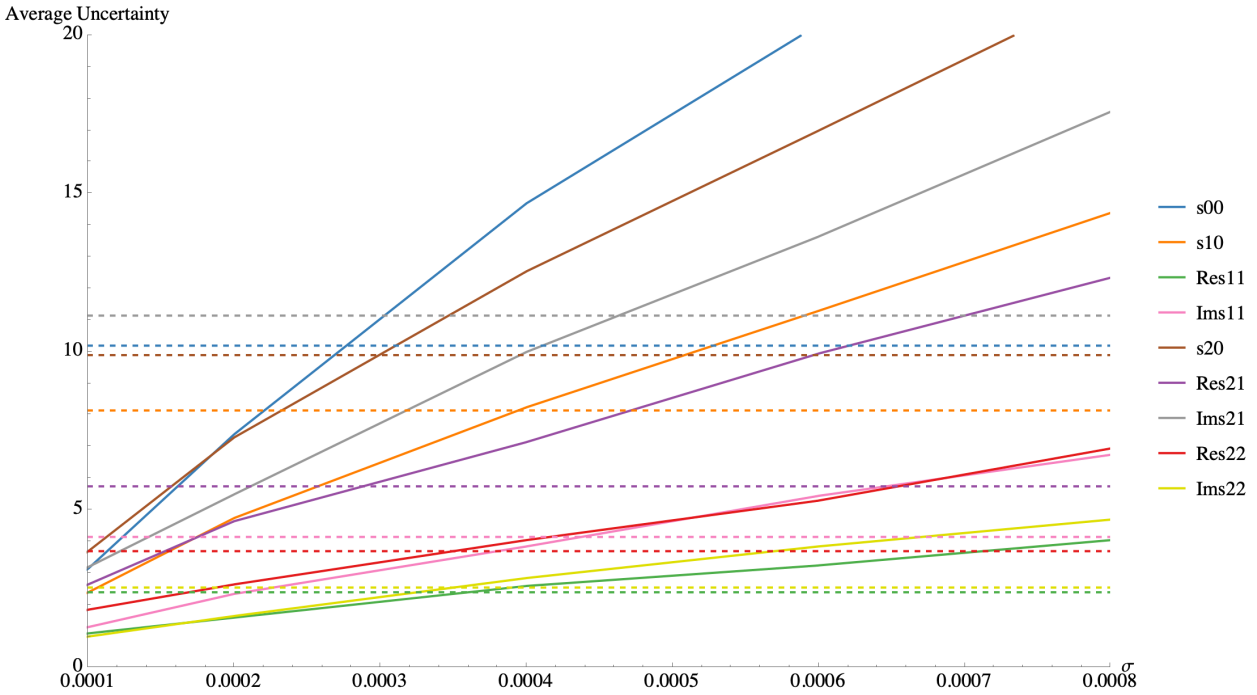


Figure 7.2: A portion of the superimposed plot of average uncertainties produced by the random draw method (dashed lines) and the Bayesian method (solid lines). From this plot, we can see $\sigma = 0.0005$ is a potential choice for σ because every solid line is marginally above the corresponding dashed line of the same color.

KDE Methods

In this section we outline a different approach from the one in the previous section, to perform Bayesian Hierarchical Inference of the SME coefficients from GW data. In this method, instead of smearing out the delta function in $\Pi(\Delta v', \theta, \phi|\bar{s}_{lm})$ with a Gaussian, we approximate the marginalized posterior of $\Delta v, \theta, \phi$ given GW data, for each event, as a fast evaluating function of these quantities, from their single event PE samples via Gaussian KDE.

The KDE approximation of the posterior is constructed by fitting a multi-variate Gaussian around each posterior sample and then writing the estimate of the posterior as a sum of these individual Gaussians. The covariance matrix of each of the Gaussians is approximated from the sample covariance matrix of the posterior samples themselves up to a constant of proportionality. The constant of proportionality is known as the bandwidth of the estimator and is computed, under reasonable assumptions regarding the true distribution being estimated(see (Scott, 1992)). We use SciPy's Gaussian KDE algorithm to obtain our estimate of the marginalized posterior as a fast evaluating function $p_{\mathrm{KDE},\,i}(\Delta v' = \Delta v(\theta, \phi, \bar{s}_{lm}), \theta, \phi)$ of the relevant parameters (Virtanen et al., 2020). We then perform the $\Delta v'$ integral of Eq. (7.18) analytically using the delta function and compute the remaining two integrals (over θ, ϕ) numerically using the trapezoidal rule. We loop over multiple events by multiplying the value of the integral obtained using the KDE corresponding to each event, to evaluate $L(\bar{s}_{lm})$:

$$L(\bar{s}_{lm}) \approx \prod_{i \epsilon \{events\}} \int p_{KDE,i}(\Delta v' = \Delta v(\theta, \phi, \bar{s}_{lm}), \theta, \phi) d\theta d\phi. \tag{7.27}$$

We then sample from it using the same MCMC method described in the previous section to constrain the SME coefficients. To incorporate selection effects in the KDE method, we estimate $f_{\mathrm{det}}^{\mathrm{marg}}(\Delta v', \theta, \phi | det)$ by performing a KDE on the samples of $(\Delta v', \theta, \phi)$ for which the simulated events are detectable given our detection criteria. By restricting our KDE to only these parameters and ignoring other parameters that characterize a simulated event, we effectively marginalize over those other parameters thus implicitly performing the integral in Eq. (7.23):

$$f_{\mathrm{det}}^{\mathrm{marg}} \approx p_{\mathrm{KDE},\,\mathrm{det}}(\Delta v', \theta, \phi), \tag{7.28}$$

where the subscript det in $p_{KDE,det}$ represents the fact that this KDE was performed on only those samples for which the simulated events are detectable given our detection criteria. Substituting

into Eq. (7.24) we get:

$$\beta_{\mathrm{det}}(\bar{s}_{lm}) \approx \int p_{\mathrm{KDE,\,det}}(\Delta v' = \Delta v(\theta, \phi, \bar{s}_l m), \theta, \phi) d\theta d\phi. \tag{7.29}$$

We note that the KDE's bandwidth acts like a control parameter with potential room for user-controlled fine-tuning in the computation of its value, somewhat analogous to the σ of the narrow Gaussian method. However, unlike the narrow Gaussian method where σ can in principle be chosen to be anything, the bandwidth of the KDE is computed directly from the properties of the samples (such as the number of samples and dimensionality of the parameter space) under reasonable assumptions regarding the true density. Thus the user's choice is restricted to a number of discrete such assumptions (for example Scott's rule (Scott, 1992), Silverman's rule (Silverman, 1986) etc.). Furthermore, the effects of choosing a bandwidth on the estimated density (and hence the remainder of the inference) is limited in the sense that the covariance matrix of the Gaussians is determined from the samples themselves with the bandwidth only acting as a scaling parameter that is usually of order unity. This additionally restricts the effects of user-controlled fine tuning on the inference as compared to the narrow Gaussian method wherein the width of the Gaussian that approximates the delta function is completely determined by the user's choice. A more detailed discussion of this comparison between the two methods in the context of the MDC can be found in Sec. 7.2.3.6 . For our chosen bandwidth approximation scheme (Scott's rule) the KDE method can be seen to perform extremely well in the MDC. For these reasons, we choose the KDE method for our final results on the SME constraints.

We present the results of the KDE method upon its use in the analysis of the events marked * listed on tables I through III (except for GW170817 for which we do not use the fixed posteriors due to the inability of the KDE to estimate very narrow densities) in Fig. 7.4 as our final result.

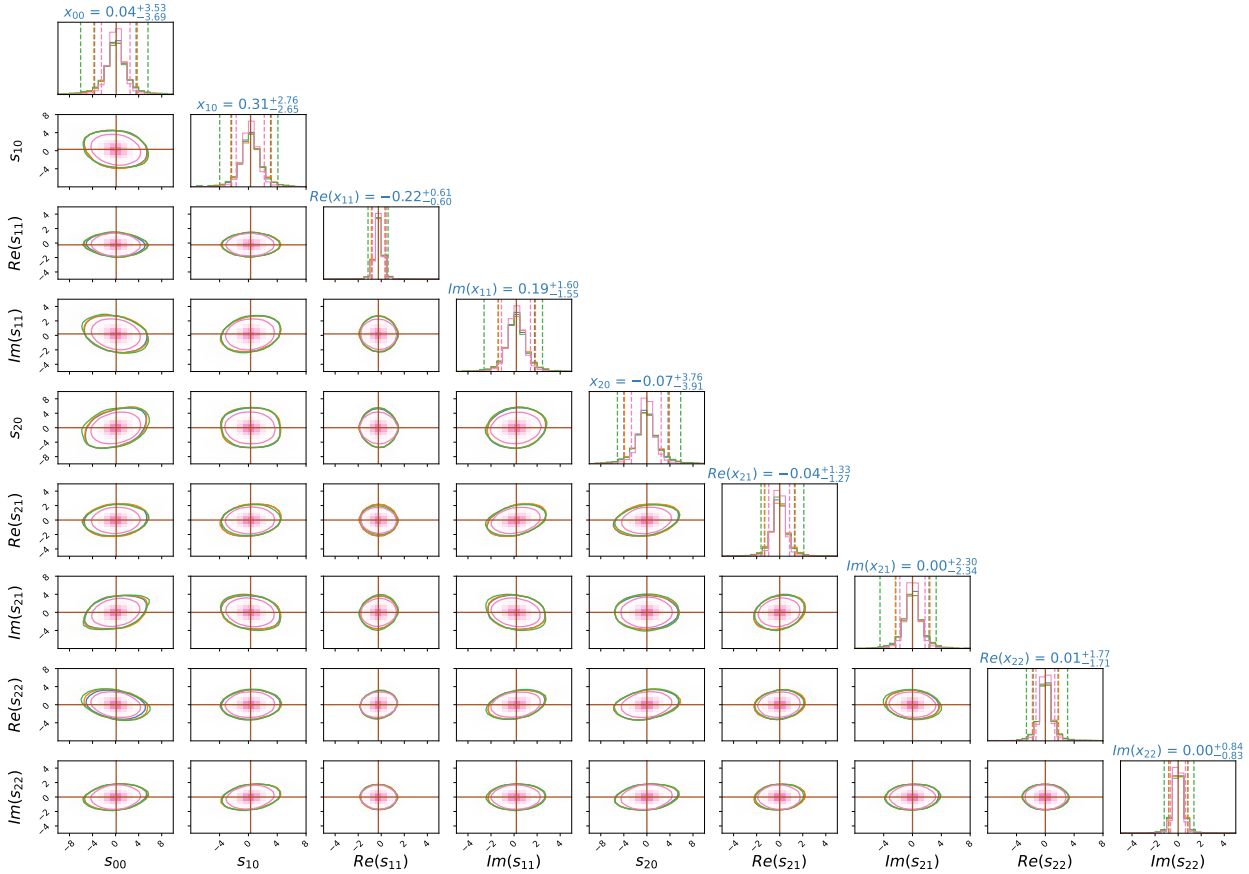


Figure 7.3: Distribution of all 9 $\bar{s}_{lm}$ for the Mock data. The Exact Bayesian, KDE, Narrow Gaussian and SVD chi-square methods are color-coded blue, orgnge, green, and magenta respectively. The mock true values of $\bar{s}_{lm}$ (which were chosen to be either equal to or very close to zero) are marked by orange lines. Numbers above the plots show median values with 90% equal tail credible intervals. The displayed constraints were obtained from the inferred $\bar{s}_{lm}$ samples using the python package. CORNER (Foreman-Mackey, 2016).

Note that these mock events are “zero noise" in the sense that the mean values of the Gaussian in Eq. (7.30) are chosen to be equal to the true values and not perturbed by another random draw.

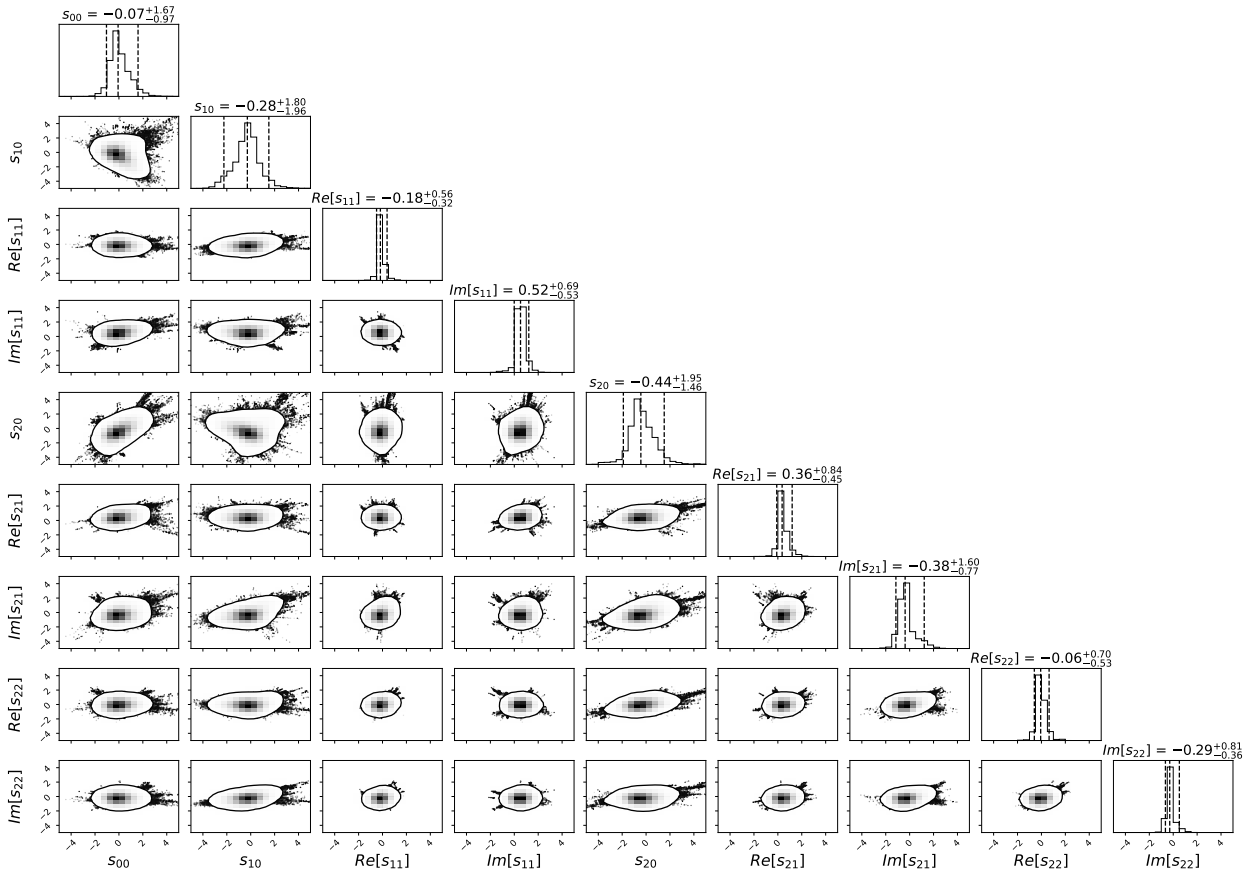


Figure 7.4: Distribution of all 9 $\bar{s}_{lm}$ using the KDE method from the 24 chosen GW events in Tables 7.2, 7.1, and 7.3. The numbers above the plots show median values with 90% equal tail credible intervals. The displayed constraints were obtained from the inferred $\bar{s}_{lm}$ samples using the python package CORNER (Foreman-Mackey, 2016).

7.2.3.5 Discussion

The random draw method, which was originally presented in (Liu et al., 2020), is very time efficient. However, the number of events that can be used in the analysis is limited by the number of $\bar{s}_{lm}$ to be estimated. Hence, it is only useful in the scenario where in we have exactly the same number of events available to be used in the analysis as the number of $\bar{s}_{lm}$ coefficients being simultaneously measured. On the other hand, Unlike any of the following methods, this method does not involve any approximation of the delta distribution. Thus, this method can be used as a quick feasibility test. For example, we used this method to estimate the size of σ for the narrow Gaussian approach.

With the use of SVD, we were able to take data from a larger set of events. However, the presence of one "bad" event, i.e. a low-significance event with biased posteriors, can disturb the entire analysis since all events are given equal weights. With the SVD chi-square method, this problem is solved by weighting each event with its uncertainty in the v_g measurement. However, this leads to artificially narrower bounds with ambiguity in the meaning of those bounds.

The Bayesian methods are free of all aforementioned pathologies that plague the other methods. They can efficiently handle a large number of events and is unaffected by a small number of bad events if any. Furthermore, the Bayesian credible intervals have the clear and unambiguous meaning of being regions of the parameter space that contain the true value of said parameters with a certain posterior probability given data.

Among the two approximate Bayesian techniques described in this paper, the Narrow Gaussian method has the following issue. The process to find σ is cumbersome and somewhat subjective as it involves partitioning the set of events into subsets and estimating σ from plots, or developing an algorithm that needs to be compared to an independent method. On the other hand, the KDE method has less user-controlled finetuning than the narrow Gaussian method as it estimates its control parameter, i.e. the bandwidth, quantitatively from properties of the posterior samples

themselves, with its variation having a much more restrictive effect on the estimated density. Hence we claim that the Bayesian analysis implemented by the KDE method produces the most trustworthy measurements of the SME coefficients.

7.2.3.6 Mock Data Challenge and Comparison of Methods

In this section, we describe the MDC that was set up to compare the different methods of SME measurements from GW data in order to verify our claims regarding them that were made in the previous section. To construct the MDC, we choose a fiducial value of the SME coefficients as their true values, say $\bar{s}_{lm,tr}$ and generate data for 15 mock events. The true sky positions of the mock events are chosen to be the mean values of the (θ, ϕ) samples of the real events.

We choose 15 of the "best" real events, i.e. the ones with the most precise sky localizations and v_g measurements, to be represented by our mock events in the MDC. We then calculate the true value of Δv for each mock event from the true values of their sky positions and those of the SME coefficients. We then generate mock posterior samples of $(\theta, \phi, \Delta v)$ by adding un-correlated Gaussian fluctuations to the true sky positions and true Δv's. The width of the fluctuations, for each mock event is chosen to be the standard deviations of the posterior samples of the corresponding real event.

This allows us to create a controlled numerical experiment wherein we know the true answer. For Gaussian distributions of $\theta, \phi, \Delta v$ about known true values, one can write down the exact functional form of the likelihood of these parameters given mock data:

$$\begin{aligned} L_{\rm mdc}(d_i|\Delta v', \theta, \phi) &= \frac{1}{(2\pi)^{3/2}\sigma_{\theta,i}\sigma_{\phi,i}\sigma_{\Delta v,i}} \exp -\frac{1}{2}\{\frac{(\theta-\theta_{tr,i})^2}{\sigma^2_{\theta,i}} \\ &\quad + \frac{(\phi-\phi_{tr,i})^2}{\sigma^2_{\phi,i}} + \frac{(\Delta v'-\Delta v_{tr,i})^2}{\sigma^2_{\Delta v,i}}\}, \end{aligned} \tag{7.30}$$

where $(\theta_{tr,i}, \phi_{tr,i})$ are the true values of the sky positions of the ith mock event, $\sigma_{\theta,i}, \sigma_{\phi,i}, \sigma_{\Delta v,i}$

are the widths of the Gaussian fluctuations used to generate the mock posterior samples of the ith mock event and $\Delta v_{tr,i} = \Delta v(\bar{s}_{lm,tr}, \theta_{tr,i}, \phi_{tr,i})$. Knowledge of these quantities allows us to exactly write down and evaluate Eq. (7.30) as a function of θ, ϕ without any smoothing approximations.

We can then substitute $L_{\rm mdc}(d_i|\Delta v', \theta, \phi)$ in place of $L(d_i|\Delta v', \theta, \phi)$ in Eq. (7.15) and carry out the integral numerically to obtain the "exact Bayesian" likelihood of our mock data given the SME coefficients:

$$\begin{aligned} L_{\rm mdc}(\bar{s}_{lm}) \quad = \quad & \prod_i \frac{1}{(2\pi)^{3/2}\sigma_{\theta,i}\sigma_{\phi,i}\sigma_{\Delta v,i}} \int \exp -\frac{1}{2}\{\frac{(\theta-\theta_{tr,i})^2}{\sigma^2_{\theta,i}} \\ & + \frac{(\phi-\phi_{tr,i})^2}{\sigma^2_{\phi,i}} \\ & + \frac{(\Delta v(\bar{s}_{lm},\theta,\phi) - \Delta v(\bar{s}_{lm,tr},\theta_{tr,i},\phi_{tr,i}))^2}{\sigma^2_{\Delta v,i}}\} d\theta d\phi. \end{aligned} \tag{7.31}$$

We can then sample the likelihood in Eq. (7.31), after applying suitable priors on $\bar{s}_{lm}$, using the MCMC techniques described above and obtain what can be thought of as the "true" posterior distribution of the SME coefficients given the mock data. We can then compute the constraints obtained from the approximate methods being applied to the mock posterior samples and compare those results with the true posterior.

The results of this comparison is displayed in Fig. 7.3. We can see that the KDE method agrees remarkably well with the exact Bayesian method, while the narrow Gaussian method deviates from it slightly for some coefficients. We note that a different choice of σ for the narrow Gaussian method leading to better agreement with the exact Bayesian result is possible. However, we conclude that the KDE method's agreement with the exact Bayesian method, independent of any external finetuning, justifies its use on the real data for producing our final result. We also note that our claim regarding the SVD chi-square method producing artificially narrower bounds is also verified by this comparison.

7.2.3.7 *Final Results*

With the 24 chosen GW events in Tables 7.2, 7.1, and 7.3, we are able to constrain all nine $\bar{s}_{lm}$ coefficients. We obtain the results shown in Fig. 7.4.

Note that the measurements of $\bar{s}_{lm}$ shown in Fig.7.4 are consistent with zero. Given that zero lies within the 90% credible intervals for all coefficients, we consider these results to be consistent with existing constraints on $\bar{s}_{lm}$ (Kostelecký & Russell, 2008).

These limits are considerably weaker than some found in the literature. However, they are valuable as independent tests. Moreover, this is also the first attempt to simultaneously constrain all $\bar{s}_{lm}$ using gravitational-wave measurements, thus putting direct limits on the full potential anisotropy of the speed of gravitational waves. Additionally, this method can theoretically incorporate as many events as available and thus improve in precision as additional high-quality events become available.

7.2.4 Conclusion

In our study, we select 52 high-SNR gravitational-wave events that were detected by at least two detectors from the first three observing runs of Advanced LIGO and Advanced Virgo. We use `LALINFERENCE_NEST` and `LALINFERENCE_MCMC` to construct posterior distributions of the speeds of GWs for each event. We find the 90% credible interval of the combined v_g posterior distribution to be $0.99^{+0.01}_{-0.02}c$. This interval is narrower than the similar one constructed with O1 and O2 events in previous studies, suggesting a more precise measurement of v_g (Liu et al., 2020). However, even with the inclusion of a high-SNR BNS event GW170817 with its pinpoint sky localization, we were only able to narrow the 90% credible interval to $0.99^{+0.01}_{-0.02}c$. We then explore multiple methods of extracting SME constraints from v_g-like data. Based on the conclusions of that investigation, we use hierarchical Bayesian inference implemented with KDE methods to simultaneously constrain all nine coefficients for Lorentz violation in the SME framework. The resultant

constraints do not exhibit evidence for Lorentz violation. We are optimistic about the possibility of further improvements in speed of gravity and associated Lorentz violation measurements in the future. The search is likely to be aided by the combination of additional detectors at additional locations around the Earth and by the possibility of combining results achieved by the methods presented here with those from additional multimessenger events.

7.3 SUMMARY

In this chapter, I summarized my contribution to the development of a robust and fully Bayesian analysis method that can constrain the Lorentz violating GW propagation in the presence of selection effects, a first-of-its-kind framework. We have obtained the first simultaneous measurements of all 9 coefficients of non-birefringent and non-dispersive yet Lorentz violating theories of gravitation from GW data alone. In addition, we have compared our new method with existing ones and demonstrated why this method should be used over others for constraining SME coefficients in general. The results of our investigation are communicated through the prepint Ray et al. (2023b) which is currently submitted to a journal.

Part V

Conclusions

Chapter 8

Summary and Future Work

Nothing beside remains. Round the decay
Of that colossal Wreck, boundless and bare
The lone and level sands stretch far away.

Percy Bysshe Shelley

In this dissertation, we have explored several topics in gravitational wave science, namely, astrophysics, cosmology, and the nature of nuclear and gravitational interactions. In this chapter, I will provide a brief summary of the main investigations conducted, the results found, and the planned follow-up explorations.

8.1 TRIGGER CLASSIFICATION AND RATE ESTIMATION

In chapter 3, I summarized the development of a robust trigger classification and rate-estimation framework that plays three highly crucial roles in GW science and astronomy, which are listed as follows.

(1) Classify triggers into astrophysical and terrestrial categories at the end of an observing run to construct GW catalogs and decide whether or not to carry out costly PE runs on particular data segments.

(2) Classify astrophysical triggers into multiple CBC classes in real time, amidst an ongoing observing run, to optimize the search for multi-messenger counterparts.

(3) Estimate the astrophysical merger rate of rarely observed CBCs while accounting for the contributions of marginal candidates in the data.

For (1), (2), and (3), I replaced the pre-existing classifiers which were based on found injections, with semi-analytical template weights that prevented underestimation of rates and $\mathrm{P}_{\mathrm{astro}}$ values while also increasing classification accuracy. For (2), I have implemented further improvements by employing a rate approximation scheme that accounts for sensitivity changes between observing runs and thereby prevents underestimation of $\mathrm{P}_{\mathrm{astro}}$ values in online trigger classification. In addition, I have contributed to developing a pipeline for updating the search-based classification probabilities of (2) using `rapid-PE` based mass estimates, which has led to a significant increase in the accuracy of real-time GW trigger classification. These calculations can be improved even further by means of two major developments which are planned for the near future, and are summarized as follows.

8.1.1 Improving the template weights calculation

The template weight calculations of Fong (2018) and the one presented in chapter 3 section 3.2.2.3, and appendix A.2, are approximate in the sense that they do not model the noise-induced template migration realistically enough. In particular, they do not account for the marginalization of unknown signal phase which makes the distribution of the matched template parameters different from a Gaussian, and assume the template bank to be a euclidean surface in parameter space which it is not (Hanna et al., 2023). Furthermore, they classify CBCs by applying a source frame population distribution on the detector frame template parameters. All three of these approximations can be relaxed to obtain more realistic template weights.

8.1.1.1 Accointing for two-phased templates and non-Euclidean bank geometries

A GW signal is actually a superposition of two time series, one corresponding to each polarization, that approximately differ by only a phase term. The unknown signal phase can be marginalized analytically during matched filtering, which needs to be correctly implemented in the semi-analytic

template weight calculation. We start by considering two-phased templates and signals:

$$\text{signal: } \vec{s} = (\rho \cos\phi_0 \vec{t_j}, \rho \sin\phi_0 \vec{t_j}) \tag{8.1}$$

$$\text{recovered template: } \vec{d} = (\rho_{o,k} \cos\phi \vec{t_k}, \rho_{o,k} \sin\phi \vec{t_k}) = \vec{s} + \vec{n} \tag{8.2}$$

where $\vec{n} = (\vec{n_1}, \vec{n_2})$ and each $\vec{n}_i$ are large-dimensional independent Gaussians (the bank is whitened). Let θ_k be the signal parameters corresponding to template t_k. The distribution of matched template parameters now takes the following form:

$$p(\rho_{o,k}, \vec{t_k}, \phi|\rho, \vec{t_j}, \phi_0) \propto e^{-\frac{1}{2}\{(\rho_{o,k}\cos\phi\vec{t_k} - \rho\cos\phi_0\vec{t_j})^2 + (\rho_{o,k}\sin\phi\vec{t_k} - \rho\sin\phi_0\vec{t_j})^2\}} \tag{8.3}$$

$$\propto e^{-\frac{1}{2}\{\rho_{o,k}^2 + \rho^2 - 2\rho\rho_{o,k}\vec{t_j}\cdot\vec{t_k}\cos(\phi-\phi_0)\}}. \tag{8.4}$$

Marginalizing over the unknown signal phase, we get:

$$p(\rho_{o,k}, \vec{t_k}|\rho, \vec{t_j}, \phi_0) = \int_0^{2\pi} d\phi e^{-\frac{1}{2}\{\rho_{o,k}^2 + \rho^2 - 2\rho\rho_{o,k}\vec{t_j}\cdot\vec{t_k}\cos(\phi-\phi_0)\}} = e^{-\frac{1}{2}(\rho^2+\rho_{o,k}^2)} \mathrm{I}_0(\rho\rho_{\mathrm{o,k}}\tilde{\mathrm{t}}_{\mathrm{j}} \cdot \tilde{\mathrm{t}}_{\mathrm{k}}) \tag{8.5}$$

Where I_0 is the modified Bessel function of the first kind of order 0. Following Fong (2018), we compute $P(\text{candidate in } \vec{t_k}|\rho, \vec{t_j})$ as the integral of $p(\rho_{o,k}, \vec{t_k}|\rho, \vec{t_j})$ over an N+1 dimensional *conical* volume around the kth template which resides in an N-parameter template bank. (For example, an m_1, m_2, χ_{eff} bank would be N=3 dimensional). Note that the bank has curvature and hence the angular volume of the cone is taken to be the volume element of the bank in mismatch space at the location of the kth template: $\Delta\theta_k$. This is different from Fong (2018)'s approach since they took the volume element to be constant. The probability of template migration then becomes:

$$P(\text{candidate in } \vec{t_k}|\rho, \vec{t_j}) \propto \Delta\theta_k \int_0^\infty \rho_{o,k}^N d\rho_{o,k} e^{-\frac{1}{2}(\rho^2+\rho_{o,k}^2)} \mathrm{I}_0(\rho\rho_{\mathrm{o,k}}\tilde{\mathrm{t}}_{\mathrm{j}} \cdot \tilde{\mathrm{t}}_{\mathrm{k}}). \tag{8.6}$$

The integral in this equation can be carried out analytically by expanding the modified Bessel function of the first kind as an infinite series: $I_0(x) = \sum_n \frac{x^{2n}}{(n!)^2 4^n}$. Each term in this series can be integrated using $\int_0^\infty x^m e^{-\frac{1}{2}x^2} dx = \Gamma(\frac{m+1}{2})$. Once the integrations are carried out we get:

$$P(\text{candidate in } \vec{t_k}|\rho, \vec{t_j}) \propto \Delta\theta_k 2^{\frac{N-1}{2}} \Gamma\left(\tfrac{N+1}{2}\right) e^{-\frac{1}{2}\rho^2} \sum_{n=0}^{\infty} \frac{\left(\frac{N+1}{2}\right)^{(n)} \times \left(\frac{(\rho\vec{t_k}\cdot\vec{t_j})^2}{2}\right)^n}{1^{(n)} n!} \tag{8.7}$$

Where $a^{(n)} = \frac{\Gamma(a+n)}{\Gamma(a)}$ is the rising factorial (note, $1^{(n)} = n!$). This series is the expansion of Kummer's confluent Hypergeometric function of the first kind. In other words,

$$P(\text{candidate in } \vec{t_k}|\rho, \vec{t_j}) \quad \propto \quad 2^{\frac{N-1}{2}} \Delta\theta_k \Gamma(\tfrac{1+N}{2}) e^{-\frac{1}{2}\rho^2} {}_1\mathrm{F}_1\left(\tfrac{1+\mathrm{N}}{2}, 1, \tfrac{(\rho\tilde{\mathrm{t}}_\mathrm{k}\cdot\tilde{\mathrm{t}}_\mathrm{j})^2}{2}\right), \tag{8.8}$$

where ${}_1\mathrm{F}_1$ is Kumer's confluent hypergeometric function of the first kind. This expression for noise-induced template-migration is expected to be more accurate than both Fong (2018)'s approach and the one presented in appendix A.2. Further testing is necessary to fully establish statements about its superior accuracy which is part of an ongoing investigation.

8.1.1.2 Source frame population model

Now that we have $P(\text{candidate in }, t_k|\rho, t_j)$, the template weights:

$$P(t_k, \rho|\mathrm{pop}, \Omega) = \sum_{\mathrm{j}} \mathrm{P}(\text{candidate in }, \mathrm{t_k}|\rho, \mathrm{t_j})\mathrm{p}(\theta_\mathrm{j}, \rho|\mathrm{pop}, \Omega)\mathrm{V_j} \tag{8.9}$$

can be constructed once the source-frame population model

$$p(\theta_j, \rho|\mathrm{pop}, \Omega) = \mathrm{p}(\chi_{\mathrm{eff,j}}, \mathrm{m}^{\mathrm{d}}_{1\mathrm{j}}, \mathrm{m}^{\mathrm{d}}_{2\mathrm{j}}, \rho|\mathrm{pop}, \Omega) \tag{8.10}$$

is correctly applied to detector frame template parameters $\theta_j = (\chi_{eff,j}, m^d_{1j}, m^d_{2j})$. This is achieved by converting detector frame parameters to source frame using a fixed cosmological model (Ω) and the horizon distance of each template at SNR=1 (D^1_H), such as:

$$z(m^d_{1j}, m^d_{2j}, \rho) = F_z(\frac{D^1_H(m^d_{1j}, m^d_{2j})}{\rho\langle\Theta\rangle}, \Omega) = F_z(\frac{D^1_H(m^d_{1j}, m^d_{2j})}{2.26\rho}, \Omega), \tag{8.11}$$

$$m^s_1(m^d_{1j}, m^d_{2j}, \rho) = \frac{m^d_{1j}}{1+z(m^d_{1j}, m^d_{2j}, \rho)}, \tag{8.12}$$

and

$$m^s_2(m^d_{1j}, m^d_{2j}, \rho) = \frac{m^d_{2j}}{1+z(m^d_{1j}, m^d_{2j}, \rho)}, \tag{8.13}$$

with these transformations introducing the following jacobian in the probability distribution:

$$J = \left|\frac{\partial(m^s_1, m^s_2, z)}{\partial(m^d_{1j}, m^d_{2j}, \rho)}\right| = \left[(1+z)^2 \frac{1}{D_L(z)} \frac{\partial D_L}{\partial z}\right]^{-1}_{z=F_z(\frac{D^1_H(m^d_{1j}, m^d_{2j})}{2.26\rho}, \Omega)}. \tag{8.14}$$

Here, F_z is redshift as a function of distance and cosmology, and $\langle\Theta\rangle$ is the average value of the ratio of effective distance to luminosity distance for a single detector (Finn & Chernoff, 1993). Putting everything together, we get the following expression for the source frame population model:

$$\begin{aligned} p(\theta_j, \rho|\mathrm{pop}, \Omega) = \Bigg[& p_{\mathrm{source}}(\chi_{eff,j}|\tfrac{m^d_{1j}}{1+z}, \tfrac{m^d_{2j}}{1+z}, pop) p_{\mathrm{source}}(\tfrac{m^d_{1j}}{1+z}, \tfrac{m^d_{2j}}{1+z}, z|pop) \\ & \times \left\{(1+z)^2 \frac{1}{D_L(z)} \frac{\partial D_L}{\partial z}\right\}^{-1} \Bigg]_{z(m^d_{1j}, m^d_{2j}, \rho)}. \end{aligned} \tag{8.15}$$

This expression represents the correct way of applying a source frame population model to detector-frame quantities. As with the improvements described in the previous subsection, testing this new template weights calculation is part of a planned investigation.

8.1.2 Distance re-construction in `rapid-PE`

Similar to the case of search-pipeline template-weights, any mass-based classifier is supposed to be maximally accurate when the source frame mass-population is correctly applied on detector-frame observables, i.e. after converting the latter into source-frame quantities using some cosmological model. In the context of the `rapie-PE`-based classification scheme, this can in principle be achieved by reconstructing posterior samples of the luminosity distance from the marginalized likelihood at each intrinsic grid point.

For analytically marginalized likelihoods, the likelihood as a function of distance corresponding to a particular value of the intrinsic and other extrinsic parameters can be constructed from the matched filter and optimal SNR at said value (Thrane & Talbot, 2019). For `rapid-PE` which implements distance and phase marginalization analytically and marginalization over other extrinsic parameters stochastically, the marginalized likelihood as a function of distance at the ith intrinsic grid-point can be constructed by:

$$\mathcal{L}_i(D_L|d) \propto \sum_q w_q \int dt I_0(\kappa_{i,q}(t)\frac{D_{ref}}{D_L}) e^{-\frac{1}{2}\rho^2_{i,q}\frac{D^2_{ref}}{D^2_L}+\kappa_{i,q}(t)\frac{D_{ref}}{D_L}}, \tag{8.16}$$

where $\kappa_{i,q}(t)$ is the matched-filter SNR time series and $\rho_{i,q}$ the optimal SNR, both evaluated at the ith intrinsic grid point, the qth extrinsic sample and a reference distance D_{ref}, w_q is the ratio of the prior on the extrinsic parameters to the distribution from which their samples were drawn to carry out the Monte Carlo marginalization, and I_0 is the modified Bessel function of the first kind and order 0. Note that κ and ρ^2 are the first and second sums in Eq (24) of Pankow et al. (2015), which are computed by `rapid-PE` and can in principle be stored to disk. In the absence of analytic phase marginalization, as is implemented in the current low-latency configuration of `rapid-PE`,

the above likelihood reduces to the following simpler form:

$$\mathcal{L}_i(D_L|d) \propto \sum_q w_q \int dt e^{-\frac{1}{2}\rho_{i,q}^2 \frac{D_{ref}^2}{D_L^2} + \kappa_{i,q}(t)\frac{D_{ref}}{D_L}}, \tag{8.17}$$

with the only difference being that the signal phase is now one of the components being marginalized over stochastically leading to a change in w_q. Using this likelihood, one can attach a distance grid at each intrinsic grid point with the likelihood as a function of masses and distance now known everywhere on the three-dimensional grid. Generalizing `rapid-PE`'s current Gaussian interpolation to three dimensions, the likelihood of Eq. (3.39) becomes:

$$p(d|m_1, m_2, d_L) = \sum_i^{N_{grid}} \bar{\mathcal{L}}_i(d) \exp\left\{-\frac{1}{2\sigma^2_{\tau_1,i}}[\tau_1(m_1,m_2) - \tau_{1,i}]^2 - \frac{1}{2\sigma^2_{\tau_2,i}}[\tau_2(m_1,m_2) - \tau_{2,i}]^2 \right.$$
$$\left. -\frac{1}{2\sigma^2_{D_L,i}}[D_L - D_{L,i}]^2\right\}, \tag{8.18}$$

where $\bar{\mathcal{L}}_i$ is calculable from either Eq. (8.16) or Eq. (8.17). Using this likelihood and the importance sampling scheme described before, the category specific evidence of Eq. (3.40) becomes:

$$Z_a = \frac{1}{N_s}\sum_i^{N_{grid}}\sum_j^{N_s}(2\pi)^{3/2}\sigma_{\tau_1,i},\sigma_{\tau_2,i}\sigma_{D_{L,i}}\bar{\mathcal{L}}_i(d)\left[\frac{p(\frac{m_1(\tau_1^{ij},\tau_2^{ij})}{1+z},\frac{m_2(\tau_1^{ij},\tau_2^{ij})}{1+z},z|\mathcal{H}_\alpha)J^{ij}}{\frac{dD_L}{dz}\frac{1}{(1+z)^2}}\right]_{z=f_z(D_L^{ij},\Omega)}, \tag{8.19}$$

where $p(m_1, m_2, z|\mathcal{H}_\alpha)$ is the population distribution of source-frame masses and redshifts for the α'th CBC category,

$$(\tau_1^{ij}, \tau_2^{ij}, D_L^{ij}) \sim \frac{1}{(2\pi)^{3/2}\sigma_{\tau_1,i}\sigma_{\tau_2,i}\sigma_{D_L,i}} \exp\left\{-\frac{1}{2\sigma^2_{\tau_1,i}}[\tau_1(m_1,m_2) - \tau_{1,i}]^2 - \frac{1}{2\sigma^2_{\tau_2,i}}[\tau_2(m_1,m_2) - \tau_{2,i}]^2 \right.$$
$$\left. -\frac{1}{2\sigma^2_{D_L,i}}[D_L - D_{L,i}]^2\right\} \tag{8.20}$$

are samples drawn from the Gaussians at the i'th grid, and J^{ij} is the same as Eq. (3.42).

Upon implementation, this scheme can be expected to further increase the accuracy of the `rapid-PE` based source classification. A detailed investigation regarding the same is planned for the near future.

8.2 ASTROPHYSICAL AND COSMOLOGICAL INFERENCE USING DATA-DRIVEN POPULATION MODELS

In chapter 4, I described a data-driven population inference method for GW sources that can elucidate the mysteries of compact binary formation and the cosmological expansion beyond the limitations of existing parametric and non-parametric approaches. The developed population model is easily scalable to higher dimensions making it a unique non-parametric probe of correlations in the astrophysical compact binary population. It can further be used to explore the cosmic expansion history using GW data alone while being free of the uncertainties regarding the mechanisms of compact binary formation. Using this model, I attempted to answer the following key questions regarding BBH formation and cosmological expansion:

(1) Can we constrain the astrophysical processes that are predicted to cause a redshift evolution of the BBH mass-spectrum?

(2) Are their imprints of multiple formation channels in the observable BBH population? If so, can they tell us about the origin of the $35M_{\odot}$ peak in the astrophysical BBH mass-spectrum?

(3) Can we obtain independent measurements of the Hubble parameter from GW data alone, while being agnostic of the underlying astrophysics of compact binary formation?

In the context of (1), analyzing GWTC-3 data using our newly developed model led to the first non-parametric constraint on the redshift evolution of the BBH mass spectrum. We found no evidence of such a mass-redshift correlation in contemporary data due to large measurement

uncertainties. We found that data from more events is necessary to make confident statements regarding such correlations.

In the context of (2), our analysis revealed hints that the data favors a dynamical origin of the $35M_\odot$ feature. In particular, we found that the BBH mass spectrum, when conditioned on different ranges of the effective inspiral spin, only displays a sharp peak near $30 - 40M_\odot$ for very small $(0.05 - 0.1)$ effective spin values. Furthermore, we found that BBHs with at least one component in the $30 - 40M_\odot$ range have a more symmetric effective spin distribution about zero than all other BBHs. We interpreted these results as hints of a dynamically formed subpopulation of BBHs since only dynamical formation in globular clusters can explain an excess near $30M_\odot$ as well as symmetric effective spin distributions.

In the investigations of (3), I obtained the first astrophysics agnostic measurements of the Hubble parameter using GW data alone. Using our non-parametric model for the CBC mass and redshift distributions, we were able to constrain the cosmological parameters simultaneously with the shapes of these distributions, thereby marginalizing over our uncertainties regarding compact binary formation. In the O5 and 3G era, our method will enable precise and unbiased cosmological measurements, potentially elucidating the Hubble tension and the nature of dark energy.

The immediate follow-ups to all three of these investigations involve re-running the corresponding analyses on O4a data which I have already started doing. This is particularly exciting for (2) since the hints we have found from GWTC-3 will either be solidified or further reduced in significance, both conclusions being astrophysically impactful. However, the BGP method itself needs further developments and generalizations, to go beyond its current limitations and reveal deeper insights into the astrophysics of compact binary formation and the nature of the cosmological expansion.

8.2.1 Exploring higher dimensional correlations in the BBH population

As mentioned before, the population-level distribution of BBH spins and its conditional dependence on masses and redshifts is a powerful tracer of the evolutionary pathways of BBH progenitor systems and their environments. However, the astrophysical implications of the inferred spin distribution are particularly sensitive to modeling choices with different models, for example, of the spin magnitude and tilt distributions yielding conflicting conclusions regarding the relative abundance of isolated and dynamically formed BBH sub-populations (Callister et al., 2022). On the other hand, Biscoveanu et al. (2022) found that population models designed to expect only a mass-spin correlation can falsely infer its existence from a true population with underlying spin-redshift correlations. Similarly, the conclusions of Biscoveanu et al. (2022) might themselves be biased given the fact that they assume the marginal mass-redshift distributions to be independent.

For fully probing the mysteries of BBH formation, it is therefore necessary to obtain a generic and fully data-driven characterization of the joint distribution of BBH masses, component spins, and redshifts. At the very least, a generic model should be able to characterize the joint distributions of component masses, redshift, and the best-measured combinations of spin parameters such as effective inspiral and effective precessing spins. This can be achieved through 4 and 5-dimensional generalizations of our BGP model which is a major planned development for this analysis method that I aim to pursue in the near future.

8.2.2 Optimizing the complexity of flexible population models

Pre-selecting the number and location of the bin edges is the only remaining source of user-controlled fine-tuning in the BGP analysis method. Even though we have demonstrated the robustness of results against variations in binning choices once a certain level of resolution is achieved in the inferred distributions, the current methodology for heuristically reaching said resolution is sub-optimal. Unbinned GP-based population inference methods like that of Callister & Farr (2024)

are free of any such fine-tuning and therefore represent fully non-parametric approaches towards population inference. However, the lack of scalability of such methods to higher dimensional population distributions justifies the necessity to bin up the CBC parameter space when attempting to constrain correlations in addition to the marginal distributions. The ideal approach towards data-driven characterizations of the joint distribution of CBC parameters is therefore to implement a BGP inference that can self-consistently optimize the complexity of the binned model.

In other words, we intend to vary both the number and location of bin-edges by treating them as hyper-parameters and imposing suitable priors on them. Varying the bin locations for a fixed number of bins is more straightforward in the sense that the dimensionality of the hyperparameter space remains unchanged throughout the inference and hence can be implemented within the stochastic sampling framework currently used to explore the hyper-posterior distribution. On the other hand, varying the number of bins leads to additional complications and requires more sophisticated sampling techniques such as reversible jump Markov Chain Monte Carlo (Green, 1995; Toubiana et al., 2023), which can simultaneously optimize the complexity of the model being inferred. Implementing this fully non-parametric BGP inference that simultaneously infers the optimal number and location of bins with the rate-density in each bin, is another major development that I intend to pursue in the future.

8.2.3 Accounting for Background contamination

Growing GW catalogs will contain an increasing number of marginally significant CBC events which will lead to a low-level contamination of the astrophysical and cosmological inferences by potential false-positives (Galaudage et al., 2020). This effect will become particularly important in the case of 3G detectors. In order to account for potential biases resulting from this contamination, the hierarchical inference of population and cosmology from GW data needs to be modified so that each trigger is treated as a realization of mixed Poisson processes, one of which represents the

occurrence of terrestrial events above the threshold used to select GW candidates for participation in mentioned analyses (Gaebel et al., 2019; Galaudage et al., 2020).

This is achieved by inferring the BGP model using the general hierarchical likelihood of Eq. (2.42), instead of the significant catalog likelihood of Eq. (2.49). In particular, the likelihood of the Eq. (2.42), when evaluated for the binned population model, takes the following form:

$$p(\vec{d}|\vec{n}, N_{B,det}) \propto N_{B,det}^{N} e^{-N_{B,det}-\sum_{\gamma} n^{\gamma}\langle VT\rangle^{\gamma}} \prod_{i} \left(1 + \frac{P_{astro,i}(N_{F,0})}{1-P_{astro,i}(N_{F,0})} \frac{1}{N_{F,0}} \sum_{\gamma} n^{\gamma} w^{\gamma}(d_i)\right), \quad (8.21)$$

where $P_{astro,i}$ is the search-pipeline's $\mathrm{P_{astro}}$ for the i'th event, and $N_{F,0}$ is the astrophysical count corresponding to the fixed population distribution used to compute search $\mathrm{P_{astro}}$ values. Using this likelihood, a GP prior on $\vec{n}$, and some broad prior on $N_{B,det}$, it is possible to constrain the population distribution of CBC parameters from a dataset like the extended one from GWTC-3, while self-consistently accounting for potential contamination by marginal candidates. Such an investigation is currently ongoing with the results shortly imminent. This framework can also be extended to the BGP-based spectral siren method of cosmological inference, which constitutes another planned investigation.

8.2.4 Correlated models for cosmological inference

While all of the improvements in the BGP-based population model and inference scheme discussed so far are beneficial to both its astrophysical and cosmological applications, there are a couple of cosmology-specific developments I intend to pursue in the future. For example, a particular caveat of the non-parametric spectral-siren siren approach presented in chapter 5 is the use of an uncorrelated mass-redshift distribution. Even though we have argued why such an assumption is good enough for contemporary datasets, Ezquiaga & Holz (2022) and Pierra et al. (2023) have shown that for large catalogs, incorrect assumptions regarding mass-redshift correlations in the CBC population can significantly bias the cosmological measurements. The immediate generalization to the

spectral siren methods of chapter 5 would therefore be to use the correlated BGP model of chapter 4, section 4.3.2.1, in place of the uncorrelated one currently implemented. Of particular interest is a study of the systematic uncertainties in cosmological measurements that appear with growing catalogs. Such investigations are also planned for the near future.

8.2.5 Probing beyond the standard model of cosmology

The spectral-siren methods based on both parametric and non-parametric population models, traditionally attempt to constrain the parameters that characterize the standard model of cosmological expansion by relying on GW-based distance measurements and features in the CBC mass-spectrum. These investigations are therefore restricted to probing the cosmic expansion history only within the assumptions of the standard model. Such a restriction can in principle be removed from the hierarchical inference framework of the spectral-siren methodology, thereby constructing a spectral-siren-based probe of the standard model of cosmology itself. This can be achieved in one of the following two ways.

The BGP-based spectral siren methodology can be generalized beyond the standard model by constructing a new four-dimensional BGP over CBC masses, redshifts, and luminosity distances. While this approach is agnostic of the astrophysics of compact binary formation and the assumptions built into the standard model of cosmology, one can presume that a model of such high complexity cannot be constrained informatively without a 3G-sized CBC catalog. For contemporary datasets, one might instead restrict to a parametric population model and construct a BGP or even an unbinned GP for the redshift-luminosity distance functional relation. Constraining such a model from GW catalogs will amount to probing the cosmological expansion beyond the limitations of the standard model. Developing both of these algorithms and testing them on real and simulated data constitutes another planned investigation I aim to pursue in the future.

8.3 EOS INFERENCE FROM GW DATA

In chapter 6, I delineated the development of a scalable EoS inference pipeline that can tractably constrain the nuclear EoS from multiple BNS and NSBH observations while being free of the problem of dimensional reduction. As more such detections are made, our method will play a vital role in pinning down the universal properties of degenerate nuclear matter and hence in elucidating the mysterious nature of low-energy strong nuclear interactions. However. there are several generalizations to the developed inference framework, which can be implemented to exploit the full scientific potential of astrophysical NS observations. I list some of them as follows.

8.3.1 Generalizing to multi-probe datasets

As mentioned before, the universality of the EoS can be leveraged to combine data from multiple observations across different probes such as the LVK, NICER, and radio telescopes to obtain more precise measurements than what either probe could have offered individually. The KDE-based inference framework can be straightforwardly generalized to multiprobe datasets by modifying the likelihoods of Eqs. (6.23), and (6.29) in the following way:

$$p(\vec{\gamma}|\{d\}_{GW},\{d\}_{XR},\{d\}_{R},I) \propto p(\vec{\gamma},\mathbf{I})\left(\prod_{i=1}^{N_{GW}}\int_0^1 K_i^{GW}(q,\Lambda_1(\bar{\mathcal{M}}_i,q,\vec{\gamma}),\Lambda_2(\bar{\mathcal{M}}_i,q,\vec{\gamma}))dq\right)\times$$
$$\left(\prod_j^{N_{XR}}\int_{m_{min}(\vec{\gamma})}^{m_{max}(\vec{\gamma})} dmK_j^{XR}(m,r(m,\vec{\gamma}))\right)\times\left(\prod_l^{N_R}\int_{m_{min}(\vec{\gamma})}^{m_{max}(\vec{\gamma})} dmK_l^{R}(m)\right), \tag{8.22}$$

where K_i^{GW} represents a KDE of the mass-ratio and tidal deformability posterior given data from the i'th GW event, K_j^{XR} represents a KDE of the mass and radius posterior given data from the j'th Pulsar X-ray observation by NICER, K_l^R represents the mass posterior of the l'th pulsar radio observation and $m_{min,max}(\vec{\gamma})$ represent the minimum and maximum masses of the EoS family determined by $\vec{\gamma}$ respectively. Note that K_j^{XR} is computed from mass-radius posterior samples

that have been re-weighted to a uniform radius prior.

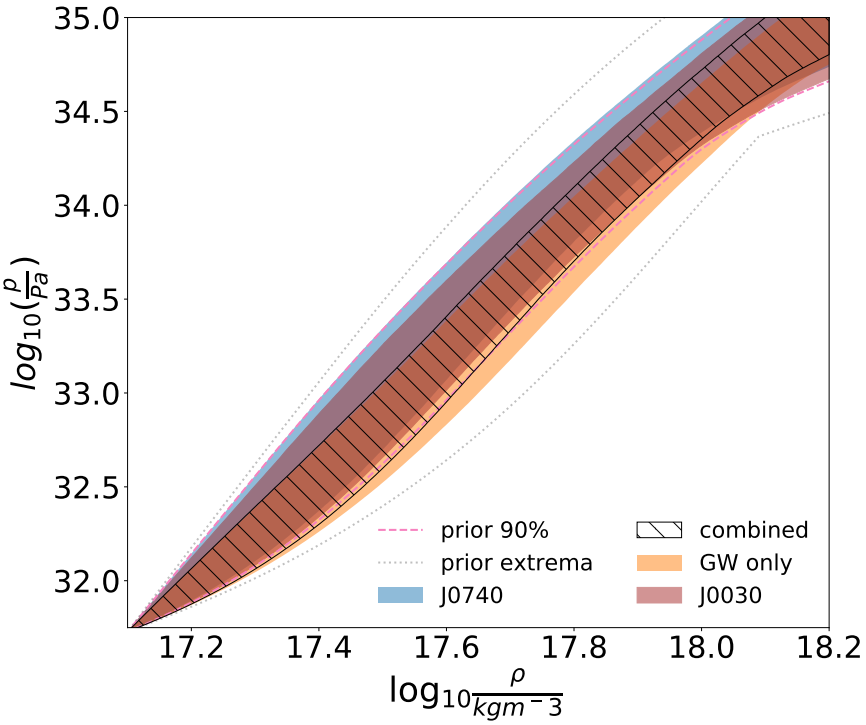


Figure 8.1: Extending our method to pulsar X-ray observations.

A preliminary result obtained by constructing the posterior distribution of Eq. (8.22) from the data corresponding to the two BNS mergers observed to date (Abbott et al., 2017c, 2020b) and the NICER observations of the pulsars J0030 (Riley et al., 2019) and J0740 (Riley et al., 2021) is displayed in figure 8.1. It can be seen that the joint EoS measurement is more precise than both the two individual probe measurements. This result can be expected to improve with the inclusion of new BNS detections and Pulsar X-ray observations in the dataset. In addition to parametric EoS inference, the same framework can be used to conduct multi-probe model comparison using the tabulated set of known EoSs. Of particular interest is the inclusion of the latest mass-radius measurement of J0030 which is significantly different from the earlier data-releases (Vinciguerra et al., 2024). These investigations are currently ongoing with new results expected shortly.

8.3.2 Simultaneous population inference

NS systems observable through different channels often belong to distinct astrophysical populations, each characterized by unique and uncertain distributions of NS masses. For example, the galactic double neutron star population probed via radio observations is potentially in tension with the extragalactic NS population observed by the LVK (Landry & Read, 2021). Similarly, there is no reason to expect that the NS systems observable by NICER will follow either of the two double NS population distributions. Incorrect assumptions about population properties such as the NS mass distribution can lead to significant biases in the inferred EoS constraints given a large enough (O(30)) number of observations (Wysocki et al., 2020).

Therefore, to obtain robust and unbiased EoS measurements from larger multi-probe datasets, it is necessary to simultaneously fit for the NS mass distribution while allowing for probe-specific subpopulations of NSs each characterized by potentially different mass distributions. This is achieved by re-weighting the Bayesian prior on the NS masses of each event to one that is informative on the ensemble properties of its probe-specific subpopulation. The likelihood conditional on the EoS and population hyper-parameters can be used to measure them simultaneously by constructing their joint posterior distribution.

Such a multiprobe exploration of the NS EoS and mass distribution that accounts for the possibility of distinct probe-specific subpopulation of NSs, has never been attempted before and can be implemented in principle by sampling the following posterior distribution:

$$
\begin{aligned}
p(\vec{\gamma}, \vec{\lambda}_{GW}, \vec{\lambda}_{XR}, \vec{\lambda}_{R} | \{d\}_{GW}, \{d\}_{XR}, \{d\}_{R}, I) \propto\!\!\propto\; & p(\vec{\gamma}, \mathbf{I}) p(\vec{\lambda}_{GW}, \vec{\lambda}_{XR}, \vec{\lambda}_{R} | \mathbf{I}) \times \\
& \left(\prod_{i=1}^{N_{GW}} \frac{1}{\beta(\vec{\lambda}_{GW})} \int_0^1 K_i^{GW}(q, \Lambda_1(\bar{\mathcal{M}}_i, q, \vec{\gamma}), \Lambda_2(\bar{\mathcal{M}}_i, q, \vec{\gamma})) \frac{p(\bar{\mathcal{M}}_i, q | \vec{\lambda}_{GW})}{p_{PE,GW}(\bar{\mathcal{M}}_i, q)} dq \right) \times \\
& \left(\prod_{j}^{N_{XR}} \int_{m_{min}(\vec{\gamma})}^{m_{max}(\vec{\gamma})} dm K_j^{XR}(m, r(m, \vec{\gamma})) \frac{p(m | \vec{\lambda}_{XR})}{p_{PE,XR}(m)} \right) \times
\end{aligned}
$$

$$\left(\prod_{l}^{N_R}\int_{m_{min}(\vec{\gamma})}^{m_{max}(\vec{\gamma})} dm K_l^R(m)\frac{p(m|\vec{\lambda}_R)}{p_{PE,R}(m)}\right), \tag{8.23}$$

where $\vec{\lambda}_\alpha$, $p(..|\vec{\lambda}_\alpha)$, and $p_{PE,\alpha}$ are the population hyper-parameters, hierarchical prior and PE prior for the α'th probe, and $\beta(\vec{\lambda}_{GW})$ is the fraction of detectable BNS mergers as a function of population hyper-parameters, which quantifies selection biases in GW observations.

This framework can be used to conduct population-marginalized EoS inference from multi-probe datasets, implementable through both parameter estimation and model comparison. In addition to real data, it is necessary to analyze simulated multi-probe datasets to understand the systematics of how probe-specific sub-populations can affect EoS measurements and explore the variation of any potential biases with the number of observations and the properties of the sub-population in each channel. This investigation will facilitate a unique multi-probe exploration of dense-matter properties while being free of potential biases arising from incorrect assumptions regarding the NS population distribution, one that I intend to pursue in the future.

8.3.3 Exploring the existence of NS phase transitions

Conventional parametric EoS models such as the ones described in chapter 6 do not account for the potential existence of QCD phase transitions in NS cores. With baryon number density in typical NS cores larger than the inverse of average nucleon volume, deconfined quark matter is expected to be more stable than the baryonic phase (Bodmer, 1971; Witten, 1984; Holdom et al., 2018). In such scenarios, hybrid stars with coexisting quark and hadronic matter phases are often expected to lie on different branches of the EoS mass-radius curve joined by a cusp (Lindblom, 1998). The existence of multiple branches on the mass-radius curve further necessitates simultaneous population and EoS inference since near the cusp, NSs with similar masses can have widely different radii.

Given enough observations with precise mass-radius measurements, it is possible to infer

composite EoS models simultaneously with the NS mass population and hence the existence of QCD phase transitions in NS cores (Chatziioannou & Han, 2020). The number (50+) of observations (Chatziioannou & Han, 2020) required to clearly identify the presence and absence of hybrid star branches in the mass-radius plane and the intractability of traditional EoS inference from large datasets necessitate the framework described in the previous section.

However, the fast TOV solvers used in the case of traditional parametric models are still under development and currently lack the functionality necessary for handling composite EoS models that allow for phase transitions. Generalizing `LALSimulation`'s TOV solver to make it compatible with hybrid star EoSs would be the first step toward this investigation which I plan to pursue in the near future. I further aim to use the developed solver in conjunction with the inference framework described in the previous section to conduct a multi-probe search for the existence of NS phase transitions.

8.3.4 BNS science in the 3G era

GW observations of BNS systems by 3G detectors are expected to be powerful probes of fundamental physics. Each observation will provide direct measurements of their detector frame masses, luminosity distances, and tidal deformabilities. 3G detectors are expected to observe about O(1000-100000) BNSs per year and up to very high redshifts O(7-10) (Baibhav et al., 2023). Given a 3G-sized catalog, the nuclear EoS which determines a unique relationship between source frame masses and tidal deformabilities of NSs, a cosmological model that determines the scaling of redshifted masses with luminosity distances, and the population level distributions of BNS masses and redshifts can all be self-consistently measured with unforeseen precision.

The possibility of ultra-precise measurements necessitates the use of robust and accurate hierarchical modeling that accounts for all possible physical scenarios whose imprints are expected to reflect in the measurements, so as to avoid systematic biases in the inferred conclusion. For

example, generic EoS models that allow for quark matter phase transitions in NS cores, and BNS population models that account for redshift evolution of the shape of the mass spectrum are necessary for such inferences, given that measurement uncertainties in the 3G era are expected to be much smaller than the systematic errors incurred by ignoring such effects should they occur in reality.

To exploit the powerful scientific potential of 3G BNS observations, a self-consistent and scalable hierarchical inference framework can be derived by varying the cosmological parameters used to convert detector frame masses into source frame ones, in the EoS and population inference frameworks described so far. The posterior distribution of the NS EoS, population, and cosmological hyperparameters then takes the following form:

$$p(\vec{\gamma}, \lambda, \Omega|\vec{d}, \mathbf{I}) \propto p(\vec{\gamma}, \lambda, \Omega|\mathbf{I}) \prod_i \int K_i(q, \tilde{\Lambda}(\frac{\bar{\mathcal{M}}_{z,i}}{1+z}, q, \vec{\gamma}), d_L(z, \Omega)) \frac{p(\frac{\bar{\mathcal{M}}_{z,i}}{1+z}, q, z|\vec{\lambda}, \vec{\Omega})}{p_{PE}(\bar{\mathcal{M}}_{z,i}, q, d_L(z,\Omega))(1+z)} dq dz, \tag{8.24}$$

where K_i is now a density estimate of the posterior distribution of $q, \tilde{\Lambda}$ and the luminosity distance d_L, $p(\mathcal{M}, q, z|\vec{\lambda}, \vec{\Omega})$ is the population level distribution of BNS masses and redshifts and $\bar{\mathcal{M}}_{z,i}$ is the mean of the posterior samples of the detector frame chirp mass corresponding to the i'th event. Note that, similar to the case of EoS-only inference, we have assumed that the posterior distribution of the detector-frame chirp mass is a highly peaked function about its mean.

Implementing this analysis will require the use of much more scalable density estimation techniques than KDEs such as the one of Lo (2023) based on masked autoregressive flows, along with efficient GPU parallelization, both of which I plan to pursue in the near future. Afterward, I intend to use this framework and conduct a detailed simulation study by constructing mock 3G-sized datasets and exploring the level of precision with which we can constrain the NS EoS, population, and the cosmological expansion using 3G BNS observations.

8.4 SEARCHING FOR DEVIATIONS FROM GR IN GW PROPAGATIONS

In chapter 7 I described a fully Bayesian methodology for constraining the coefficients of Lorentz violating GW propagation from growing GW catalogs in the presence of selection effects, a first-of-its-kind endeavor. Using this method and GWTC-3 data we simultaneously constrained all 9 coefficients of Lorentz violating GW propagation in the non-dispersive and non-birefringent limit of the SME's gravity sector. I further compared our method with others in existing literature, demonstrating its several advantages. In addition to communicating new constraints on modified theories of gravity, this study will guide future attempts at probing the SME's gravity sector using GW data, with our hierarchical inference framework applicable directly to explorations of birefringent and dispersive GW propagation. I aim to implement these applications in the future. These studies will lead to robust conclusions on spacetime symmetry-breaking theories of modified gravity, many of which would have otherwise remained unconstrained

8.4.1 Birefringence and dispersion

In birefringent GW propagation, the plus and cross polarizations are predicted to be inter-mingled by means of a frequency-dependent rotation matrix characterized by the angle βf^2 (Haegel et al., 2023; O'Neal-Ault et al., 2021). Constraining the parameter β from GW data amounts to testing GR which predicts β to be 0. For each CBC in a catalog, the rotated waveform model can be used to estimate β simultaneously with other CBC parameters such as sky location and distance. The coefficients in the spherical harmonic decomposition of β over the sky correspond to Lorentz and CPT violating terms in the SME lagrangian. Haegel et al. (2023) attempted to constrain these coefficients using a combination of Bayesian PE and chi-squared fitting while also ignoring selection effects. As we demonstrated in the context of measuring the speed of gravity, a fully bayesian method of hierarchical inference that accounts for selection effects, is necessary to obtain robust and unambiguous measurements of these coefficients. I plan to apply our newly developed method

on the analysis of symmetry-breaking birefringent GW propagation and re-do the investigations of Haegel et al. (2023), both for current and upcoming datasets.

Similarly, in the case of dispersive GW propagation, the dispersion relation of GWs is expected to contain additional frequency dependence than the predictions of GR: $\omega = |\vec{p}|(1 + A_\alpha \omega^{\alpha-2})$ (eg, O'Neal-Ault et al., 2021). Here A_α is the deviation from GR parameter which can be constrained by conducting Bayesian PE with waveform models account for such a dispersion relation as a modification of the GR-predicted phase evolution (Gong et al., 2023). Again, the coefficients appearing in the spherical harmonic decomposition of A_α map to those in the Lorentz and CPT violating terms of the SME lagrangian. Previous studies have attempted to measure these coefficients using suboptimal inference frameworks that are unable to simultaneously constrain all coefficients in a fully Bayesian way, while also ignoring selection effects (Gong et al., 2023). Measuring these coefficients using our method and growing GW catalogs is another investigation I intend to pursue in the future.

8.5 SUMMARY

In this chapter, I summarized my thesis work's main results and findings and proposed several follow-up projects and investigations on every topic explored that I will pursue in the near future. GW and multi-messenger astronomy has an exciting future ahead with current and future observations unveiling new insights into the mysteries of the Universe. The research conducted as part of this thesis will likely contribute significantly to these endeavors.

BIBLIOGRAPHY

Abac, A. G., et al. (2024). Observation of Gravitational Waves from the Coalescence of a $2.5 - 4.5\ M_\odot$ Compact Object and a Neutron Star.

Abbott, B. P., et al. (2016a). Observation of Gravitational Waves from a Binary Black Hole Merger. Phys. Rev. Lett., 116(6), 061102.

Abbott, B. P., et al. (2016b). Properties of the Binary Black Hole Merger GW150914. Phys. Rev. Lett., 116(24), 241102.

Abbott, B. P., et al. (2017a). A gravitational-wave standard siren measurement of the hubble constant. Nature, 551(7678), 85–88.

Abbott, B. P., et al. (2017b). Gravitational waves and gamma-rays from a binary neutron star merger: Gw170817 and grb 170817a. The Astrophysical Journal Letters, 848(2), L13.

Abbott, B. P., et al. (2017c). GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral. Phys. Rev. Lett., 119(16), 161101.

Abbott, B. P., et al. (2017d). Multi-messenger observations of a binary neutron star merger. Astrophys. J, 848, 2.

Abbott, B. P., et al. (2018a). GW170817: Measurements of neutron star radii and equation of state. Phys. Rev. Lett., 121(16), 161101.

Abbott, B. P., et al. (2018b). Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA. Living Rev. Rel., 21(1), 3.

Abbott, B. P., et al. (2019a). Binary black hole population properties inferred from the first and second observing runs of Advanced LIGO and Advanced Virgo. Astrophys. J., 882(2), L24.

Abbott, B. P., et al. (2019b). GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs. Phys. Rev. X, 9(3), 031040.

Abbott, B. P., et al. (2019c). Low-latency Gravitational-wave Alerts for Multimessenger Astronomy during the Second Advanced LIGO and Virgo Observing Run. Astrophys. J., 875(2), 161.

Abbott, B. P., et al. (2019d). Properties of the binary neutron star merger GW170817. Phys. Rev. X, 9(1), 011001.

Abbott, B. P., et al. (2019e). Tests of General Relativity with GW170817. Phys. Rev. Lett., 123(1), 011102.

Abbott, B. P., et al. (2020a). A guide to LIGO–Virgo detector noise and extraction of transient gravitational-wave signals. Class. Quant. Grav., 37(5), 055002.

Abbott, B. P., et al. (2020b). GW190425: Observation of a Compact Binary Coalescence with Total Mass $\sim 3.4 M_{\odot}$. Astrophys. J. Lett., 892(1), L3.

Abbott, B. P., et al. (2020c). Model comparison from LIGO–Virgo data on GW170817's binary components and consequences for the merger remnant. Class. Quant. Grav., 37(4), 045006.

Abbott, B. P., et al. (2021a). A Gravitational-wave Measurement of the Hubble Constant Following the Second Observing Run of Advanced LIGO and Virgo. Astrophys. J., 909(2), 218.

Abbott, R., et al. (2020d). GW190521: A binary black hole merger with a total mass of $150\ M_{\odot}$. Phys. Rev. Lett., 125, 101102.

Abbott, R., et al. (2020e). GW190814: Gravitational Waves from the Coalescence of a 23 Solar Mass Black Hole with a 2.6 Solar Mass Compact Object. Astrophys. J. Lett., 896(2), L44.

Abbott, R., et al. (2021b). Constraints on the cosmic expansion history from GWTC-3.

Abbott, R., et al. (2021c). GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run. Phys. Rev. X, 11, 021053.

Abbott, R., et al. (2021d). GWTC-2.1: Deep extended catalog of compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run.

Abbott, R., et al. (2021e). Observation of Gravitational Waves from Two Neutron Star–Black Hole Coalescences. Astrophys. J. Lett., 915(1), L5.

Abbott, R., et al. (2021f). Open data from the first and second observing runs of Advanced LIGO and Advanced Virgo. SoftwareX, 13, 100658.

Abbott, R., et al. (2021g). Population Properties of Compact Objects from the Second LIGO-Virgo Gravitational-Wave Transient Catalog. Astrophys. J. Lett., 913(1), L7.

Abbott, R., et al. (2021h). Tests of general relativity with gwtc-3.

Abbott, R., et al. (2023a). Constraints on the Cosmic Expansion History from GWTC–3.

Abbott, R., et al. (2023b). GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo during the Second Part of the Third Observing Run. Phys. Rev. X, 13(4), 041039.

Abbott, R., et al. (2023c). Open data from the third observing run of LIGO, Virgo, KAGRA and GEO.

Abbott, R., et al. (2023d). Population of Merging Compact Binaries Inferred Using Gravitational Waves through GWTC-3. Phys. Rev. X, 13(1), 011048.

Abbott, R., et al. (2023e). Search for gravitational-lensing signatures in the full third observing run of the LIGO-Virgo network.

Acernese, F., et al. (2014). Advanced Virgo: a second-generation interferometric gravitational wave detector. Class. Quantum Grav., 32(2), 024001.

Adams, T., Buskulic, D., Germain, V., Guidi, G. M., Marion, F., Montani, M., Mours, B., Piergiovanni, F., & Wang, G. (2016). Low-latency analysis pipeline for compact binary coalescences in the advanced gravitational wave detector era. Classical and Quantum Gravity, 33(17), 175012.

Ade, P. A. R., et al. (2016). Planck 2015 results. XIII. Cosmological parameters. Astron. Astrophys., 594, A13.

Aghanim, N., Akrami, Y., Ashdown, M., Aumont, J., Baccigalupi, C., Ballardini, M., Banday, A., Barreiro, R., Bartolo, N., Basak, S., et al. (2020). Planck 2018 results-vi. cosmological parameters. Astronomy & Astrophysics, 641, A6.

Akmal, A., Pandharipande, V. R., & Ravenhall, D. G. (1998). Equation of state of nucleon matter and neutron star structure. Phys. Rev. C, 58, 1804–1828.

Akutsu, T., et al. (2021). Overview of KAGRA: Calibration, detector characterization, physical environmental monitors, and the geophysics interferometer. PTEP, 2021(5), 05A102.

Allen, B. (2005). χ^2 time-frequency discriminator for gravitational wave detection. Phys. Rev. D, 71, 062001.

Allen, B., Anderson, W. G., Brady, P. R., Brown, D. A., & Creighton, J. D. E. (2012). Findchirp: An algorithm for detection of gravitational waves from inspiraling compact binaries. Phys. Rev. D, 85, 122006.

Antoniadis, J., et al. (2013). A massive pulsar in a compact relativistic binary. Science, 340(6131), 1233232.

Antonini, F., Gieles, M., Dosopoulou, F., & Chattopadhyay, D. (2023). Coalescing black hole binaries from globular clusters: mass distributions and comparison to gravitational wave data from GWTC-3. Mon. Not. Roy. Astron. Soc., 522(1), 466–476.

Antonini, F., & Perets, H. B. (2012). Secular evolution of compact binaries near massive black holes: Gravitational wave sources and other exotica. Astrophys. J., 757, 27.

Antonini, F., Toonen, S., & Hamers, A. S. (2017). Binary Black Hole Mergers from Field Triples: Properties, Rates, and the Impact of Stellar Evolution. , 841(2), 77.

Arun, K. G., Buonanno, A., Faye, G., & Ochsner, E. (2009). Higher-order spin effects in the amplitude and phase of gravitational waveforms emitted by inspiraling compact binaries: Ready-to-use gravitational waveforms. Phys. Rev. D, 79, 104023.

Ashton, G., et al. (2019). BILBY: A User-friendly Bayesian Inference Library for Gravitational-wave Astronomy. Apjs, 241(2), 27.

ATNF (2021). The atnf pulsar database. https://www.atnf.csiro.au/people/pulsar/psrcat/.

Aubin, F., et al. (2021). The mbta pipeline for detecting compact binary coalescences in the third ligo–virgo observing run. Classical and Quantum Gravity, 38(9), 095004.

Baibhav, V., Doctor, Z., & Kalogera, V. (2023). Dropping Anchor: Understanding the Populations of Binary Black Holes with Random and Aligned-spin Orientations. Astrophys. J., 946(1), 50.

Bailey, Q. G., & Kostelecký, V. A. (2006). Signals for Lorentz violation in post-Newtonian gravity. , 74(4), 045001.

Barkat, Z., Rakavy, G., & Sack, N. (1967). Dynamics of supernova explosion resulting from pair formation. Phys. Rev. Lett., 18, 379–381.

Bartos, I., Kocsis, B., Haiman, Z., & Márka, S. (2017). Rapid and Bright Stellar-mass Binary Black Hole Mergers in Active Galactic Nuclei. Astrophys. J., 835(2), 165.

Bavera, S. S., et al. (2020). The origin of spin in binary black holes. Predicting the distributions of the main observables of Advanced LIGO. , 635, A97.

Bavera, S. S., et al. (2023). The formation of merging black holes with masses beyond 30 $M_{\odot}$ at solar metallicity. Nature Astronomy, 7, 1090–1097.

Belczynski, K., Doctor, Z., Zevin, M., Olejak, A., Banerje, S., & Chattopadhyay, D. (2022). Black Hole-Black Hole Total Merger Mass and the Origin of LIGO/Virgo Sources. , 935(2), 126.

Belczynski, K., Dominik, M., Bulik, T., O'Shaughnessy, R., Fryer, C., & Holz, D. E. (2010). The effect of metallicity on the detection prospects for gravitational waves. The Astrophysical Journal Letters, 715(2), L138.

Belczynski, K., et al. (2016). The effect of pair-instability mass loss on black-hole mergers. Astronomy & Astrophysics, 594, A97.

Bellovary, J. M., Mac Low, M.-M., McKernan, B., & Ford, K. E. S. (2016). Migration Traps in Disks around Supermassive Black Holes. , 819(2), L17.

Benacquista, M. J., & Downing, J. M. B. (2013). Relativistic Binaries in Globular Clusters. Living Rev. Rel., 16, 4.

Bini, D., Damour, T., & Faye, G. (2012). Effective action approach to higher-order relativistic tidal interactions in binary systems and their effective one body description. Phys. Rev. D, 85, 124034.

Biscoveanu, S., Callister, T. A., Haster, C.-J., Ng, K. K. Y., Vitale, S., & Farr, W. M. (2022). The Binary Black Hole Spin Distribution Likely Broadens with Redshift. Astrophys. J. Lett., 932(2), L19.

Bodmer, A. R. (1971). Collapsed nuclei. Phys. Rev. D, 4, 1601–1606.

Bogdanov, S., et al. (2019). Constraining the neutron star mass–radius relation and dense matter equation of state with NICER. II. emission from hot spots on a rapidly rotating neutron star. The Astrophysical Journal, 887(1), L26.

Bohé, A., Faye, G., Marsat, S., & Porter, E. K. (2015). Quadratic-in-spin effects in the orbital dynamics and gravitational-wave energy flux of compact binaries at the 3pn order. Class. Quantum Grav., 32(19), 195010.

Bohé, A., Marsat, S., & Blanchet, L. (2013). Next-to-next-to-leading order spin–orbit effects in the gravitational wave flux and orbital phasing of compact binaries. Class. Quantum Grav., 30(13), 135009.

Briel, M. M., Stevance, H. F., & Eldridge, J. J. (2023). Understanding the high-mass binary black hole population from stable mass transfer and super-Eddington accretion in BPASS. , 520(4), 5724–5745.

Broekgaarden, F. S., Stevenson, S., & Thrane, E. (2022). Signatures of Mass Ratio Reversal in Gravitational Waves from Merging Binary Black Holes. Astrophys. J., 938(1), 45.

Brott, I., de Mink, S. E., Cantiello, M., Langer, N., de Koter, A., Evans, C. J., Hunter, I., Trundle, C., & Vink, J. S. (2011). Rotating massive main-sequence stars. Astronomy & Astrophysics, 530, A115.

Buonanno, A., Iyer, B. R., Ochsner, E., Pan, Y., & Sathyaprakash, B. S. (2009). Comparison of post-newtonian templates for compact binary inspiral signals in gravitational-wave detectors. Phys. Rev. D, 80, 084043.

Callister, T. A., & Farr, W. M. (2023). Data Release: "A parameter-free tour of the binary black hole population". https://doi.org/10.5281/zenodo.7616096.

Callister, T. A., & Farr, W. M. (2024). Parameter-Free Tour of the Binary Black Hole Population. Phys. Rev. X, 14(2), 021005.

Callister, T. A., Haster, C.-J., Ng, K. K. Y., Vitale, S., & Farr, W. M. (2021). Who Ordered That? Unequal-mass Binary Black Hole Mergers Have Larger Effective Spins. Astrophys. J. Lett., 922(1), L5.

Callister, T. A., Miller, S. J., Chatziioannou, K., & Farr, W. M. (2022). No evidence that the majority of black holes in binaries have zero spin. The Astrophysical Journal Letters, 937(1), L13.

Canizares, P., Field, S. E., Gair, J., Raymond, V., Smith, R., & Tiglio, M. (2015). Accelerated gravitational wave parameter estimation with reduced order modeling. Phys. Rev. Lett., 114, 071104.

Cannon, K., Hanna, C., & Peoples, J. (2015). Likelihood-ratio ranking statistic for compact binary coalescence candidates with rate estimation.

Cannon, K., et al. (2021). Gstlal: A software framework for gravitational wave discovery. SoftwareX, 14, 100680.

Canton, T. D., et al. (2021). Real-time search for compact binary mergers in advanced ligo and virgo's third observing run using pycbc live. The Astrophysical Journal, 923(2), 254.

Carney, M. F., Wade, L. E., & Irwin, B. S. (2018). Comparing two models for measuring the neutron star equation of state from gravitational-wave signals. Phys. Rev. D, 98, 063004.

Chatterjee, D., Ghosh, S., Brady, P. R., Kapadia, S. J., Miller, A. L., Nissanke, S., & Pannarale, F. (2020). A Machine Learning Based Source Property Inference for Compact Binary Mergers. Astrophys. J., 896(1), 54.

Chattopadhyay, D., Stegmann, J., Antonini, F., Barber, J., & Romero-Shaw, I. M. (2023). Double black hole mergers in nuclear star clusters: eccentricities, spins, masses, and the growth of massive seeds.

Chatziioannou, K., & Han, S. (2020). Studying strong phase transitions in neutron stars with gravitational waves. Phys. Rev. D, 101, 044019.

Chaudhary, S. S., et al. (2023). Low-latency alert products and their performance in anticipation of the fourth ligo-virgo-kagra observing run. (in prep).

Chen, H.-Y., Ezquiaga, J. M., & Gupta, I. (2024). Cosmography with next-generation gravitational wave detectors.

Chen, H.-Y., Fishbach, M., & Holz, D. E. (2018). A two per cent Hubble constant measurement from standard sirens within five years. Nature, 562(7728), 545–547.

Cheng, A. Q., Zevin, M., & Vitale, S. (2023). What You Don't Know Can Hurt You: Use and Abuse of Astrophysical Models in Gravitational-wave Population Analyses. Astrophys. J., 955(2), 127.

Collaboration, T., et al. (2022). LIGO-T2000012-v2: Noise curves used for Simulations in the update of the Observing Scenarios Paper — dcc.ligo.org. https://dcc.ligo.org/LIGO-s̄T2000012/public. [Accessed 22-03-2024].

Colladay, D., & Kostelecký, V. A. (1998). Lorentz-violating extension of the standard model. Phys. Rev. D, 58, 116002.

Contopoulos, I., Kazanas, D., & Fendt, C. (1999). The axisymmetric pulsar magnetosphere. The Astrophysical Journal, 511(1), 351–358.

Cornfield, J. (1969). The bayesian outlook and its application. Biometrics, 25(4), 617.

Cornish, N., Blas, D., & Nardini, G. (2017). Bounding the speed of gravity with gravitational wave observations. Phys. Rev. Lett., 119, 161102.

Coulter, D. A., et al. (2017). Swope Supernova Survey 2017a (SSS17a), the optical counterpart to a gravitational wave source. Science, 358(6370), 1556–1558.

Creighton, J. (2017). Certain identities in FGMC. Tech. rep., University of Wisconsin Milwaukee.

Creighton, J. D. E., & Anderson, W. G. (2011). Gravitational-wave physics and astronomy: An introduction to theory, experiment and data analysis.

Creutz, M. (1988). Global monte carlo algorithms for many-fermion systems. Phys. Rev. D, 38, 1228–1238.

Cutler, C., & Flanagan, E. E. (1994). Gravitational waves from merging compact binaries: How accurately can one extract the binary's parameters from the inspiral waveform? Phys. Rev. D, 49, 2658–2697.

Damour, T., & Nagar, A. (2009). Relativistic tidal properties of neutron stars. Phys. Rev. D, 80, 084035.

Davies, G. S., et al. (2020). Extending the PyCBC search for gravitational waves from compact binary mergers to a global network. Phys. Rev. D, 102, 022004.

de Mink, S. E., & Mandel, I. (2016). The chemically homogeneous evolutionary channel for binary black hole mergers: rates and properties of gravitational-wave events detectable by advanced LIGO. , 460(4), 3545–3553.

Del Pozzo, W. (2012). Inference of the cosmological parameters from gravitational waves: application to second generation interferometers. Phys. Rev. D, 86, 043011.

Del Pozzo, W., Li, T. G. F., & Messenger, C. (2017). Cosmological inference using only gravitational wave observations of binary neutron stars. Phys. Rev. D, 95, 043502.

Di Carlo, et al. (2019). Merging black holes in young star clusters. , 487(2), 2947–2960.

Di Carlo, et al. (2020). Binary black holes in young star clusters: the impact of metallicity. , 498(1), 495–506.

Diaz, C. C., & Mukherjee, S. (2022). Mapping the cosmic expansion history from LIGO-Virgo-KAGRA in synergy with DESI and SPHEREx. Mon. Not. Roy. Astron. Soc., 511(2), 2782–2795.

Díaz, J. S., Kostelecký, V. A., & Mewes, M. (2014). Testing relativity with high-energy astrophysical neutrinos. Phys. Rev. D, 89, 043005.

Dietrich, T., et al. (2020). Multimessenger constraints on the neutron-star equation of state and the hubble constant. Science, 370(6523), 1450–1453.

Dominik, M., et al. (2015). Double compact objects. iii. gravitational-wave detection rates. The Astrophysical Journal, 806(2), 263.

Douchin, F., & Haensel, P. (2001). A unified equation of state of dense matter and neutron star structure. Astronomy & Astrophysics, 380(1), 151–167.

Edelman, B., Doctor, Z., & Farr, B. (2021). Poking Holes: Looking for Gaps in LIGO/Virgo's Black Hole Population. Astrophys. J. Lett., 913(2), L23.

Edelman, B., Doctor, Z., Godfrey, J., & Farr, B. (2022a). Ain't no mountain high enough: Semi-parametric modeling of ligo–virgo's binary black hole mass distribution. The Astrophysical Journal, 924(2), 101.

Edelman, B., Farr, B., & Doctor, Z. (2022b). Cover your basis: Comprehensive data-driven characterization of the binary black hole population. https://doi.org/10.5281/zenodo.7566301.

Edelman, B., Farr, B., & Doctor, Z. (2023). Cover Your Basis: Comprehensive Data-driven Characterization of the Binary Black Hole Population. Astrophys. J., 946(1), 16.

Einstein, A. (1918). Über Gravitationswellen. Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.), 1918, 154–167.

El-Badry, K., Quataert, E., Weisz, D. R., Choksi, N., & Boylan-Kolchin, M. (2018). The formation and hierarchical assembly of globular cluster populations. Monthly Notices of the Royal Astronomical Society, 482(4), 4528–4552.

Essick, R., Farah, A., Galaudage, S., Talbot, C., Fishbach, M., Thrane, E., & Holz, D. E. (2022). Probing extremal gravitational-wave events with coarse-grained likelihoods. The Astrophysical Journal, 926(1), 34.

Essick, R., & Farr, W. (2022). Precision requirements for monte carlo sums within hierarchical bayesian inference.

Essick, R., Landry, P., & Holz, D. E. (2020). Nonparametric inference of neutron star composition, equation of state, and maximum mass with GW170817. Phys. Rev. D, 101(6), 063007.

Ewing, B., et al. (2024). Performance of the low-latency gstlal inspiral search towards ligo, virgo, and kagra's fourth observing run. Phys. Rev. D, 109, 042008.

Ezquiaga, J. M. (2021). Hearing gravity from the cosmos: GWTC-2 probes general relativity at cosmological scales. Phys. Lett. B, 822, 136665.

Ezquiaga, J. M., & Holz, D. E. (2022). Spectral Sirens: Cosmology from the Full Mass Distribution of Compact Binaries. Phys. Rev. Lett., 129(6), 061102.

Farah, A., Fishbach, M., Essick, R., Holz, D. E., & Galaudage, S. (2022). Bridging the gap: Categorizing gravitational-wave events at the transition between neutron stars and black holes. The Astrophysical Journal, 931(2), 108.

Farah, A., et al. (In prep.). "no need to know: gravitational-wave cosmology without astrophysical assumptions". In preparation.

Farah, A. M., Edelman, B., Zevin, M., Fishbach, M., Ezquiaga, J. M., Farr, B., & Holz, D. E. (2023). Things that might go bump in the night: Assessing structure in the binary black hole mass spectrum.

Farmer, R., Renzo, M., de Mink, S. E., Fishbach, M., & Justham, S. (2020). Constraints from gravitational-wave detections of binary black hole mergers on the 12c(,)16o rate. The Astrophysical Journal Letters, 902(2), L36.

Farmer, R., Renzo, M., de Mink, S. E., 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.

Farr, W. M. (2019). Accuracy requirements for empirically measured selection functions. Research Notes of the AAS, 3(5), 66.

Farr, W. M., Fishbach, M., Ye, J., & Holz, D. (2019). A Future Percent-Level Measurement of the Hubble Expansion at Redshift 0.8 With Advanced LIGO. Astrophys. J. Lett., 883(2), L42.

Farr, W. M., Gair, J. R., Mandel, I., & Cutler, C. (2015). Counting and confusion: Bayesian rate estimation with multiple populations. Phys. Rev. D, 91, 023005.

Farrell, E., Groh, J. H., Hirschi, R., Murphy, L., Kaiser, E., Ekström, S., Georgy, C., & Meynet, G. (2020). Is GW190521 the merger of black holes from the first stellar generations? Monthly Notices of the Royal Astronomical Society: Letters, 502(1), L40–L44.

Farrow, N., Zhu, X.-J., & Thrane, E. (2019). The mass distribution of galactic double neutron stars. The Astrophysical Journal, 876(1), 18.

Finn, L. S. (1992). Detection, measurement, and gravitational radiation. Phys. Rev. D, 46, 5236–5249.

Finn, L. S. (1997). Issues in gravitational wave data analysis. In 2nd Edoardo Amaldi Conference on Gravitational Waves, (pp. 180–191).

Finn, L. S. (1998). Gravitational radiation sources and signatures. eConf, C9808031, 07.

Finn, L. S., & Chernoff, D. F. (1993). Observing binary inspiral in gravitational radiation: One interferometer. Phys. Rev. D, 47, 2198–2219.

Fishbach, M., Doctor, Z., Callister, T., Edelman, B., Ye, J., Essick, R., Farr, W. M., Farr, B., & Holz, D. E. (2021). When are ligo/virgo's big black hole mergers? The Astrophysical Journal, 912(2), 98.

Fishbach, M., Farr, W. M., & Holz, D. E. (2020). The most massive binary black hole detections and the identification of population outliers. The Astrophysical Journal Letters, 891(2), L31.

Fishbach, M., & Holz, D. E. (2017). Where are ligo's big black holes? The Astrophysical Journal Letters, 851(2), L25.

Fishbach, M., & Holz, D. E. (2020). Picky Partners: The Pairing of Component Masses in Binary Black Hole Mergers. Astrophys. J. Lett., 891(1), L27.

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.

Fishbach, M., et al. (2019). A Standard Siren Measurement of the Hubble Constant from GW170817 without the Electromagnetic Counterpart. Astrophys. J. Lett., 871(1), L13.

Flowers, N. A., Goodge, C., & Tasson, J. D. (2017). Superconducting-gravimeter tests of local Lorentz invariance. Phys. Rev. Lett., 119(20), 201101.

Fong, H. K. Y. (2018). From simulations to signals: Analyzing gravitational waves from compact binary coalescences. Ph.D. thesis, Toronto U. https://tspace.library.utoronto.ca/handle/1807/91831.

Foreman-Mackey, D. (2016). corner.py: Scatterplot matrices in python. J. Open Source Softw., 1(2), 24.

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. (2013). emcee: The MCMC Hammer. Publications of the Astronomical Society of the Pacific, 125(925), 306.

Foreman-Mackey, D., Hogg, D. W., & Morton, T. D. (2014). Exoplanet population inference and the abundance of earth analogs from noisy, incomplete catalogs. The Astrophysical Journal, 795(1), 64.

Fortune, S. (1997). Voronoi Diagrams and Delaunay Triangulations, (p. 377–388). USA: CRC Press, Inc.

Fowler, W. A., & Hoyle, F. (1964). Neutrino Processes and Pair Formation in Massive Stars and Supernovae. , 9, 201.

Friedman, J. L., & Stergioulas, N. (2013). Rotating Relativistic Stars. Cambridge Monographs on Mathematical Physics. Cambridge University Press.

Fryer, C. L., Belczynski, K., Wiktorowicz, G., Dominik, M., Kalogera, V., & Holz, D. E. (2012). Compact remnant mass function: Dependence on the explosion mechanism and metallicity. The Astrophysical Journal, 749(1), 91.

Gaebel, S. M., Veitch, J., Dent, T., & Farr, W. M. (2019). Digging the population of compact binary mergers out of the noise. Mon. Not. Roy. Astron. Soc., 484(3), 4008–4023.

Galaudage, S., Talbot, C., & Thrane, E. (2020). Gravitational-wave inference in the catalog era: evolving priors and marginal events. Phys. Rev. D, 102(8), 083026.

Gallegos-Garcia, M., Berry, C. P. L., Marchant, P., & Kalogera, V. (2021). Binary Black Hole Formation with Detailed Modeling: Stable Mass Transfer Leads to Lower Merger Rates. , 922(2), 110.

Gardner, J. R., Pleiss, G., Bindel, D. S., Weinberger, K. Q., & Wilson, A. G. (2018). Gpytorch: Blackbox matrix-matrix gaussian process inference with gpu acceleration. In Neural Information Processing Systems.

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

Ghosh, R., Nair, S., Pathak, L., Sarkar, S., & Sengupta, A. S. (2023a). Test of the Second Postulate of Relativity from Gravitational Wave Observations.

Ghosh, S. (2020). GWXtreme documentation. https://gwxtreme.readthedocs.io/en/latest/. [Online; accessed 05-Sep-2022].

Ghosh, S., Liu, X., Creighton, J., Hernandez, I. M. n., Kastaun, W., & Pratten, G. (2021). Rapid model comparison of equations of state from gravitational wave observation of binary neutron star coalescences. Phys. Rev. D, 104, 083003.

Ghosh, T., More, S., Bera, S., & Bose, S. (2023b). Bayesian framework to infer the Hubble constant from cross-correlation of individual gravitational wave events with galaxies.

Godfrey, J., Edelman, B., & Farr, B. (2023). Cosmic Cousins: Identification of a Subpopulation of Binary Black Holes Consistent with Isolated Binary Evolution.

Goldreich, P., & Julian, W. H. (1969). Pulsar Electrodynamics. , 157, 869.

Goldstein, A., et al. (2017). An ordinary short gamma-ray burst with extraordinary implications: Fermi -GBM detection of GRB 170817a. The Astrophysical Journal, 848(2), L14.

Golomb, J., Isi, M., & Farr, W. (2023). Physical Models for the Astrophysical Population of Black Holes: Application to the Bump in the Mass Distribution of Gravitational Wave Sources.

Golomb, J., & Talbot, C. (2022). Hierarchical inference of binary neutron star mass distribution and equation of state with gravitational waves. The Astrophysical Journal, 926(1), 79.

Golub, G. H., & Reinsch, C. (1970). Singular value decomposition and least squares solutions. Numer. Math., 14(5), 403–420.

Gong, C., Zhu, T., Niu, R., Wu, Q., Cui, J.-L., Zhang, X., Zhao, W., & Wang, A. (2023). Gravitational wave constraints on nonbirefringent dispersions of gravitational waves due to Lorentz violations with GWTC-3 events. Phys. Rev. D, 107, 124015.

Goodman, J., & Weare, J. (2010). Ensemble samplers with affine invariance. Communications in Applied Mathematics and Computational Science, 5(1), 65–80.

Górski, K. M., Hivon, E., Banday, A. J., Wandelt, B. D., Hansen, F. K., Reinecke, M., & Bartelmann, M. (2005). HEALPix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere. , 622, 759–771.

Gray, R., Magaña Hernandez, I., Qi, H., Sur, A., et al. (2019). Cosmological Inference using Gravitational Wave Standard Sirens: A Mock Data Challenge.

Gray, R., Messenger, C., & Veitch, J. (2022). A pixelated approach to galaxy catalogue incompleteness: improving the dark siren measurement of the Hubble constant. Mon. Not. Roy. Astron. Soc., 512(1), 1127–1140.

Gray, R., et al. (2023). Joint cosmological and gravitational-wave population inference using dark sirens and galaxy catalogues. JCAP, 12, 023.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika, 82(4), 711–732.

H. Pihan-le Bars, et al. (2019). New Test of Lorentz invariance using the MICROSCOPE space mission. Phys. Rev. Lett., 123, 231102.

Haegel, L., O'Neal-Ault, K., Bailey, Q. G., Tasson, J. D., Bloom, M., & Shao, L. (2023). Search for anisotropic, birefringent spacetime-symmetry breaking in gravitational wave propagation from GWTC-3. Phys. Rev. D, 107, 064031.

Hanna, C., et al. (2020). Fast evaluation of multidetector consistency for real-time gravitational wave searches. Phys. Rev. D, 101, 022003.

Hanna, C., et al. (2023). A binary tree approach to template placement for searches for gravitational waves from compact binary mergers. Phys. Rev. D, 108, 042003.

Hannam, M., Schmidt, P., Bohé, A., Haegel, L., Husa, S., Ohme, F., Pratten, G., & Pürrer, M. (2014). Simple model of complete precessing black-hole-binary gravitational waveforms. Phys. Rev. Lett., 113, 151101.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. Biometrika, 57(1), 97–109.

Heger, A., Fryer, C. L., Woosley, S. E., Langer, N., & Hartmann, D. H. (2003). How massive single stars end their life. The Astrophysical Journal, 591(1), 288.

Heger, A., & Woosley, S. E. (2002). The nucleosynthetic signature of population iii. The Astrophysical Journal, 567(1), 532.

Heinzel, J., Biscoveanu, S., & Vitale, S. (2023). Probing Correlations in the Binary Black Hole Population with Flexible Models.

Hendriks, D. D., van Son, L. A. C., Renzo, M., Izzard, R. G., & Farmer, R. (2023). Pulsational pair-instability supernovae in gravitational-wave and electromagnetic transients.

Hernandez Vivanco, F., Smith, R., Thrane, E., Lasky, P. D., Talbot, C., & Raymond, V. (2019). Measuring the neutron star equation of state with gravitational waves: The first forty binary neutron star merger observations. Phys. Rev. D, 100, 103009.

Hernandez Vivanco, F., Smith, R., Thrane, E., & Lasky, P. D. (2020). A scalable random forest regressor for combining neutron-star equation of state measurements: a case study with GW170817 and GW190425. Monthly Notices of the Royal Astronomical Society, 499(4), 5972–5977.

Hinderer, T. (2008). Tidal love numbers of neutron stars. The Astrophysical Journal, 677(2), 1216–1220.

Holdom, B., Ren, J., & Zhang, C. (2018). Quark matter may not be strange. Phys. Rev. Lett., 120, 222001.

Holz, D. E., & Hughes, S. A. (2005). Using gravitational-wave standard sirens. Astrophys. J., 629, 15–22.

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

Hu, H., Kramer, M., Wex, N., Champion, D. J., & Kehl, M. S. (2020). Constraining the dense matter equation-of-state with radio pulsars. Monthly Notices of the Royal Astronomical Society, 497(3), 3118–3130.

Hulse, R. A., & Taylor, J. H. (1975). Discovery of a pulsar in a binary system. Astrophys. J. Lett., 195, L51–L53.

Husa, S., Khan, S., Hannam, M., Pürrer, M., Ohme, F., Forteza, X. J., & Bohé, A. (2016). Frequency-domain gravitational waves from nonprecessing black-hole binaries. i. new numerical waveforms and anatomy of the signal. Phys. Rev. D, 93, 044006.

J. Aasi, et al. (2015). Advanced LIGO. Class. Quantum Grav., 32(7), 074001.

Kapadia, S. J., et al. (2020). A self-consistent method to estimate the rate of compact binary coalescences with a poisson mixture model. Classical and Quantum Gravity, 37(4), 045007.

Karathanasis, C., Mukherjee, S., & Mastrogiovanni, S. (2023). Binary black holes population and cosmology in new lights: signature of PISN mass and formation channel in GWTC-3. Mon. Not. Roy. Astron. Soc., 523(3), 4539–4555.

Kasen, D., Metzger, B., Barnes, J., Quataert, E., & Ramirez-Ruiz, E. (2017). Origin of the heavy elements in binary neutron-star mergers from a gravitational-wave event. Nature, 551(7678), 80–84.

Khan, S., Husa, S., Hannam, M., Ohme, F., Pürrer, M., Forteza, X. J., & Bohé, A. (2016). Frequency-domain gravitational waves from nonprecessing black-hole binaries. ii. a phenomenological model for the advanced detector era. Phys. Rev. D, 93, 044007.

Kiendrebeogo, R. W., et al. (2023). Updated Observing Scenarios and Multimessenger Implications for the International Gravitational-wave Networks O4 and O5. Astrophys. J., 958(2), 158.

Kim, C., Kalogera, V., & Lorimer, D. R. (2003). The probability distribution of binary pulsar coalescence rates. I. double neutron star systems in the galactic field. Astrophys. J., 584, 985–995.

Kimball, C., et al. (2021). Evidence for Hierarchical Black Hole Mergers in the Second LIGO–Virgo Gravitational Wave Catalog. Astrophys. J. Lett., 915(2), L35.

Kinugawa, T., Nakamura, T., & Nakano, H. (2020). Formation of binary black holes similar to GW190521 with a total mass of 150 M from Population III binary star evolution. Monthly Notices of the Royal Astronomical Society: Letters, 501(1), L49–L53.

Kostelecký, V. A. (2004). Gravity, Lorentz violation, and the standard model. , 69, 105009.

Kostelecký, V. A., & Mewes, M. (2016). Testing local Lorentz invariance with gravitational waves. Phys. Lett. B, 757, 510–514.

Kostelecký, V. A., & Russell, N. (2008). Data tables for Lorentz and CPT violation.

Kostelecký, V. A., & Tasson, J. D. (2015). Constraints on Lorentz violation from gravitational Čerenkov radiation. Phys. Lett. B, 749, 551–559.

Kudritzki, R.-P., & Puls, J. (2000). Winds from hot stars. Annual Review of Astronomy and Astrophysics, 38(1), 613–666.

Kumar, P., Pürrer, M., & Pfeiffer, H. P. (2017). Measuring neutron star tidal deformability with Advanced LIGO: a Bayesian analysis of neutron star - black hole binary observations. Phys. Rev. D, 95(4), 044039.

Lackey, B. D., & Wade, L. (2015). Reconstructing the neutron-star equation of state with gravitational-wave detectors from a realistic population of inspiralling binary neutron stars. Phys. Rev. D, 91, 043002.

Landry, P., & Essick, R. (2019). Nonparametric inference of the neutron star equation of state from gravitational wave observations. Phys. Rev. D, 99, 084049.

Landry, P., Essick, R., & Chatziioannou, K. (2020). Nonparametric constraints on neutron star matter with existing and upcoming gravitational wave and pulsar observations. Phys. Rev. D, 101, 123007.

Landry, P., & Read, J. S. (2021). The Mass Distribution of Neutron Stars in Gravitational-wave Binaries. Astrophys. J. Lett., 921(2), L25.

Lau, K. N., & Seifert, M. D. (2017). Direct-coupling lensing by antisymmetric tensor monopoles. Phys. Rev. D, 95(2), 025023.

Li, S.-S., Mao, S., Zhao, Y., & Lu, Y. (2018). Gravitational lensing of gravitational waves: a statistical perspective. Monthly Notices of the Royal Astronomical Society, 476(2), 2220–2229.

Li, Y.-J., Wang, Y.-Z., Tang, S.-P., & Fan, Y.-Z. (2023). Resolving the stellar-collapse and hierarchical-merger origins of the coalescing black holes.

LIGO Scientific Collaboration (2018). LIGO Algorithm Library.

LIGO Scientific Collaboration And Virgo Collaboration And KAGRA Collaboration (2021). Gwtc-3: Compact binary coalescences observed by ligo and virgo during the second part of the third observing run — candidate data release. https://zenodo.org/record/5546665.

Lindblom, L. (1992). Determining the Nuclear Equation of State from Neutron-Star Masses and Radii. , 398, 569.

Lindblom, L. (1998). Phase transitions and the mass-radius curves of relativistic stars. Phys. Rev. D, 58, 024008.

Lindblom, L. (2010). Spectral representations of neutron-star equations of state. Phys. Rev. D, 82, 103011.

Liu, X., He, V. F., Mikulski, T. M., Palenova, D., Williams, C. E., Creighton, J., & Tasson, J. D. (2020). Measuring the speed of gravitational waves from the first and second observing run of Advanced LIGO and Advanced Virgo. Phys. Rev. D, 102, 024028.

Lo, R. K. L. (2023). denmarf: a python package for density estimation using masked autoregressive flow.

Loredo, T. J. (1992). Promise of Bayesian inference for astrophysics. In Statistical Challenges in Modern Astronomy, (pp. 275–297).

Loredo, T. J. (2004). Accounting for source uncertainties in analyses of astronomical survey data. AIP Conf. Proc., 735(1), 195–206.

LVC (2018). GW170817: Measurements of Neutron Star Radii and Equation of State. LIGO-P1800115-v12:.

LVC (2020). Parameter estimation sample release for GW190425. LIGO-P2000026-v2.

Magaña Hernandez, I., & Ray, A. (2024). Beyond Gaps and Bumps: Spectral Siren Cosmology with Non-Parametric Population Models.

Maggiore, M., et al. (2020). Science Case for the Einstein Telescope. JCAP, 03, 050.

Manchester, R. N., Hobbs, G. B., Teoh, A., & Hobbs, M. (2005). The australia telescope national facility pulsar catalogue. The Astronomical Journal, 129(4), 1993–2006.

Mandel, I., & de Mink, S. E. (2016). Merging binary black holes formed through chemically homogeneous evolution in short-period stellar binaries. Mon. Not. Roy. Astron. Soc., 458(3), 2634–2647.

Mandel, I., & Farmer, A. (2022). Merging stellar-mass binary black holes. Phys. Rept., 955, 1–24.

Mandel, I., Farr, W. M., Colonna, A., Stevenson, S., Tiňo, P., & Veitch, J. (2017). Model-independent inference on compact-binary observations. Mon. Not. Roy. Astron. Soc., 465(3), 3254–3260.

Mandel, I., Farr, W. M., & Gair, J. R. (2019). Extracting distribution parameters from multiple uncertain observations with selection biases. Mon. Not. R. Astron. Soc., 486(1), 1086–1093.

Mapelli, M. (2020a). Astrophysics of stellar black holes. Proc. Int. Sch. Phys. Fermi, 200, 87–121.

Mapelli, M. (2020b). Binary Black Hole Mergers: Formation and Populations. Front. Astron. Space Sci., 7, 38.

Mapelli, M., Bouffanais, Y., Santoliquido, F., Arca Sedda, M., & Artale, M. C. (2022). The cosmic evolution of binary black holes in young, globular, and nuclear star clusters: rates, masses, spins, and mixing fractions. , 511(4), 5797–5816.

Mapelli, M., Bouffanais, Y., Santoliquido, F., Sedda, M. A., & Artale, M. C. (2022). The cosmic evolution of binary black holes in young, globular, and nuclear star clusters: rates, masses, spins, and mixing fractions. Mon. Not. Roy. Astron. Soc., 511(4), 5797–5816.

Mapelli, M., Giacobbo, N., Santoliquido, F., & Artale, M. C. (2019). The properties of merging black holes and neutron stars across cosmic time. Monthly Notices of the Royal Astronomical Society, 487(1), 2–13.

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

Marchant, P., Renzo, M., Farmer, R., Pappas, K. M. W., Taam, R. E., de Mink, S. E., & Kalogera, V. (2019). Pulsational pair-instability supernovae in very close binaries. The Astrophysical Journal, 882(1), 36.

Martinez, M. A. S., Rodriguez, C. L., & Fragione, G. (2022). On the Mass Ratio Distribution of Black Hole Mergers in Triple Systems. , 937(2), 78.

Mastrogiovanni, S., Laghi, D., Gray, R., Santoro, G. C., Ghosh, A., Karathanasis, C., Leyde, K., Steer, D. A., Perries, S., & Pierra, G. (2023). Joint population and cosmological properties inference with gravitational waves standard sirens and galaxy surveys. Phys. Rev. D, 108(4), 042002.

Mastrogiovanni, S., Leyde, K., Karathanasis, C., Chassande-Mottin, E., Steer, D. A., Gair, J., Ghosh, A., Gray, R., Mukherjee, S., & Rinaldi, S. (2021). On the importance of source population models for gravitational-wave cosmology. Phys. Rev. D, 104(6), 062009.

Mastrogiovanni, S., Pierra, G., Perriès, S., Laghi, D., Caneva Santoro, G., Ghosh, A., Gray, R., Karathanasis, C., & Leyde, K. (2024). ICAROGW: A python package for inference of astrophysical population properties of noisy, heterogeneous, and incomplete observations. Astron. Astrophys., 682, A167.

McKernan, B., Ford, K. E. S., O'Shaugnessy, R., & Wysocki, D. (2020). Monte Carlo simulations of black hole mergers in AGN discs: Low χ_{eff} mergers and predictions for LIGO. , 494(1), 1203–1216.

Messenger, C., & Read, J. (2012). Measuring a cosmological distance-redshift relationship using only gravitational wave observations of binary neutron star coalescences. Phys. Rev. Lett., 108, 091101.

Messick, C., et al. (2017). Analysis framework for the prompt discovery of compact binary mergers in gravitational-wave data. Phys. Rev. D, 95, 042001.

Mészáros, P., Fox, D. B., Hanna, C., & Murase, K. (2019). Multi-Messenger Astrophysics. Nature Rev. Phys., 1, 585–599.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. J. Chem. Phys., 21(6), 1087–1092.

Metzger, B. D. (2019). Kilonovae. Living Reviews in Relativity, 23(1).

Mikóczi, B., Vasúth, M., & Gergely, L. A. (2005). Self-interaction spin effects in inspiralling compact binaries. Phys. Rev. D, 71, 124043.

Miller, M. C., et al. (2019). PSR j00300451 mass and radius from NICER data and implications for the properties of neutron star matter. The Astrophysical Journal, 887(1), L24.

Miller, M. C., et al. (2021). The radius of PSR j07406620 from NICER and XMM-newton data. The Astrophysical Journal Letters, 918(2), L28.

Mohite, S. (2022). Data-driven population inference from gravitational-wave sources and electromagnetic counterparts. https://dc.uwm.edu/etd/2926/.

Morisaki, S., & Raymond, V. (2020). Rapid parameter estimation of gravitational waves from binary neutron star coalescence using focused reduced order quadrature. Phys. Rev. D, 102, 104020.

Mukherjee, S. (2022). The redshift dependence of black hole mass distribution: is it reliable for standard sirens cosmology? Mon. Not. Roy. Astron. Soc., 515(4), 5495–5505.

Mukherjee, S., Wandelt, B. D., Nissanke, S. M., & Silvestri, A. (2021a). Accurate precision Cosmology with redshift unknown gravitational wave sources. Phys. Rev. D, 103(4), 043520.

Mukherjee, S., Wandelt, B. D., & Silk, J. (2021b). Testing the general theory of relativity using gravitational wave propagation from dark standard sirens. Mon. Not. Roy. Astron. Soc., 502(1), 1136–1144.

Nair, R., Bose, S., & Saini, T. D. (2018). Measuring the Hubble constant: Gravitational wave observations meet galaxy clustering. Phys. Rev., D98(2), 023502.

Neal, R. M. (1993). Probabilistic inference using markov chain monte carlo methods.

Neal, R. M. (2011). MCMC Using Hamiltonian Dynamics. Chapman and Hall/CRC.

Neijssel, C. J., Vigna-Gómez, A., Stevenson, S., Barrett, J. W., Gaebel, S. M., Broekgaarden, F. S., de Mink, S. E., Szécsi, D., Vinciguerra, 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.

Nitz, A. H., Dal Canton, T., Davis, D., & Reyes, S. (2018). Rapid detection of gravitational waves from compact binary mergers with pycbc live. Phys. Rev. D, 98, 024050.

Niu, R., Zhu, T., & Zhao, W. (2022). Testing Lorentz invariance of gravity in the Standard-Model Extension with GWTC-3. J. Cosmol. Astropart. Phys., 2022(12), 011.

O'Neal-Ault, K., Bailey, Q. G., Dumerchat, T., Haegel, L., & Tasson, J. (2021). Analysis of Birefringence and Dispersion Effects from Spacetime-Symmetry Breaking in Gravitational Waves. Universe, 7(10), 380.

Oppenheimer, J. R., & Volkoff, G. M. (1939). On massive neutron cores. Phys. Rev., 55, 374–381.

Oriol, A.-P., et al. (2023). Pymc: A modern and comprehensive probabilistic programming framework in python. PeerJ Computer Science, 9, e1516.

Paczynski, B. (1976). Common Envelope Binaries. In P. Eggleton, S. Mitton, & J. Whelan (Eds.) Structure and Evolution of Close Binary Systems, vol. 73, (p. 75).

Palmese, A., et al. (2020). A statistical standard siren measurement of the Hubble constant from the LIGO/Virgo gravitational wave compact object merger GW190814 and Dark Energy Survey galaxies. Astrophys. J. Lett., 900(2), L33.

Pankow, C., Brady, P., Ochsner, E., & O'Shaughnessy, R. (2015). Novel scheme for rapid parallel parameter estimation of gravitational waves from compact binary coalescences. Phys. Rev. D, 92, 023002.

Payne, E., Banagiri, S., Lasky, P. D., & Thrane, E. (2020). Searching for anisotropy in the distribution of binary black hole mergers. Phys. Rev. D, 102, 102004.

Payne, E., & Thrane, E. (2023). Model exploration in gravitational-wave astronomy with the maximum population likelihood. Phys. Rev. Res., 5, 023013.

Perna, R., Lazzati, D., & Farr, W. (2019). Limits on electromagnetic counterparts of gravitational-wave-detected binary black hole mergers. The Astrophysical Journal, 875(1), 49.

Petrovich, C., & Antonini, F. (2017). Greatly Enhanced Merger Rates of Compact-object Binaries in Non-spherical Nuclear Star Clusters. , 846(2), 146.

Pierra, G., Mastrogiovanni, S., Perriès, S., & Mapelli, M. (2023). A Study of Systematics on the Cosmological Inference of the Hubble Constant from Gravitational Wave Standard Sirens.

Planck Collaboration (2018). Planck 2018 results. VI. Cosmological parameters.

Postnov, K., & Yungelson, L. (2006). The Evolution of Compact Binary Star Systems. Living Rev. Rel., 9, 6.

Quiñonero-Candela, J., & Rasmussen, C. E. (2005). A unifying view of sparse approximate gaussian process regression. Journal of Machine Learning Research, 6(65), 1939–1959.

Radice, D., Perego, A., Zappa, F., & Bernuzzi, S. (2018). GW170817: Joint Constraint on the Neutron Star Equation of State from Multimessenger Observations. The Astrophysical Journal Letters, 852(2), L29.

Ray, A., Camilo, M., Creighton, J., Ghosh, S., & Morisaki, S. (2023a). Rapid hierarchical inference of neutron star equation of state from multiple gravitational wave observations of binary neutron star coalescences. Phys. Rev. D, 107, 043035.

Ray, A., Fan, P., He, V. F., Bloom, M., Yang, S. M., Tasson, J. D., & Creighton, J. D. E. (2023b). Measuring Gravitational Wave Speed and Lorentz Violation with the First Three Gravitational-Wave Catalogs.

Ray, A., Magaña Hernandez, I., Breivik, K., & Creighton, J. (2024). Searching for binary black hole sub-populations in gravitational wave data using binned Gaussian processes.

Ray, A., Magaña Hernandez, I., Mohite, S., Creighton, J., & Kapadia, S. (2023c). Nonparametric Inference of the Population of Compact Binaries from Gravitational-wave Observations Using Binned Gaussian Processes. Astrophys. J., 957(1), 37.

Ray, A., et al. (2023d). When to Point Your Telescopes: Gravitational Wave Trigger Classification for Real-Time Multi-Messenger Followup Observations.

Read, J. S., Lackey, B. D., Owen, B. J., & Friedman, J. L. (2009). Constraints on a phenomenologically parametrized neutron-star equation of state. Phys. Rev. D, 79, 124032.

Reitze, D., et al. (2019). Cosmic Explorer: The U.S. Contribution to Gravitational-Wave Astronomy beyond LIGO. Bull. Am. Astron. Soc., 51(7), 035.

Renzo, M., Farmer, R., Justham, S., Götberg, Y., de Mink, S. E., Zapartas, E., Marchant, P., & Smith, N. (2020). Predictions for the hydrogen-free ejecta of pulsational pair-instability supernovae. Astronomy & Astrophysics, 640, A56.

Riess, A. G., Casertano, S., Yuan, W., Macri, L. M., & Scolnic, D. (2019). Large Magellanic Cloud Cepheid Standards Provide a 1% Foundation for the Determination of the Hubble Constant and Stronger Evidence for Physics beyond ΛCDM. Astrophys. J., 876(1), 85.

Riess, A. G., Wenlong, Y., Macri Lucas, M., Dan, S., Dillon, B., Stefano, C., Jones David, O., Yukei, M., Louise, B., Brink Thomas, G., et al. (2022). A comprehensive measurement of the local value of the hubble constant with 1 km sˆ-1 mpcˆ-1 uncertainty from the hubble space telescope and the sh0es team. Astrophys. J. Lett, 934, L7.

Riley, T. E., et al. (2019). A iNICER/i view of PSR j00300451: Millisecond pulsar parameter estimation. The Astrophysical Journal, 887(1), L21.

Riley, T. E., et al. (2021). A NICER view of the massive pulsar PSR j07406620 informed by radio timing and XMM-newton spectroscopy. The Astrophysical Journal Letters, 918(2), L27.

Rinaldi, S., Del Pozzo, W., Mapelli, M., Medina, A. L., & Dent, T. (2023). Evidence for the evolution of black hole mass function with redshift.

Rinaldi, S., & Del Pozzo, W. (2021). (H)DPGMM: a hierarchy of Dirichlet process Gaussian mixture models for the inference of the black hole mass function. Monthly Notices of the Royal Astronomical Society, 509(4), 5454–5466.

Rodriguez, C. L., Amaro-Seoane, P., Chatterjee, S., Kremer, K., Rasio, F. A., Samsing, J., Ye, C. S., & Zevin, M. (2018). Post-Newtonian Dynamics in Dense Star Clusters: Formation, Masses, and Merger Rates of Highly-Eccentric Black Hole Binaries. Phys. Rev. D, 98(12), 123005.

Rodriguez, C. L., & Loeb, A. (2018). Redshift evolution of the black hole merger rate from globular clusters. The Astrophysical Journal Letters, 866(1), L5.

Rodriguez, C. L., Zevin, M., Amaro-Seoane, P., Chatterjee, S., Kremer, K., Rasio, F. A., & Ye, C. S. (2019). Black holes: The next generation—repeated mergers in dense star clusters and their gravitational-wave properties. Physical Review D, 100(4).

Rodriguez, C. L., et al. (2022). Modeling Dense Star Clusters in the Milky Way and beyond with the Cluster Monte Carlo Code. Astrophys. J. Supp., 258(2), 22.

Romani, R. W., Kandel, D., Filippenko, A. V., Brink, T. G., & Zheng, W. (2022). Psr j09520607: The fastest and heaviest known galactic neutron star. The Astrophysical Journal Letters, 934(2), L17.

Romero-Shaw, I. M., Kremer, K., Lasky, P. D., Thrane, E., & Samsing, J. (2021). Gravitational waves as a probe of globular cluster formation and evolution. Monthly Notices of the Royal Astronomical Society, 506(2), 2362–2372.

Romero-Shaw, I. M., et al. (2020). Bayesian inference for compact binary coalescences with bilby: validation and application to the first LIGO–Virgo gravitational-wave transient catalogue. Mon. Not. Roy. Astron. Soc., 499(3), 3295–3319.

Rose, C., Ray, A., Valsan, V., Brady, P., et al. (In prep.). "updating gravitational-wave source classification estimates with rapid parameter estimation". In preparation.

Rose, C. A., Valsan, V., Brady, P. R., Walsh, S., & Pankow, C. (2022). Supplementing rapid bayesian parameter estimation schemes with adaptive grids.

Roulet, J., Venumadhav, T., Zackay, B., Dai, L., & Zaldarriaga, M. (2020). Binary black hole mergers from ligo/virgo o1 and o2: Population inference combining confident and marginal events. Phys. Rev. D, 102, 123022.

Sachdev, S., et al. (2019). The gstlal search analysis methods for compact binary mergers in advanced ligo's second and advanced virgo's first observing runs.

Sadiq, J., Dent, T., & Wysocki, D. (2022). Flexible and fast estimation of binary merger population distributions with an adaptive kernel density estimator. Phys. Rev. D, 105, 123014.

Safarzadeh, M., & Farr, W. M. (2019). The impact of metallicity evolution of the universe on the maximum mass of ligo binary black holes. The Astrophysical Journal Letters, 883(1), L24.

Sakon, S., et al. (2023). Template bank for compact binary mergers in the fourth observing run of advanced ligo, advanced virgo, and kagra.

Salpeter, E. E. (1955). The Luminosity Function and Stellar Evolution. The Astrophysical Journal, 121, 161.

Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. (2016). Probabilistic programming in python using PyMC3. PeerJ Computer Science, 2, e55.

Samsing, J. (2018). Eccentric black hole mergers forming in globular clusters. Physical Review D, 97(10).

Santoliquido, F., Mapelli, M., Bouffanais, Y., Giacobbo, N., Carlo, U. N. D., Rastello, S., Artale, M. C., & Ballone, A. (2020). The cosmic merger rate density evolution of compact binaries formed in young star clusters and in isolated binaries. The Astrophysical Journal, 898(2), 152.

Savchenko, V., et al. (2017). INTEGRAL Detection of the First Prompt Gamma-Ray Signal Coincident with the Gravitational-wave Event GW170817. The Astrophysical Journal Letters, 848(2), L15.

Schmidt, P., Ohme, F., & Hannam, M. (2015). Towards models of gravitational waveforms from generic binaries: Ii. modelling precession effects with a single effective precession parameter. Phys. Rev. D, 91, 024043.

Schutz, B. F. (1986). Determining the Hubble Constant from Gravitational Wave Observations. Nature, 323, 310–311.

Scott, D. W. (1992). Multivariate Density Estimation. Wiley.

Silva, H. O., Holgado, A. M., Cárdenas-Avendaño, A., & Yunes, N. (2021). Astrophysical and theoretical physics implications from multimessenger neutron star observations. Phys. Rev. Lett., 126, 181101.

Silverman, B. W. (1986). Density Estimation for Statistics and Data Analysis. London: Chapman & Hall.

Skilling, J. (2006). Nested sampling for general Bayesian computation. Bayesian Analysis, 1(4), 833 – 859.

Smith, R., Field, S. E., Blackburn, K., Haster, C.-J., Pürrer, M., Raymond, V., & Schmidt, P. (2016). Fast and accurate inference on gravitational waves from precessing compact binaries. Phys. Rev. D, 94, 044031.

Smith, R. J. E., Ashton, G., Vajpeyi, A., & Talbot, C. (2020). Massively parallel Bayesian inference for transient gravitational-wave astronomy. Monthly Notices of the Royal Astronomical Society, 498(3), 4492–4502.

Soares-Santos, M., et al. (2019). First Measurement of the Hubble Constant from a Dark Standard Siren using the Dark Energy Survey Galaxies and the LIGO/Virgo Binary–Black-hole Merger GW170814. Astrophys. J. Lett., 876(1), L7.

Speagle, J. S. (2020). dynesty: a dynamic nested sampling package for estimating Bayesian posteriors and evidences. Mon. Not. Roy. Astron. Soc., 493(3), 3132–3158.

Spitkovsky, A. (2006). Time-dependent force-free pulsar magnetospheres: Axisymmetric and oblique rotators. The Astrophysical Journal, 648(1), L51–L54.

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'. , 464(1), 946–954.

Suwa, Y., Yoshida, T., Shibata, M., Umeda, H., & Takahashi, K. (2018). On the minimum mass of neutron stars. Mon. Not. Roy. Astron. Soc., 481(3), 3305–3312.

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, & The KAGRA Collaboration (2021). GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run — Parameter estimation data release. https://doi.org/10.5281/zenodo.5546663.

The LIGO Scientific Collaboration, Virgo Collaboration, & KAGRA Collaboration (2023). GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run — O1+O2+O3 Search Sensitivity Estimates. https://doi.org/10.5281/zenodo.7890398.

Thrane, E., & Talbot, C. (2019). An introduction to bayesian inference in gravitational-wave astronomy: Parameter estimation, model selection, and hierarchical models. Publ. Astron. Soc. Austral., 36, e010.

Tiwari, V. (2018). Estimation of the sensitive volume for gravitational-wave source populations using weighted monte carlo integration. Classical and Quantum Gravity, 35(14), 145009.

Tiwari, V. (2021). VAMANA: modeling binary black hole population with minimal assumptions. Class. Quant. Grav., 38(15), 155007.

Tiwari, V. (2022). Exploring features in the binary black hole population. The Astrophysical Journal, 928(2), 155.

Tiwari, V., & Fairhurst, S. (2021). The emergence of structure in the binary black hole mass distribution. The Astrophysical Journal Letters, 913(2), L19.

Toubiana, A., Katz, M. L., & Gair, J. R. (2023). Is there an excess of black holes around 20 M$\odot$? Optimizing the complexity of population models with the use of reversible jump MCMC. Mon. Not. Roy. Astron. Soc., 524(4), 5844–5853.

Tsukada, L., , et al. (2023). Improved ranking statistics of the gstlal inspiral search for compact binary coalescences. Phys. Rev. D, 108, 043004.

Valentino, D., et al. (2021). In the realm of the Hubble tension—a review of solutions. Class. Quant. Grav., 38(15), 153001.

van den Heuvel, E. P. J., Portegies Zwart, S. F., & de Mink, S. E. (2017). Forming short-period Wolf–Rayet X-ray binaries and double black holes through stable mass transfer. Mon. Not. Roy. Astron. Soc., 471(4), 4256–4264.

van Son, L. A. C., de Mink, S. E., Callister, T., Justham, S., Renzo, M., Wagg, T., Broekgaarden, F. S., Kummer, F., Pakmor, R., & Mandel, I. (2022a). The redshift evolution of the binary black hole merger rate: A weighty matter. The Astrophysical Journal, 931(1), 17.

van Son, L. A. C., de Mink, S. E., Chruslinska, M., Conroy, C., Pakmor, R., & Hernquist, L. (2022b). The locations of features in the mass distribution of merging binary black holes are robust against uncertainties in the metallicity-dependent cosmic star formation history.

van Son, L. A. C., de Mink, S. E., Renzo, M., Justham, S., Zapartas, E., Breivik, K., Callister, T., Farr, W. M., & Conroy, C. (2022c). No peaks without valleys: The stable mass transfer channel for gravitational-wave sources in light of the neutron star–black hole mass gap. The Astrophysical Journal, 940(2), 184.

Varma, V., Field, S. E., Scheel, M. A., Blackman, J., Gerosa, D., Stein, L. C., Kidder, L. E., & Pfeiffer, H. P. (2019). Surrogate models for precessing binary black hole simulations with unequal masses. Phys. Rev. Res., 1, 033015.

Veitch, J., & Vecchio, A. (2010). Bayesian coherent analysis of in-spiral gravitational wave signals with a detector network. Phys. Rev. D, 81, 062003.

Veitch, J., et al. (2015). Parameter estimation for compact binaries with ground-based gravitational-wave observations using the lalinference software library. Phys. Rev. D, 91, 042003.

Vinciguerra, S., et al. (2024). An Updated Mass–Radius Analysis of the 2017–2018 NICER Data Set of PSR J0030+0451. Astrophys. J., 961(1), 62.

Vines, J., Flanagan, E. E., & Hinderer, T. (2011). Post-1-newtonian tidal effects in the gravitational waveform from binary inspirals. Phys. Rev. D, 83, 084051.

Vink, J. S., Higgins, E. R., Sander, A. A. C., & Sabhahit, G. N. (2021). Maximum black hole mass across cosmic time. Monthly Notices of the Royal Astronomical Society, 504(1), 146–154.

Virtanen, P., et al. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in python. Nature Methods, 17, 261–272.

Vitale, S., Gerosa, D., Farr, W. M., & Taylor, S. R. (2020). Inferring the properties of a population of compact binaries in presence of selection effects, (pp. 1–60). Singapore: Springer Singapore.

Voronoi, G. (1908). Nouvelles applications des paramètres continus à la théorie des formes quadratiques. premier mémoire. sur quelques propriétés des formes quadratiques positives parfaites. Journal für die reine und angewandte Mathematik (Crelles Journal), 1908(133), 97–102.

Wade, L., Creighton, J. D. E., Ochsner, E., Lackey, B. D., Farr, B. F., Littenberg, T. B., & Raymond, V. (2014). Systematic and statistical errors in a bayesian approach to the estimation of the neutron-star equation of state using advanced gravitational wave detectors. Phys. Rev. D, 89, 103012.

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. Cogn. Psychol., 60(3), 158–189.

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.

Watts, A. L., et al. (2016). Colloquium: Measuring the neutron star equation of state using x-ray timing. Rev. Mod. Phys., 88, 021001.

Weatherford, N. C., Fragione, G., Kremer, K., Chatterjee, S., Ye, C. S., Rodriguez, C. L., & Rasio, F. A. (2021). Black hole mergers from star clusters with top-heavy initial mass functions. The Astrophysical Journal Letters, 907(2), L25.

Weiss, R. (2022). Republication of: Electromagnetically coupled broadband gravitational antenna. Gen. Rel. Grav., 54(11), 153.

Wen, L. (2003). On the eccentricity distribution of coalescing black hole binaries driven by the Kozai mechanism in globular clusters. Astrophys. J., 598, 419–430.

Witten, E. (1984). Cosmic separation of phases. Phys. Rev. D, 30, 272–285.

Wong, K. W. K., Breivik, K., Kremer, K., & Callister, T. (2021). Joint constraints on the field-cluster mixing fraction, common envelope efficiency, and globular cluster radii from a population of binary hole mergers via deep learning. Phys. Rev. D, 103(8), 083021.

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

Woosley, S. E. (2019). The Evolution of Massive Helium Stars, Including Mass Loss. , 878(1), 49.

Woosley, S. E., & Heger, A. (2015). The Deaths of Very Massive Stars, (pp. 199–225). Cham: Springer International Publishing.

Woosley, S. E., Heger, A., & Weaver, T. A. (2002). The evolution and explosion of massive stars. Rev. Mod. Phys., 74, 1015–1071.

Wysocki, D., Lange, J., & O'Shaughnessy, R. (2019). Reconstructing phenomenological distributions of compact binaries via gravitational wave observations. Phys. Rev. D, 100, 043012.

Wysocki, D., O'Shaughnessy, R., Wade, L., & Lange, J. (2020). Inferring the neutron star equation of state simultaneously with the population of merging neutron stars.

Yang, Y., Bartos, I., Haiman, Z., Kocsis, B., Márka, S., & Tagawa, H. (2020). Cosmic evolution of stellar-mass black hole merger rate in active galactic nuclei. The Astrophysical Journal, 896(2), 138.

Ye, C., & Fishbach, M. (2022). Inferring the Neutron Star Maximum Mass and Lower Mass Gap in Neutron Star–Black Hole Systems with Spin. Astrophys. J., 937(2), 73.

Zevin, M., Bavera, S. S., Berry, C. P. L., Kalogera, V., Fragos, T., Marchant, P., Rodriguez, C. L., Antonini, F., Holz, D. E., & Pankow, C. (2021). One channel to rule them all? constraining the origins of binary black holes using multiple formation pathways. The Astrophysical Journal, 910(2), 152.

Zhu, X., Thrane, E., Osłowski, S., Levin, Y., & Lasky, P. D. (2018). Inferring the population properties of binary neutron stars with gravitational-wave measurements of spin. Phys. Rev. D, 98, 043002.

Zhu, X.-J., & Ashton, G. (2020). Characterizing astrophysical binary neutron stars with gravitational waves. The Astrophysical Journal Letters, 902(1), L12.

Zonca, A., Singer, L., Lenz, D., Reinecke, M., Rosset, C., Hivon, E., & Górski, K. (2019). Healpy: equal area pixelization and spherical harmonics transforms for data on the sphere in Python. J. Open Source Softw., 4(35), 1298.

Özel, F., & Freire, P. (2016). Masses, radii, and the equation of state of neutron stars. Annual Review of Astronomy and Astrophysics, 54(1), 401–440.

Özel, F., Psaltis, D., Güver, T., Baym, G., Heinke, C., & Guillot, S. (2016). The dense matter equation of state from neutron star radius and mass measurements. The Astrophysical Journal, 820(1), 28.

Appendix A

Details of low-latency trigger classification analysis

A.1 DERIVATION OF BAYES FACTORS FROM GSTLAL'S RANKING STATISTICS

In Sec. 3.2.2.3, we mentioned that among the various quantities assigned to a GW candidate by a search pipeline, the trigger data relevant to Bayes factor calculation can be thought to comprise only three quantities, namely the ranking-statistic, the representative template, and the SNR. For the GstLAL search pipeline, Eq. (3.29) suggests that it is indeed the case. However, in order to derive Eq. (3.29) itself, the details of how the ranking statistic is computed from additional pipeline-assigned quantities will become relevant.

For GstLAL, the full dataset corresponding to a particular trigger comprises additional quantities to t_k and ρ_k such as the signal-based veto parameter $\vec{\xi^2}$ at each detector. The other independent quantities that comprise the full dataset are measurements of arrival time and coalescence phase $(\vec{\Delta}t_k, \vec{\Delta}\phi_k)$ by different detectors expressed relative to the corresponding measurements by some reference detector $(t_{\mathrm{ref,k}}, \phi_{\mathrm{ref,k}})$. The log-likelihood ratio used by GstLAL to rank triggers can be expressed in terms of the probability distributions of these quantities Hanna et al. (2020); Tsukada et al. (2023) in the following way:

$$e^{x_k} \propto \frac{p(\vec{\xi^2}_k,\vec{\Delta}t_k,\vec{\Delta}\phi_k,t_{\mathrm{ref,k}},\phi_{\mathrm{ref,k}},\vec{O}_k|\rho_k,t_k,H_1)}{p(\vec{\xi^2}_k,\vec{\Delta}t_k,\vec{\Delta}\phi_k,t_{\mathrm{ref,k}},\phi_{\mathrm{ref,k}},\vec{O}_k|\rho_k,t_k,H_0)} \times \frac{P(t_k,\rho_k|H_1)}{P(t_k,\rho_k|H_0)} \tag{A.1}$$

where $\vec{O}$ is the subset of detectors that found the trigger with a signal-to-noise ratio (SNR) above threshold. While we do describe in Sec. 3.2.2.3 the calculation of $P(t_k, \rho_k|H_1)$, details regarding the other individual probabilities in Eq. (A.1) can be found in ref. Hanna et al. (2020). On the other hand, in terms of the same trigger-data used for calculating x_k, the Bayes factors required for $\mathrm{P}_{\mathrm{astro}}$ can be written as:

$$K_a(d_k) = \frac{p(\vec{\xi^2}_k,\vec{\Delta}t_k,\vec{\Delta}\phi_k,t_{\mathrm{ref,k}},\phi_{\mathrm{ref,k}},\vec{O}_k|\rho_k,t_k,H_a)}{p(\vec{\xi^2}_k,\vec{\Delta}t_k,\vec{\Delta}\phi_k,t_{\mathrm{ref,k}},\phi_{\mathrm{ref,k}},\vec{O}_k|\rho_k,t_k,H_0)} \times \frac{P(t_k,\rho_k|H_a)}{P(t_k,\rho_k|H_0)} \tag{A.2}$$

. We note that template t_k is the only parameter that determines which astrophysical source category the signal might have originated from in the sense that the only difference between astrophysical source categories is the population-level distributions of the template parameters corresponding to each category. Hence the probabilities of the remaining quantities $(\vec{\xi^2}_k, \vec{\Delta}t_k, \vec{\Delta}\phi_k, t_{\mathrm{ref,k}}, \phi_{\mathrm{ref,k}}, \vec{O}_k)$ conditional on H_a can be expected to equal the probabilities of the same quantities conditioned instead on H_1. Under this consideration, we can use Eq. (A.1) to rewrite Eq. (A.2) in the following way:

$$K_a(d_k) = Ae^{x_k} \frac{P(t_k,\rho_k|H_a)}{P(\rho_k,t_k|H_1)} \tag{A.3}$$

Since the category-specific signal hypotheses (H_a) correspond to disjoint regions of the template bank $(H_a \cap H_b = \emptyset \quad for \quad a \neq b)$ that together span the entire bank $(\cup_a H_a = H_1)$, we can

re-write the signal model in the denominator to be $P(\rho_k, t_k|H_1) \approx \sum_a P(t_k, \rho_k|H_a)P(H_a|H_1)$, where $P(H_a|H_1) \approx \frac{1}{3}$ for $a \in \{\text{BBH}, \text{BNS}, \text{NSBH}\}$. Under this consideration, Eq. (A.3) can be expressed as follows:

$$K_a(d_k) = Ae^{x_k} \frac{P(t_k,\rho_k|H_a)}{(\frac{1}{3})\sum_b P(t_k,\rho_k|H_b)} \tag{A.4}$$

To obtain an expression for the normalization constant A, we can make use of GstLAL's estimates of the probability distributions of its ranking statistic. The probability distribution of obtaining some value for the ranking statistic x can be obtained from the same probabilities used to calculate the ranking statistic itself conditional on either hypothesis H_1 *or* H_0, as follows:

$$p(x|H_{(0,1)}) = \sum_j \int_{\Sigma(x)} d\Sigma^{N-1} p(\vec{\xi_k^2}, \vec{\Delta t}_k, \vec{\Delta\phi}_k, t_{\text{ref,k}}, \phi_{\text{ref,k}}, \vec{O}_k|\rho_k, t_k, H_{(0,1)}) P(t_j, \rho_k|H_{(0,1)}) \tag{A.5}$$

where $\Sigma(x)$ is a surface in the space of the parameters $(\rho_k, \vec{\xi_k^2}, \vec{\Delta t}_k, \vec{\Delta\phi}_k, t_{\text{ref,k}}, \phi_{\text{ref,k}}, \vec{O}_k)$ corresponding to a fixed value of the ranking statistic x, as defined by Eq. (A.1). For details regarding how the integrals are implemented by means of importance sampling, see ref. Cannon et al. (2015); Messick et al. (2017). Using Eq. (A.1), we can re-write Eq. (A.5) for the ranking statistic distribution conditional on the signal hypothesis in terms of that conditional on the noise hypothesis in the following way:

$$p(x|H_1) = Ae^x \sum_j \int_{\Sigma(x)} d\Sigma^{N-1} p(\vec{\xi_k^2}, \vec{\Delta t}_k, \vec{\Delta\phi}_k, t_{\text{ref,k}}, \phi_{\text{ref,k}}, \vec{O}_k|\rho_k, t_k, H_0) P(t_j, \rho|H_0) \tag{A.6}$$

$$\implies p(x|H_1) = Ae^x p(x|H_0) \tag{A.7}$$

Now demanding that both $p(x|H_1)$ and $p(x|H_0)$ is normalized, we can find an expression for A in terms of the noise pdf:

$$A = \left[\int_{x_{min}}^{\infty} dx e^x p(x|H_0)\right]^{-1} \tag{A.8}$$

where x_{min} is the ranking statistic corresponding to the FAR threshold above which P_{astro} values are calculated for triggers and uploaded.

A.2 DERIVATION OF $P(t_k, \rho_k|t_j, \rho_k)$

In this appendix we derive Eq. (3.28) which evaluates the probability of a template t_k being triggered by a GW signal that matches exactly with template t_j. The GW strain data in a detector can be represented by a time-series d which can be expressed as a superposition of a potential signal s and noise n in the following way:

$$d = s + n \tag{A.9}$$

. Matched filtering involves computing the noise averaged inner product of this time-series data with the time-series corresponding to a template waveform. If the true signal is a perfect match to the jth template then the matched filter corresponding to the kth template can be expressed in the following way:

$$\rho_k = \rho_j t_j \cdot t_k + n_k \tag{A.10}$$

,where $\rho_k = t_k \cdot d$ is the matched filter SNR corresponding to template t_k, $n_k = t_k \cdot n$ and ρ_j is the SNR of the true signal s. As mentioned in Sec. 3.2.2.3, we are interested in quantifying how noise fluctuations might cause a true signal t_j to be highest SNR match with template t_k. Hence, as a first approximation, we can set $\rho_j = \rho_k$. Under this approximation and the assumption that the noise is stationary and Gaussian, we can compute the probability of template migration:

$$P(t_k, \rho_k | t_j, \rho_j) \propto \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}|n_k|^2} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\rho_k^2\left[1-2t_k\cdot t_j+(t_j\cdot t_k)^2\right]} \tag{A.11}$$

A.3 TEMPLATE WEIGHTS $P(t_k, \rho_k|H_a)$: VARIATION WITH TEMPLATE PARAMETERS AND SNR

In this appendix, we summarize how the template weights vary as a function of the template parameters and SNR. Given the form of the migration probability in Eq. (3.28), we expect there to be significant confusion among source categories that map onto neighboring regions of the template-bank at low SNR. As the SNR increases, we expect the probability of miss-classification to decrease with perfect classification at infinite SNR.

For the Salpeter function mass-model in Eq. (3.31) that shall be used by GstLAL to compute $\mathrm{P_{astro}}$ values in low-latency for O4, the template weights as a function of template masses are displayed in Fig. A.1. We find that the variation with SNR is exactly what we expect them to be.

Similarly, for the population model corresponding to the MDC injections (summarized in Eq. (3.32)), the template weights as a function of component masses is displayed in Fig. A.2 We also over plot the recovered parameters of the injections from category a found by the pipeline above threshold with SNR ρ, atop the plots of the template weights $P(t_k, \rho|H_a)$.

It can be seen that the density of blue points at various regions matches with the heat map representing $\log P(t_k, \rho|H_a)$. For example, in some SNR ranges wherein BNS injections are recovered in the NSBH region, $P(t_k, \rho|H_{\mathrm{BNS}})$ in that region corresponding to the SNR in question is also found to have non-zero value. This is implicative of the fact our template weights correctly model the mismatch between the recovered template and the true signal due to noise fluctuations, provided the population distribution corresponding to each category is exactly known.

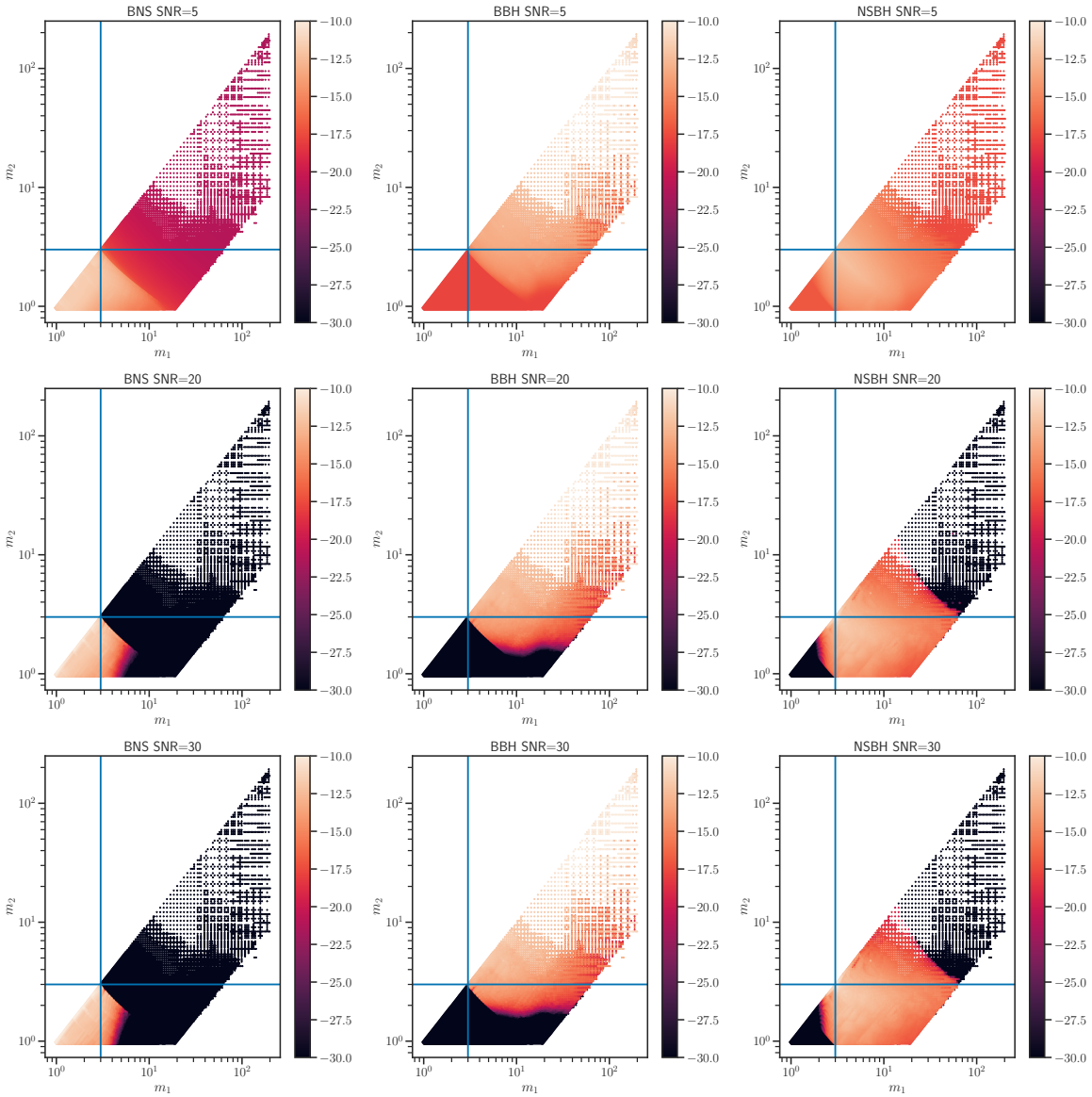


Figure A.1: Variation of the template weights corresponding to the Salpeter population distribution (i.e. the ones to be used in O4), as a function of component masses and SNR. The colorbars represent $\log P(t_k, \rho_k|H_a)$ with a and ρ specified at the top of each panel.

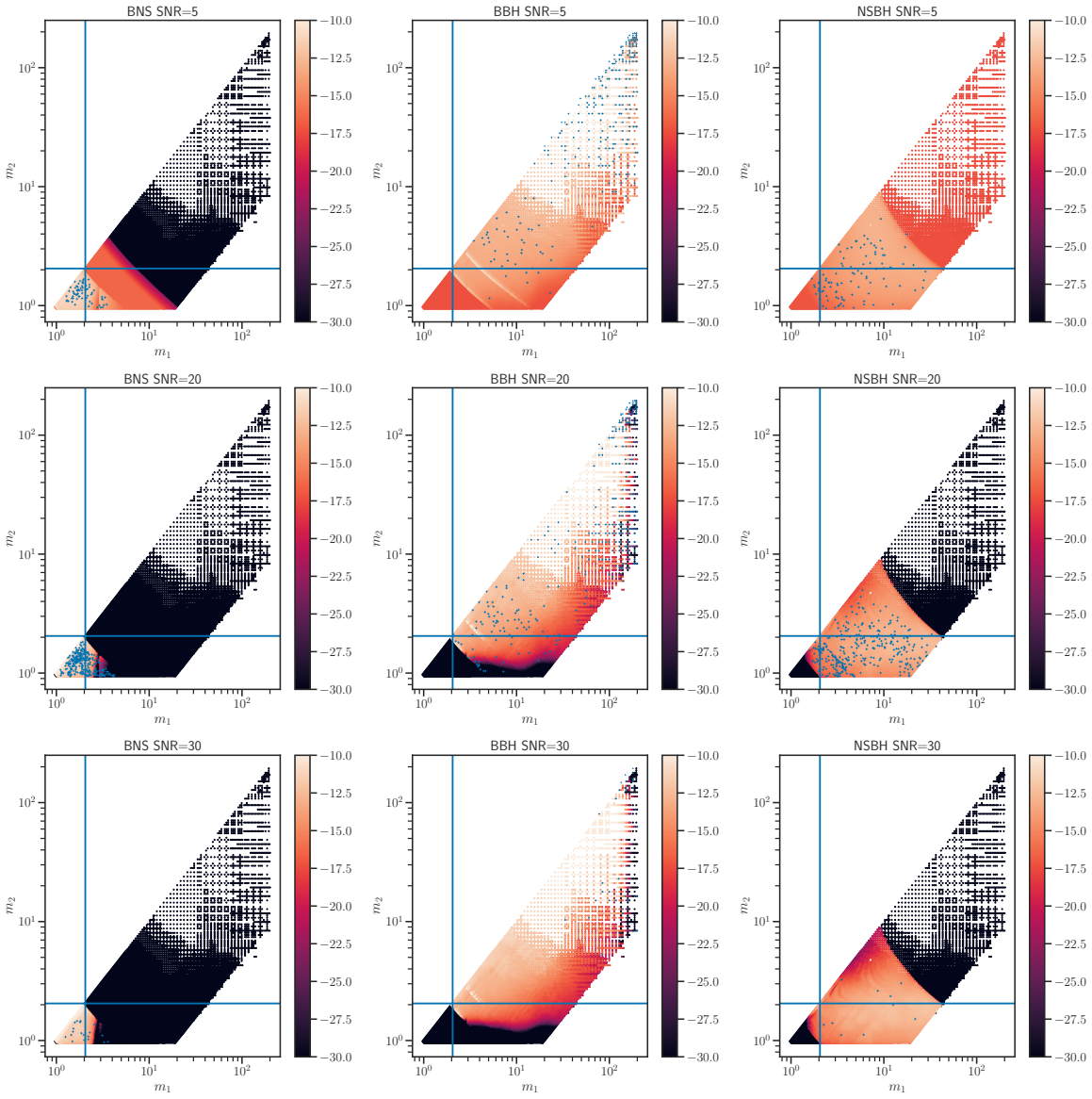


Figure A.2: Variation of the template weights corresponding to the population distribution of injections, as a function of component masses and SNR. The colorbars represent $\log P(t_k, \rho_k|H_a)$ with a and ρ specified at the top of each panel. The blue dots in each panel correspond to recovered masses of injections whose true source category coincides with the one specified on top of the panel. The SNRs of the injections in each panel lie within a narrow range about the SNR specified on top of the pannel

Appendix B

Data-driven population inference: analysis details

B.1 THE POSTERIOR

In this appendix, we summarize in detail how the various quantities required to evaluate the likelihood in Eq. (4.3) are computed from the GW posterior samples. We also provide an expression for the population posterior as an explicit function of the rate densities, and the hyper-parameters. For the ith GW observation being analyzed, single event parameter estimation (carried out for generating GW transient catalogs) yields posterior samples of masses and redshifts (converted from luminosity distance samples using a cosmological model), that can be used to compute *posterior weights* in each bin (Mohite, 2022):

$$w_{d_i}^{\gamma} = \frac{1}{N_{\text{samples},i}} \sum_{(m_{1j},m_{2j},z_j)\sim p(m_1,m_2,z|d_i)} \frac{\delta^{\gamma,\alpha(m_{1j},m_{2j},z_j)} \frac{dV_c}{dz}|_{z=z_j} \frac{1}{1+z_j}}{p_{\text{PE}}(m_{1j},m_{2j},z_j) m_{1j} m_{2j}} \tag{B.1}$$

where $\alpha(m_1, m_2, z)$ is the index of the bin inside which m_1, m_2, z lies, $p_{\text{PE}}(m_{1j}, m_{2j}, z_j)$ the fiducial prior used during single event PE, $N_{\text{samples},i}$ the number of posterior samples drawn for the ith observation and $\delta^{\gamma,\gamma'}$ the Kroneker delta function. Explicitly, the fiducial PE priors used throughout this study are uniform in detector frame masses and luminosity distance squared. We downsample single-event PEs which have a larger number of samples so as to enforce that each event has an exactly identical number of posterior samples prior to weight computation. These weights can be pre-computed for each event corresponding to every bin for computational efficiency. Similarly, the set of detectable simulated events generated for computing N_{det}, can be used to compute the sensitive spacetime hypervolume corresponding to each bin in the following way (Mohite, 2022):

$$\langle \text{VT} \rangle^{\gamma} = \sum_{j\sim(\text{det,draw})} \frac{\delta^{\gamma,\alpha(m_{1j},m_{2j},z_j)} T_{\text{obs}}}{K_{draw}} \frac{\frac{dV_c}{dz}|_{z=z_j} \frac{1}{1+z_j} p(\vec{\lambda}_j|\text{draw})}{p(m_{1j},m_{2j},z_j,\vec{\lambda}_j|\text{draw}) m_{1j} m_{2j}} \tag{B.2}$$

where $\vec{\lambda}$ are parameters that characterize a GW waveform in addition to masses and luminosity distance (i.e. the parameters that are not binned up), $p(...|\text{draw})$ the fiducial distribution from which simulations are drawn, $j \sim$ (det,draw) signifies simulated events that pass the detection threshold (Farr, 2019). The sensitive hypervolumes, like the posterior weights, can also be pre-computed for efficiency.

The posterior weights and the hypervolumes can be used to evaluate the likelihood as a fast-evaluating function of the merger rate densities:

$$p(\vec{d}|\vec{n}) = e^{-\sum_{\gamma} n^{\gamma} \langle \text{VT} \rangle^{\gamma}} \prod_i \sum_{\gamma} w_{d_i}^{\gamma} n^{\gamma} \tag{B.3}$$

To obtain an expression for the posterior as an explicit function of the rate densities and the hyper-parameters, we must do so for the GP prior as well. The explicit functional form of the log GP prior for an exponential quadratic kernel looks like the following:

$$\log p(\vec{n}|\vec{\mu},\sigma,\lambda) = -\tfrac{1}{2}\log|\Sigma(\sigma,\lambda)| - \tfrac{1}{2}\sum_{\gamma\gamma'}\left\{(\log n_\gamma - \mu_\gamma)(\log n_{\gamma'} - \mu_{\gamma'})\Sigma^{-1}_{\gamma,\gamma'}(\sigma,\lambda)\right\} \tag{B.4}$$

$$\Sigma_{\gamma\gamma'}(\sigma,\lambda) = \sigma^2 \exp\left[-\sum_{\theta\in\{m_1,m_2,z\}} \frac{(c_{\gamma,\theta}-c_{\gamma',\theta})^2}{2\lambda_\theta^2}\right] \tag{B.5}$$

Where, $|A|$ is the determinant of the matrix A. In terms of the likelihood in Eq. (B.3) and the explicit form of the GP prior, the log of the posterior in Eq. (4.7) can be written as:

$$\log p(\vec{n},\vec{\mu},\sigma,\vec{\lambda}|\{w_d\},\langle \mathrm{VT}\rangle) = \log p(\vec{\mu},\sigma,\vec{\lambda}) - \tfrac{1}{2}\log|\Sigma(\sigma,\lambda)| - \tfrac{1}{2}\sum_{\gamma,\gamma'}\{(\log n_\gamma - \mu_\gamma)\times$$
$$(\log n_{\gamma'} - \mu_{\gamma'})\Sigma^{-1}_{\gamma\gamma'}(\sigma,\lambda)\} - \sum_\gamma n^\gamma \langle \mathrm{VT}\rangle^\gamma + \sum_i \log\left(\sum_\gamma w^\gamma_{d_i} n^\gamma\right) + const. \tag{B.6}$$

Note that while $\{w_d\}$ replaces the conditional dependence of the posterior on $\vec{d}$, the conditional dependence on $\langle \mathrm{VT}\rangle$ was not explicitly mentioned before. However, since $\langle \mathrm{VT}\rangle$ is computed empirically from a dataset of simulated sources, our likelihood and hence posterior was always in fact conditional on $\langle \mathrm{VT}\rangle$. The reason for righting it explicitly at this point of the derivation will become clear in the next appendix wherein we discuss the Monte Carlo uncertainties in $\langle \mathrm{VT}\rangle$. However, prior to investigating the Monte Carlo uncertainties in the selection function we first describe the convergence of the event-specific Monte Carlo integrals in Eq. (B.1).

To check for the convergence of event-specific Monte Carlo integrals we first compute the *variance* of the posterior weight corresponding to each bin:

$$\mathrm{Var}\left[\mathrm{w}_\mathrm{i}^\gamma\right] = \frac{1}{N_{\mathrm{samples},i}} \sum_{(m_{1j},m_{2j},z_j)\sim p(m_1,m_2,z|d_i)} \frac{\delta^{\gamma,\alpha(m_{1j},m_{2j},z_j)}\left(\frac{dV_c}{dz}|_{z=z_j}\frac{1}{1+z_j}\right)^2}{p^2_{\mathrm{PE}}(m_{1j},m_{2j},z_j)m^2_{1j}m^2_{2j}} \tag{B.7}$$

Using these variances one can construct metrics for testing the convergence of single-event Monte Carlo integrals, such as the number of effective samples:

$$N_{eff,i} = \frac{\sum_\gamma (n^\gamma)^2 \mathrm{Var}\left[\mathrm{w}_\mathrm{i}^\gamma\right]}{\left(\sum_\gamma n^\gamma w_i^\gamma\right)^2} \tag{B.8}$$

Following Callister & Farr (2024), to ensure convergence of the event-specific Monte Carlo integrals, we demand that for each draw of $\vec{n}$ from the hyper-posterior, the condition $\min_i \log_{10} N_{eff,i} > 0.6$ be satisfied. However, given the size of the bins chosen for our study and the hyper-priors on the covariance amplitude, we expect these conditions to hold for most draws from the hyper-posterior.

Hence, instead of penalizing the likelihood upon violation of said condition, we verify in post-processing that for all rate-density samples, this condition is automatically satisfied. On the other hand, for future studies that aim to implement inference at a much higher bin resolution, for efficient sampling, we propose that this condition is imposed during sampling in the form of a steep penalty on the likelihood, similar to what is done in Callister & Farr (2024).

B.2 UNCERTAINTIES IN $\langle \mathrm{VT} \rangle$ ESTIMATION

Empirically estimated $\langle \mathrm{VT} \rangle$s are subject to Monte Carlo uncertainties (Farr, 2019). According to the central limit theorem, the sum over samples in Eq. (B.2) can be interpreted as the realization of a Gaussian distribution centered around the true value of the integral being approximated by the mentioned sum. Since we only have a finite number of samples, the Gaussian can be expected to have a finite width as well:

$$\langle \mathrm{VT} \rangle^{\gamma} \sim \mathcal{N}(\mu_{\mathrm{VT}}^{\gamma}, \sigma_{\mathrm{VT}}^{\gamma}) \tag{B.9}$$

. Large Monte Carlo uncertainties in $\langle \mathrm{VT} \rangle$ in some bins due to the sparsity of simulated sources can lead to inaccurate estimation of the astrophysical rate densities in said bins. To get around this problem, it is possible to marginalize our posterior distribution over the mentioned uncertainties in $\langle \mathrm{VT} \rangle$ estimation (Farr, 2019).

In order to carry out such a marginalization, we first need to estimate $\mu_{\mathrm{VT}}^{\gamma}$ and $\sigma_{\mathrm{VT}}^{\gamma}$. Following Farr (2019), we approximate these two quantities from the single realization of samples used to compute the Monte Carlo sum, as in:

$$\mu_{\mathrm{VT}}^{\gamma} \approx \sum_{j \sim (\mathrm{det,draw})} \frac{\delta^{\gamma,\alpha(m_{1j},m_{2j},z_j)} T_{\mathrm{obs}}}{K_{draw}} \frac{\frac{dV_c}{dz}|_{z=z_j} \frac{1}{1+z_j} p(\vec{\lambda}_j|\mathrm{draw})}{p(m_{1j},m_{2j},z_j,\vec{\lambda}_j|\mathrm{draw}) m_{1j} m_{2j}} \tag{B.10}$$

$$(\sigma_{\mathrm{VT}}^{\gamma})^2 \approx \sum_{j \sim (\mathrm{det,draw})} \frac{\delta^{\gamma,\alpha(m_{1j},m_{2j},z_j)} T_{\mathrm{obs}}^2}{K_{\mathrm{draw}}^2} \left[\frac{\frac{dV_c}{dz}|_{z=z_j} \frac{1}{1+z_j} p(\vec{\lambda}_j|\mathrm{draw})}{p(m_{1j},m_{2j},z_j,\vec{\lambda}_j|\mathrm{draw}) m_{1j} m_{2j}} \right]^2 - \frac{\left(\mu_{\mathrm{VT}}^{\gamma}\right)^2}{K_{draw}} \tag{B.11}$$

With these estimates of $\mu_{\mathrm{VT}}^{\gamma}$ and $\sigma_{\mathrm{VT}}^{\gamma}$ it is possible to compute the posterior distribution of rate densities and hyper-parameters that have been marginalized over the Monte Carlo uncertainties in $\langle \mathrm{VT} \rangle$ estimation:

$$\begin{aligned} p(\vec{n}, \vec{\mu}, \sigma, \vec{\lambda} | \{w_d\}, \{\mu_{\mathrm{VT}}^{\gamma}\}, \{\sigma_{\mathrm{VT}}^{\gamma}\}) = \int & p(\vec{n}, \vec{\mu}, \sigma, \vec{\lambda} | \{w_d\}, \{\langle \mathrm{VT} \rangle\}) \\ & \times \left(\prod_{\gamma} d \langle \mathrm{VT} \rangle^{\gamma} \, p(\langle \mathrm{VT} \rangle^{\gamma} | \mu_{\mathrm{VT}}^{\gamma}, \sigma_{\mathrm{VT}}^{\gamma}) \right) \end{aligned} \tag{B.12}$$

where $p(\langle \mathrm{VT} \rangle^{\gamma} | \mu_{\mathrm{VT}}^{\gamma}, \sigma_{\mathrm{VT}}^{\gamma})$ is the normal distribution in Eq. (B.9). The integral in Eq. (B.12) can be

evaluated analytically to obtain:

$$\log p(\vec{n}, \vec{\mu}, \sigma, \vec{\lambda} | \{w_d\}, \{\mu^\gamma_{\rm VT}\}, \{\sigma^\gamma_{\rm VT}\}) = \log p(\vec{\mu}, \sigma, \vec{\lambda}) - \tfrac{1}{2} \log |\Sigma(\sigma, \lambda)| - \\ \tfrac{1}{2} \sum_{\gamma, \gamma'} \left\{ (\log n_\gamma - \mu_\gamma)(\log n_{\gamma'} - \mu_{\gamma'}) \Sigma^{-1}_{\gamma\gamma'}(\sigma, \lambda) \right\} \sum_\gamma n^\gamma \mu^\gamma_{\rm VT} + \\ \tfrac{1}{2} \sum_\gamma (n^\gamma \sigma^\gamma_{\rm VT})^2 + \sum_i \log \left(\sum_\gamma w^\gamma_{d_i} n^\gamma \right) + const. \tag{B.13}$$

It can be seen in Eq. (B.13) that the marginalized posterior is not normalizable due to the $+(n^\gamma)^2$ term. This implies that expectation values of the hyper-parameters $\vec{\mu}, \sigma, \lambda$ with respect to the marginalized posterior become arbitrarily large. This is corrected for by imposing an additional constraint on $\vec{n}^\gamma$ during sampling, which is:

$$n^\gamma \mu^\gamma_{\rm VT} \leq 2 \left(\frac{\mu^\gamma_{\rm VT}}{\sigma^\gamma_{\rm VT}} \right)^2 \implies N^\gamma_{\rm det} \leq 2 N^\gamma_{\rm eff} \tag{B.14}$$

. This condition is similar to the one derived in Farr (2019) for parametric population inference. Here $N^\gamma_{\rm eff}$ is the effective number of independent samples of simulated events in the γth bin. Hence the condition in Eq. (B.14) is implicative of the fact that the effective number of simulated events in a bin that are required for accurate VT estimation should be higher than the expected number of detectable events in said bin. It can be imposed by rejecting samples of n^γ for which the condition is not satisfied.

Given the small number of observations analyzed for each study described in Sec. 4.3.3 and the correspondingly large number of simulated events used in VT estimation, this condition and the marginalized posterior in Eq. (B.13) were not implemented. On the other hand, it was verified in post-processing that for each sample of n^γ, the condition $N^\gamma_{\rm det} << N^\gamma_{\rm eff}$ was satisfied automatically. However, as summarized in this appendix, future studies that might attempt to use a sparse dataset of simulated events in conjunction with a large number of observations can straightforwardly implement the condition in Eq. (B.14) as well as the marginalized posterior in Eq. (B.13) to prevent inaccurate VT and hence rate estimation.

On the other hand, Essick & Farr (2022) note that marginalizing over Monte Carlo uncertainties in the manner described so far might lead to biases in the inferred population, that in the case of parametric modeling, can only be resolved in computationally expensive ways. Their argument is based on the fact that Monte Carlo uncertainties in the selection function are correlated between different points in the space of population hyper-parameters. The marginalization procedure described in Farr (2019), on which ours is based, assumes that the Monte Carlo uncertainties corresponding to different values of population hyper-parameters are uncorrelated which leads to biases in the inferred distribution as compared to using the point estimates directly without marginalization (Essick & Farr, 2022). Hence, Essick & Farr (2022) argue that the resolution of this bias requires one to either use a much larger number of samples as compared to the scenario wherein the point estimates are used directly or to account for the aforementioned correlations by evaluating the point estimates on a multi-dimensional grid of population hyper-parameters. Given that

both these methods are expensive to implement for parametric models, they recommend the use of point estimates directly instead of marginalization.

However, these concerns are not applicable in the context of our population model due to the following reason. In our case, the "population hyperparameters" are the rate densities themselves. As described before, the mean and variance of the number of detectable events, as a function of the rate densities, are just sums of the rate densities raised to sum power, weighted by *pre-computable* Monte Carlo integrals. Hence the first concern raised by Essick & Farr (2022) is not applicable to our model since the Monte Carlo sums over detectable simulations are only computed once and unlike parametric models, not for every draw from the hyper-posterior, leading to the net computational cost of our analyses remaining unchanged irrespective of how many detectable samples are used.

Similarly, accounting for correlations while marginalizing over Monte Carlo uncertainties in our model can be implemented without any increase in computational cost. In the context of our model, correlations in the selection function between two points in the space of population hyper-parameters translate to correlations between the $\langle \mathrm{VT} \rangle$ estimates of different bins. Hence, unlike parametric models, we need not compute Monte Carlo sums on a multi-dimensional grid of population hyper-parameters for each draw from the hyper-posterior. Instead we need only compute them once, that too only for all possible bin pairs.

For example, to account for these correlations, we need to replace Eq. (B.9) with a multi-variate correlated Gaussian:

$$\{\langle \mathrm{VT} \rangle\} \sim \mathcal{N}(\{\mu_{\mathrm{VT}}\}, \Sigma_{\mathrm{VT}}) \tag{B.15}$$

where Σ_{VT} is the covariance matrix of point estimates among the different bins. Following Essick & Farr (2022), we can estimate this covariance matrix from the Monte Carlo samples in the following way:

$$\begin{aligned}\Sigma^{\gamma,\gamma'}_{\mathrm{VT}} = \frac{1}{K_{draw}} \Bigg[\frac{K_{\mathrm{det}}}{(K_{\mathrm{det}}-1)K_{draw}} \sum_{j\sim(\mathrm{det,draw})} \Bigg\{ \Bigg(\delta^{\gamma,\alpha(m_{1j},m_{2j},z_j)} \frac{T_{\mathrm{obs}} \frac{dV_c}{dz}|_{z=z_j} \frac{1}{1+z_j} p(\vec{\lambda}_j|\mathrm{draw})}{p(m_{1j},m_{2j},z_j,\vec{\lambda}_j|\mathrm{draw}) m_{1j} m_{2j}} - \\ \frac{K_{det}}{K_{draw}} \mu^{\gamma}_{\mathrm{VT}} \Bigg) \times \Bigg(\delta^{\gamma',\alpha(m_{1j},m_{2j},z_j)} \frac{T_{\mathrm{obs}} \frac{dV_c}{dz}|_{z=z_j} \frac{1}{1+z_j} p(\vec{\lambda}_j|\mathrm{draw})}{p(m_{1j},m_{2j},z_j,\vec{\lambda}_j|\mathrm{draw}) m_{1j} m_{2j}} - \frac{K_{det}}{K_{draw}} \mu^{\gamma'}_{\mathrm{VT}} \Bigg) \Bigg\} + \\ \frac{K_{draw}-K_{\mathrm{det}}}{K_{det}} \mu^{\gamma}_{\mathrm{VT}} \mu^{\gamma'}_{\mathrm{VT}} \Bigg]\end{aligned} \tag{B.16}$$

Note that Eq. (B.16) is pre-computable given samples of simulated events. Once the covariances are estimated one can proceed to carry out the marginalization in Eq. (B.12) with the correlated multivariate Gaussian of Eq. (B.15):

$$p(\vec{n}, \vec{\mu}, \sigma, \vec{\lambda}|\{w_d\}, \{\mu^{\gamma}_{\mathrm{VT}}\}, \{\sigma^{\gamma}_{\mathrm{VT}}\}) = \int p(\vec{n}, \vec{\mu}, \sigma, \vec{\lambda}|\{w_d\}, \{\langle \mathrm{VT} \rangle\})$$

$$\times\, p(\{\langle \mathrm{VT}\rangle\}\,|\,\{\mu_{\mathrm{VT}}\}\,,\Sigma_{\mathrm{VT}})\left(\prod_{\gamma} d\,\langle \mathrm{VT}\rangle^{\gamma}\right) \tag{B.17}$$

. As before the marginalization integral can be carried out analytically to obtain the following expression of the posterior:

$$\begin{aligned}\log p(\vec{n},\vec{\mu},\sigma,\vec{\lambda}|\{w_d\},\{\mu^{\gamma}_{\mathrm{VT}}\},\Sigma^{\gamma\gamma'}_{\mathrm{VT}}) &= \log p(\vec{\mu},\sigma,\vec{\lambda}) - \tfrac{1}{2}\log|\Sigma(\sigma,\lambda)| \\ &- \tfrac{1}{2}\sum_{\gamma,\gamma'}\left\{(\log n_{\gamma}-\mu_{\gamma})(\log n_{\gamma'}-\mu_{\gamma'})\Sigma^{-1}_{\gamma\gamma'}(\sigma,\lambda)\right\}\sum_{\gamma} n^{\gamma}\mu^{\gamma}_{\mathrm{VT}} \\ &+ \tfrac{1}{2}\sum_{\gamma,\gamma'}\left(n^{\gamma}n^{\gamma'}\Sigma^{\gamma\gamma'}_{\mathrm{VT}}\right) + \sum_{i}\log\left(\sum_{\gamma} w^{\gamma}_{d_i} n^{\gamma}\right) + const.\end{aligned} \tag{B.18}$$

Similar to the case of uncorrelated marginalization, the posterior in Eq. (B.18) is not normalizable unless we impose the following condition:

$$n^{\gamma}\mu^{\gamma}_{\mathrm{VT}} \;\leq\; 2\sum_{\gamma'}\left(\mu^{\gamma}_{\mathrm{VT}}\mu^{\gamma'}_{\mathrm{VT}}\left[\Sigma^{-1}_{\mathrm{VT}}\right]^{\gamma\gamma'}\right) \tag{B.19}$$

in the form of a likelihood penalization. Note that Eq. (B.19) holds the same meaning as Eq. (B.14) with the only difference being that the number of effective samples of simulated events in each bin is now calculated while accounting for correlations with other bins. As before, we do not impose this condition during sampling in the form of a likelihood cut due to the small number of observations and the correspondingly large number of detectable simulations. However, we do verify in post-processing that this condition is automatically satisfied for all the rate density samples.

To summarize, marginalization over Monte Carlo uncertainties is implementable straightforwardly within our framework. The concerns raised by Essick & Farr (2022) are not applicable to our binned model since, correlated or otherwise, the Monte Carlo sums required to evaluate our marginalized posterior are pre-computable. Hence, unlike parametric models, we can correctly implement marginalization over Monte Carlo uncertainties in the manner recommended by Essick & Farr (2022) while suffering no increase in computational cost.

B.3 EFFECTS OF CHANGING THE CHOICE OF BINNING FOR MASS-ONLY INFERENCE

In this appendix, we summarize the results referred to in Sec. 4.3.2.6 to demonstrate that different bin choices yield consistent results. Given the scalability of the current implementation with the dimensionality of the parameter space and the number of bins along each parameter dimension, we carry out this study in the context of the two-dimensional version of our model that was used by Abbott et al. (2023d) to infer non-parametric constraints on the CBC mass-spectrum. The three-dimensional models discussed in this work reduce to the mentioned two-dimensional one upon

fixing the redshift evolution of the CBC population to a function that is uniform in co-moving volume and completely uncorrelated with the masses (Abbott et al., 2023d; Mohite, 2022).

For this study, we generate a simplistic population of simulated BBHs that is a smoothed truncated power-law in primary mass and mass ratio. The redshift evolution of the merger rate is chosen to be uniform resulting in a redshift population that matches exactly with the assumptions of the above-mentioned two-dimensional model. The functional form of the underlying population and the corresponding hyperparameters that characterize said function are specified in Eq. (B.20) and table B.1 respectively. We generate two realizations of this mock population each comprising a set of 100 BBH events. We simulate measurement uncertainty in the BBH parameters and detection sensitivity following the methodology described in (Farah et al., 2023; Fishbach et al., 2020, 2018) using the exact same procedure and PSDs summarized in Sec. 4.3.3.1.

$$\frac{dR}{dm_1 dm_2}(z) \quad \propto \quad m_1^{-\alpha} \left(\frac{m_2}{m_1}\right)^{\beta_q} \tag{B.20}$$

Hyper-Parameter	True value
α	3.14
β_q	1.7
$m_{\rm min}$	$4.5M_\odot$
$m_{\rm max}$	$55M_\odot$

Table B.1: True values for the hyper-parameters characterizing the underlying population in Eq. (B.20)

We analyze the aforementioned realizations of this simulated BBH population using the two-dimensional BGP model corresponding to two different binning choices, with one containing twice as many bins as the other. The resulting inference is summarized in Figs. Figs. B.1 and B.2.

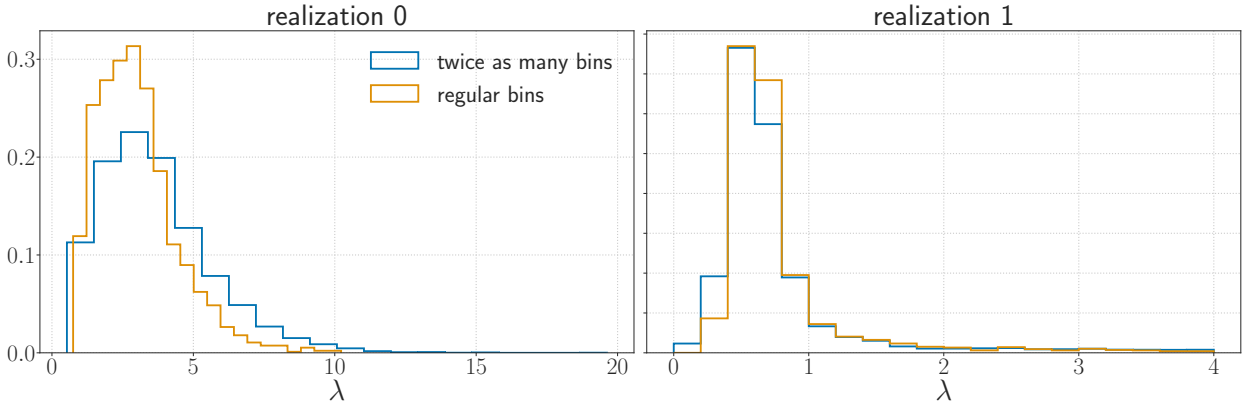


Figure B.1: Effects of doubling the number of bins on the inferred lengthscale posterior

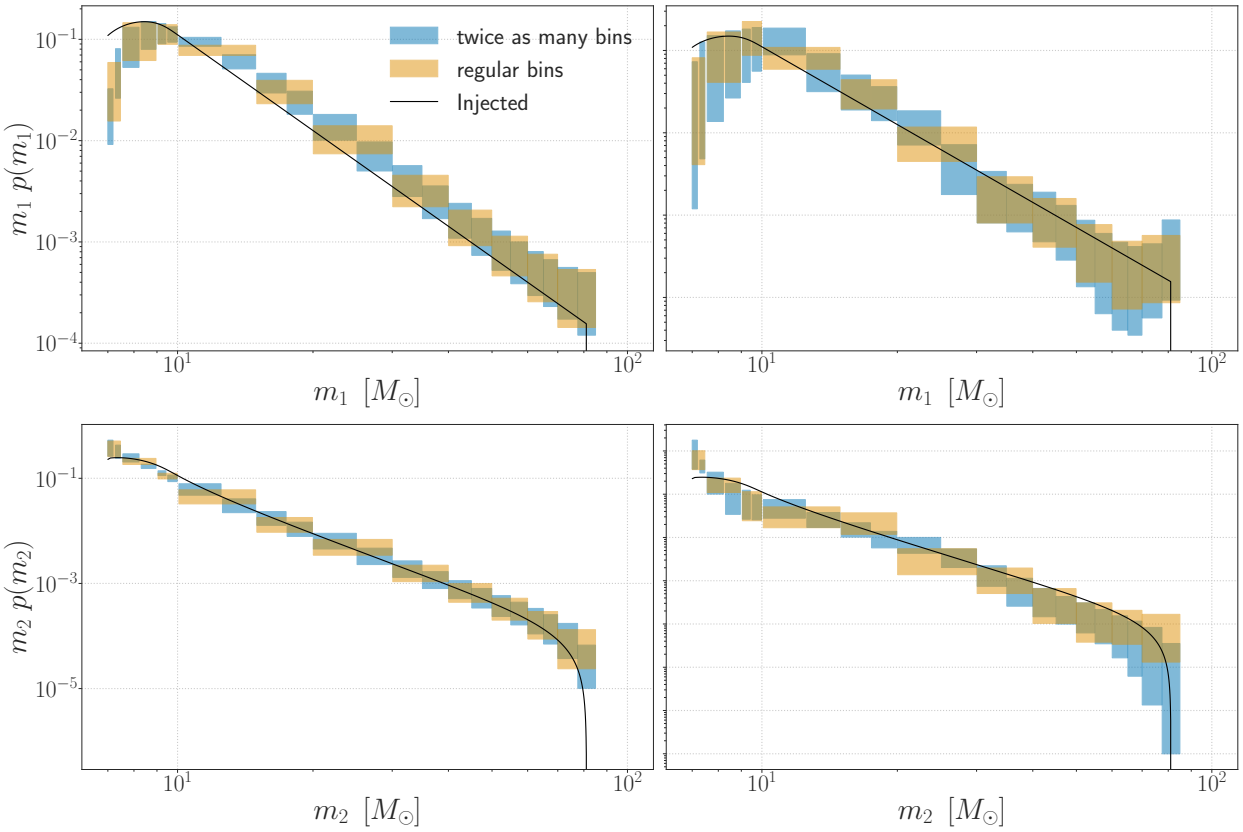


Figure B.2: Effects of doubling the number of bins on the inferred population constraints

It can be seen in Fig. B.1 that upon doubling the number of bins, the inferred posterior of the GP length scale remains unchanged which is implicative of the GP correlating twice as many bins within the same interval of log component masses. This leads to the inferred population constraints displayed in Fig. B.2, corresponding to the two different choices of binning, to be fully consistent with each other as well with the injected population up to measurement uncertainties. We thus conclude that increasing bin resolution beyond what is needed to identify features in the underlying population leads to constraints that are consistent with those obtained from a lower-resolution inference, provided that the latter already captures all the features in the underlying population.

B.4 UPDATING THE BEST-FIT POPULATION

Once the rate densities are inferred from the data comprising say N_{obs} events, the best-fit population distribution can be re-constructed from the expectation values of their posterior samples:

$$p_{N_{obs}}(\vec{\theta}) \quad = \quad \int d\vec{n} d\vec{\mu} d\sigma d\vec{\lambda} p(\vec{n}, \vec{\mu}, \sigma, \vec{\lambda} | \{w_d\}^{N_{obs}}, \{\mu_{\rm VT}^{\gamma}\}, \{\sigma_{\rm VT}^{\gamma}\}) \frac{dN}{d\vec{\theta}}(\vec{n}) \quad \text{(B.21)}$$

$$\implies p_N(m_1, m_2, z) \propto \sum_{\gamma} \frac{\langle n^{\gamma} \rangle_{N_{obs}} \delta^{\gamma, \alpha(m_1, m_2, z)}}{m_1 m_2} \frac{dV}{dz} \frac{1}{1+z} \tag{B.22}$$

where $\{w_d\}^{N_{obs}}$ is the posterior weights for the N_{obs} observed events and $\langle n^{\gamma} \rangle_{N_{obs}}$ is the expectation value of n^{γ} with respect to the posterior $p(\vec{n}, \vec{\mu}, \sigma, \vec{\lambda} | \{w_d\}^{N_{obs}}, \{\mu^{\gamma}_{\rm VT}\}, \{\sigma^{\gamma}_{\rm VT}\})$. In various scenarios, when a new event is found, existing population fits are used to re-weight the estimated parameters of said event *after* updating the existing fit with data from the new event. This is often carried out to obtain population-informed measurements of the parameters corresponding to an interesting event such as a BNS or an NSBH. If the new event in question is found early on during an ongoing observing run then the updated population fit is often computed under the assumption that there has been no change in the sensitivity of the detectors between the past and ongoing observing runs. In such a scenario, our best-fit population distribution can be updated without re-running the entire analysis of the combined dataset comprising the existing events and the new one.

Assuming no change in detector sensitivity, the posterior distribution of the rate densities given the combined dataset can be expressed in terms of the posterior distribution obtained from past events and the posterior weights of the new event, as in:

$$p(\vec{n}, \vec{\mu}, \sigma, \vec{\lambda} | \{w_d\}^{N_{obs}+1}, \{\mu^{\gamma}_{\rm VT}\}, \{\sigma^{\gamma}_{\rm VT}\}) = A p(\vec{n}, \vec{\mu}, \sigma, \vec{\lambda} | \{w_d\}^{N_{obs}}, \{\mu^{\gamma}_{\rm VT}\}, \{\sigma^{\gamma}_{\rm VT}\}) \sum_{\gamma} n^{\gamma} w^{\gamma}(d_{N_{obs}+1}) \tag{B.23}$$

where $d_{N_{obs}+1}$ denotes the data corresponding to the new event. The normalization constant A can be obtained given the assumption that the existing posterior distribution $p(\vec{n}, \vec{\mu}, \sigma, \vec{\lambda} | \{w_d\}^{N_{obs}}, \{\mu^{\gamma}_{\rm VT}\}, \{\sigma^{\gamma}_{\rm VT}\})$ is normalized, as in:

$$A = \frac{1}{\sum_{\gamma} \langle n^{\gamma} \rangle w^{\gamma}(d_{N_{obs}+1})} \tag{B.24}$$

The expectation values of the rate densities with respect to the updated posterior can thus be obtained in the following way:

$$\langle n^{\gamma} \rangle_{N_{obs}+1} = \langle n^{\gamma} \rangle_{N_{obs}} + \frac{\sum_{\gamma'} cov_{N_{obs}}(n^{\gamma}, n^{\gamma'}) w^{\gamma'}(d_{N_{obs}+1})}{\sum_{\gamma'} \langle n^{\gamma'} \rangle w^{\gamma'}(d_{N_{obs}+1})} \tag{B.25}$$

, where $cov_{N_{obs}}(n^{\gamma}, n^{\gamma'})$ is the covariance matrix of the rate densities with respect to the existing population fit. Since both $cov_{N_{obs}}(n^{\gamma}, n^{\gamma'})$ and $\langle n^{\gamma} \rangle$ are obtainable from the existing posterior samples of the rate densities, the updated expectation values in Eq. (B.25) can be computed without re-running the population inference on the updated dataset. Once obtained, these updated expectation values can be used to update the best-fit population distribution through Eq. (B.25), for the purpose of re-weighting the new events PE as well as to study the impact of the new event on existing population constraints.

B.5 VARIATION OF THE REDSHIFT EVOLUTION PARAMETER IN THE MASS-SPIN INFERENCE

In this appendix we demonstrate that our results for GWTC-3 are robust against variations in the redshift evolution parameter κ, which we had originally fixed to be the median value found by Abbott et al. (2023d) using the POWERLAW+PEAK model from the same dataset ($\kappa = 2.9$). Here we compare our results for the the joint mass-spin distributions for two additional choices of κ which are taken to be the 90% lower ($\kappa = 1.1$) and upper ($\kappa = 4.6$) bounds on it as reported by Abbott et al. (2023d).

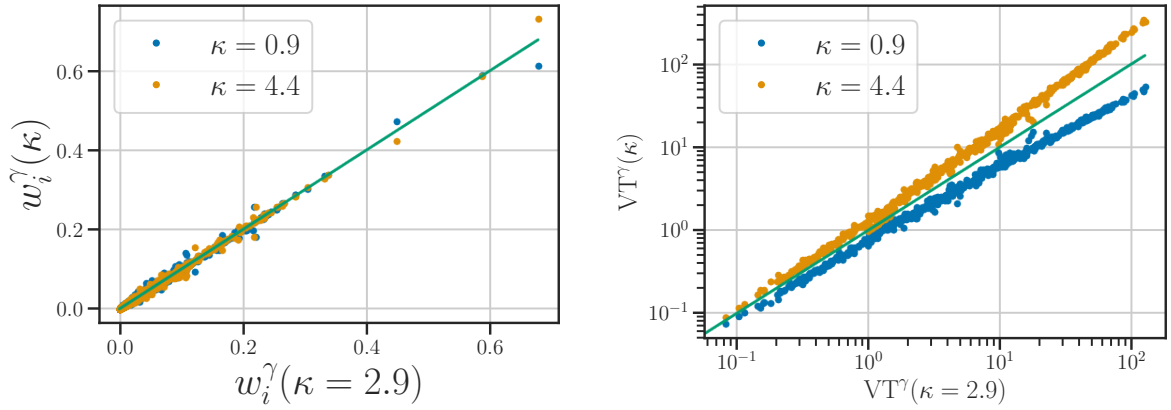


Figure B.3: Variation of posterior weights (w_i^γ, *left*) and detectable time-volumes ($\mathrm{VT}_\mathrm{i}^\gamma$,*right*) of Eq. (4.15) in each bin with κ.

In figure B.3, we first show how the posterior weights (w_i^γ) and detectable time-volumes ($\mathrm{VT}_\mathrm{i}^\gamma$) of Eq. (4.15) required to compute our population likelihood vary with κ. In Figures. B.4, and B.5, we show that the trends we find in the astrophysical mass and spin distributions of GWTC-3 BBHs are robust against variations in κ with the upper and lower bounds of its measured values that were reported by Abbott et al. (2023d). We therefore conclude that our conclusions regarding the mass-spin sub-populations are not biased by our restrictions on κ.

B.6 VARIATION OF BINNING CHOICES: MASS-SPIN INFERENCE

In this appendix, we show that the trends we find in the mass-spin population of GWTC-3 BBHs are robust against changes in choices of binning, for both masses and effective inspiral spins. We change both the location and number of bins which is done in two ways. In the first case we change the mass bins to while keeping the same χ_eff bins as in Sec. 4.4.3. In the second case we change the χ_eff bins while keeping the mass bins the same as they were in Sec. 4.4.3.

In both scenarios, we recover the same trends as in Sec. 4.4.3. This is expected given that the GP hyper-parameters such as the correlation lengths are themselves inferred from the data, with

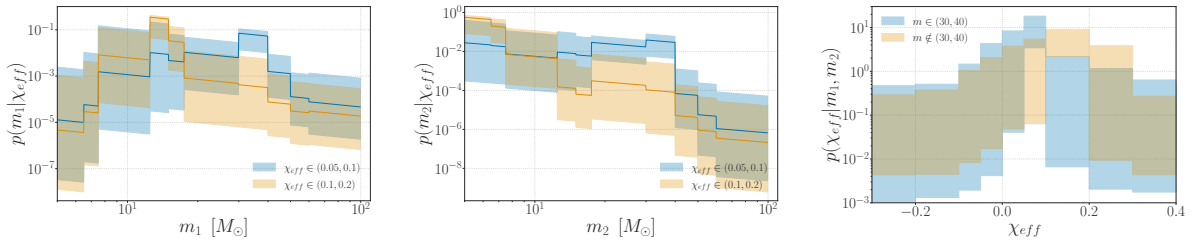


Figure B.4: GWTC-3 results for $\kappa = 1.1$

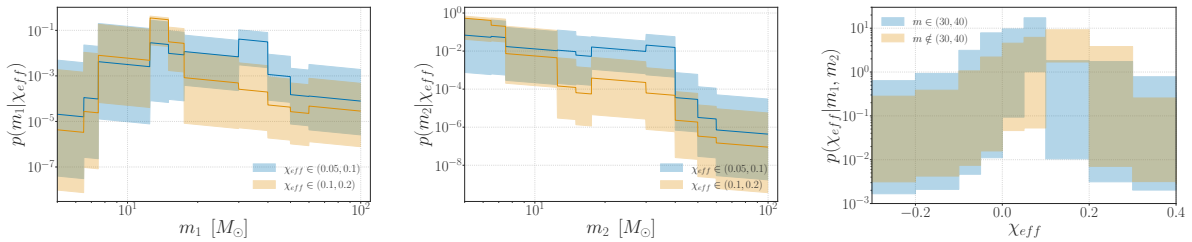


Figure B.5: GWTC-3 results for $\kappa = 4.4$

their priors determined by the distribution of bin centers (Ray et al., 2023c). In Figure B.6 we show that the posterior distributions of the GP length scales along both the mass and spin dimensions remain consistent for all three choices of binning. This is indicative of the GP correlating more or less bins over the same range of parameter space depending on how many bins reside within that range, ultimately leading to consistency in the inferred shapes of the population disribution (Ray et al., 2023c), as shown in Figures B.7 and B.8.

We note that increasing the number of bins makes the constraints broader, which is to be expected given that we are trying to fit a larger number of hyper-paramaters from the same dataset. Furthermore, the increased resolution does not reveal any additional features in the dataset which justifies our choice of binning used in Sec. 4.4.3.

B.7 MORE ON SCALABILITY

Previous implementations of the BGP method had scalability issues which limited the resolution that could be achieved within a tractable inference run (Ray et al., 2023c). This was mostly due to the cubic time-complexity of GP engines which in our case previously translated to $O(\prod_\theta N^3_{bins,\theta})$, where $N_{bins,\theta}$ are the number of bins along dimension θ, and was bound to become prohibitive with the addition of more parameters or an increase in resolution. Specifically this particular form of complexity arises from the Cholesky decomposition of the GP's covariance kernel, which takes

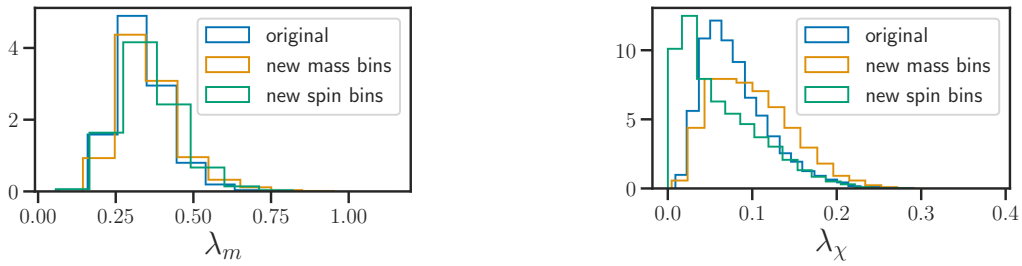


Figure B.6: Posterior distributions of the GP lengthscales along the mass (per log mass bin center,*left*) and spin (per spin bin center *right*) dimensions.

.

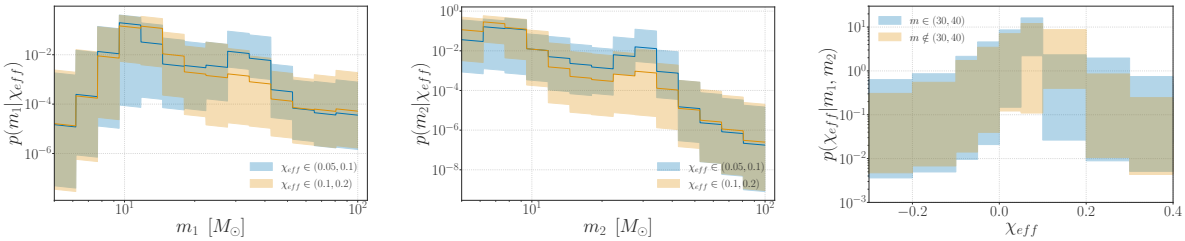


Figure B.7: GWTC-3 results for changed mass bins

the following form:

$$K^{\gamma\gamma'}(\lambda,\sigma) = \sigma^2 e^{-\sum_\theta \frac{(c_\theta^\gamma - c_\theta^{\gamma'})^2}{2\lambda_\theta^2}} \tag{B.26}$$

$$\mathbf{K} = \mathbf{L}\mathbf{L}^T, \tag{B.27}$$

where L is a lower triangular matrix. Given that $\gamma \in \{1, 2, ..., \prod_\theta N_{bins,\theta}\}$, and the cubic complexity of Cholesky decomposition algorithms, the total computational cost is expected to scale as $O(\prod_\theta N_{bins,\theta}^3)$.

Fortunately, the structure of the kernel in Eq. (B.26) provides a nice work around to this problem. For example, the matrix in Eq. (B.26) can be expressed as a Kronecker product of smaller

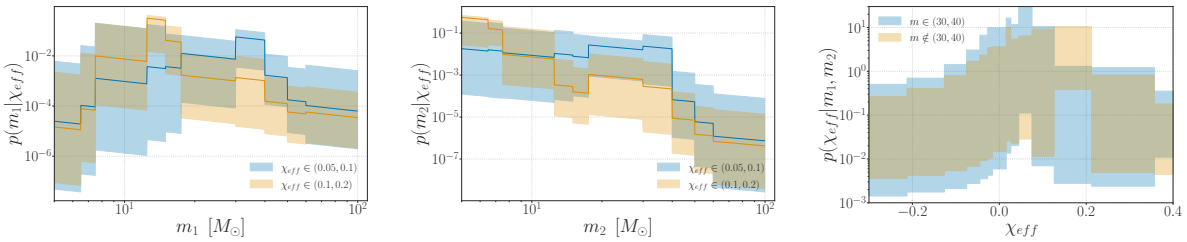


Figure B.8: GWTC-3 results for changed $\chi_{\rm eff}$ bins

matrices (one for each BBH parameter):

$$\mathbf{K} = \bigotimes_{\theta} \mathbf{K}_{\theta} \tag{B.28}$$

$$K_{\theta}^{\gamma\gamma'} = \sigma^{\frac{2}{n_\theta}} e^{-\frac{(c_\theta^\gamma - c_\theta^{\gamma'})^2}{2\lambda_\theta^2}}, \tag{B.29}$$

where n_θ is the total number of BBH parameters considered in the population model. Exploiting the fact that the Cholesky decomposition of Kronecker product is the Kornecker product of Cholesky decompositions, as in:

$$\mathbf{L} = \bigotimes_{\theta} \mathbf{L}_{\theta} \tag{B.30}$$

$$\mathbf{K}_{\theta} = \mathbf{L}_{\theta}\mathbf{L}_{\theta}^{T}, \tag{B.31}$$

we are able to implement an inference whose cost scales as $O(\sum_\theta N_{bins,\theta}^3)$. This leads to higher dimensional generalizations of our models straight forward to implement in a scalable manner. Similarly, an increase in bin resolution along a particular dimension can also be implemented tractably, upto the scenario wherein we have around 50+ bins along each dimension, a level of resolution that is likely unnecessary for contemporary and near-future datasets.

Appendix C

Scalable Equation of State Inference from Binary Neutron Star Mergers

C.1 VALIDATION STUDY: THE PIECEWISE-POLYTROPIC PARAMETRIZATION

In this appendix, we demonstrate that our algorithm is compatible with potentially any EoS parameterization by reproducing our results with the 4-parameter piecewise-polytropic parameterization instead of the spectral one. We run our algorithm with the piecewise parameterization on the same data for which we displayed the spectral results in Sec. 6.2.4. The piecewise polytropic parameterization is based on the assumption that the EoS pressure density relation can be accurately represented by multiple polytropes (power-law relations of the form $p = K\rho^{\Gamma}$), stitched together at fixed joining densities. If the joining densities ρ_i are fixed then such a parameterization containing n_p polytropes can be completely characterized by $n_p + 1$ parameters: the adiabatic indices Γ_i of each polytrope and the pressure at the first joining density p_1.

For a four parameter model, three polytropes are stitched together resulting in an EoS characterized by the free parameters: $\vec{\gamma} = \{\log p_1, \Gamma_1, \Gamma_2, \Gamma_3\}$. We impose priors on these parameters that are physically equivalent to the priors imposed on the spectral parameters as described in Sec. 6.2.2.3. We impose the thermal stability condition by imposing a uniform prior on the parameters: $\log_{10}(p/\mathrm{dyn\,cm^{-2}}) \in [33.6, 35.4]$, $\Gamma_1 \in [2, 4.5]$,$\Gamma_2 \in [2.0, 4.5]$ and $\Gamma_2, \Gamma_3 \in [1.1, 4.5]$, which are also broad enough to accurately fit a large number of candidate EoS models. In addition to these uniform bounds, we further impose the causality of sound speed and the observational consistency of maximum NS mass prior exactly similar to the spectral case, as described in 6.2.2.3. These choice of priors are consistent with previous works such as Lackey & Wade (2015) and Carney et al. (2018).

We again use LALSIMULATION to solve the TOV equations for this four parameter realization of the piecewise-polytropic parameterization to convert pressure-density relation into a mass and tidal parameter relation. As implemented in LALSIMULATION, we choose the joining densities ρ_1 and ρ_2 to be $\rho_1 = 1.8\rho_{\mathrm{nuc}}$ and $\rho_2 = 3.6\rho_{\mathrm{nuc}}$ respectively, where $\rho_{\mathrm{nuc}} \approx 2.8 \times 10^{15}\,\mathrm{g\,cm^{-3}}$ is the nuclear saturation density. As in the case of the spectral parameterization, we anchor the high density parameterized EoS with a fixed low density EoS, which is chosen to be the SLY. This implementation of the piecewise polytropic parameterization is consistent with previous works. Under these considerations, with the priors described above, running our algorithm for the simulated events yields EoS constraints that are completely consistent with the spectral results as can be seen in Figs. C.1 to C.2.

C.1.1 Comparison with the spectral parameterization

While these results in Fig. C.1 through Fig. C.3 are in complete agreement with the injected EoS and what we expect in terms of the variability of the EoS bounds with mass and SNR, we can see

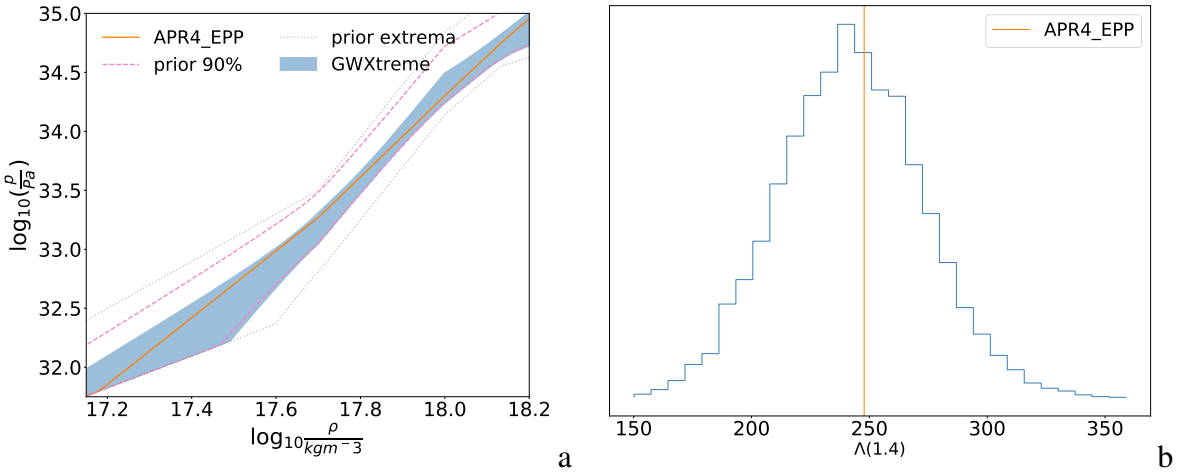


Figure C.1: EoS constraints obtained using the piecewise polytrope for the 16 events drawn from the galactic population distribution.

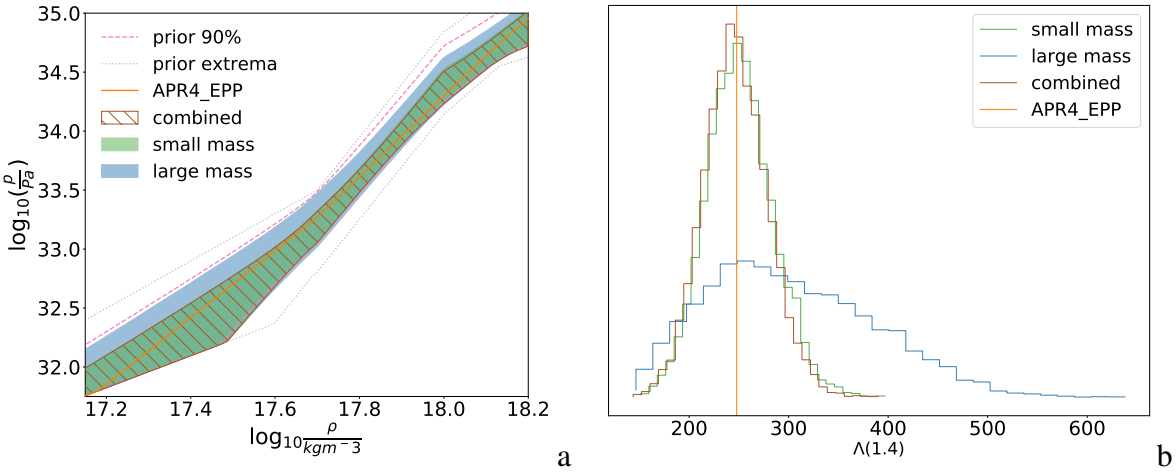


Figure C.2: EoS constraints obtained using the piecewise polytrope for simulated events drawn in the 23 to 25 SNR bin

sharp changes in the EoS constraints near the joining densities of the polytrope pieces. Near those densities, we can also identify broadening of the EoS constraints which can be attributed to the non-differentiability of the EoS at those points and hence interpreted as artificial. This pathology is inherent to the piecewise parameterization, and is absent in the spectral parameterization.

Due to these aforementioned issues with the piecewise parameterizations we chose the spectral EoS for our main result, leaving the results of the piecewise polytropic parameterization as a validation study, that demonstrates the compatibility of our algorithm with multiple EoS parameterizations. The exact origin of these issues with the piecewise parameterization is an interesting academic exercise that we do not pursue in greater detail since the spectral EoS, which is free of

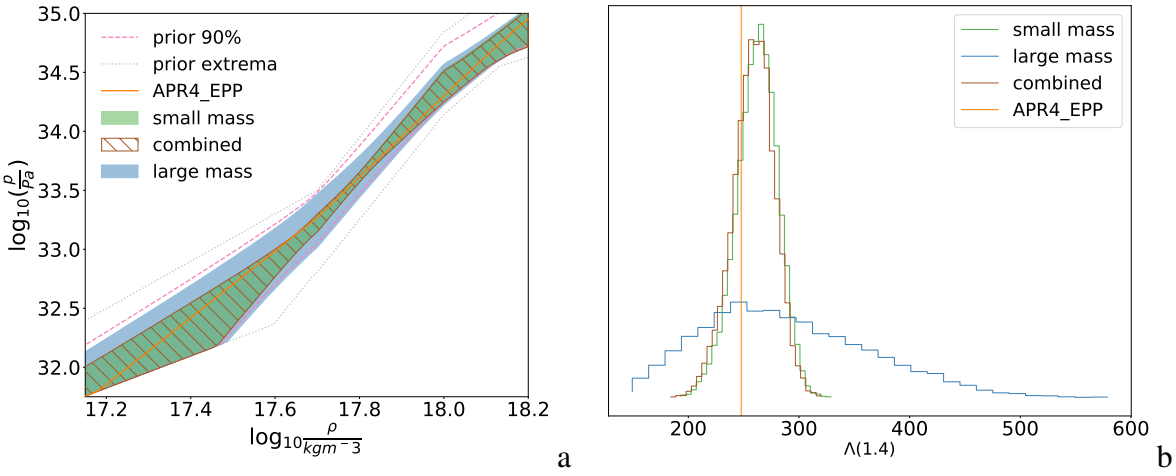


Figure C.3: EoS constraints obtained using the piecewise polytrope for events drawn in the 33 to 35 SNR bin

these issues, has been shown to produce consistent results in conjunction with our algorithm.

C.2 NUMBER OF SIMULATED EVENTS

The number of events one might expect to observe in O4 can be estimated from the number of events seen up until now and the ratio of spacetime hypervolume that will be surveyed in O4 to the hypervolume previously surveyed. Due to the small number of BNSs observed to date, this estimated number will have a large uncertainty. Here a simple calculation using Poisson statistics is used to make reasonable guesses for the expected number of events, within counting uncertainties. Let the observed rate of BNS events during a past observing be μ_0. Then the likelihood of observing n_0 number of events during a past observing run given μ_0 can be modeled by a Poisson distribution:

$$p(n_0|\mu_0) = \frac{1}{n_0!}\mu_0^{n_0}e^{-\mu_0} \tag{C.1}$$

Using Bayes theorem and some uninformative prior on μ_0 such as a powerlaw, $p(\mu_0) = \mu_0^{-\beta}$, one can write the posterior distribution of the true rate given the number of past observations n_0, after the correct normalization becomes

$$p(\mu_0|n_0) = \frac{1}{\Gamma(n_0+1-\beta)}\mu_0^{n_0-\beta}e^{-\mu_0} \tag{C.2}$$

If the new observing run is α times more sensitive than the past one then the rate of observed events in the new run relates to that of the older run via $\mu = \alpha\mu_0$. Substituting into Eq. (C.2) allows us to compute the probability of the observed rate during the new run given the number of observations in the older run: $p(\mu|n_0)$. Noting that the expected number of events n to be observed during the new observing run given μ is another Poisson distribution like Eq. (C.1): $p(n|\mu) = \mu^n e^{-\mu}/n!$, one

can marginalize over μ and find the probability of observing n given the observed n_0:

$$p(n|n_0) = \int_0^\infty p(n|\mu)p(\mu|n_0)d\mu = \frac{\Gamma(n+n_0+1-\beta)}{\Gamma(n+1)\Gamma(n_0+1-\beta)} \frac{\alpha^n}{(1+\alpha)^{n+n_0+1-\beta}} \tag{C.3}$$

A sanity check of this equation is to show that the total probability, when summed over n, is unity, or, equivalently, that

$$(1+\alpha)^{n_0+1-\beta} = \sum_{n=0}^{\infty} \frac{1}{n!} \frac{\Gamma(n+n_0+1-\beta)}{\Gamma(n_0+1-\beta)} \left(\frac{\alpha}{1+\alpha}\right)^n. \tag{C.4}$$

To show this, consider the function $(1-z)^{-a}$ for real $a > 0$, which is analytic in the domain $|z| < 1$. Its Maclauren series is

$$\frac{1}{(1-z)^a} = \sum_{n=0}^{\infty} \frac{1}{n!} \frac{\Gamma(a+n)}{\Gamma(a)} z^n. \tag{C.5}$$

Now let $z = \alpha/(1+\alpha)$ (note: $|z| < 1$ for $\alpha > 0$) and $a = n_0 + 1 - \beta$ (note: $a > 0$ for $n_0 \geq 0$ and $\beta < 1$), which completes the proof. As a second sanity check, the expected number of events, $\langle n \rangle$, is computed by differentiating both sides of

$$1 = \sum_{n=0}^{\infty} \frac{\Gamma(n+n_0+1-\beta)}{\Gamma(n+1)\Gamma(n_0+1-\beta)} \frac{\alpha^n}{(1+\alpha)^{n+n_0+1-\beta}} \tag{C.6}$$

with respect to α and then multiplying by $\alpha(1+\alpha)$; the result is

$$\langle n \rangle = \sum_{n=0}^{\infty} np(n|n_0) = (n_0 + 1 - \beta)\alpha \tag{C.7}$$

which, as expected, is proportional to α. Repeating the procedure on the above equation results in $\mathrm{Var}\, n = \langle n^2 \rangle - \langle n \rangle^2 = (\alpha+1)\langle n \rangle$.

We note that the ratio α that relates the observed rates of two different runs is essentially the ratio between the spacetime hypervolume $\langle VT \rangle$ to which these runs are sensitive. Since no data is available for computing $\langle VT \rangle$ of O4, we approximate α from the effective BNS ranges of the observing runs and their run times:

$$\alpha = \frac{d^3_{\mathrm{BNS,O4}} T_{\mathrm{O4}}}{d^3_{\mathrm{BNS,O1}} T_{\mathrm{O1}} + d^3_{\mathrm{BNS,O2}} T_{\mathrm{O2}} + d^3_{\mathrm{BNS,O3}} T_{\mathrm{O3}}} \tag{C.8}$$

where $d_{\mathrm{BNS},r}$ is the horizon distance of the detectors for a $(1.4\,M_\odot, 1.4\,M_\odot)$ optimally oriented BNS corresponding to an SNR threshold of 8 given the sensitivity of the rth observing run and T_r is the run-time of the rth observing run. With $d_{\mathrm{BNS},r}$ and T_r chosen from Abbott et al. (2018b) we are left with $\alpha = 2.34$. Then choosing $n_0 = 2$ and $\beta = 0$, we find the expected number of events to be $n = 6^{+10}_{-5}$. These bounds on n are computed by drawing samples from the distribution in

Eq. (C.3) and computing the equal tail 90% confidence intervals of those samples. Thus we draw 16 simulated events from the Galactic BNS population mentioned above for our simulation study so that the results of the study can also serve as a forecast of optimistic EoS constraints from GW data one can expect to achieve at the end of O4.

Appendix D

Waveform Systematics of Speed of Gravity measurements

D.1 ROBUSTNESS AGAINST CHOICE OF WAVEFORM APPROXIMANTS

In this section, we elaborate on the insensitivity of our results to the choice of waveform approximants. All measurements presented in this work rely on the posterior samples of v_g, α, δ that are obtained from single event parameter estimation runs. Samples of these parameters are implicitly computed from those of direct observables such as the time delay between the arrival of signals at different detectors, the effective distance to the source as measured by each detector, and the coalescence phase. In figure D.1, we show that parameter estimation results for such observables are fully consistent between different choices of the waveform approximant for a randomly chosen event in O3. The samples are obtained from LVK's public data release. We note that fixing $v_g = c$ (as was done for the publicly released samples) or allowing it to vary does not affect these observables since v_g is a derived quantity while the observables are directly measurable.

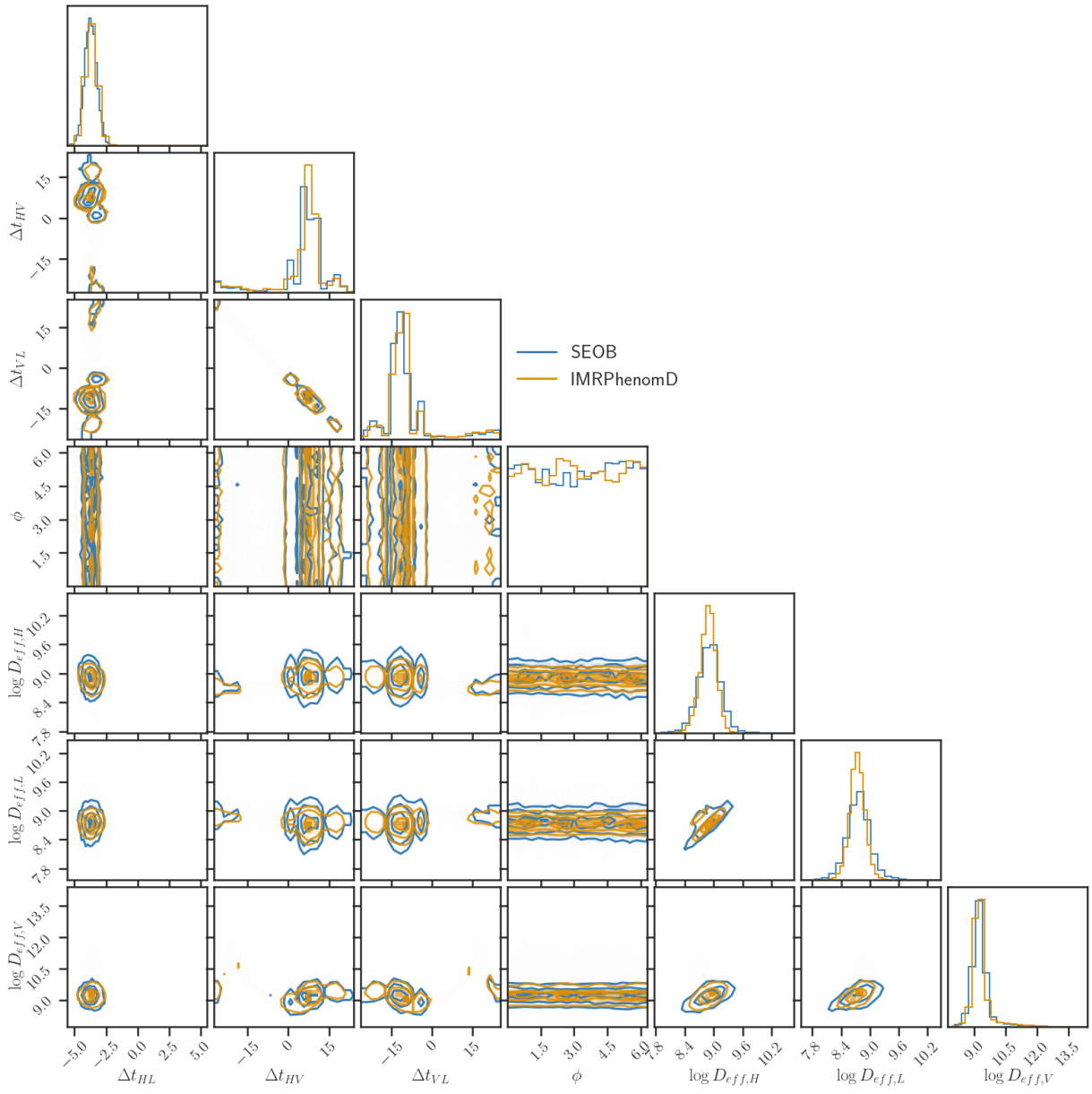


Figure D.1: Insensitivity of observables to choice of waveform approximant for a randomly chosen event

Anarya Ray

Citizenship: India

Education

University of Wisconsin-Milwaukee — Milwaukee, WI
Ph.D., Physics — *September 2019 - May 2024*

Presidency University — Kolkata, India
M.S., Physics (valedictorian) — *September 2017 - September 2019*

Presidency University — Kolkata, India
B.S., Physics — *September 2014 - September 2017*

Research Interests

Gravitational wave astrophysics and cosmology: Characterizing the population of gravitational wave sources using data-driven methods, to learn about their formation and to measure the cosmological parameters.

Third generation (3G) GW detectors: Self-consistent, ultra-precise inference of the nuclear equation of state, the BNS mass and redshift distributions, and the cosmological parameters from 3G BNS observations.

Extreme matter physics: Constraining the nuclear equation of state using gravitational wave and multiprobe observations of neutron stars.

Modified gravitational wave propagation: Searching for imprints of Lorentz violation in gravitational wave propagation.

Multi-messenger astronomy: Classifying gravitational wave sources in low-latency to assist in Multi-messenger follow-up campaigns.

Awards & Fellowships

The Papastamatiou scholarship for outstanding performance in theoretical physics 2024
University of Wisconsin Milwaukee

The Parthasarathi Gupta scholarship and gold medal for securing highest GPA in M.S. 2019
Presidency University

Research Positions

LIGO, University of Wisconsin-Milwaukee *Research Assistant, May, 2020 - August 2024*

- My main work is on constraining the astrophysics of compact binary formation, the nuclear equation of state, and modified theories of gravity by combining multiple gravitational wave observations.
- I am one of the lead developers of `GPpop`[1], a data-driven population inference tool for exploring compact binary formation using a population of compact binary coalescence observations.
- I am one of the lead developers of `GWXtreme`[2], a package for implementing the rapid hierarchical equation of state inference from multiple observations of neutron star-containing systems.
- I am a developer of `bilby-tgr`[3], a pipeline used for estimating deviations from general relativity. I am leading the parameter estimation analyses that use birefringent waveforms.
- My additional work is on the classification of gravitational wave triggers found by the `GstLAL` search pipeline, both in low-latency and post-processing, for assisting in the optimized searches of their multimessenger counterparts, and the generation of catalogs.
- I am one of the lead developers of `GWSCI-PASTRO`[4], a package that calculates the probability of astrophysical origin of gravitational wave triggers found by the `GstLAL` search pipeline in real-time, **that is sent out as part of public alerts to gravitational wave candidates**.
- I am one of the lead developers of parameter-estimation based source-classification modules in the `pesummary`[5] package and the `rapid-pe-rift`[6] pipeline.

[1] https://github.com/AnaryaRay1/gppop
[2] https://github.com/shaonghosh/GWXtreme
[3] https://pypi.org/project/bilby-tgr/
[4] https://pypi.org/project/gwsci-pastro/
[5] https://git.ligo.org/lscsoft/pesummary/-/blob/new_p_astro/pesummary/gw/classification.py
[6] https://git.ligo.org/rapidpe-rift/rapidpe-rift-pipe/-/blob/main/src/rapidpe_rift_pipe/pastro.py

Short author publications

1. *Searching for binary black hole sub-populations in gravitational wave data using binned Gaussian processes*
   Ray, A., Magaña, I., Breivik, K., and Creighton, J.
   (arxiv:2404.03166) 2 citations
2. *Beyond Gaps and Bumps: Spectral Siren Cosmology with Non-Parametric Population Models*
   Magaña, I., and **Ray, A.**
   (arxiv:2404.02522) 1 citation
3. *Updating gravitational-wave source classification estimates with rapid parameter estimation*
   Rose, C., **Ray, A.**, Valsan, V., Brady, P., *et. al.*
   (In preparation)
4. *Non-parametric inference of the population of compact binaries from gravitational wave observations using binned Gaussian processes**
   Ray, A., Magaña, I., Mohite, S., Crehighton, J., Kapadia, S.
   ApJ 957 37 (2023) (arXiv:2304.08046 , 15 citations)
5. *Measuring Gravitational Wave Speed and Lorentz Violation with the First Three Gravitational-Wave Catalogs**
   Ray, A., Fan, P., He, V. F., Bloom, M. *et. al.*
   (arXiv:2307.13099) 1 citation
6. *When to Point Your Telescopes: Gravitational Wave Trigger Classification for Real-Time Multi-Messenger Followup Observations**
   Ray, A., Niu, W., Sakon, S., Ewing R., *et. al.*
   (arXiv:2304.08046v1) 2 citations
7. *Rapid Hierarchical Inference of Neutron Star Equation of State from multiple Gravitational Wave Observations of Binary Neutron Star Coalescences**
   Ray, A., Camilo, M., Creighton, J., Ghosh, S., Morisaki, S.
   Phys. Rev. D 107, 043035 (2023) (arXiv:2211.06435 , 2 citations)
8. *Holographic bound on area of a compact binary merger remnant*
   Majumdar, P. and **Ray, A.**
   Phys. Rev. D, 104, 044036 (2021) (arXiv:2008.13425)
9. *Gravitational Waves with Orbital Angular Momentum*
   Baral P., **Ray, A.**, Majumdar, P. and Koley, R.
   Eur. Phys. Jour. C, 80, 326 (2020) (arXiv:1901.08804 , 8 citations)
10. *Maxwell Electrodynamics in Terms of Physical Potentials*
    Majumdar, P. and **Ray, A.**
    Symmetry 11, 915 (2019) (arXiv:1905.13748 , 6 citations)
11. *Background gravity correction to the limiting mass of white dwarfs*
    Ray, A., Maity, P. and Majumdar, P.
    Eur. Phys. J. C 79 97 (2019) (arXiv:1708.02730, 3 citations)

LIGO-Virgo Publications to which I highly contributed:

1. *Observation of Gravitational Waves from the Coalescence of a 2.5-4.5 $M_\odot$ Compact Object and a Neutron Star*
   The LIGO Scientific, Virgo and KAGRA Collaborations (including **Ray, A.**)
   Submitted to APJL (, arXiv:2404.04248 citations) **Was on the editorial team, wrote and led the broad-population rate-estimation, parametric EoS inference, and flexible population inference analyses**

2. *GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run*
   The LIGO Scientific, Virgo and KAGRA Collaborations (including **Ray, A.**)
   Phys. Rev. X 13, 041039 (2023) (arXiv:2111.03606, 382 citations) **Led and wrote the analysis of source classification for GstLAL triggers**
3. *The population of merging compact binaries inferred using gravitational waves through GWTC-3*
   The LIGO Scientific, Virgo and KAGRA Collaborations (including **Ray, A.**)
   Phys. Rev. X 13, 011048 (2023) (arXiv:2111.03634, 186 citations) **Led and wrote the appendix on merger rates including sub-threshold triggers**

4. *GWTC-2.1: Deep Extended Catalog of Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run*
The LIGO Scientific, Virgo and KAGRA Collaborations (including **Ray, A.**)
Phys. Rev. D 109, 022001 (2023) (arXiv:2111.03634, 108 citations) **Led and wrote the analysis of rate estimation and source classification for GstLAL triggers**

5. *Observation of gravitational waves from two neutron star-black hole coalescences*
The LIGO Scientific, Virgo and KAGRA Collaborations (including **Ray, A.**)
ApJL, 915, L5 (2021) (arXiv:2106.15163, 562 citations) **Led the analysis of low-threshold and broad population rate estimation**

Teaching Positions

University of Wisconsin-Milwaukee *Teaching Assistant, September, 2019 - May 2024*
I have taught four undergraduate physics courses at the 100 level, including two labs.

Student Supervision

- **Yu-Ping Teng** University of Wisconsin-Milwaukee
PhD student *January 2024 - Present*
- **Yu-Kuang Chu** University of Wisconsin-Milwaukee
PhD student *January 2024 - Present*
- **Tamal Roychoudhury** University of Wisconsin-Milwaukee
PhD student *January 2024 - Present*
- **Michael Camilo** Mont. Clair State University
Undergraduate student *September 2022 - Present*

Invited Seminars

1. **CGCA seminar, University of Wisconsin Milwaukee** April 2024
Astrophysics and Cosmology with gravitational waves: a data-driven perspective
2. **PUG seminar, Pennsylvania State University** December 2023
Blackboard talk, Background gravity correction to the limiting mass of white-dwarfs
3. **OzGrav Videocon** April 2023
Nuclear physics, modified gravity and astrophysics with Multi-probe observations
4. **University of Wisconsin Milwaukee** December 2022
Measuring Gravitational Wave Speed and Lorentz Violation with the First Three Gravitational-Wave Catalogs

Talks at Conferences

1. **APS April Meeting** April 2024
Searching for binary black hole subpopulations in gravitational wave data using binned gaussian processes
2. **APS April Meeting** April 2023
Non-parametric inference of the population of compact binaries from gravitational wave observations using binned Gaussian processes
3. **Midwest Relativity meeting** October 2022
Measuring Gravitational Wave Speed and Lorentz Violation with the First Three Gravitational-Wave Catalogs
4. **Midwest Relativity meeting** November 2021
Rapid Hierarchical Inference of Neutron Star Equation of State from multiple Gravitational Wave Observations of Binary Neutron Star Coalescences

Posters

1. **LVK Meeting: North Western University** March 2023
Non-parametric inference of the population of compact binaries from gravitational wave observations using binned Gaussian processes

Additional Training

1. **Cyberinfrastructure Comprehensive, Applied and Tangible Summer School** May 2023
University of Wisconsin Milwaukee

Skills

- **Programming skills:**
  - Proficiency with: Python, C, Fortran, Mathematica.
  - Familiar with: C++, SQL, Java Script
  - LVK related software: GstLAL, Bilby, LALSuite
- **Data Analysis:**
  - Bayesian hierarchical inference, model selection, parameter estimation and non-parametric methods, Monte Carlo methods.

References

Jolien Creighton (PhD Supervisor)
University of Wisconsin-Milwaukee
jolien[at]uwm.edu

Shaon Ghosh
Montclair State University
ghoshs[at]montclair.edu

Chad Hanna
Pennsylvania State University
crh184[at]psu.edu

Jay Tasson
Carleton College
jtasson[at]carleton.edu