INFERENCE AND MODEL COMPARISON IN GRAVITATIONAL WAVE ASTRONOMY

By

Steven Reyes

B.A., University of Chicago

Dissertation
Submitted in Partial Fulfillment of the Requirements
for the degree of
Doctor of Philosophy in Physics

Syracuse University
September 2019

INFERENCE AND MODEL COMPARISON IN GRAVITATIONAL WAVE ASTRONOMY

By

Steven Reyes

B.A., University of Chicago

Dissertation
Submitted in Partial Fulfillment of the Requirements
for the degree of
Doctor of Philosophy in Physics

Syracuse University
September 2019

Approved : _____
Prof. Duncan A. Brown

Date : _____

ABSTRACT

In this thesis, we explore the limitations and possibility of astrophysical modeling on detected gravitational waves from the Laser Interferometer Gravitational wave Observatory and Virgo. First we discuss the statistical inference that are possible on sources that have not yet been detected, techniques for evaluating the statistical significance of gravitational wave candidates, and finally modeling detected gravitational waves through different hypotheses on the parameters that may characterize the signal. Finally, we move towards evaluation fitting-and-overfitting models to signals, looking for an efficient set of parameters that accurately characterize the signal. With the success of gravitational wave observatories, scientists have, for the first time means to test and evaluate various physical theories on the parameters that may characterize binary black hole models and systems with neutron stars. We take a first look at the difficulties and solutions towards efficiently evaluating these models through model selection and comparison techniques.

ACKNOWLEDGEMENTS

I would like to first thank my advisor, Duncan Brown, for his support and mentorship throughout my PhD. Learning to be a scientist and think scientifically is a challenging endeavor and I am grateful for Duncan's guidance through the entire process. His experience and suggestions have been a constant help in attaining my research goals.

I am grateful to the entire Syracuse University Gravitational wave group for their unique perspectives and for their diverse expertise. Science, I believe, is best done in the context of diversity, community, and passion, and I believe the group deeply embodies this ethic in a way that few places could rival.

To Peter Saulson and Stefan Ballmer, I am grateful for your helpful perspectives, critiques, and advice in the progress and development of the research I have conducted here at Syracuse University. It is always helpful to gain the perspective of experts and I have grown significantly in my knowledge and confidence in the field of gravitaitonal wave astronomy due to their help.

I would like to thank my coworkers and officemates: Thomas Vo, Jaysin Lord, TJ Massinger, Soumi De, Chris Biwer, Lorena Magaña Zertuche, Swetha Bhagwat, Daniel Finstad, Chaitanya Afle, Derek Davis, Daniel Vander-Hyde, Nick Didio, Fabian Magaña-Sandoval, Ari Pederson, Eric Muñiz, Gaby , and my virtual-officemate Alex Urban. Thank you for your friendship and constant encouragement. Graduate school wouldn't have been nearly as enjoyable, enlightening, or even possible without all of you. I hope each of you continues to cultivate the same culture of inclusiveness and openness that you have helped foster here. You all have been great fun to be around.

To the PyCBC group, of which there are perhaps too many to list, I thank you for your guidance, encouragement, and expertise. I have benefited enormously from being in the presence of such world-class scientists who have helped me in every which way to complete this PhD. Thank you.

I would like to thank my committee members, A, B, C, D, E, and Duncan Brown, for taking time out of their schedules to serve on my thesis defense committee.

To the other (under)graduate students at Syracuse University who have befriended me over the past years, I would like to thank you as well for the fun I've shared with you all. I hope you all continue to be excellent.

To my parents, my siblings, my cousins, my aunts, my uncles, and my entire family, you've given me every opportunity possible to become anything that I could ever wish to be. Thank you for everything.

Contents

List of Tables

List of Figures

Preface

What goes here?

To my family.

Chapter 1

Introduction to Gravitational Wave Astronomy

1.1 Introduction

1.1.1 General Relativity and Gravitational Waves

The early 20th century brought about a large change to scientists' understanding of physics in many different ways. In 1914, Albert Einstein developed a general theory of relativity that changed the way that scientists think about space, time, energy, and mass. This theory was a new, geometric theory of gravity that improved up on Newton's theory of gravitation. The new theory was able to successfully predict the bending of light due to gravitation in a 1919 experiment. One immediate theoretical prediction of Einstein's new general theory of relativity is that plane wave gravitational wave radiation should be possible. Below we will go through a simplified overview of the theory of general relativity and gravitational waves. It is not intended to be an exhaustive nor comprehensive explanation of much of the theory as that extends far beyond the scope of this work. Useful introductions to the theory of general relativity and to gravitational waves can be found in (insert texts).

Einstein's theory of general relativity requires us to consider the following set of differential equations expressed in tensor notation as,

$$G_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}. \tag{1.1}$$

Here the Greek indices μ and ν are indices of a rank-2 tensor where μ and ν are

permitted to be integer values between 0 and 3. From here on out we will use this convention for all other Greek indices as well. This means that Eqn. 1.1 represents a set of 16 equations for every index value that μ and ν can take on. To better illustrate this, we can write Eqn. 1.1 as:

$$\begin{pmatrix} G_{00} & G_{01} & G_{02} & G_{03} \\ G_{10} & G_{11} & G_{12} & G_{13} \\ G_{20} & G_{21} & G_{22} & G_{23} \\ G_{30} & G_{31} & G_{32} & G_{33} \end{pmatrix} = \frac{8\pi G}{c^4} \begin{pmatrix} T_{00} & T_{01} & T_{02} & T_{03} \\ T_{10} & T_{11} & T_{12} & T_{13} \\ T_{20} & T_{21} & T_{22} & T_{23} \\ T_{30} & T_{31} & T_{32} & T_{33} \end{pmatrix} \tag{1.2}$$

The term $G_{\mu\nu}$, also known as the Einstein tensor, on the left-hand-side of Eqn. 1.1 represents the geometric structure and curvature of spacetime. On the right hand side of Eqn. 1.1 is the rank-2 tensor $T_{\mu\nu}$, which is the stress-energy tensor, representing the momentum flux through a surface of spacetime. The units of the stress-energy tensor in the International System of Units (SI units) are kilograms-meters-squared-per-second-squared ($\frac{kg\,m^2}{s^2}$), and the term $\frac{G}{c^4}$ provides the exact reciprocal units to provide a unitless right-hand-side. The unit G is Newton's gravitational constant and c is the speed of light. They will retain these definitions throughout the text. The 8π in the right-hand-side of Eqn. 1.1 provides an equality to Newtonian gravity in the weak-gravitational field limit.

We now expand out the Einstein tensor, $G_{\mu\nu}$ of Eqn. 1.1, into the following expressions from differential geometry

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} R\, g_{\mu\nu}. \tag{1.3}$$

Here, $R_{\mu\nu}$ is the Ricci tensor, representing the curvature or differential derivatives of the spacetime. It is a *contracted* form of the Riemann tensor ($R_{\mu\nu\gamma\delta}$) which describes the deviation of a vector under parallel transport through the manifold. Following this is, R which is the Ricci scalar, which itself is a *contraction* of the Ricci tensor over μ and ν at a particular point in spacetime. The last term $g_{\mu\nu}$ represents the spacetime metric tensor which describes the geometric relationships between coordinates in the spacetime. More explicitly the metric is used to measure distances in the spacetime (a spacetime interval, ds^2) using the following expression:

$$ds^2 = g_{\mu\nu}\, dx^\mu\, dx^\nu. \tag{1.4}$$

Repeated raised and lowered indices in Eqn. 1.4 refer to an implied Einstein summation. Length scales can be calculated from Eqn. 1.4 using the following expression:

$$L = \int_\lambda ds = \int \sqrt{g_{\mu\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau}}d\tau. \tag{1.5}$$

The length L along the worldline of an observer can be described as the integral of the spacetime interval along an affine parametrization λ describing the manner of traversing two points in the spacetime. This can be done by re-expressing Eqn. 1.4 into an expression with four-velocity $\frac{dx^\mu}{d\tau}$, where τ represents the proper time of an observer traversing the worldline. The differential equation nature of Eqn. 1.1 is made explicit by considering an expression for the Ricci tensor,

$$R_{\mu\nu} = \partial_\rho \Gamma^\rho_{\ \mu\nu} - \partial_\beta \Gamma^\rho_{\ \rho\alpha} + \Gamma^\rho_{\ \rho\lambda}\Gamma^\lambda_{\ \mu\nu} - \Gamma^\rho_{\ \nu\lambda}\Gamma^\lambda_{\ \rho\mu}, \tag{1.6}$$

and the Christoffel symbol $\Gamma^\rho_{\ \mu\nu}$ is an expression of derivatives of the metric, defined as:

$$\Gamma^\rho_{\ \mu\nu} = \frac{1}{2}g^{\rho\sigma}\left(\partial_\mu g_{\nu\rho} + \partial_\nu g_{\mu\rho} - \partial_\rho g_{\mu\nu}\right) \tag{1.7}$$

Now that we have all of the formalism, we move on to consider whether plane waves are possible in general relativity. We consider a spacetime represented by the Minkowski metric, $\eta_{\mu\nu}$, with a small perturbation $h_{\mu\nu}$:

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} = \begin{pmatrix} -c^2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} + h_{\mu\nu}. \tag{1.8}$$

We consider the case of a vacuum universe where $T_{\mu\nu} = 0$. Also, since we consider very small $h_{\mu\nu}$ and only linear contributions of $h_{\mu\nu}$, we can use $\eta_{\mu\nu}$ to raise and lower indices of tensors. Plugging in Eqn. 1.8 into the Einstein field equations, Eqn. 1.1, and only keeping linear (first-order terms) in h yields the following expression:

$$G_{\mu\nu} = \frac{1}{2}\left(\partial_\sigma\partial_\nu h^\sigma_{\ \mu} + \partial_\sigma\partial_\mu h^\sigma_{\ \mu} - \partial_\mu\partial_\nu h - \eta_{\mu\nu}\partial_\rho\partial_\lambda h^{\rho\lambda} + \eta_{\mu\nu}\partial_\alpha\partial^\alpha h\right) = 0 \tag{1.9}$$

Here, $h \equiv h^\mu_{\ \mu}$. If we then only consider $h_{\mu\nu}$ in the transverse-traceless (TT) gauge, Eqn. 1.9 is simplified to:

$$\partial_\alpha\partial^\alpha h^{TT}_{\mu\nu} = 0. \tag{1.10}$$

Furthermore, under this choice of coordinates we have, $h^{TT}_{t\nu} = h^{TT} = 0$. Here $h^{TT}_{t\nu}$ refers to choosing $\mu = t$, the time component of the perturbation in the first index. Thus, Eqn. 1.10 describes a differential equation whose solution is a plane wave in spacetime. If we choose standard Cartesian coordinates with the above conditions and consider that the wave propagates in the z-direction we can consider the wave as:

$$h^{TT}_{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & h_+ & h_\times & 0 \\ 0 & h_\times & -h_+ & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} cos\left[\omega\left(t - z/c\right)\right]. \tag{1.11}$$

The h_+ and $h_\times$ give the property of a traceless tensor, but they also describe the two degrees of polarization for the gravitational wave. Here ω describes the angular frequency of the wave. Thus, we can see that gravitational waves can induce changes in measured distances by combining Eqn. 1.11 with Eqn. 1.8 and Eqn. 1.4. For a wave with purely plus polarized gravitational wave this gives the expression:

$$ds^2 = -c^2dt^2 + \left\{1 + h_+ cos\left[\omega\left(t - z/c\right)\right]\right\} dx^2 + \left\{1 - h_+ cos\left[\omega\left(t - z/c\right)\right]\right\} dy^2 + dz^2. \tag{1.12}$$

While for a purely cross polarized wave this gives the expression:

$$ds^2 = -c^2dt^2 + 2\left\{h_\times cos\left[\omega\left(t - z/c\right)\right]\right\} dxdy + dx^2 + dy^2 + dz^2. \tag{1.13}$$

If we consider a plus-polarized gravitational wave we can combine equations Eqns. 1.12 with the Eqn. 1.5 to describe relative length changes in the x direction, which we place below:

$$\Delta x = \int \sqrt{g_{\mu\nu}\left(\frac{dx}{d\tau}\right)^2} d\tau \approx \left(1 + \frac{h_+}{2}\right) cos(\omega t - z/c). \tag{1.14}$$

And in the y direction:

$$\Delta y = \int \sqrt{g_{\mu\nu}\left(\frac{dy}{d\tau}\right)^2} d\tau \approx \left(1 - \frac{h_+}{2}\right) cos(\omega t - z/c). \tag{1.15}$$

The approximation in Eqns. 1.14 and 1.15 comes from taking the first order term in a Taylor series expansion of the square root. A similar approach can be taken for a cross-polarized gravitational wave, where the result will be similar to the plus-polarized gravitational wave except that the relative change in lengths will be rotated

by 45° in the plane that is perpendicular to propagation (in this case, the propagation direction is in the z coordinate). Fig. X displays the effect of changes to a ring of freely falling particles under the influence of purely plus and purely cross polarized gravitational waves.

1.1.2 The Possibility of Gravitational Wave Astronomy

The prospect of gravitational waves that impact measurable distances provide some hope for the plausibility of detecting them. It is useful to consider the analogy with electromagnetic waves in that plane wave electromagnetic waves can be generated by accelerating electric charges. In classical electromagnetic theory, dipole, quadrupole, octopole and higher order moments generate electromagnetic radiation. A similar examination of gravitational charges (matter-energy) might also yield similar sources of gravitational waves. Below we only consider radiation in the far-field regime, where the distance to the source is much larger than the wavelength of the radiation. The near-field regime for gravitational wave physics is outside of the scope of this work.

The possibility of electromagnetic waves is covered extensively in [] and we only provide a brief recap here. Electromagnetic theory provides a similar wave solution as a linearized general relativity theory provided. In electromagnetic theory, the solution to the wave equation can be given in terms of an electromagnetic wave with scalar and vector potential fields:

A quick consideration of matter provides an interesting analogy. The local conservation of energy provided in () prevents the amount of matter-energy from changing in a similar manner that conservation of charge prevents electromagnetic monopole radiation. This means that in general relativity there is no gravitational radiation from matter-monopoles. The next leading order in the power series expansion of the matter distribution is dipole matter distributions. However, here the conservation of momentum prevents gravitational radiation from dipole contributions. This similarly implies that when angular momentum is conserved, no gravitational waves will be emitted. However, there are no conservation laws that prevent the quadrupole moment of a matter distribution (or higher order moments) from generating gravitational waves.

In the far field limit for a weak gravitational wave we can write the leading order

strain tensor in terms of this quadrupole moment:

$$h_{ij}^{TT}(t) = \frac{1}{r}\frac{2G}{c^4}\Lambda_{ij,kl}(\hat{n})\,\ddot{\mathcal{I}}^{kl}(t - r/c). \tag{1.16}$$

Here, r represents the distance to the source in SI units. The term $\Lambda_{ij,kl}$ represents a projection operator that projects from the inertia tensor coordinate system into the coordinates that describe the plane wave as travelling in the unit direction $\hat{n}$. Colloquially, this is projecting from the source frame into the radiation frame. This term is expressed as:

$$\Lambda_{ij,kl}(\hat{n}) \equiv P_{ik}P_{jl} - \frac{1}{2}P_{ij}P_{kl} = (\delta_{ik} - \hat{n}_i\hat{n}_k)\,(\delta_{jl} - \hat{n}_j\hat{n}_l) - \frac{1}{2}\,(\delta_{ij} - \hat{n}_i\hat{n}_j)\,(\delta_{kl} - \hat{n}_k\hat{n}_l)\,. \tag{1.17}$$

Here δ_{ij} represents the Kronecker delta, and $\hat{n}_i$ represents the unit vector x_i/r. The term in Eqn. 1.16, $\ddot{\mathcal{I}}^{kl}(t - r/c)$, represents the second time derivative of the spatial components of the quadrupole moment tensor. The spatial quadrupole inertia moment tensor in this regime can be defined as:

$$\mathcal{I}^{kl}(t) = \int c^2\rho(t - r/c, \vec{x})\left(x^k x^l - \frac{1}{3}r^2\delta^{kl}\right)d^3\vec{x}. \tag{1.18}$$

In discrete form for i particles with mass m_i this can be given as:

$$\mathcal{I}^{kl}(t) = \sum_i^N m_i(t - r/c, \vec{x})\left(x^k x^l - \frac{1}{3}r^2\delta^{kl}\right). \tag{1.19}$$

This is sufficient to describe the expected gravitational wave radiation from a non-zero time-varying quadrupole moment tensor in the far-field and weak-field regime.

To gain some intuition regarding this gravitational wave from a time-varying quadrupole distribution of matter we consider a simplification of Eqn. 1.16 as:

$$h \sim \frac{G}{c^4}\frac{\ddot{\mathcal{I}}}{r}. \tag{1.20}$$

We consider a body with mass $M \sim 10\ M_\odot$ solar masses ($\approx 2 \times 10^{31}$ kilograms), at a distance of 300 megaparsecs ($\approx 9.2 \times 10^{24}$ meters). If the moment of inertia of the body is roughly MR^2, where R describes the moment arm about the motion of the body, then $\ddot{\mathcal{I}} \sim Mv^2$ where v is some non-spherically symmetric velocity. Now let v to be approximately 10% the speed of light ($\approx 3 \times 10^7$ meters per second). This would give a strain amplitude of $h \sim 1.6 \times 10^{-23}$. Note that the metric, and hence the change to measured distances, is quite small. We will explore possibilities from astronomy that could potentially create gravitational waves of this magnitude.

Compact Binary Coalescence

One possible source for generating large perturbations in spacetime would be gravitational waves from the mergers of astronomically massive binaries, hence forth called compact binary coalescence. In order to do so we consider two binary objects with masses m_1 and m_2 in Newtonian orbit about their center of mass such that in the frame of reference of the center of mass we can describe the coordinates of each binary as:

$$\vec{r_1} = \frac{m_1 m_2}{m_1 (m_1 + m_2)} a \begin{pmatrix} cos(\omega t) \\ sin(\omega t) \\ 0 \end{pmatrix} \tag{1.21}$$

$$\vec{r_2} = -\frac{m_1 m_2}{m_2 (m_1 + m_2)} a \begin{pmatrix} cos(\omega t) \\ sin(\omega t) \\ 0 \end{pmatrix} \tag{1.22}$$

Here the binaries orbit with orbital frequency ω and are separated by a distance a. The orbital frequency is given by Kepler's Law:

$$\omega = \sqrt{\frac{G(m_1 + m_2)}{a^3}}. \tag{1.23}$$

Computing the inertia tensor from Eqn. 1.19 gives us:

$$\mathcal{I}^{ij} = a^2 \frac{m_1 m_2}{m_1 + m_2} \begin{pmatrix} cos^2(\omega t) & sin(2\omega t) & 0 \\ sin(2\omega t) & sin^2(\omega t) & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{1.24}$$

Passing this expression through Eqn. 1.16 then gives us the expression for the metric perturbation in the TT gauge:

$$h_{ij}^{TT}(t, \iota, \psi) = \frac{4}{r} \frac{a^2 \omega^2 G}{c^4} \frac{m_1 m_2}{m_1 + m_2} \begin{pmatrix} -cos(2\omega t + 2\psi) \left(\frac{1+cos^2 \iota}{2}\right) & 2sin(2\omega t + 2\psi) cos(\iota) & 0 \\ -2sin(2\omega t + 2\psi) cos(\iota) & cos(2\omega t + 2\psi) \left(\frac{1+cos^2 \iota}{2}\right) & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{1.25}$$

The interesting result from this expression is that the gravitational wave frequency is twice the orbital frequency of the binary. Here the terms ι and ψ represent the spatial angles of incidence from the source to the observer in the radiation frame from

application of Eqn. 1.17. And so ι represents the inclination angle of the binary's plane of orbit relative to a distant observer, and ψ is the polarization angle. For the purposes of detection from we will later have to use the Euler angles to project from this reference frame into a detector reference frame. For now this suffices as an introduction to gravitational waves from compact binary coalescence.

Burst Signals, Continuous Waves, and the Stochastic Background

Compact binary coalescence are a promising source of gravitational waves but there are other possible sources of gravitational waves from astronomical sources. We will not describe these sources in depth in this work but we will briefly describe some possible sources of gravitational waves.

Gravitational Wave Interferometers and the Advanced LIGO Gravitational Wave Interferometer

Now that we have demonstrated that compact binary coalescence is a plausible source of gravitational waves as well as other possible sources of gravitational wave we will investigate a method for measuring these changes to the spacetime metric. To do so we introduce a means for projecting from the radiation frame of the gravitational wave into a detector reference frame.

To do so, we apply the Euler angle projection angles on Eqn. 1.25 to express the gravitational wave perturbation in the reference frame of a length-measuring detector. The Euler angle projection can be described using the expression:

$$\mathcal{R}(\theta,\phi) = \begin{pmatrix} cos(\theta)\,cos(\phi) & sin(\phi) & cos(\phi)\,sin(\theta) \\ -cos(\theta)\,sin(\phi) & cos(\phi) & -sin(\theta)\,sin(\phi) \\ -sin(\theta) & 0 & cos(\theta) \end{pmatrix}. \tag{1.26}$$

We get the gravitational wave strain in a detector frame then as $h'_{ij}(\iota,\psi,\theta,\phi) = \mathcal{R}^T h_{ij}^{TT} \mathcal{R}$, where $\mathcal{R}^T$ is the transpose of the matrix in Eqn. 1.26. Recomposing the product we arrive at the following expression:

$$h'(\iota,\psi,\theta,\phi) = F_+(\theta,\phi)\,h_+(\iota,\psi) + F_\times(\theta,\phi)\,h_\times(\iota,\psi) \tag{1.27}$$

For which F_+ and $F_\times$ can be interpreted as an antenna pattern or sensitivity of a

length-measuring detector. We express them below as:

$$F_+ (\theta, \phi) \equiv \frac{1}{2} [] \tag{1.28}$$

$$F_\times (\theta, \phi) \equiv \frac{1}{2} \tag{1.29}$$

A plot of the antenna patterns F_+ and $F_\times$ in Fig. X show that any length-measuring detector such as a gravitational wave interferometer will have blind-spots to incoming gravitational waves. The net sensitivity to gravitational waves in an idealized gravitational wave interferometer is then the quadrature sum, $\mathcal{F} = F_+^2 + F_\times^2$. A network of N detectors, $N_{\text{detectors}}$,can improve coverage over the entire sky and this sensitivity can be expressed as:

$$\mathcal{F}^2_{\text{network}} = \sum_i^{N_{\text{detectors}}} F^2_{+,i} + F^2_{\times,i} \tag{1.30}$$

In practical application the antenna patterns of a specific gravitational wave interferometer requires a precise coordinate location on Earth relative to conventional choices in astronomy such as right ascension, α, and declination, δ. These conventions can be found in LALsuite, etc.

The current gravitational wave interferometer network known as aLIGO are located in Hanford, Washington and Livingston, Lousiana. In blank the Virgo interferometer joined the network of gravitational wave detectors. More about gravitational wave observatories in O1 and O2 and beyond.

Elements of noise in the detector. Spectral Noise density. Limits to the strain sensitivity.

Chapter 2

Introduction to Probability, Statistics

2.1 Rules of Probability

Here we simply outline a few simple rules for probability that will help make the following discussion simpler to understand. Perhaps the simplest rule of probability is that given a series of possible outcomes, the sum of the probabilities of the outcome must equal unity. This is expressed as:

$$\sum_{i=1}^{N} p_i = 1. \tag{2.1}$$

Here p_i represents the probability mass function, or more simply, the probability of the i^{th} outcome, given N possible events. If the random variable is continuous then we simply express this as an integral:

$$\int p\left(x\right) dx = 1. \tag{2.2}$$

Where $p(x)$ represents the probability density function of a particular outcome x.

Finally, we describe a few rules of adding probabilities, multiplying probabilities, and conditional probability. The probability of two events occurring is $p(A \text{ and } B)$:

$$p\left(A \text{ and } B\right) \equiv p(A, B) = p(A)\, p(B|A) = p(B)\, p(A|B) \tag{2.3}$$

This new term here $p(B|A)$ is to be interpreted as the probability that event B occurs given that A has occurred, and similarly, $p(A|B)$ means the probability that event A occurs given that B has occurred.

This last expression from Eq. 2.3 motivates the theorem known as Bayes Theorem, which we will express as follows:

$$p(H|D) = \frac{p(H)\,p(D|H)}{p(D)}. \tag{2.4}$$

In this formulation we have written, the probability of the hypothesis given the data, $p(H|D)$, is sometimes called the posterior probability. The probability of the hypothesis being true is $p(H)$, and is often called the prior probability since it is what we believe prior to looking at the data. The probability of the data given the hypothesis, $p(D|H)$, which is called the likelihood. And finally we have the probability of obtaining the data, $p(D)$. We will devote a large amount of time in this work towards Bayes Theorem and its usefulness in conducting statistical inferences.

2.2 A Brief Introduction to Frequentist Statistics

Frequentist statistics is the perspective that probabilities represent the frequency in which a random process will generate a particular result in the long-term. Frequentist statistics is a perspective that may be taken in the event that one expects to be able to examine the long term behavior of a particular random process. Furthermore, Frequentist statistical inference will place emphasis on the likelihood term, $p(D|H)$, from Eq. 2.4. A choice on the probability of $p(H)$ is not made explicitly.

2.2.1 Null Hypothesis Significance Testing: Statistical Significance

In Frequentist statistics we are often concerned with evaluating whether our particular data can be well-explained by a particular hypothesis. Notably, in science we are concerned whether the data we observe have a high probability of being generated by a null hypothesis, often that the data are not significantly different from what we might expect from noise processes alone.

The p-value represents the probability that the data (datum) is generated by the data-generating process in the null hypothesis. If we choose some confidence threshold α that is sufficiently small, we can gain confidence in rejecting the null hypothesis if the p-value is smaller than this α. This term α is sometimes called the significance level. Typically the p-value concerns the value of an estimator x being greater (less) than some critical value, where the critical value is chosen so as to produce the right

(left)-tailed probability α. In the case of a right-tailed p-value we can write this in the following notation:

$$p(x \geq X|H_0) < \alpha. \tag{2.5}$$

Typical choices of α are chosen to be small so as to reduce an incorrect rejection (or retention) of the null hypothesis H_0. One of the downsides of the p-value is that it states the probability of obtaining some value of x given that the null hypothesis is true. This requires that the null hypothesis is a well-thought-out hypothesis and that it is an adequate model to test against in an experiment. In many areas of particle physics an α of 10^{-7} is often chosen as an acceptable threshold to guard against false rejection of the null hypothesis. In general however, a choice of statistical significance threshold α is arbitrary.

Controlling Error Rates

In light of choosing a particular α threshold by which we choose to reject or retain the null hypothesis, we inevitably open ourselves up to error rates if we happen across a very improbable datum or data set that **is** generated by the null hypothesis.

Incorrect inferences in Frequentist statistical significance hypothesis testing are typically called *Type I* (false positive) errors and *Type II* (false negative) errors. Correct inferences are usually then called true positive inferences and true negative inferences. If the null hypothesis is true, and we set our significance threshold to α, and we reject the null hypothesis, then we are making a false positive error. The probability of doing this is:

$$p\left(\text{rejecting } \mathrm{H}_0|\mathrm{H}_0 = \text{True}\right) \equiv \alpha \tag{2.6}$$

If we assume that the null hypothesis is true and choose α threshold for our p-value significance test, then the probability of failing to reject the null hypothesis when the null hypothesis is true is a true negative inference. The probability of doing this is:

$$p\left(\text{failing to reject } \mathrm{H}_0|\mathrm{H}_0 = \text{True}\right) = 1 - \alpha \tag{2.7}$$

If we fail to reject the null hypothesis, but the null hypothesis is false, then we make a false negative error. We can ascribe the probability of doing this as:

$$p\left(\text{rejecting } \mathrm{H}_0|\mathrm{H}_0 = \text{False}\right) \equiv \beta \tag{2.8}$$

The probability of rejecting the null hypothesis when the null hypothesis is false is given as:

$$p(\text{failing to reject } \mathrm{H}_0 | H_0 = \text{False}) = 1 - \beta \tag{2.9}$$

This probability, $1-\beta$ is sometimes called the *power* of the statistical inference, since it tells us how efficacious the hypothesis test is in rejecting the null hypothesis when we should reject it (e.g. there is a gravitational wave present in the data and we find it). Hence we see that choosing an α level for rejecting the null hypothesis permits us to protect from Type I errors, false positives, and gives us some understanding on the probability of correctly assessing true negatives. However choosing an α level does not offer much in the way of protecting from Type II errors, false negatives, nor much control over assessing true positives.

Finally, the last topic we describe is the problem of analyzing the data multiple times for statistical significance. This is sometimes called the problem of multiple comparisons or the look-elsewhere-effect. The problem can be simply explained under the following example. Say that one conducts a null-hypothesis statistical significance test at an $\alpha = 0.05$ threshold. . This opens the possibility of reporting a statistically significant result without properly attributing the number of trials conducted on the data. This is a form of misreporting in statistics and is sometimes called data-dredging or p-hacking. To correct for this possibility, a trials factor can be applied to a statistical significance test. One method for coherently assessing p-values after multiple comparisons is the Bonferroni correction wherein the α significance required to reject the null hypothesis is modified by the number of trials, n, performed. This can be expressed, to first order, as (Include derivation? Full term is used in most recent LIGO IMBH serach):

$$p\left(\text{rejecting } \mathrm{H}_0 \text{ due to any trial} | \mathrm{H}_0 = \text{True}\right) = \frac{\alpha}{n}. \tag{2.10}$$

Implicit in the Bonferroni correction is that each new search for the parameter of interest is independent of the other. This provides the most conservative case for correcting against false positives, but it opens the possibility of increasing the rate of false negatives. In practice, multiple comparisons are not always statistically independent and so other significance level adjustments for multiple comparisons are possible.

2.3 A Brief Introduction to Bayesian Statistics

Another interpretation of probability and statistical inference is Bayesian inference, which relies more heavily on all of the aspects of Eq. 2.4. In Bayesianism, inference is performed by stating our assumptions at the outset of our experiment. We must outright state the probability distributions that we take as given in order for our inference to be complete. Probabilities here reflect our implicit ignorance or our level of belief in a particular hypothesis. We will describe in some detail the steps of statistical inference in Bayesian statistics below.

2.3.1 Bayesian Inference

Before conducting the experiment we need to start at the outset and describe our prior beliefs about plausible measurements that we could arrive at from our experiment. In general, there is no "correct" method for assigning prior probabilities to the range of plausible measurements, however we provide some helpful guidelines below. Let us recall that the prior probability distribution must obey the standard rules of probability as outlined in Eqs. 2.1, 2.2.

In order to better illustrate choices of prior distribution let us consider the example for the data generating process, Eq. **??** from the previous section.

The first prior distribution that we suggest is the uniform prior distribution. A uniform prior in x would suggest that we assign a uniform probability to all possible values of x possible. The expression for the probability density of a uniform continuous distribution in some interval, $b < x < a$ is given as:

$$\pi(x, H) = \frac{1}{b - a} \tag{2.11}$$

Here we have substituted the notation $p(H)$ for the prior distribution with $\pi(x, H)$, the probability distribution function of x given this particular hypothesis choice, H. In this particular case our data generating procedure extends over all real values of x. Formally extending π to all values of x would present a mathematical problem. However, a reasonable choice of prior range can be chosen, such as $-10 < x < 10$, with, in this case, relatively little penalty.

The second prior distribution that we suggest is the informed prior distribution. Choosing an informed prior distribution is merely using our experimental expertise

regarding the data generating procedure to provide a prior belief. Say for example, that the data generating procedure had been examined closely just prior to a recalibration of the machinery that generated the data. In a previous experiment, scientists estimated that the mean of a sample of the data was 0.12, and with a variance of 1.02. We might choose a prior distribution that used a Gaussian distribution with mean, $\mu = 0.12$, and variance, $\sigma^2 = 1.02$ for our inference.

A third useful prior distribution to consider is a conjugate prior distribution. Conjugate prior distributions are sometimes available to certain kinds of likelihood distributions. If one has chosen a particular likelihood distribution and then the conjugate prior distribution to this likelihood distribution, then the posterior distribution will be in the same family of distributions as the conjugate prior distribution. This is advantageous if our prior beliefs on the family of the probability distribution that the data should belong to should not change with respect to the data. One example of a conjugate prior distribution is the conjugate prior distribution when choosing a Poisson likelihood distribution to measure the data. The conjugate prior distribution to the Poisson likelihood distribution is the Gamma distribution. The posterior distribution is then proportional to another Gamma distribution.

There are many other methods of selecting a prior distribution function, of which we will discuss in further chapters.

Finally, after a prior distribution is chosen, the likelihood can be measured from the data and inference on the probability distribution of x can be estimated as the product of the prior and the likelihood. This product is sometimes called the joint probability distribution and it is proportional to the posterior probability distribution function. The posterior probability distribution reflects our updated beliefs about plausible values of x.

Finally, the term $P(D)$ is sometimes called the evidence or the prior predictive. We will use the short hand, $\mathcal{E}$ for evidence from now on. It is called the evidence because its value represents the level of credibility we should assign our inference relative to other prior hypothesis choices. In a similar vein, the term prior predictive gives us an idea of how well of a job we did in predicting the data with our choice of prior. Lastly, we can compute the evidence by computing the marginal likelihood. Formally, this is done via,

$$\mathcal{E} = \int \pi(\vec{\theta})\mathcal{L}(\vec{\theta})d\vec{\theta}, \tag{2.12}$$

where we have used $\pi(\vec{\theta})$ for the prior distribution over all parameters $\vec{\theta}$, and likelihood distribution $\mathcal{L}(\vec{\theta})$. The marginal likelihood, or evidence, will be useful in Bayesian hypothesis testing. Many times the marginal likelihood is a difficult multi-dimensional integral to compute. Although oftentimes we are only interested in the marginal posterior probability distribution function for parameters, leading us to achieve adequate statistical inference without computing Eq. 2.12. Computation of the marginal likelihood is not often required for parameter estimation.

2.3.2 Parameter Estimation and Credible Intervals

Bayesian inference covers a method for performing parameter estimation of a particular data set. In particular, we aim to extract meaningful inference on the parameters of our data from the posterior distribution. This is in contrast to Frequentist statistical inference which only relies on the likelihood distribution for statistical inference.

To do so, first we introduce the concept of marginalization of the posterior probability distribution function. Marginalization of a probability distribution is the process of finding a probability distribution of a given parameter, A, by integrating the joint distribution of A with the other parameters, over all the values of the other parameters. This approach is taken in finding the marginal likelihood in Eq. 2.12, however it can be done for any parameter of interest. Marginalization for a continuous probability distribution can be expressed as,

$$p(x) = \int p(x|\vec{\theta}')\, p(\vec{\theta}')\, d\vec{\theta}', \tag{2.13}$$

where the integration occurs over $d\vec{\theta}'$, all variables in the parameter set excluding x. This procedure is also known as integrating out *nuisance parameters*, or can be thought of as the expectation value of the probability distribution of x after averaging over all other parameters. This procedure of marginalization provides a useful and consistent method for estimating plausible parameter values for a data set.

It is often useful to summarize the results of our inference through summary statistics such as the mean, median, standard deviation, or interquartile range. One such method is to design a credible interval based around a mean, median value, a maximum likelihood estimate (MLE), or a maximum marginal posterior probability (in *Latin* this is *maximum a posteriori*, hence it is usually termed the MAP) estimate.

A credible interval can be designed by determining some confidence level α wherein we desire that α percent of the posterior probability distribution is contained between some surface or interval. For a 1-dimensional marginalized probability distribution, $\mathcal{P}(x)$ this takes the form:

$$\alpha = \int_{x_{low}}^{x_{high}} \mathcal{P}(x)dx. \tag{2.14}$$

Thus, one can say that the random variable x is believed to have an α percent probability of being between the interval x_{low} and x_{high}. This procedure of credible intervals is not required to be 1-dimensional, and can pertain to credible surface contours. Choosing where to set x_{low} and x_{high} is somewhat subjective, although choosing the 10^{th} and 90^{th} percentiles of the posterior probability is common practice. This is called an equal-tailed interval, which we will use for the remainder of this work. Other examples of credible interval include the highest posterior density interval and choosing an interval that centers around the mean parameter values or MAP values 1995.

2.3.3 Bayesian Hypothesis Testing

The Bayes Factor

Another essential aspect of Bayesian inference is the evaluation of the statistical signficance of hypothesis choices. This occurs through evaluating the effectiveness of the choice in prior probability distribution. The marginal likelihood, $\mathcal{E}$, is the main driver behind establishing the level of evidence or support that the data has for a particular prior distribution choice. Simply put, the prior distribution that results in the largest evidence value is the model that has the most support.

Calculation of the odds for support of one hypothesis, H_1, over another hypothesis, H_2, is encapsulated in the following expression for the posterior odds ratio:

$$\mathcal{O}^{H_1}_{H_2} = \mathcal{B}^{H_1}_{H_2} \times \frac{\pi(H_1)}{\pi(H_2)}. \tag{2.15}$$

In this equation, $\mathcal{O}^{H_1}_{H_2}$ represents the posterior odds that hypothesis 1 is preferred over hypothesis 2. The ratio of the evidences, $\mathcal{B}^{H_1}_{H_2} \equiv \frac{\mathcal{E}_{H_1}}{\mathcal{E}_{H_2}}$, between the two models is known as the Bayes factor. The Bayes factor provides an intuition for the relative support of one hypothesis over the other. The ratio $\frac{\pi(H_1)}{\pi(H_2)}$ represents our prior odds

ratio, that is, how much more did we believe that hypothesis 1 was preferred over hypothesis 2 prior to our analysis. Said in another way, the prior odds ratio gives us a statement of what level of Bayes factor we would require before we begin to change our minds about the odds of hypothesis 2 being better supported in the data than hypothesis 1. When testing new physics, one may set the prior odds ratio to unity if one is fundamentally unsure about what hypotheses the data may support.

The posterior odds ratio then gives us a method for making a decision about whether to accept one hypothesis over the other hypothesis. One advantage to Bayesian hypothesis testing is that it gives us a straightforward method for testing hypotheses other than the null hypothesis that is commonly tested in Frequentist statistical inference. The downside however is that effectively and consistently computing Bayes factors remains an open area of research because of how difficult it can be to calculate the marginal likelihood. A conventional choice for hypothesis decision making is given to us by Jeffreys, and an alternative by Kass and Raftery 1995, see Fig. X.

An odds ratio can be converted into a probability of one hypothesis over another hypothesis through the following expression:

$$p^{H_1}_{H_2} = \frac{\mathcal{O}^{H_1}_{H_2}}{1 + \mathcal{O}^{H_1}_{H_2}}. \tag{2.16}$$

As such, a plot of the $\log_{10} \mathcal{O}^{H_1}_{H_2}$ can be made to suggest decision rules for odds ratios similar to choices on p-values in Frequentist statistics. As we can see in the plot below, when the odds ratio is 1 ($\log_{10} \mathcal{O} = 0$) the probability of one hypothesis versus another is 0.5. Furthermore, we can map this probability to a ranking statistic that is more familiar to Frequentists. That is the one-tailed z-score which states the integrated probability density from $-\infty$ to a particular multiple of the standard deviation of a Gaussian function. A z-score of 0σ indicates a 50% probability, while a z-score of 5σ is $\sim 1 - 10^{-7}$ probability. We place a plot of this below for convenience.

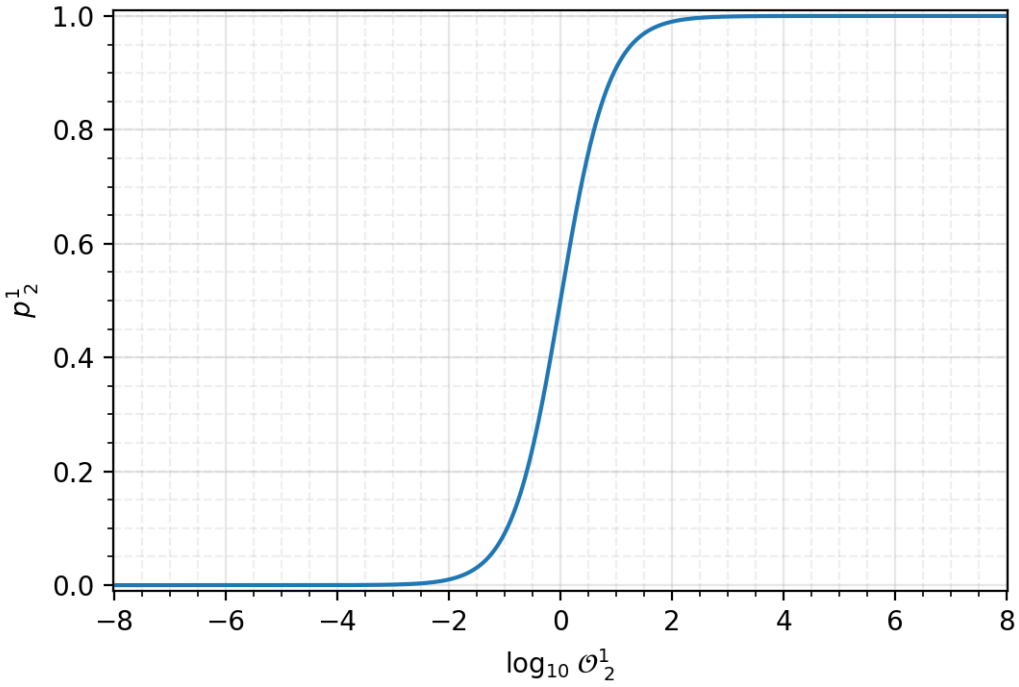

Figure 1: The probability of hypothesis 1 being favored over hypothesis 2 when considering the $\log_{10} \mathcal{O}$. When $\log_{10} \mathcal{O} = 0$, the probability for each hypothesis is 50%. At odds ratios close to 100 (0.01) the evidence becomes heavily stacked towards one hypothesis or another.

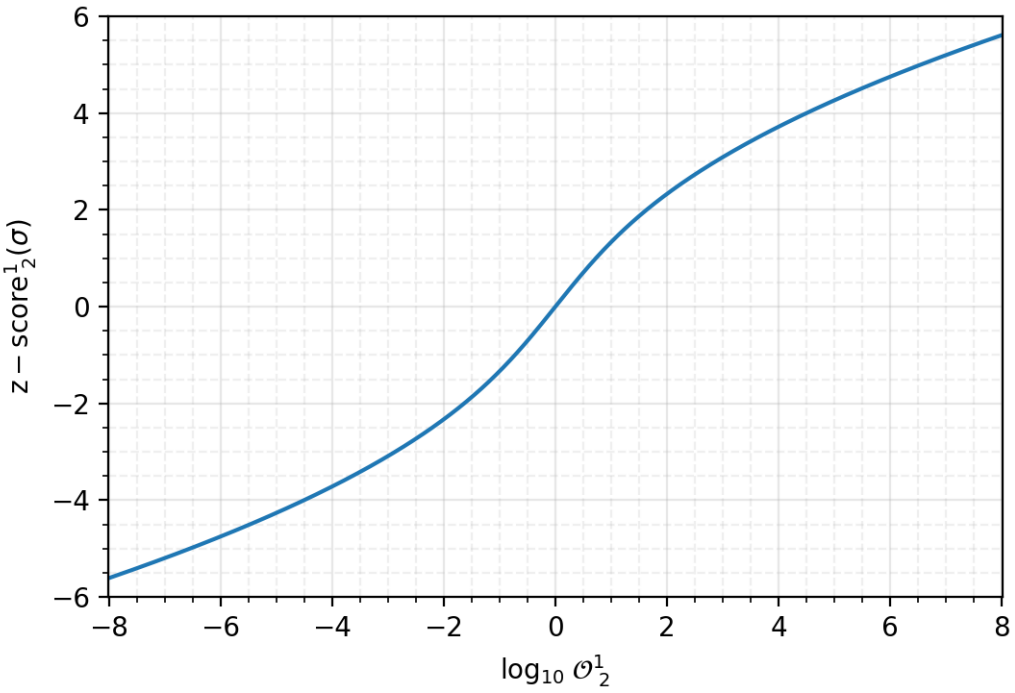

Figure 2: The Frequentist z-score pertaining to the same level of probability for hypothesis 1 being favored over hypothesis 2 when considering the $\log_{10}\mathcal{O}$. When $\log_{10}\mathcal{O} = 0$, the z-score is 0σ and the probability for each hypothesis is 50%. A z-score of $> 5\sigma$ has the same probability value as an odds ratio of $> 10^7$.

Chapter 3

Upper Limits on the Estimated Rate of Mergers of Systems with a Neutron Star

The content of this chapter are primarily taken from "Upper Limits on the Rates of Binary Neutron Star and Neutron-StarBlack-Hole Mergers from Advanced LIGOS First Observing Run" in 2016. The main focus is on the results from the PyCBC offline search analysis conducted in this study.

Between September 12^{th}, 2015 and January 19^{th}, 2016, the two aLIGO detectors in Hanford, Washington and Livingston, Lousiana conducted their first observing period. During the first observing run two binary black hole coalescences were confidently observed by the offline PyCBC gravitational wave search pipeline with a p-value of being generated by noise of less than 10^{-7}. A third candidate, termed LVT151012, was also observed by PyCBC, although with a smaller p-value. The masses of these binaries were assessed to not be compatible with the expected masses of neutron stars, and thus must be black hole binaries. This study assessed upper limits on the expected rate of mergers of neutron star binaries and neutron-star-black-hole binaries.

3.1 The PyCBC Offline Search in the First Observing Run

Below we briefly describe the methodology of the PyCBC Offline Search Pipeline in the first observing period.

The PyCBC Offline search analysis (cite) is a matched filter search for compact binary coalescence...

Derive matched-filter-snr...

3.1.1 A Compact Binary Coalescence Template Bank

The PyCBC Offline search is a modeled-search for gravitational waves from compact binary coalescence. And so we construct a bank of potential templates for astrophysical gravitational wave signals. This bank is referred to as the template bank and holds a catalog of gravitational wave signals that could be discovered. To do so we are required to generate templates in such a way as to span the parameter space of potential signals. In the first observing run, the template bank was a four-dimensional (parameter) bank. The four parameters were component masses of the binaries, m_1, m_2, and the component angular-momentum aligned-spin of the binary χ_1^z and χ_2^z. The masses are drawn such that $m_1 > m_2$, which helps reduce the size of the template bank. The aligned-spin components are drawn to reflect the expected astrophysical properties of the binaries in that respective region of the mass parameter space.

The template bank is segmented into three sections, intended to find BNS, NSBH, and BBH signals. The template bnak in the first observing run consisted of approximately $250,000$ templates.

For n detectors the harmonic mean of the detector's power spectral densities is chosen as the reference power spectral density when designing a common template bank for the Hanford and Livingston detectors. Previous searches used independent template banks for independent detectors, however this was not done in the first-observing run.

As the run progresses the template bank is validated using injected software signals to ensure that the noise properties of the detector have not changed to a sufficient degree as to make the template bank and template placement ineffectual in recovering expected signals. To do so, a population of simulated signals are expected to have a minimal match of 97% with any particular template in the bank. This is often called the Fitting Factor of the template bank. It is often a desired trait that less than 1% of the simulated signals exhibit less than a 3% loss in anticipated signal-to-noise-ratio due to the discreteness of the template bank (or the change in the detector's power spectral density). The template bank was validated several times across the entire

first observing run and only one template bank was necessary for the duration of the analysis.

3.1.2 Data Quality and Conditioning

It is well-known in the gravitational wave community that the output of the detector is not Gaussian and non-stationary over long periods of times and that the quality of the data from the interferometers is not always suitable for gravitational-wave analysis (cite). For this reason, the data undergo data quality checks and data conditioning prior to being analyzed for gravitational wave events.

In particular, there are several data quality vetoes

Gating

3.1.3 The Ranking Statistic: Signal-to-Noise-Ratio and NewSNR

Derive SNR from Gaussian distribution

Maximize over amplitude

Maximize over phase

However, the noise in the detector is decidedly not Gaussian, nor stationary and so the a new ranking statistic was developed as a further signal consistency check. The signal-to-noise-ratio statistic was made subject to a reduced χ^2 consistency check, hence labeled as χ^2_r. The χ^2_r test separates the signal into k frequency bins and measures the accumulated signal-to-noise-ratio in each frequency bin. For a true signal, we expect a steady contribution to the signal-to-noise-ratio from each frequency bin. Thus, the χ^2_r test for a true signal embedded in Gaussian noise returns a value ≤ 1, while a χ^2_r value > 1 indicates a poor fit to the data across all frequency bins. The χ^2_r test can be expressed as:

$$\chi^2_r = \frac{k}{2k-2} \sum_{i=1}^{k} \left(\rho_i - \frac{\rho_i}{k}\right)^2 \tag{3.1}$$

Here ρ_i represents the signal-to-noise-ratio in the i^{th} frequency bin, and k is the number of frequency bins.

Prior to the first observing run, $k = 16$ frequency bins were used, however, in the first observing run the number of frequency bins used was dependent on the peak frequency of the template. The formulation in the first observing run was,

$k = 0.4\,[f_{\rm peak}/{\rm Hz}]^{2/3}$, and k is rounded to the nearest integer. Next, the reduced chi^2 values are used to re-rank the signal-to-noise-ratio ranking statistic through an empirically designed ranking statistic called newSNR. When $\chi^2_r > 1$, newSNR can be calculated as:

$$\hat{\rho} = \frac{\rho}{\left\{\left[1 + (\chi^2_r)^3\right]/2\right\}^{1/6}} \tag{3.2}$$

For values of $\chi^2_r \leq 1$, NewSNR is equal to the signal-to-noise-ratio ($\hat{\rho} = \rho$), while for values of $\chi^2_r > 1$, Eq. 3.2 is applied and newSNR is less than the signal-to-noise-ratio ($\hat{\rho}_c < \rho$). Finally then, for the Livingston and Hanford detectors the ranking statistic of a candidate trigger is $\hat{\rho}^2_{\rm network} = \hat{\rho}^2_{\rm L1} + \hat{\rho}^2_{\rm H1}$. For a more detailed description of the first observing run's ranking statistics and the χ^2_r please consult N, N.

3.1.4 Evaluation of the Statistical Significance of Events

Description of timeslides

After coincident triggers are collected across five days of coincident analysis time between the two detectors. These coincident triggers are the foreground event triggers and are potentially of astrophysical origin. In order to assess the statistical significance of these foreground triggers we need an estimate of the rate at which non-astrophysical or background noise could potentially generate triggers of similar magnitude in the ranking statistic. Unfortunately, gravitational waves cannot be shielded from as electromagnetic waves can be, and so we use an empirical method to estimate the false alarm rate of triggers at a specific loudness in ranking statistic.

Description of inclusive and exclusive background

Description of p-value evaluation for single events

Since $n = 3$ independent PyCBC offline searches were carried out in the first observing run for different compact binary coalescences, a Bonferroni correction is applied to the measured p-values from each search. The Bonferroni correction, or trials factor, states that if one conducts n trials in searching for an effect, the threshold α on the p-value should also be divided by the number of trials n. This provides a conservative method for re-ranking the statistical significance of events provided multiple instances of searches. Since the different CBC search categories are weakly correlated (e.g. GW150914 was seen in the BBH search and the BBH-NSBH search), a better p-value correction could be considered to be less than 3. Quantifying this

level of correlation however, was beyond the scope of the study.

The p-value of a particular event can then be converted to a z-score, or the number of standard deviations in a one-tailed Gaussian probability integral via:

$$z = -\sqrt{2}\,\mathrm{erf}^{-1}\left[1 - 2\,(1 - p)\right]. \tag{3.3}$$

Here z represents the z-score, erf^{-1} is the inverse error function, and p is shorthand for the p-value (e.g. 0.05). It is standard practice in many subfields of physics to require a z-score of 5 (sometimes called 5σ) level of confidence before rejecting the null hypothesis that noise generated the gravitational wave signals. This corresponds to a p-value of $< 10^{-7}$.

Additional data quality checks are then used to verify that the gravitational wave candidate is more consistent with a signal hypothesis than with the null (noise) hypothesis. The probability that the event is of astrophysical origin cannot be gathered from the p-value itself but requires the use of Bayes Rule to calculate the probability of being of astrophysical origin, p_{astro}. P-values are statements about the probability of observing data greater than some critical value under the assumption that the null hypothesis is true, while the probability of the event being astrophysical is a probability of a hypothesis being true (that the signal model better *explains* the candidate than a noise model).

3.1.5 Results of the Search

The PyCBC offline search analysis found two confident gravitational wave events, GW150914 and GW151226, with statistical significance $> 5\sigma$ (p-value $< 10^{-7}$) and a gravitational wave candidate, LVT151012, with a statistical significance of $\sim 2\sigma$. These statistical significance values are quoted after a trials factor of 3 was applied to the analysis. All three triggers were found to be consistent with stellar mass black hole binaries, although LVT151012 was not confidently claimed as a detection.

No statistically significant BNS or NSBH events were found. And so we turn towards trying to determine an upper limit on the rate of mergers of BNS and NSBH events.

3.2 Bayesian Rates Estimation

Since no events were discovered in the first observing run, we seek to try to model the expected number of events for BNS and NSBH events. The expected number of events Λ can be expressed as:

$$\Lambda = R \langle VT \rangle. \tag{3.4}$$

Here, the rate of mergers is given by R, and $\langle VT \rangle$ describes the sensitive-spacetime volume averaged over space, observation time, and the parameters of the source population of interest. The units of a rate are henceforth in $\mathrm{Gpc}^{-3}\mathrm{yr}^{-1}$. For reference the core-collapse supernova rate is $\sim 10^5\,\mathrm{Gpc}^{-3}\,\mathrm{yr}^{-1}$ (Cappellaro et al. 2015).

Here we assume that gravitational wave events are exceedingly rare, and so we can model the likelihood for finding an observation of a gravitational wave event in the data D as a Poisson distribution. The likelihood of seeing s_i independent events after N observations, given some Λ can be expressed as:

$$p\left(s|\Lambda\right) = \prod_{i=1}^{N} \frac{1}{s_i!} \Lambda^{s_i} e^{-\Lambda} \tag{3.5}$$

In the case that we have zero observations, in one observing period, this likelihood can be written as:

$$p\left(s|\Lambda\right) = e^{-\Lambda}. \tag{3.6}$$

Using Bayes theorem we can find a posterior probability on the number of expected rates as:

$$p\left(\Lambda|s\right) \propto p\left(\Lambda\right) p\left(s|\Lambda\right). \tag{3.7}$$

Here $p\left(\Lambda\right)$ is our prior belief on Λ, the number of expected mergers. The most straightforward prior distribution to use is a conjugate prior distribution. The advantage of a conjugate prior distribution is that our posterior distribution will be of the same kind or type of distribution as the prior distribution. For a Poisson likelihood the conjugate prior is proportional to a Gamma probability distribution, in our case we express this as:

$$p\left(\Lambda\right) \propto \Lambda^{\alpha}. \tag{3.8}$$

A conjugate prior when multiplied with its respective likelihood distribution gives a posterior distribution that is within the same family of functions as the prior distribution. Setting $\alpha = 0$ gives a uniform prior distribution, while setting $\alpha = 1/2$

gives $p(\Lambda) \propto 1/\sqrt{\Lambda}$. This is the Jeffreys prior for the Poisson likelihood in that it is an uninformative prior. A Jeffreys prior is sometimes useful in that it gives a equal prior probability weight to all possible values of the parametrization of the Bayesian inference.

Prior information from previous LIGO data were not used in a prior on Λ due to changes in the detector sensitivity and the large changes to $\langle VT \rangle$.

The likelihood $e^{-\Lambda}$ is estimated by empirically measuring $\langle VT \rangle$ through the use of software injections of gravitational wave events into the aLIGO data set. These software injections involve generating gravitational wave strain data for different astrophysical objects in the radiation frame and then projecting them into the detector frame. These simulated signals are then added in to aLIGO data for detection.

Simulated signals are considered to be recovered when they are detected with an IFAR of $> 100\,yrs$. Since only a finite set of software injections can be processed and recovered, we use Monte-Carlo integration techniques to estimate the volume of injections recovered and a variance estimate of the volume recovered. The volume of injections recovered is compared then to the spacetime volume that the injections were injected from. Furthermore, the uncertainty in the calibration of the detector plays a role in the uncertainty in the recovered volume, since the calibration of the data plays a role in recovered signal-to-noise-ratios of simulated signals. It was found that the calibration added an 18% uncertainty to the estimated sensitive volume $\langle VT \rangle$. There is an additional uncertainty due to waveform modelling between injected simulated signals and waveform models used for recovery. This is additionally folded into the uncertainty on $\langle VT \rangle$.

It is convenient to consider the posterior on Λ as a joint-posterior on the variables that compose Λ, that being R and VT. To do so we can refactor the prior distribution on Λ into a joint prior distribution on R and $\langle VT \rangle$.

$$p(R, \langle VT \rangle) = p(R|\langle VT \rangle)\, p(\langle VT \rangle). \tag{3.9}$$

In this publication, the prior $p(R|\langle VT \rangle)$ was chosen to be uniform in R, or a Jeffreys prior given by, $1/\sqrt{R\langle VT \rangle}$. In keeping with previous Refs, (Abbott et al. 2016f,m,c), the prior distribution on $\langle VT \rangle$ is given as a log-normal distribution given by:

$$p(\langle VT \rangle) = ln\,\mathcal{N}\left(\mu, \sigma^2\right). \tag{3.10}$$

Here μ is the average value for $\langle VT \rangle$ obtained from the offline search, and σ represents the 18% uncertainty in $\langle VT \rangle$ due to calibration uncertainty.

Thus the posterior of Λ in the new parametrization becomes:

$$p\left(R, \langle VT \rangle\right) \propto p\left(R, \langle VT \rangle\right) \, e^{-R\langle VT \rangle}. \tag{3.11}$$

To obtain a posterior distribution on the rate R of mergers of a certain class we are required to marginalize the joint-posterior in R and $\langle VT \rangle$ over $\langle VT \rangle$. This can be expressed as:

$$p\left(R|s\right) = \int p\left(R, \langle VT \rangle|s\right) d\langle VT \rangle. \tag{3.12}$$

Finally then, the upper limit on the rate at the 90% credible level is possible by integrating the marginalized posterior distribution of R from 0 to an upper-value of R_{critical} that grants a 90% probability.

$$0.9 = \int_0^{R_{\text{critical}}} p\left(R|s\right) dR. \tag{3.13}$$

For a uniform prior in $p\left(R, \langle VT \rangle|s\right)$ and no uncertainty in $\langle VT \rangle$ the 90% credible limit is given as:

$$R_{\text{critical}} = \frac{-ln\left(1 - 0.9\right)}{\langle VT \rangle} = \frac{2.303}{\langle VT \rangle}. \tag{3.14}$$

The expression in Eq. 3.14 is also the Frequentist's 90% confidence interval(cite Jolien), although one should be careful in that the interpretation of the intervals are distinct in Frequentism and Bayesianism. Under the Jeffreys prior on $p\left(R, \langle VT \rangle|s\right)$ the upper limit can be expressed as:

$$R_{\text{critical}} = \frac{\left[\text{erf}^{-1}\left(0.9\right)\right]^2}{\langle VT \rangle} = \frac{1.353}{\langle VT \rangle}. \tag{3.15}$$

Using these expressions, we can thus model the rate of mergers R of a particular class of binaries by conducting software injections for that class of binaries that model our expectations of the astrophysical population for that binary.

3.2.1 Astrophysical Populations of BNS and NSBH

There are thousands of identified neutron stars, of which the majority are detected as pulsars (). Of these there are an estimated 70 neutron stars identified as binary neutron stars. Mass estimates of neutron stars range between $\approx 1M_{\odot}$ to $3M_{\odot}$.

Eight candidate neutron star binaries permitted measurement of the binary component masses (), of which an estimate of the distribution of events is consistent with $\mathcal{N}(\mu = 1.35, \sigma^2 = 0.13^2)$.

With respect to the spin characteristics of neutron stars, the fastest spinning pulsar was measured to have a rotational frequency of $716\,Hz$. For reasonable estimates of the mass and moment of inertia for this pulsar, this corresponds to a dimensionless spin $\chi = c|S|/(Gm^2) \sim 0.4$, where m is the mass of the pulsar, c is the speed of light, G is Newton's gravitational constant, and $|S|$ is the angular momentum. However, the fastest spinning neutron star in a binary system is estimated to have a $\chi \leq 0.04$ (Brown et al). A neutron star in a binary can be spun up to larger rotational frequencies if its binary companion accretes matter and angular momentum on to the neutron star. This is called a recycled neutron star and estimates of the spin of a possible candidate BNS pulsar J1807-2500B is $\chi \sim 0.2$.

With these things in mind, our simulated astrophysical population for neutron star binaries contains two distinct populations in the mass parameters that we simulate. The first BNS population is drawn uniform in component masses between $[1, 3]\,M_\odot$, while another population is drawn with mass distribution, $\mathcal{N}(\mu = 1.35, \sigma^2 = 0.13^2)$. Each mass distribution is also subject to two distinct component spin distributions. The first spin distribution is an isotropic component-spin distribution such that $|\chi_i| < 0.05$. The second spin distribution is an isotropic component-spin distribution with $|\chi| < 0.4$. This second spin distribution is drawn with considerations from (Nitz 2015) that it is not necessary to create a template bank modeling the spin distribution of BNS mergers above 0.4 since BNS mergers with spin > 0.4 can still be well recovered by templates with $\chi_i < 0.4$.

Neutron-star-black-hole systems are thought to be efficiently formed in one of two ways: either through the stellar evolution of field binaries or through dynamical capture of a NS by a BH (Grindlay et al. 2006; Sadowski et al. 2008; Lee et al. 2010; Benacquista and Downing 2013). Though no NSBH systems are known to exist, one likely progenitor has been observed, Cyg X-3 (Belczynski et al. 2013).

NSBH mass distribution

NSBH spin distribution

Each population is also simulated according to a uniform distribution in the binary orientation, that is to say that no binary orientation relative to the detector is

preferred. This means that binaries are drawn from a probability distribution that is uniform in ι, inclination angle, and Ψ polarization angle. The binaries are also distributed uniformly across the sky, which is uniform in α right ascension and δ declination.

Finally, binaries are drawn uniformly in comoving volume, which we briefly describe below following (Hogg 1999). A comoving volume in terms of cosmology means the volume of the universe assuming a frame of reference which expands with the universe. In particular we measure a comoving volume out to a particular luminosity distance or redshift z. The redshift z is defined via:

$$1 + z = \frac{a(t_o)}{a(t_e)}. \tag{3.16}$$

The redshift z is defined via the ratio of $a(t_0)$, the size of the universe (scale factor) at the time of observation, and $a(t_e)$, the size of the universe at the time of emission. As the name implies the redshift factor relates how light (or gravitational waves) are redshifted in frequency due to the expansion of the universe. Those who are interested in the explicit definition of comoving volume can see (Hogg 1999) as the formal definition in terms of measured and inferred cosmological constants is quite involved. The comoving volume is dependent on H_0, Ω_M, Ω_Λ and Ω_k measured or inferred values from Planck2015.

Here H_0 is the Hubble constant which relates the present-time recessional velocity of distant stars relative to the distance of the distant stars in the form $v = H_0 d$. The Hubble constants is measured in experiment. We have the term $\Omega_M \equiv \frac{8\pi G\rho_0}{3H_0^2}$, where G is Newton's gravitational constant and ρ_0 represents the measured mass density of the universe at the present time. We also have the term $\Omega_\Lambda \equiv \frac{\Lambda c^2}{3H_0^2}$, where Λ is the cosmological constant representing the energy density of space, and c is the speed of light. Finally then $\Omega_k \equiv 1 - \Omega_\Lambda - \Omega_M$, defining a homogeneous, isotropic, and matter dominated universe, and is responsible for describing the spacetime curvature of the universe. We take the mean values from Planck2015 without any associated variation on the parameters.

Due to cosmological distances being involved the masses of binaries will be measured in the detector frame, $M_{\rm det}$, relative to the source frame mass, $M_{\rm src}$ under the relation:

$$M_{\rm det} = (1 + z)\, M_{\rm src} \tag{3.17}$$

Waveform models used BNS TaylorT4 NSBH

The software injections are generated in the radiation frame using time domain waveforms. They are assigned a geocentric time of arrival (coalescence) at which point they are projected from the radiation frame onto a specific detector frame for the Livingston and Hanford detectors respectively. This strain is then added into the strain of the data at the time of arrival.

3.2.2 Rates Inference Results

Using the equations in the previous section we found improved upper limit results on the expected rate of mergers of BNSs and NSBHs.

The posterior probability density on the rate of merger of BNSs for a Gaussian-like distribution of masses can be found in Fig. 3 for the two different prior distribution choices on the rate of mergers. Following a uniform prior on Λ we found that the 90% credible level on the upper limit on BNS merger rate is $\sim 1.2 \times 10^4\, Gpc^3\, yr^1$ for low-spin and high-spin populations. This is approximately a factor of 10 improvement over (Abedie et al. 2012a). The posterior rates inference for a uniform mass distribution can be found in Fig. 4 and does not differ significantly from the inference results presented in Fig. 3.

The 90% posterior probabilty upper limits on the rate of mergers for NSBHs can be summarized in Table **??**. We found that the search is less sensitive to populations of NSBHs with isotropic spin distributions when compared to (anti-) aligned spin distributions. One factor that contributes to this increased sensitivity to (anti-) aligned spin NSBHs is the fact that the template bank is constructed using (anti-) aligned spin templates and not isotropic-spin distributed templates.

3.3 Astrophysical Interpretation and Future Results

The inferred upper limit estimation on the rate of binary mergers for compact binary mergers involving neutron stars permit us to explore if any population models for the genesis of these binaries can be ruled out by the available data. For BNS populations we compare our posterior rate estimations with 10 other studies. While for the NSBH populations we compare our posterior rate estimations with 9 other studies.

For BNS populations we consider the population models from : During the first

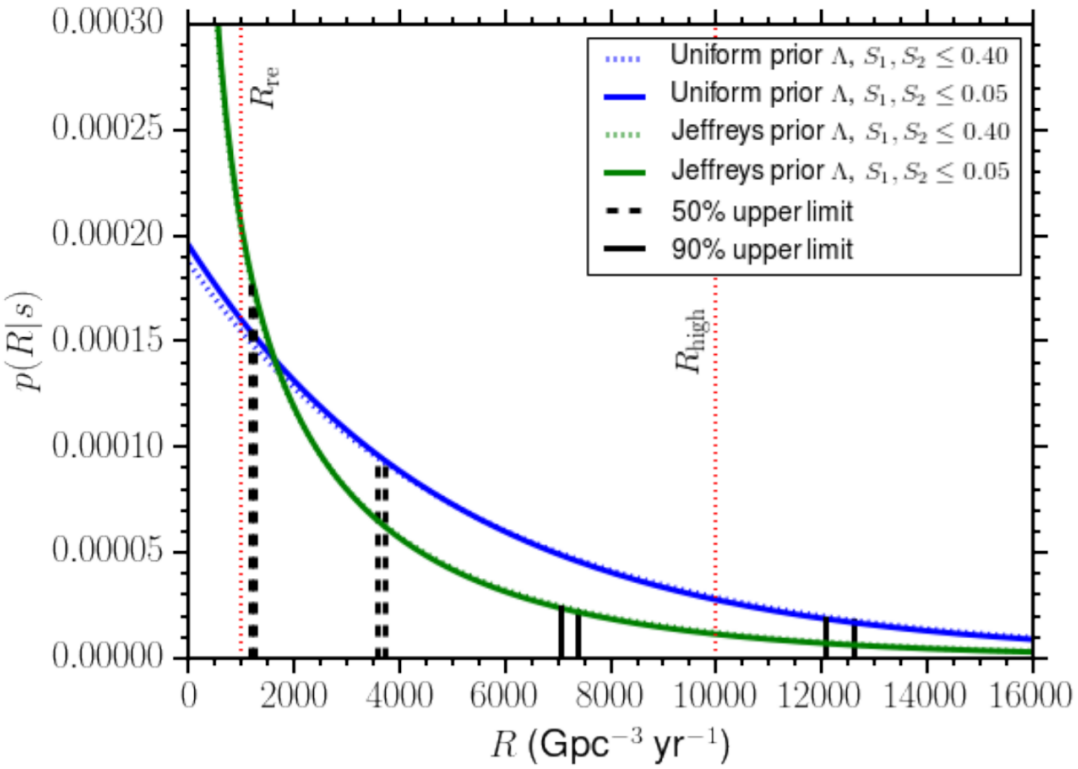

Figure 3: Posterior probability density on the rate of BNS mergers. Blue curves represent a uniform prior on Λ, while green curves represent a Jeffreys prior on Λ. The solid (low spin population) and dotted (high spin population) posteriors almost overlap. The vertical dashed and solid lines represent the 50% and 90% credible upper limits respectively for each choice of prior on Λ. For each pair of vertical lines, the left line is the upper limit for the low spin population and the right line is the upper limit for the high spin population. Also shown are the realistic $R_{\rm re}$ and high end $R_{\rm high}$ of the expected BNS merger rates identified in Ref. (Abadie et al. 2010).

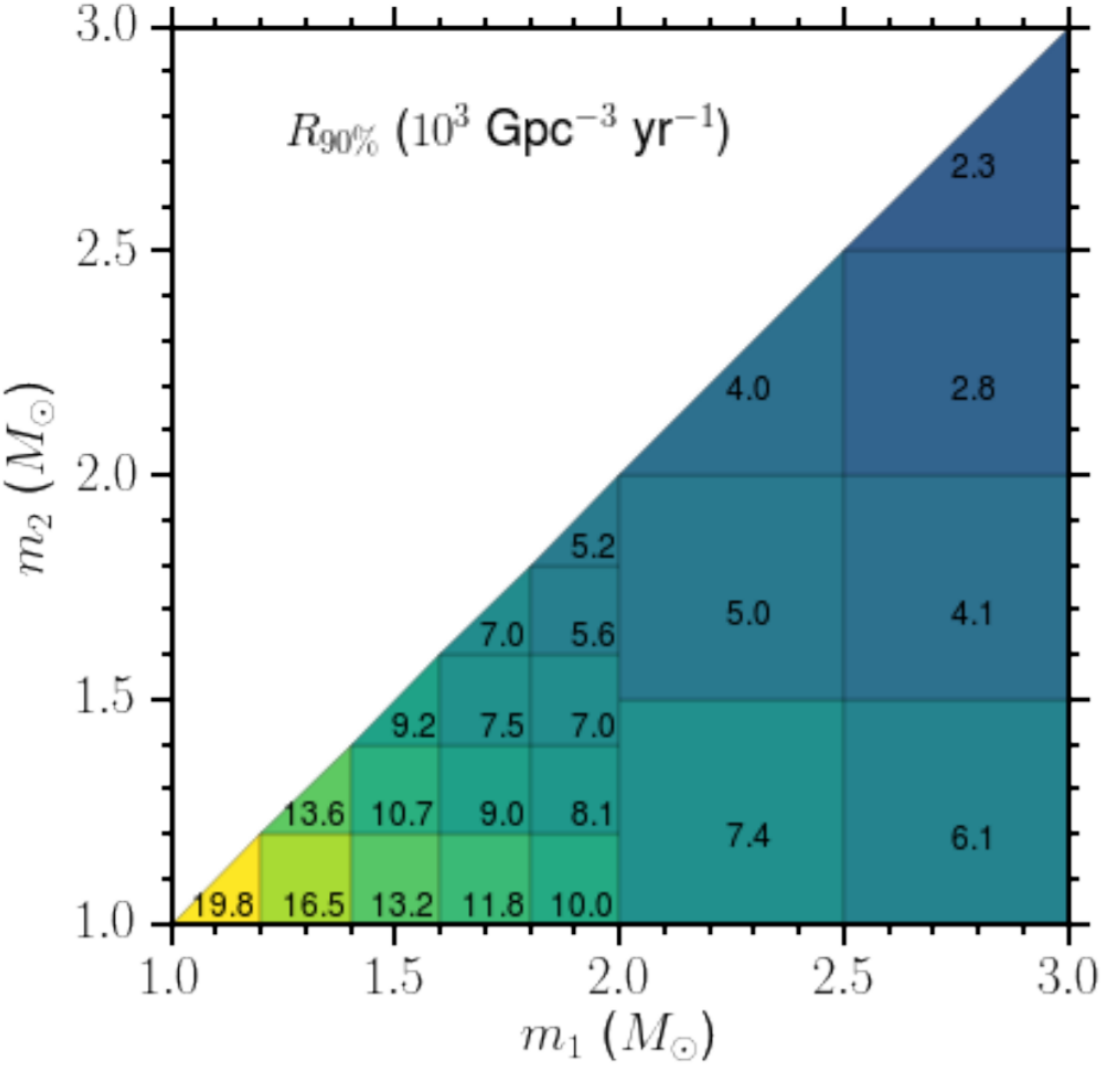

Figure 4: The 90% credible upper limit on the BNS merger rate as a function of the two component masses. Here the upper limit for each bin is obtained assuming a BNS population with masses distributed uniformly within the limits of each bin, considering isotropic spin direction and dimensionless spin magnitudes uniformly distributed in $[0, 0.05]$.

NS Mass ($M_\odot$)	BH Mass ($M_\odot$)	Spin Distribution	$R_{90\%}$	$(Gpc^{-3}\,yr^{-1})$
1.4	5	Isotropic	3,600	
1.4	5	Aligned	3,210	
1.4	10	Isotropic	2,530	
1.4	10	Aligned	1,850	
1.4	15	Isotropic	2,300	
1.4	15	Aligned	1,280	

Table 1: The 90% credible upper limit for NSBH systems with isotropic and aligned spin distributions. The NS spin magnitudes are in the range [0,0.04] and the BH spin magnitudes are in the range [0,1].

observing run our upper limit estimations on the rate of BNS mergers were not in conflict with any of the population models considered. It was estimated that assuming estimates of O2 and O3 aLIGO detector sensitivities that continued non-detections of BNS mergers in aLIGO's second observing run and third observing run that our upper limits could begin to rule out some of the population models considered. Fortunately, however GW170817, a BNS merger event, was discovered in the second observing run.

For NSBH poppulations we consider the population models from:

BNS and NSBH mergers are considered to be possible sources for short-hard gamma ray bursts (). We can also explore a geometrical argument regarding the beaming angle of binary neutron stars and neutron star-black hole binaries if they are assumed to be the progenitors of short-hard gamma ray burst events.

Rate Implications on population Models

Explain short-hard gamma ray bursts... Assuming that every merger from a BNS or an NSBH produces a gamma ray burst the rate of short-hard gamma ray bursts detected follows from a geometric relationship of the beaming angle of the gamma ray burst jet and the expected rate of mergers of these binaries. The relation can be given as follows:

$$R_{\rm GRB} = [1 - cos\,(\theta_{\rm jet})]\,R_{\rm merger}. \tag{3.18}$$

The relation can be understood as an axially symmetric jet whose solid angle covers a fraction of a unit sphere that envelops the binary. Hence, for some rate of binary

mergers, $R_{\rm merger}$, a fraction of these mergers will be seen as gamma ray bursts from which we can infer a rate of gamma ray bursts, $R_{\rm GRB}$. This fraction is related to the opening angle of the gamma ray jet, $\theta_{\rm jet}$. In blank we assumed a rate of gamma ray bursts of $R_{\rm GRB} = 10^{+20}_{-7}{\rm Gpc}^{-3}{\rm yr}^{-1}$ from (Nakar et al. 2006; Coward et al. 2012). The inferred beaming angles can then be constructed by using the estimates on $R_{\rm merger}$ from this study. The study found that we could infer at 90% confidence that the gamma-ray burst jet angle must be greater than $2.^{\circ}3^{1.7}_{-1.1}$ if short-hard gamma ray bursts are exclusively caused by BNS mergers and if each BNS merger produces a short-hard gamma ray burst. This inference comes from the low-spin, Gaussian mass distribution population of BNS. If NSBH mergers from $5M_{\odot}$ and $1.4M_{\odot}$ with aligned spin distributions are considered as the main progenitors of short-hard gamma ray bursts then the 90% confidence interval is $\theta_{\rm jet}^{\rm lower\,limit} = 4.^{\circ}3^{+3.1}_{-1.9}$.

3.3.1 GWTC-1: Inferences from the 1st and 2nd Observing Run

The LIGO Scientific Collaboration and Virgo Collaboration (LVC) found in [] that from the confident BNS merger, GW170817, that they were able to place a posterior estimate on the rate of BNS mergers in the universe to be, at the 90% credible interval to be between $110 - 3840\,{\rm Gpc}^{-3}{\rm yr}^{-1}$, and with no confident NSBH merger detections a 90% confidence upper limit on the rate of NSBH mergers of $610\,{\rm Gpc}^{-3}{\rm yr}^{-1}$. Note that the 90th percentile matches relatively closely with the 50th percentile estimate for the uniform prior distribution on Λ in Fig. 3. The procedure for generating an upper limit on the rates given a detection event is somewhat different from the procedure outlined in this chapter, however the procedure in [] for estimating the upper limit on the rate of NSBH mergers is mostly identical to the procedure outlined in this chapter, with the exception of some changes to the PyCBC offline search pipeline in terms of ranking statistic of candidate triggers.

Chapter 4

1-OGC

4.1 Introduction

The following chapter is taken from work presented in [4].

The Advanced LIGO gravitational wave observatories [5] performed their first observing run (O1) from September 12, 2015 to January 19, 2016. This provided a total of 51.5 days of coincident observations from the two detectors located in Hanford, WA and Livingston, LA. The binary black hole mergers observed in this observing run have been reported by the LIGO and Virgo Collaborations (LVC) in [6, 7, 1]. These binary black hole detections have been independently studied by [8, 9, 10].

Since the publication of the results by [1, 11], improvements to the data-analysis methods used [12] have been implemented [13, 14, 15]. Using these improvements, we re-analyze the O1 data and provide—for the first time—a full catalog of candidate events from a matched filter search for compact binary coalescences using the O1 data, which we call 1-OGC. This catalog provides estimates of the significance of previously known events and a ranked list of sub-threshold candidates. Although not significant by themselves, these sub-threshold candidates can be correlated with archival data or transient events found by other astronomical observatories to provide constraints on the population of compact-object mergers [16, 17].

Our catalog is based entirely on public, open data and software. We use the LIGO data available from the Gravitational Wave Open Science Center [18], and analyze the data using the open source PyCBC toolkit [19, 20, 21].This toolkit was also used by one of the two analyses described in [12]. The lowest mass sources targeted in our

search are neutron star binaries with total mass $m_1 + m_2 = 2\,M_\odot$. The search space extends to binary black hole systems that produce gravitational waveforms longer than 0.15 s from 20 Hz. This corresponds to a total mass up to $500 M_\odot$ for sources with high mass ratios and spins where the component aligned with the orbital angular momentum is positive and large. For binaries with negligible spin, this corresponds to total mass $\lesssim 200 M_\odot$. The search space also includes neutron star–black hole binaries. After applying cuts for data quality [22, 23], a total of 48.1 days of coincident data are searched for signals.

The three most significant signals in the catalog correspond to GW150914 [6], LVT151012 [6, 1], and GW151226 [7], respectively. No other astrophysically significant signals are observed. In the analysis of [1], LVT151012 was the third-most significant event, but it was not sufficiently significant to be labeled as an unambiguous detection. With the improved methods employed here, the false alarm rate of this candidate improves by an order of magnitude and it should be considered a true astrophysical event. The analyses of [1, 11] restricted the astrophysical search space to binaries with a total mass less that $100\,M_\odot$. Our analysis extends this target space to higher mass signals. No additional signals are detected in this region of parameter space, consistent with the results of [24].

A second observing run (O2) of the Advanced LIGO detectors took place from November 30, 2016 to August 25, 2017 [25]. The Virgo gravitational wave detector also collected data for part of this period, starting from August 1, 2017. The detections reported in this second observing run thus far include nary black hole coalescence even three additional binary black hole coalescence events [26, 27, 28], and a binary neutron star merger [29]. However, the full O2 data set has not yet been released. The catalog presented here is therefore restricted to the first observing run, O1.

Our paper is organized as follows: In Sec. 4.2 and Sec. 4.3, we summarize our analysis methods, including the parameter space searched, the detection statistic used for ranking candidate events, and our method for calculating the statistical significance of events. The search results are summarized in Sec. 5.5. Our full catalog and released data are described in Sec. **??** and are available online as supplementary materials (`www.github.com/gwastro/1-ogc`). In this paper, we focus on the detection of compact objects. Since no new astrophysical events have been observed, we do

not consider measurement of the signals' parameters and refer to [1, 30] for discussion of the detected events' source-frame properties. Consequently, we quote binary mass parameters in the detector frame in this work.

4.2 Search Methodology

To search for gravitational waves from compact-object mergers, we use matched filtering [31] implemented in the open-source PyCBC library [19, 20, 21]. Our methods improve on the analyses of [1, 11, 12] by imposing a phase, amplitude and time delay consistency on candidate signals, an improved background model, and a larger search parameter space [13, 14, 15].

4.2.1 Target Search Space

A discrete bank of gravitational-wave template waveforms [32, 33, 34] is used to target binary neutron star, neutron star–black hole, and binary black hole mergers with total mass from $2 - 500 M_\odot$ [15]. The templates are parameterized by their component masses $m_{1,2}$ and their dimensionless spins $\vec{\chi}_{1,2} = c\vec{S}_{1,2}/Gm_{1,2}^2$, where $\vec{S}_{1,2}$ are the spin vectors of each compact object. For compact objects with component masses greater than $2M_\odot$, the template bank covers a wide range of spins, with $\chi_{(1,2)z} \in [\pm 0.998]$, where $\chi_{(1,2)z}$ are the components aligned with the orbital angular momentum. For compact objects with masses less than $2M_\odot$, the spin is restricted to $\chi_{(1,2)z} \in [\pm 0.05]$ [34]. Templates that correspond to sources with a signal duration less than 0.15 seconds (starting from 20 Hz) are excluded due to the difficulty in separating candidates arising from these templates from populations of instrumental glitches [15]. Consequently, the total mass boundary of the search depends strongly on the "effective spin" [35, 36],

$$\chi_{\rm eff} = \frac{\chi_{1z} m_1 + \chi_{2z} m_2}{m_1 + m_2}. \tag{4.1}$$

This dependence is visible in the distribution of the approximately 400,000 templates required to cover the space shown in Fig. 5. A dotted line in Fig. 5 denotes the upper boundary of the O1 analysis performed in [1]. For binaries with total mass greater than $4\,M_\odot$, we use the spinning effective-one-body model (SEOBNRv4) [37, 38] as template gravitational waveforms. For sources with total masses less than $4M_\odot$ we

use TaylorF2 post-Newtonian waveforms with phasing accurate to 3.5 post-Newtonian order and the dominant amplitude evolution [39, 40, 41, 42]. Our choice of template bank discretization causes less than a 10% loss in detection rate for any source within the boundaries of the template bank. Our search assumes that the source can be adequately described by only the dominant gravitational-wave mode, two component masses, non-precessing spins, and negligible eccentricity.

4.2.2 Creation and Ranking of Candidate Events

For each template and each detector, we calculate the matched filter signal-to-noise ratio (SNR) as a function of time $\rho(t)$ [31]. The template bank is divided into 15 equal sized sub-banks based on the chirp mass $\mathcal{M} = (m_1 m_2)^{3/5}/(m_1 + m_2)^{1/5}$ of each template. A single-detector "trigger" is a peak in the SNR time series that is greater than 4 and larger than any other peaks within 1s. For each sub-bank, the loudest 100 triggers (by ρ) are recorded in $\sim$ 1s fixed time windows. This method has been shown to improve search sensitivity, while making the rate of single-detector triggers manageable [43]. We have found this choice of sub-banks to be an effective method to ensure the analysis can concurrently record triggers from separate regions of parameter space that respond differently to instrumental noise. Other choices are possible.

We use the data-quality segments provided by the Gravitational-Wave Open Science Center to exclude triggers that occur in times when there are problems with the detectors' data quality [22, 23]. In addition, very loud transient glitches, corresponding to $> 100\sigma$ deviations from Gaussian noise, are excised from the strain data according to the procedure of [19] before calculation of the SNR time series. However, there remain many types of transient non-Gaussian noise in the LIGO data which produce triggers with large values of SNR [44, 22, 23].

For every trigger with $\rho > 5.5$ we calculate the signal consistency test, χ^2_r, introduced in [45]. The statistic χ^2_r divides the matched filter into frequency bands and checks that the contribution from each band is consistent with the expected signal. The statistic takes values close to unity when the data contains either Gaussian noise or the expected signal and larger values for many types of transient glitches. We impose the SNR limit as the χ^2_r test is generally non-informative when $\rho < 5.5$. The

χ^2_r value is used to re-weight the SNR ρ as [46]

$$\tilde{\rho} = \begin{cases} \rho & \text{for } \chi^2_r \leq 1 \\ \rho \left[\frac{1}{2}\left(1 + (\chi^2_r)^3\right)\right]^{-1/6}, & \text{for } \chi^2_r > 1. \end{cases} \tag{4.2}$$

For single-detector triggers from templates with total mass greater than $40M_\odot$ we apply an additional test, $\chi^2_{r,sg}$, that determines if the detector output contains power at higher frequencies than the maximum expected frequency content of the gravitational-wave signal [14]. This test is only applied for higher mass systems, since these templates are shorter in duration and more difficult to separate from instrumental noise. For other systems, we set $\chi^2_{r,sg} = 1$. Using this statistic, we apply a further re-weighting as

$$\hat{\rho} = \begin{cases} \tilde{\rho} & \text{for } \chi^2_{r,sg} \leq 4 \\ \tilde{\rho}(\chi^2_{r,sg}/4)^{-1/2}, & \text{for } \chi^2_{r,sg} > 4. \end{cases} \tag{4.3}$$

Candidate events are generated when single-detector triggers occur in both the LIGO Hanford and Livingston data within 12 ms (the light-travel time between the observatories extended by 2 ms for signal time-measurement error) and if the triggers are recorded in the same template in each detector [19]. Following the procedure of [13], we model the distribution of single detector triggers from each template as an exponentially decaying function, $\lambda(\hat{\rho}, \vec{\theta}^N)$, where $\vec{\theta}^N$ allows the parameters of the exponential to vary as a function of total mass, symmetric mass ratio $\eta = m_1 m_2/M^2$, and $\chi_{\rm eff}$. This fitted model allows us to rescale $\hat{\rho}$ to better equalize the rate of triggers from each template.

We improve upon the ranking of candidates in [11, 1] by also taking into account $p^S(\vec{\theta}^S)$, which is the expected distribution of SNR ρ_H and ρ_L, phase difference $\phi_{c,H} - \phi_{c,L}$, and arrival time delay $t_{c,H} - t_{c,L}$ between the two LIGO instruments for an astrophysical population [13]. No assumption is made about the distribution of intrinsic source parameters in this term. The primary benefit arises from assuming the population of sources is isotropically distributed in orientation and sky location. The final ranking statistic $\tilde{\rho}_c$ is then calculated as

$$\tilde{\rho}_c \propto \left[\log p^S(\vec{\theta}^S) - \log\left(\lambda_H(\hat{\rho}_H, \vec{\theta}^N)\lambda_L(\hat{\rho}_L, \vec{\theta}^N)\right)\right] + \text{const.} \tag{4.4}$$

This expression is normalized so that $\tilde{\rho}_c$ approximates the standard network SNR $\rho_c = (\rho_L^2 + \rho_H^2)^{1/2}$ for candidates from regions of parameter space that are not affected

by elevated rates of instrumental noise. Candidates from regions affected by elevated rates of noise triggers are down-weighted and assigned a smaller statistic value by this method. As multiple candidates, which arise from different template waveforms, may occur in response to the same signal, we select only the highest ranked candidate within ten seconds. A simpler version of this statistic where the single-detector exponential noise model is only a function of the template duration has also been employed in the analysis of data from LIGO's second observing run [47, 48, 27].

4.2.3 Statistical Significance

The statistical significance of candidate events is estimated by measuring empirically the rate of false alarms (FAR). To measure the noise background rate, we generate additional analyses by time shifting the data from one instrument with respect to the other by multiples of 100 ms. Since this time shift is greater than the maximum astrophysical time of flight between observatories, any candidates produced in these analyses are false alarms. This time shift is much greater than the auto-correlation length of our template waveforms of $\mathcal{O}(1\text{ms})$. The time-slid analyses are produced following the same procedure as the search; This is a key requirement for our analysis to produce valid statistical results [12]. The equivalent of more than 50,000 years of observing time can be generated from 5 days of data.

To provide an unbiased measure of the rate of false alarms at least as significant as a potential candidate, the single-detector triggers that compose the candidate event should be included in the background estimation [49]. However, when a real signal with a large $\tilde{\rho}_c$ is present in the data, the rate of false alarms for candidate events with smaller $\tilde{\rho}_c$ tends to be overestimated. This is due to the fact that the loud single-detector triggers from the real event in one detector form coincidences with noise fluctuations in the other detector, producing loud coincident background events. As in [1], an unbiased rate of false alarms can be achieved by a hierarchical procedure whereby a candidate with large $\tilde{\rho}_c$ is removed from the estimation of background for candidates with smaller $\tilde{\rho}_c$; we use this procedure here.

NEED TO EXPAND THIS PART

4.3 Evaluating Candidates based on the Astrophysical Population

We find two candidate events with FAR $<$ 1 per 50 000 years, corresponding to GW150914 and GW151226. Although FAR does not give the probability that an event is an astrophysical signal, we can be confident that these events were not caused by chance coincidence between the detectors. It is possible that these events were caused by a correlated source between the detectors. However, detailed followup studies of GW150914 and GW151226 found no correlated noise sources between the detectors that could be mistaken for a gravitational wave [22, 7].

We conclude that GW150914 and GW151226 are astrophysical in origin and use them to constrain the rate of real signals. A "true discovery rate" (TDR) can be constructed for less significant events. The TDR is defined as:

$$\mathrm{TDR}(\tilde{\rho}_c) = \frac{\mathcal{T}(\tilde{\rho}_c)}{\mathcal{T}(\tilde{\rho}_c) + \mathcal{F}(\tilde{\rho}_c)}, \tag{4.5}$$

where $\mathcal{T}(\tilde{\rho}_c)$ is the rate that signals of astrophysical origin are observed with a ranking statistic $\geq \tilde{\rho}_c$ (the "true alarm rate") and $\mathcal{F}(\tilde{\rho}_c)$ is the false alarm rate.

The true discovery rate is the complement of the false discovery rate [50], and can be used to estimate the fraction of real signals in a population. For example, if $\mathrm{TDR}(\tilde{\rho}_c) = 0.9$, it means that 90% of events with a ranking statistic $\geq \tilde{\rho}_c$ are expected to be real signals. The TDR is also independent of the observation time.

Note that TDR is not the probability that a particular event is a signal of astrophysical origin P_{astro}. For that, one needs to model the distribution of signals and noise at a given $\tilde{\rho}_c$. In this work, we use a simple model of these distributions as functions of the ranking statistic $\tilde{\rho}_c$. Models incorporating additional parameters are also possible, but we do not consider them here. As a function of $\tilde{\rho}_c$, P_{astro} can be computed as

$$P_{\mathrm{astro}}(\tilde{\rho}_c) = \frac{\Lambda_S P_S(\tilde{\rho}_c)}{\Lambda_S P_S(\tilde{\rho}_c) + \Lambda_N P_N(\tilde{\rho}_c)}, \tag{4.6}$$

where $P_S(\tilde{\rho}_c)$ and $P_N(\tilde{\rho}_c)$ are the probabilities of an event having ranking statistic $\tilde{\rho}_c$ given the signal and noise hypotheses respectively [51, 52, 53]. Λ_S and Λ_N are the rates of signal and noise events.

Since no binary neutron star or neutron star–black hole candidates are obtained from a search of the O1 data, here we restrict the calculation of both the TDR

and $P_{\rm astro}$ to binary black hole (BBH) observations. We include signals with total mass $M \geq 10\,\mathrm{M}_\odot$, mass ratio $m_1/m_2 < 5$ (where $m_1 \geq m_2$), and dimensionless spins $|\chi_{(1,2)z}| < 0.5$. These choices are based on a combination of what has been observed [1, 47, 48, 27] and what is expected from models of isolated binary-star evolution ("field" binaries). The mass distribution of field binaries is dependent on a number of unknown parameters, such as the metallicity of the environment [54]. Generally, it is expected that most binaries are close to equal mass, as typically less than 1 in $\mathcal{O}(10^3)$ simulated binaries have mass ratio > 5 in models of field-binary evolution [55]. The majority of observations of nearby X-ray binaries have yielded black holes with masses greater than $5\,\mathrm{M}_\odot$, which has led to speculation of a "mass gap" between 3–5 $\mathrm{M}_\odot$ [56, 57, 58]. The signals detected so far by LIGO and Virgo are consistent with this: the smaller component mass in the lowest-mass system known to date, GW170608, has an estimated mass of $7^{+2}_{-2}\,\mathrm{M}_\odot$ [27].

The spin distribution of black holes is not well constrained [59]. The component spins of the most significant binary black holes detected by LIGO and Virgo are only weakly constrained [1]. The best measured quantity related to spin is $\chi_{\rm eff}$. All of the BBH gravitational-wave signals detected so far have $|\chi_{\rm eff}| \lesssim 0.2$. A binary with low $\chi_{\rm eff}$ may still have component masses with large spin magnitudes, if the spins are anti-parallel or are purely in the plane of the binary. However, it seems unlikely that this would be the case for all of the detections made so far. Hence we include signals that have component spins with $|\chi_{(1,2)z}| < 0.5$. This is consistent with recent population synthesis models, which indicate that black holes must have low natal spin in order to obtain a distribution of $\chi_{\rm eff}$ that satisfies gravitational-wave observations [60, 61].

To estimate the rate and distribution of false alarms that arise only from the region consistent with this selected population of binary black hole mergers, we must determine which templates are sensitive to these sources. It is necessary to analyze a simulated set of signals since the template associated with a particular event is not guaranteed to share the true source parameters. We find that the region of the template bank defined by $M > 8.5\,\mathrm{M}_\odot$, $m_{1,2} > 2.7\,\mathrm{M}_\odot$, and $\chi_{\rm eff} < 0.9$ is effective at recovering this population of sources. This region is shown in Fig. 5 in red.

To estimate the true rate $\mathcal{T}$, we use the two significant events observed during O1, GW150914 and GW151226. We do not use any of the O2 events because the

full data is not yet available for analysis, making it difficult to obtain a consistent rate estimate. The total analysis time in O1 was $\sim$ 48 days, giving $\mathcal{T} \approx 15\mathrm{yr}^{-1}$. Given the uncertainty in this estimate based on only two events, we take the rate of observations as a Poisson process, and choose the lower 95% bound on $\mathcal{T}$. This yields a $\mathcal{T} \approx 2.7\mathrm{yr}^{-1}$. For the calculation of the TDR we use this value for all events, independent of their ranking statistic. This means we likely underestimate the TDR for events quieter than GW151226 and GW150914, but this is a conservative bias.

To estimate the probability that a given event is astrophysical in origin P_{astro}, we model the distribution of signals and noise as a function of $\tilde{\rho}_c$. It is reasonable to approximate the signal probability distribution $P_S(\tilde{\rho}_c)$ as $\propto \tilde{\rho}_c^{-4}$ [62, 63]. We normalize the signal number density $\Lambda_S P_S(\tilde{\rho}_c)$ so that the number of signals with $\tilde{\rho}_c$ greater than or equal to some threshold $\tilde{\rho}_c^{\dagger}$ is $\approx 2.7\mathrm{yr}^{-1}$. We make the conservative choice to place $\tilde{\rho}_c^{\dagger}$ at the value of the next largest $\tilde{\rho}_c$ value after GW150914 and GW151226.

To approximate the noise number density $\Lambda_N P_N(\tilde{\rho}_c)$, we make a histogram of the $\tilde{\rho}_c$ values of false alarms arising from our selected BBH region. We use only the false alarms which are uncorrelated with possible candidate events to ensure an unbiased estimate of the mean false alarm rate [49]. We fit an exponential decay to this histogram from $8 < \tilde{\rho}_c < 9.2$. For $\tilde{\rho}_c$ much less than 8, $\Lambda_N P_N$ is not well modeled by an exponential due to the effects of applying a threshold to single-detector triggers. We note, however, there is only a 50% chance that an event is astrophysical at $\tilde{\rho}_c \sim 8.6$, and this chance quickly becomes negligible with decreasing $\tilde{\rho}_c$. The result of this procedure is shown in Fig. 6. We caution that P_{astro} for candidates with $\tilde{\rho}_c > 9.2$ will be sensitive to the form of the model chosen since it is not constrained by empirically measured false alarms.

While we do not assess the astrophysical probabilities of sources outside our selected BBH region, we are not precluding that such sources exist. Our P_{astro} is compatible with any model of the true BBH source distribution that allows for a signal rate to be at least as high as our estimate within the chosen region. This holds irrespective of whatever other kinds of sources may also be permitted.

Table 2: Candidate events from the full search for compact binary mergers in O1 data. Candidates are sorted by FAR evaluated for the entire bank of templates. The FAR of the top two candidates is limited only by the amount of background time estimated, and only differ due to the variation in time available in their respective analyses to create background. The parameters of the template associated with each candidate are listed. Note that these are not intended as a rigorous estimation of the source parameters. Masses are given in the detector frame.

Designation	Julian Date	$FAR^{-1}(yr)$	$\tilde{\rho}_c$	ρ_H	ρ_L	m_1	m_2	$\chi_{\rm eff}$
150914+09:50:45UTC	2457279.910665	>66000	18.45	19.67	13.38	44.21	32.16	0.09
151226+03:38:53UTC	2457382.652426	>59000	11.62	10.73	7.43	14.83	8.50	0.24
151012+09:54:43UTC	2457307.913420	24	9.06	6.96	6.71	30.75	12.89	-0.0
151019+00:23:16UTC	2457314.516585	0.060	8.39	6.81	5.47	14.93	1.27	0.11
150928+10:49:00UTC	2457293.951122	0.042	8.37	6.05	6.34	2.53	1.02	-0.7
151218+18:30:58UTC	2457375.271929	0.029	8.24	7.11	5.38	31.29	2.35	-0.0
160103+05:48:36UTC	2457390.742504	0.026	8.22	6.01	6.60	9.75	7.29	0.49
151202+01:18:13UTC	2457358.554740	0.025	8.23	6.54	5.73	40.42	1.77	-0.2
160104+03:51:51UTC	2457391.661424	0.021	8.19	5.80	6.39	6.76	1.10	-0.5
151213+00:12:20UTC	2457369.508985	0.019	8.22	5.70	7.24	11.12	3.30	-0.7
150923+07:10:59UTC	2457288.799711	0.014	8.20	6.78	5.84	2.14	1.08	0.65
151029+13:34:39UTC	2457325.066149	0.014	8.21	6.83	5.23	2.19	1.07	-0.2
151206+14:19:29UTC	2457363.097291	0.013	8.17	5.80	6.37	100.60	1.64	0.98
151202+15:32:09UTC	2457359.147751	0.012	8.14	5.93	6.41	6.33	1.18	-0.5
151012+06:30:45UTC	2457307.771774	0.011	8.19	6.74	5.70	3.16	1.73	-0.1
151116+22:41:48UTC	2457343.446120	0.010	8.14	5.79	6.64	2.00	1.04	-0.4
151121+03:34:09UTC	2457347.649138	0.010	8.12	6.48	5.78	7.43	1.00	-0.8
150922+05:41:08UTC	2457287.737317	0.010	8.16	6.05	6.34	2.78	1.02	0.17
151008+14:09:17UTC	2457304.090202	0.008	8.16	5.84	6.10	46.38	1.19	0.38
151127+02:00:30UTC	2457353.584101	0.008	8.10	6.28	5.44	39.12	2.01	0.99

4.4 Results

The results presented here are generated using the data from the first observing run of Advanced LIGO which ran from September 12, 2015 to January 19, 2016. We divide the 16 kHz LIGO open data into 9 consecutive periods of time and search each time period independently so that each analysis contains roughly five days of observing time. This time interval is set by the disk and memory requirements of the search pipeline, but it is sufficient to estimate the FAR of candidate events to better than 1 in 50,000 years. It is possible to combine these time intervals during the analysis to improve this limit, but we have not done so here. Our analysis is restricted to times marked as observable by the metadata provided by the Gravitational-Wave Open Science Center. After accounting for times which are marked as not analyzable, there remain ~ 48.1 days of data when both the Hanford and Livingston LIGO instruments were operating.

The top candidate events by FAR from the full search are given in Table 2. There are three candidates which are statistically significant. These are the binary black hole mergers GW150914, LVT151012, and GW151226, which were previously reported in [1, 6, 7]. The false alarm rates for GW150914 and GW151226 of 1 per 66,000 and 1 per 59,000 years, respectively, are limits based on the amount of background time available in their respective analysis. These limits are less stringent than those reported in [1] as we have created less background time. There are no other individually convincing candidates. Fig. 7 shows candidate events with $\tilde{\rho}_c > 7.5$. The three binary black hole mergers stand out from the other candidate events and are clustered in a portion of the parameter space that is analyzed with relatively few template waveforms.

4.4.1 Binary Black Hole Candidates

Given that there are two binary black hole mergers (GW150914 and GW151226) that are well established from their statistical significance, we can estimate the rate of detecting binary black hole mergers by this analysis. Candidate events that are consistent with our selected binary black hole population are listed in Table 3. We estimate the false alarm rate of events for just this region of the analysis, and using our estimate of the true rate of detections, calculate the true discovery rate as a

Table 3: Candidate events consistent with the selected population of binary black holes. There are three binary black hole mergers above a threshold corresponding to a true discovery rate of 99.92%. The third most significant event, LVT151012, has a 97.6% probability of being astrophysical in origin. Note that the FARs indicated do not reflect the false alarm rate for the full search, but instead for the limited region of the template bank indicated in red in Fig. 5. The FARs listed for the top two events are limited by the background time generated and so are identical to those in Table 2.

Designation	Julian Date	$P_{\rm astro}$	TDR	$FAR^{-1}(yr)$	$\tilde{\rho}_c$	ρ_H	ρ_L	m_1
150914+09:50:45UTC	2457279.910665	-	-	>66000	18.45	19.67	13.38	44.21
151226+03:38:53UTC	2457382.652426	-	-	>59000	11.62	10.73	7.43	14.83
151012+09:54:43UTC	2457307.913420	0.976	0.999	446	9.06	6.96	6.71	30.75
160103+05:48:36UTC	2457390.742504	0.061	0.517	0.396	8.22	6.01	6.60	9.75
151213+00:12:20UTC	2457369.508985	0.047	0.455	0.309	8.22	5.70	7.24	11.12
151216+18:49:30UTC	2457373.284799	0.017	0.223	0.106	8.09	6.10	6.01	13.92
151222+05:28:26UTC	2457378.728506	0.012	0.169	0.075	8.03	5.67	6.46	6.86
151217+03:47:49UTC	2457373.658627	0.006	0.088	0.036	7.96	6.69	5.57	40.02
151009+05:06:12UTC	2457304.713060	0.005	0.087	0.035	7.99	5.66	5.90	25.55
151220+07:45:36UTC	2457376.823761	0.003	0.053	0.021	7.87	6.55	5.39	17.50
151104+04:12:55UTC	2457330.676062	0.003	0.053	0.021	7.91	5.94	6.33	19.25
151120+16:20:06UTC	2457347.181049	0.003	0.047	0.018	7.86	6.11	5.44	5.49
151216+09:24:16UTC	2457372.892271	0.003	0.045	0.017	7.86	5.76	5.66	58.56
151128+14:37:02UTC	2457355.109478	0.003	0.040	0.016	7.83	6.79	5.02	9.25
160109+08:08:42UTC	2457396.839798	0.003	0.035	0.014	7.82	5.24	6.23	24.29
160111+22:49:34UTC	2457399.451507	0.003	0.035	0.013	7.82	5.10	6.55	5.75
151124+11:25:19UTC	2457350.976339	0.002	0.033	0.013	7.81	5.65	6.27	98.89
150912+15:39:02UTC	2457278.152523	0.002	0.032	0.012	7.84	6.23	5.23	9.86
151006+06:06:50UTC	2457301.755168	0.002	0.031	0.012	7.89	6.77	5.47	11.59
151015+01:40:52UTC	2457310.570466	0.002	0.029	0.011	7.85	5.37	5.92	87.87

function of ranking statistic. The TDR at the ranking statistic of the fourth most

significant candidate is 0.52. This means that only 52% of candidates with $\tilde{\rho}_c$ at least as large are expected to be of astrophysical origin. For each candidate we estimate its individual probability of being astrophysical in origin, $P_{\rm astro}$. The fourth event has only a 6% chance of being astrophysical. We do not report $P_{\rm astro}$ and TDR values for the top two events since these events are assumed to be signals in the construction of these statistics.

4.4.2 Revisiting LVT151012

LVT151012 was first announced in [12], with a FAR of 1 per 2.3 years. Our improved methods yield a false alarm rate for LVT151012 of 1 per 24 years. Restricting attention to our selected BBH region, which is consistent with the other observed binary black hole mergers, gives a FAR for LVT151012 in this region alone of 1 per 446 years. We combine this FAR with our conservative estimate of the rate of detections to estimate that 99.92% of binary black hole merger candidates at least as significant as LVT151012 are astrophysical in origin. We also estimate the probability that specifically LVT151012 is astrophysical in origin to be 97.59%.

These measures both depend on our selected region of binary black hole sources and our estimate of the rate of true detections, but we believe our choices for both of these to be conservative. The FAR of 1 per 446 years is not a statistical statement about the search as a whole and is used only in comparison against the rate of real signals within this same region. Selecting different boundaries for this region would yield a different FAR. However, assuming that the false alarm rate and true alarm rate are both approximately uniform in this region, then $P_{\rm astro}$ and TDR will not change.

As data from future observing runs becomes available, it will be possible to more precisely estimate this rate in a consistent way, and improve our estimate of this event's significance. We have modeled our signal distribution and population of false alarms as being characterized by the ranking statistic $\tilde{\rho}_c$ alone. An improved model could take into account the variation over the parameter space and in time. Fig. 6 shows the probability distribution of our noise and signal models for the analysis which contains LVT151012. Compared to the $P_{\rm astro}$ reported in [1] of 87%, our analysis has improved the ranking of candidate events, the boundaries of our selected BBH distribution differ from what was used there, and we use a more conservative estimate

of the signal rate. Given a $P_{\rm astro}$ value of 97.6% we conclude that LVT151012 is astrophysical in origin. For comparison, if we had chosen the rate of observed mergers to be $\approx 15{\rm yr}^{-1}$, which is the linear extrapolation of two detections in 48 days, we would find that LVT151012 had a 99.6% probability of astrophysical origin.

We present a full catalog of gravitational-wave events and candidates from a PyCBC-based, templated, matched-filter search of the LIGO O1 open data. Our analysis represents an improvement over that of [1, 11] by using improved ranking of candidates by considering phase, amplitude and time delay consistency, an improved background model and a template bank targeting a wider range of sources [13, 14, 15]. We independently verify the discovery of GW150914 and GW151226 and report an improved significance of the candidate event LVT151012, which we claim should be viewed as a confident detection. Apart from these three signals, none of the other candidate events are individually significant in our analysis. All of these candidates are listed in our catalog available at www.github.com/gwastro/1-ogc, along with tools for exploring and using it. Complete gravitational-wave event catalogs of this nature will become important tools in multi-messenger astronomy.

A larger data set from the second observing run of LIGO and Virgo already exists. Individual detections have been published, and short periods of data around the detections are available publicly. However, the bulk of this data has not yet been released publicly. It will be possible to create a similar open catalog with the most up-to-date analysis tools when these data are released.

We thank Thomas Dent and Sumit Kumar for useful discussions and comments. We thank Stuart Anderson, Jonah Kannah, and Alan Weinstein for help accessing data from the Gravitational-Wave Open Science Center. We acknowledge the Max Planck Gesellschaft for support and the Atlas cluster computing team at AEI Hannover. Computations were also supported by Syracuse University and NSF award OAC-1541396. DAB acknowledges NSF awards PHY-1707954, OAC-1443047, and OAC-1738962 for support. SR acknowledges NSF award PHY-1707954 and OAC-1443047 for support. RW acknowledges NSF award OAC-1823378 for support. 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. LIGO is funded by the U.S. National Science Foundation. Virgo is funded by

the French Centre National de Recherche Scientifique (CNRS), the Italian Istituto Nazionale della Fisica Nucleare (INFN) and the Dutch Nikhef, with contributions by Polish and Hungarian institutes.

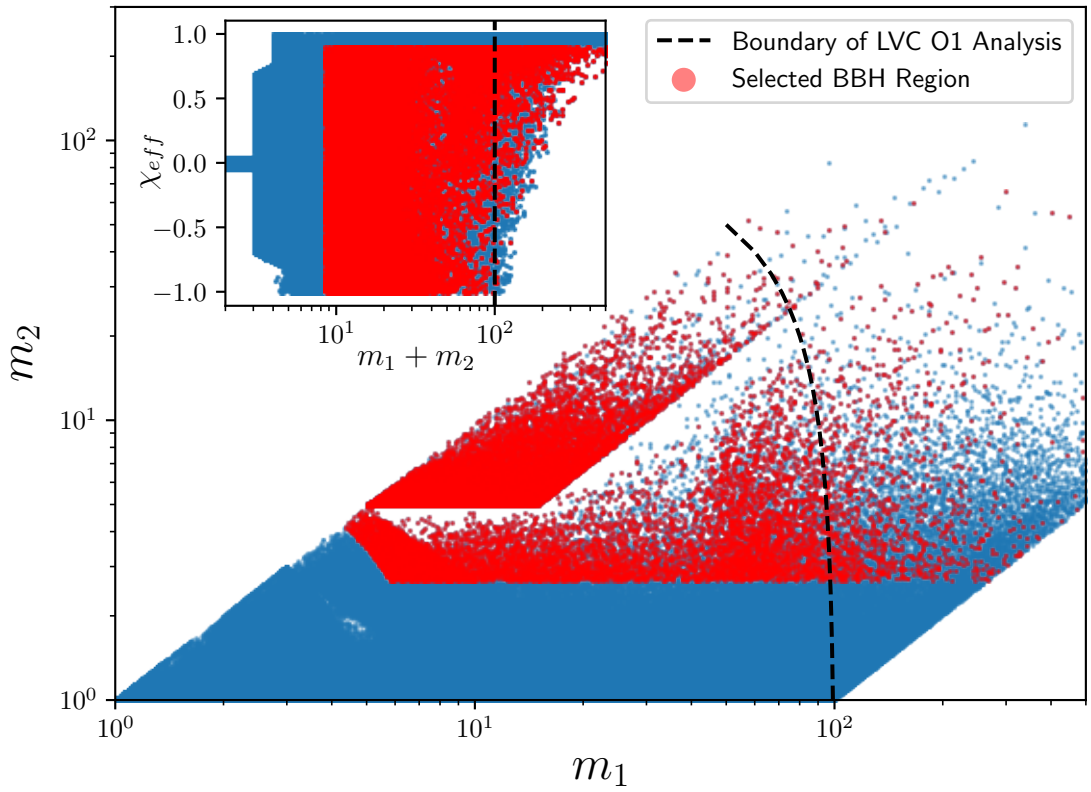

Figure 5: The component masses and spins of the templates used to search for compact binary mergers. Due to the exclusion of short duration templates, there is a dependency on the total mass searched and its effective spin. For binary black holes with negligible spin, this implies that this study only probes sources with total mass less than $200\,\mathrm{M}_\odot$. Visible artifacts due to the procedure for constructing the template bank do not impact performance. Templates which we conservatively consider to produce binary black hole (BBH) candidates consistent with known observations are shown in red as discussed in Sec. 4.3. The upper mass boundary of the analysis performed by the LVC in [1] is shown as a black dotted line.

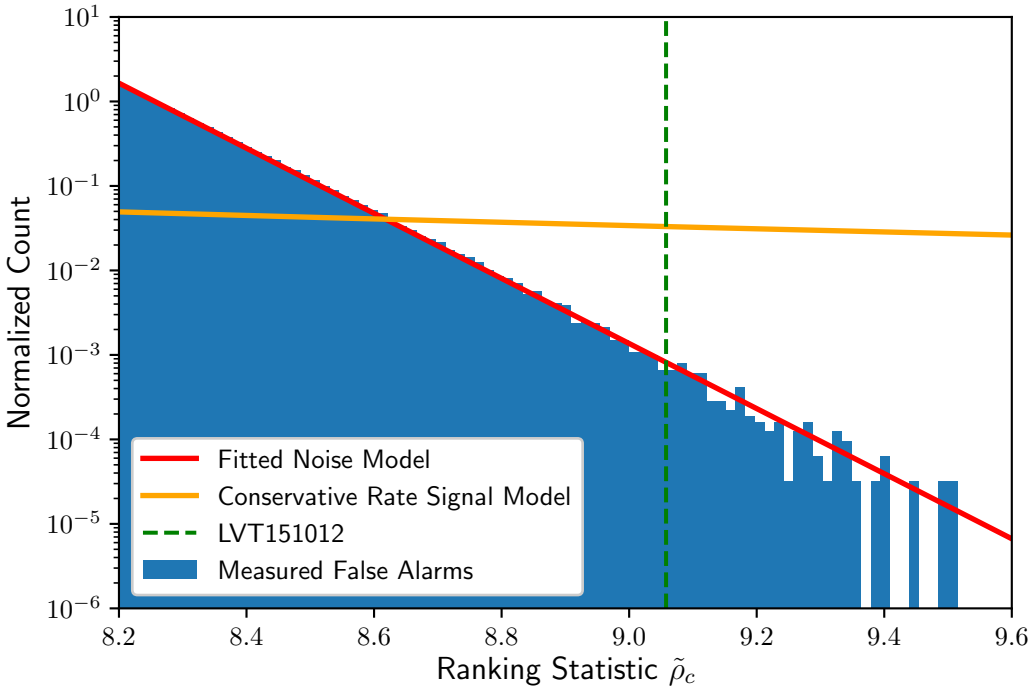

Figure 6: The scaled probability distributions of assumed signals and noise as a function of the ranking statistic $\tilde{\rho}_c$ for the analysis containing LVT151012. Blue shows the normalized histogram of empirically measured false alarms that are within our selected BBH region of the template bank, P_N. Red is the exponential decay model that has been fitted to this set of false alarms, $P_S\Lambda_S/\Lambda_N$, normalized so that the counts can be directly compared to the noise distribution. Orange shows the signal model based on our conservative rate of detections. The value of $\tilde{\rho}_c$ for LVT151012 is shown as a dotted green vertical line. The ratio of signal to noise at this value of $\tilde{\rho}_c$ strongly favors the signal model.

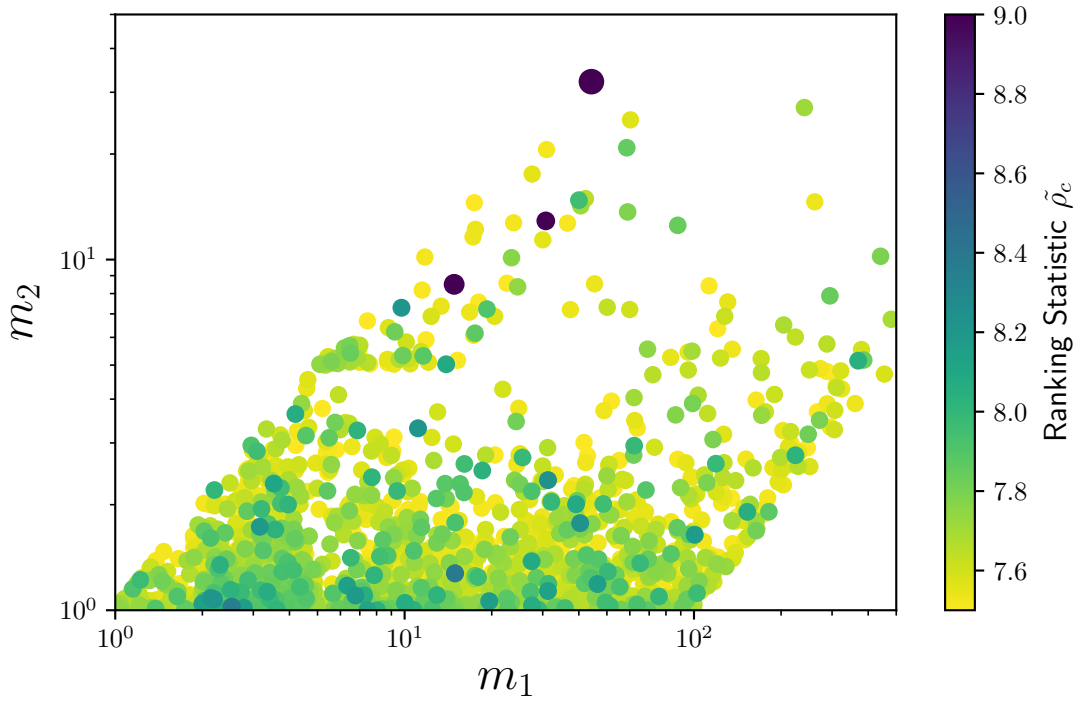

Figure 7: Candidate events with a ranking statistic $\tilde{\rho}_c > 7.5$ from the full search for compact binary mergers in O1 data. The colorbar is capped at 9. The three BBH mergers are clearly visible in the plots, while the remaining events are largely distributed according to the density of the template bank.

Chapter 5

Pressure-Gravity Mode Instability in GW170817

5.1 Introduction

The discovery of the binary neutron star merger GW170817 [29] has given us a new way to explore the physics of neutron stars. Recent studies have measured the star's tidal deformability and placed constraints on the equation of state of the neutron stars [29, 64, 65, 66, 3, 67, 68**? ? ? ?**]. [69] have suggested that the star's tidal deformation can induce nonresonant and nonlinear daughter wave excitations in p- and g-modes of the neutron stars via a quasi-static instability. This instability would remove energy from a binary system and possibly affect the phase evolution of the gravitational waves radiated during the inspiral. Although [70] concluded that there is no quasi-static instability and hence no effect on the inspiral, [71] claims that the instability can rapidly drive modes to significant energies well before the binary merges. However, the details of the instability saturation are unknown and so the size of the effect of the p-g mode coupling on the gravitational-waveform is not known [71]. The discovery of the binary neutron star merger GW170817 by Advanced LIGO and Virgo provides an opportunity to determine if there is evidence for nonlinear tides from p-g mode coupling during the binary inspiral.

Since the physics of the p-g mode instability is uncertain, [72] developed a parameterized model of the energy loss due to nonlinear tides. This model is parameterized

by the amplitude and frequency dependence of the energy loss, and the gravitational-wave frequency at which the instability saturates and the energy loss turns on. For plausible assumptions about the saturation, [72] concluded that $> 70\%$ of binary merger signals could be missed if only point-particle waveforms are used, and that neglecting nonlinear tidal dynamics may significantly bias the measured parameters of the binary. Bayesian inference can be used to place constraints on nonlinear tides during the inspiral of GW170817. An analysis by [2] computed Bayes factors that investigate whether the GW170817 signal is more likely to have been generated by a model which includes nonlinear tides or one which does not. [2] find a Bayes factor of order unity, and conclude that the GW1701817 signal is consistent with both a model that neglects nonlinear tides and with a model that includes energy loss from a broad range of p-g mode parameters. However, the prior space used in this analysis includes a large region of parameter space where the amplitude of the effect produces a gravitational-wave phase shift that is extremely small. In this case, a waveform that includes p-g mode parameters will have a likelihood that is identical to the likelihood of the waveform without the p-g mode instability. The p-g mode model extends the standard waveform model by adding additional parameters that describe the nonlinear tidal effects. However, when including new parameters in a hypothesis if the likelihood does not vary across large portions of the prior volume for these new parameters relative to the likelihood of the original model, then the Bayes factor will not penalize this additional prior volume, nor will it penalize any extraneous parameters in the model (see e.g. [73, 74]). We examine prior space of p-g model used by [2] and find that although the p-g model model contains regions that are not consistent with the standard model, there are large regions of the prior space where the likelihood is high because the p-g mode model is degenerate with the standard model. These regions of prior space dominate the evidence and hence the Bayes factor neither favors nor disfavors the inclusion of p-g mode parameters.

We investigate a variety of different prior distributions on the p-g mode parameters beginning with a prior distribution that is similar to that tested in [2] and includes large regions of the parameter space that produce a negligible gravitational-wave phase shift. When comparing the evidence for this model with the standard waveform model used by [3] we find a Bayes factor of order unity, as expected. We then investigate a prior distribution in which the p-g mode instability parameters are

constrained to induce a phase shift to the waveform that is greater than 0.1 radians. This phase shift is calculated from the time the waveform enters the sensitive band of the detector to the time when the waveform reaches the innermost stable circular orbit. We choose this threshold to exclude trivial regions of the parameter space that produce a non-measurable effect. However, we again find a Bayes factor of order unity when compared to the model hypothesis that does not model the p-g mode instability. Investigation of these results showed that this is due to parameter degeneracies between the p-g mode model and the intrinsic parameters of the standard waveform model.

Finally, we reduce the prior space to contain only the regions where the p-g mode waveform is not degenerate with the standard model by computing the fitting factor [75] of p-g signals against a set of standard waveforms. We do this to restrict the region of parameter space to that where the p-g effect is *measurably* distinct from a model that neglects nonlinear tides. We calculate the Bayes factor as a function of the fitting factor. We find that as the p-g mode parameter space is restricted to exclude regions that have a high fitting factor with standard waveforms, the Bayes factor decreases significantly. Regions of the nonlinear tide parameter space that have a fitting factor of less than 99% (98.5%) are strongly disfavored by a Bayes factor of 15 (25). While certain prior distributions of p-g mode parameters are consistent with the data, we find that these distributions are ones that contain large regions of non-measurable parameter space either because the effect produced is too small to measure, or the effect is degenerate with other parameters of the standard model. We conclude that the consistency of the GW170817 signal with the model of [72] is due to degeneracies and that regions where non-linear tides produce a measurable effect are strongly disfavored.

5.2 Waveform model

As two neutron stars orbit each other, they lose orbital energy $E_{\rm orbital}$ due to gravitational radiation $\dot{E}_{GW}$. The gravitational waveform during the inspiral is well modeled by post-Newtonian theory (see e.g. [76]). The effect of the p-g mode instability is to dissipate orbital energy by removing energy from the tidal bulge of the stars [69, 71, 72]. Once unstable, the coupled p- and g-modes are continuously driven

by the tides, giving rise to an extra energy dissipation $\dot{E}_{NL}$ for each star in the standard energy-balance equation [77]

$$\dot{E}_{\rm orbital} = -\dot{E}_{\rm GW} - \dot{E}^1_{\rm NL} - \dot{E}^2_{\rm NL}. \tag{5.1}$$

Since the details of how the nonlinear tides extract energy from the orbit is not known, [72] constructed a simple model of the energy loss and calculated plausible values for the model's parameters. In this model, the rate of orbital energy lost during the inspiral is modified by

$$\dot{E}_{\rm NL} \propto A f^{n+2} \Theta(f - f_0), \tag{5.2}$$

where A is a dimensionless constant that determines the overall amplitude of the energy loss, n determines the frequency dependence of the energy loss, and f_0 is the frequency at which the p-g mode instability saturation occurs and the effect turns on. By solving Eq. (5.1), [72] computed the leading order effect of the nonlinear tides on the gravitational-wave phase as a function of A, n, and f_0. In this analysis, they allowed each star to have independent values of A, f_0, and n, but found that the energy loss due to nonlinear tides depends relatively weakly on the binary's mass ratio. Hence, they consider a model that performs a Taylor expansion in the binary's component mass [78] and include only the leading order terms in the binary's phase evolution. Given this, we parameterize our nonlinear tide waveform with a single set of parameters A, n, and f_0, by setting $\dot{E}^1_{\rm NL} = \dot{E}^2_{\rm NL}$. We keep only the leading order nonlinear tide terms when we obtain the quantities $t(f)$ and $\phi(f)$ used to compute the stationary phase approximation [39, 40, 79]. This approach is reasonable for GW170817, since both neutron stars have similar masses and radii [3].

The dependence of A, n, and f_0 on the star's physical parameters is not known [71]. [72] estimate that plausible parameter ranges are $A \lesssim 10^{-6}$, $0 \lesssim n \lesssim 2$, and $30 \lesssim f_0 \lesssim 80$ Hz. [80] found that the frequency at which the instability begins to grow is equation-of-state dependent and can occur at gravitational-wave frequencies as high as 700 Hz. [81] suggest that the instability may only act during the late stages of inspiral, (above 300 Hz), otherwise the large energy dissipation will cause the temperature of the neutron stars to be very large.

In this paper, we compare two models for the gravitational waves radiated by GW170817. The first is the standard restricted stationary-phase approximation to the Fourier transform of the gravitational waveform $\tilde{h}(f)$, known as the TaylorF2

waveform [39]. We begin with the same waveform model used by [3], which is accurate to 3.5 PN order in the orbital phase, 2.0 PN order in spin-spin, self-spin and quadrupole-monopole interactions, 3.5 PN order in spin-orbit coupling, and includes the leading and next-to-leading order corrections from the star's tidal deformability [82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92]. We then construct a second model that adds the leading order effect of nonlinear tides computed using the model of [72]. We compute the Fourier phase for the TaylorF2 model $\Psi(f)_{\rm TaylorF2}$ and add a term that accounts for the additional energy lost due to nonlinear tides $\Psi_{\rm NL}(f)$, given by

$$\Psi_{\rm NL}(f) = -\frac{25}{768} A \left(\frac{G\mathcal{M}\pi f_{\rm ref}}{c^3}\right)^{-\frac{10}{3}} \times \begin{cases} \left(\frac{f_0}{f_{\rm ref}}\right)^{n-3} \left[\left(\frac{f}{f_0}\frac{1}{n-4}\right) - \frac{1}{n-3}\right] & f < f_0, \\ \left(\frac{f}{f_{\rm ref}}\right)^{n-3} \left(\frac{1}{n-4} - \frac{1}{n-3}\right) & f \geq f_0. \end{cases} \tag{5.3}$$

Here, $f_{\rm ref}$ is a reference frequency which we set to 100 Hz following [72], G is Newton's gravitational constant, c is the speed of light, and $\mathcal{M} = (m_1 m_2)^{3/5}/(m_1+m_2)^{1/5}$ is the chirp mass of the binary.[1] This waveform model can have a degeneracy in the gravitational wave phasing with chirp mass when $n = 4/3$. For this value of n, the Fourier phase in Eq. (5.3) for nonlinear tides is $\Psi(f) \propto f^{-5/3}$, which is the same power law dependence as the chirp mass phasing. A degeneracy occurs when f_0 is comparable or lower than the frequency at which chirp mass can be accurately measured. In this case, the p-g mode instability is degenerate with changing the chirp mass. In principle, there will be other degeneracies with other intrinsic parameters of the gravitational wave signal for other values of n.

We generate the standard TaylorF2 waveform using the LIGO Algorithm Library [93] and multiply this frequency-domain waveform by the term due to the nonlinear tides,

$$\tilde{h}_{\rm TaylorF2+NL}(f) = \tilde{h}_{\rm TaylorF2}(f) \times \exp[-i\Psi_{\rm NL}(f)]. \tag{5.4}$$

The Fourier phase for the nonlinear tides is implemented as a patch to the version of the PyCBC software [94] used by [3]. Both the standard and nonlinear tide waveform

[1] Appendix A of [72] gives the change to the gravitational-wave phase $\phi(f)$ as a function of frequency and not the change to the Fourier phase $\Psi(f)$ (see e.g. [79] for a discussion of how these differ). The former quantity is useful to compute the change in the number of gravitational-wave cycles, but the latter is required to compute the modification to the TaylorF2 waveform. The study by [2] corrects this mistake.

models are terminated when the gravitational-wave frequency reaches that of a test particle at the innermost stable circular orbit of a Schwarszchild black hole of mass $M = m_1 + m_2$. For the neutron star masses considered here, this frequency is between 1.4 kHz and 1.6 kHz.

5.3 Model Priors

Bayes theorem offers a methodology for evaluating the plausibility of models relative to a given data set, and then updating these prior model beliefs with better hypotheses. Bayes theorem states that

$$\mathcal{P}\left(\vec{\theta}|H,\mathbf{d}\right) = \frac{\pi\left(\vec{\theta}|H\right)\,\mathcal{L}\left(\mathbf{d}|H,\vec{\theta}\right)}{\mathcal{Z}\left(\mathbf{d}|H\right)}, \tag{5.5}$$

where $\mathcal{Z}\left(\mathbf{d}|H\right)$ is the evidence of the model H, $\pi\left(\vec{\theta}|H\right)$ is the is the prior distribution of the parameters given the signal model, $\mathcal{L}\left(\mathbf{d}|H,\vec{\theta}\right)$ is the likelihood of the data for a particular set of parameters $\vec{\theta}$, and $\mathcal{P}\left(\vec{\theta}|H,\mathbf{d}\right)$ is the posterior distribution of the parameters given the signal model. The likelihood used in this analysis assumes a Gaussian model of detector noise and depends upon the noise-weighted inner product between the gravitational waveform and the data from the gravitational-wave detectors [95, 96]. The choice of prior distributions on the parameters of the signal model represent the hypothesis that we want to test. The posterior distributions reflect how to update ones beliefs with respect to the likelihood and the data. Thus, by examining many different parameter hypotheses we can investigate the extent to which GW170817 is accurately modeled by p-g mode instability waveform models.

In our analysis, we fix the sky location and distance to GW170817 [97, 98] and assume that both neutron stars have the same equation of state by imposing the common radius constraint [3]. In the case of the standard TaylorF2 waveform $H_{\rm TaylorF2}$, our analysis is identical to that described in [3]. This analysis considered three prior distributions on the binary's component mass. Here, we only consider the uniform prior on each star's mass, with $m_{1,2} \sim U[1,2]\,M_\odot$, and the Gaussian prior on the component masses $m_{1,2} \sim N(\mu = 1.33, \sigma = 0.09)\,M_\odot$ [99]. For both mass priors, we restrict the chirp mass to the range $1.1876 M_\odot < \mathcal{M} < 1.2076 M_\odot$. Since our analysis

is identical to that of [3], we refer to that paper for the details of the data analysis configuration.

Given the uncertainty on the range of the nonlinear tide parameters, we follow [2] and let $n \in U[-1.1, 2.999]$, draw A from a distribution uniform in $\log_{10}$ between 10^{-10} and $10^{-5.5}$, and $f_0 \in U[10, 100]$ Hz. We use this along with a uniform prior distribution on the mass from [3].

We also consider two alternative choices of drawing f_0: we draw f_0 from a uniform distribution between 15 and 100 Hz, as used by [72], and from a uniform distribution between 15 and 800 Hz to allow for the larger values of f_0 suggested by [80] and [81]. For these choices we consider A uniform in $\log_{10}$ between 10^{-10} to 10^{-6}. For these alternative prior distributions we also consider applying a further constraint on the parameters. Since some combinations of A, n, and f_0 can produce extremely small gravitational-wave phase shifts [72], we place a cut on the gravitational-wave phase shift due to nonlinear tides

$$\delta\phi(f_{\rm ISCO}) = \frac{-25}{768}\frac{A}{n-3}\left(\frac{G\mathcal{M}\pi f_{\rm ref}}{c^3}\right)^{-10/3}\left[\left(\frac{f_0}{f_{\rm ref}}\right)^{n-3} - \left(\frac{f_{\rm ISCO}}{f_{\rm ref}}\right)^{n-3}\right], \qquad (5.6)$$

where $f_{\rm ISCO}$ is the termination frequency of the waveform (which is always larger than f_0 in our analysis). This gravitational-wave phase shift from the p-g mode instability is strictly negative, but we take the convention of using the absolute value of the phase shift for convenience. We restrict the prior space to values of $\delta\phi > 0.1$ rad. Phase shifts of $\delta\phi \approx 0.1$ rad have an overlap between the two waveform models greater than 99.98%. This cut means that the resulting priors on A, n, and f_0 are not uniform, but are biased in favor of combinations of parameters that may produce a measurable effect on the phasing of the waveform due to nonlinear tides. While $\delta\phi$ is a simple proxy for how similar or dissimilar two waveforms are, formally this is given by the match between two waveforms. A $\delta\phi$ of 1 radian may have a low overlap with a waveform if the radian is accumulated over a large bandwidth but a high overlap if the radian is accumulated near the very end of the signal. Fig. 8 shows a depiction of the prior distributions used when using a permissive prior on $\delta\phi$, similar to [2], and when using a constraint on the p-g mode parameters such that $\delta\phi > 0.1$ rad.

A stricter approach to constructing a prior distribution that considers p-g mode effects that are distinguishable from standard waveforms is to examine the fitting factor between a distribution of p-g mode waveforms and a set of comparable TaylorF2

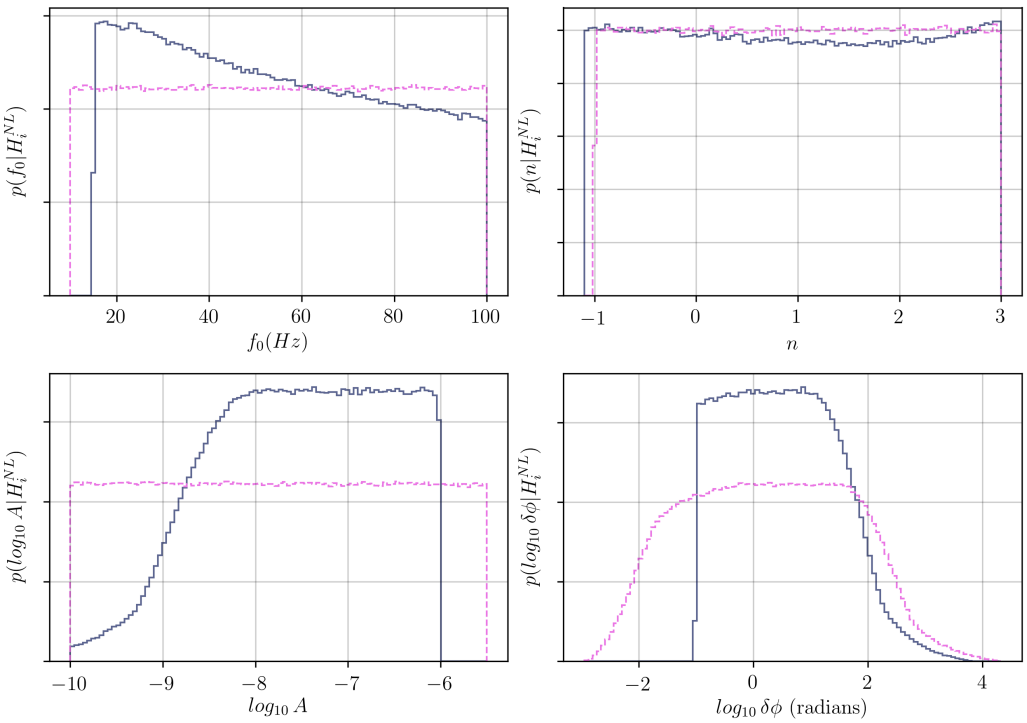

Figure 8: Prior probability distributions on the parameters (f_0, n, A) for the waveform model $\mathrm{H^{NL}} = \mathrm{H_{TaylorF2+NL}}$ and the resulting prior on the gravitational-wave phase shift $\delta\phi$ shift due to nonlinear tides. The dark blue, solid lines shows the priors when f_0 is drawn from a uniform distribution between 15 and 100 Hz with a $\delta\phi \geq 0.1$ rad constraint restricting some of the prior space. The pink, dotted lines represent prior distributions on the nonlinear tidal parameters similar to [2].

waveforms. To do so, we examine the fitting factor of our Bayesian inference analysis with respect to a template bank of non-spinning, mass-only TaylorF2 waveforms. We construct a template bank of $\sim 20,000$ non-spinning, mass-only waveforms of comparable masses to the prior distribution on the mass parameters. The template bank is constructed with component masses, $m_{(1,2)} \in (1.0, 2.0)M_\odot$, chirp masses, $\mathcal{M}_c \in (1.1826, 1.2126)M_\odot$, and a minimal match placement of 99.9%. We then place a threshold on the evidence calculation from the Bayesian analysis based on the maximum overlap with this template bank of standard waveforms. This permits an analysis of the Bayes factor for nonlinear tides where the prior distribution on *p-g* mode parameters is determined by the fitting factor with a set of standard signals.

5.4 Methods

We use the gravitational-wave strain data from the Advanced LIGO and Virgo detectors for the GW170817 event, made available through the GW Open Science Center [18, 100]. We then repeat the analysis of [3] using the waveform model $H_{\rm TaylorF2+NL}$ to compute the evidence $p(\mathbf{d}\,|\,\mathrm{H}_{\rm TaylorF2+NL})$.

We use Bayesian model selection to determine which of the two waveform models described in Sec. 5.2 is better supported by the observation of GW170817. Bayes theorem in Eq. (5.5) permits us a method for model through the ratio of the evidence from each model. This ratio of the model evidences is called the Bayes factor, which we denote as $\mathcal{B}$. A Bayes factor greater than unity indicates support for the model in the numerator, while a Bayes factor less than unity indicates support for the model in the denominator. The Bayes factor can be written as,

$$\mathcal{B} = \frac{\mathcal{Z}\left(\mathbf{d}|H_{\rm TaylorF2+NL}\right)}{\mathcal{Z}\left(\mathbf{d}|H_{\rm TaylorF2}\right)}. \tag{5.7}$$

The numerator of Eq. (5.7) is the evidence for nonlinear tides $p\left(\mathbf{d}|H_{\rm TaylorF2+NL}\right)$. For the denominator of Eq. (5.7), we use the evidence $\mathcal{Z}(\mathbf{d}|H_{\rm TaylorF2})$ provided as supplemental materials by [3].

Posterior distributions for parameters of interest can be also computed by marginalizing the posterior probability distribution over other parameters. Marginalization to obtain the posterior probabilities and the evidence is performed using Markov Chain Monte Carlo (MCMC) techniques. To compute posterior probability distributions and evidences, we use the *PyCBC Inference* software [94, 101] using the parallel-tempered *emcee* sampler [102, 103]. This sampler allows the use of multiple temperatures to sample the parameter space [102, 104, 105]. These multiple temperatures β permit the construction of tempered posterior distributions that form a slow thermodynamic transition from the prior distribution to the posterior distribution in Eq. 5.5. Tempered posteriors are called power-posteriors in [106, 107]. The power-posterior can be according to:

$$\mathcal{P}\left(\vec{\theta}|\mathbf{d}, H\right)_{\beta} \propto \pi\left(\vec{\theta}|H\right)\mathcal{L}\left(\mathbf{d}|\vec{\theta}, H\right)^{\beta}. \tag{5.8}$$

The normalization constant for a power-posterior is the evidence for that power-posterior, given as $\mathcal{Z}(\mathbf{d}|H)_{\beta} = \int \pi\left(\vec{\theta}|H\right)\mathcal{L}\left(\mathbf{d}|\vec{\theta}, H\right)^{\beta} d\vec{\theta}$.

From these power-posterior distributions we use the thermodynamic integration method [106, 107] to estimate the logarithm of the evidence, $\ln \mathcal{Z}$, given as:

$$\ln \mathcal{Z} = \int_0^1 \langle \ln \mathcal{L} \rangle_\beta \, d\beta. \tag{5.9}$$

We provide a more thorough derivation in Appendix A.2. The estimate of the evidence is determined by the integral over inverse temperatures, β, of the average untempered log likelihood, $\langle \ln \mathcal{L} \rangle_\beta$, drawn from the power-posterior corresponding to the inverse temperature β. An approximation to this integral can be made through use of trapezoid rule integration method. Following [3] we use 51 temperatures where we use a combination of geometric and logarithmic temperature placements to improve the accuracy of the integral [108].

We verify the results of the thermodynamic integration evidence calculation by comparing it with the steppingstone algorithm [109], which utilizes the same likelihoods from multi-tempering sampling as the thermodynamic integration method. Both trapezoidal rule thermodynamic integration and steppingstone methods can have some bias in the estimate of the logarithm of the Bayesian evidence due to a finite number of temperatures being used. This bias is mitigated by an increased number of temperatures [109, 110]. Additionally, this bias can be mitigated in thermodynamic integration by improving the order of the quadrature integration [111]. We also use a higher order trapezoidal rule from [111] and verify that the results are consistent.

We also estimate the error for each method of evidence calculation. The thermodynamic integration method and steppingstone algorithm both contain Monte Carlo error [112]. For the thermodynamic integration method the Monte Carlo error on the thermodynamic integral can be estimated following the methodology of [112]. We use this same uncertainty estimate for the higher order trapezoidal rule as well. In [109] there is a Monte Carlo variance estimate for the logarithm of the evidence from the steppingstone method that we also use here.

The last source of error in the evidence calculation that we consider is whether the MCMC has converged to stable likelihood values across all of the temperatures. This requires examining the stability of the evidence calculations as the MCMC progresses. Independent samples are drawn according to the $\mathrm{n_{acl}}$ method as described by [101] at various points in the run. This method takes a specific endpoint iteration,

takes half the endpoint iteration as the starting point iteration, and calculates the autocorrelation length of the samples between the starting point and the endpoint iteration. Independent samples are drawn in intervals of the maximum autocorrelation length for the samples within this segment. We divide the full run into 12 segments and calculate the evidence from each one of these segments to examine how the evidence progresses along the MCMC iterations. Gradually the evidence begins to settle towards a constant value as the MCMC progresses. We take the difference between the last two evidence estimates as the convergence error.

We estimate the total error on our evidence calculations, $\sigma_{\ln \mathcal{Z}}$, by adding the errors in quadrature according to,

$$\sigma_{\ln \mathcal{Z}} = \sqrt{\sigma_{\mathrm{MC}}^2 + \sigma_{\mathrm{convergence}}^2}\,. \tag{5.10}$$

Here, the error σ_{MC} is the Monte Carlo error and $\sigma_{\mathrm{convergence}}$ is the convergence error. Finally, to estimate the Bayes factors we model the log evidence as a normal distribution, with mean given from the log evidence calculation, and standard deviation given by the error propagation formula in Eq. (5.10). The logarithm of the Bayes factor can then be calculated from the difference in the logarithm of the evidences. The standard Bayes factor is then the exponential of the logarithm of the Bayes factor.

As a means of verifying the results from the above Bayes factor calculations we also make use of the Savage-Dickey density ratio method [113, 114, 115] for calculating the Bayes factor of the model where the *p*-*g* mode parameters were chosen independently of one another. This is the approach taken in [2].

For certain kinds nested models where prior distributions on parameters are *factorizable*, or independent from one another, there exists a method for deriving the Bayes factor for two models from one parameter estimation analysis. If there exists a parameter X for which at a critical value X_{crit} the parameter model is equivalent to a nested model that has no parameterization in X, then the Bayes factor for the model with X relative to the model without X is taken as the limit of the prior density at X_{crit} relative to the posterior density at X_{crit}. This method does not require a multi-dimensional integral to be approximated, but only requires the likelihood ratio of the prior density and posterior density at X_{crit}. In the case of the *p*-*g* mode instability the parameter that effectively turns-on-and-turns-off the instability is the amplitude factor A. We can in effect evaluate the ratio between the prior distribution density

as $A \to 0$ and the posterior distribution density as $A \to 0$. This expression thus can be written as:

$$\mathcal{B}^{\mathrm{NL}}_{!\mathrm{NL}} = \lim_{\mathrm{A}\to 0} \frac{\pi\left(\mathrm{A}|H^{NL}\right)}{\mathcal{P}\left(\mathrm{A}|\mathbf{d}, H^{NL}\right)}. \tag{5.11}$$

Formally, the parameter model is constructed such that the prior density on A is distributed uniformly in $\log_{10} A$ and so the limit cannot be strictly taken from within the data acquired in these analyses. However, when $Ais10^{-10}$, the matched-filter is not sensitive in this data set to distinguish the difference between $A = 0$ to $> 99.999\%$ overlap, indicating that the substituting $A \to 0$ for $A \to 10^{-10}$ will likely generate identical results. We provide a more rigorous introduction and derivation for the Savage-Dickey density ratio method in Appendix A.3.

This changes the problem of inference from one-dimensional numerical integration to probability density inference. In our case, we only have one model that has a prior on A that is independent of all of the other priors and so we focus on the model that is most like [2], where A is uniform in $\log_{10}$ between 10^{-10} and $10^{-5.5}$. Fortunately, the prior distributional density is analytically known to us at 10^{-10}, and so we only need to infer the probability density of the posterior distribution at 10^{-10}. There are a variety of methods for estimating the density of a probability distribution when the distribution has to be constructed from sampled data. The simplest method is to construct histograms or use a kernel density function. In [115] it is recommended that the logspline-density package in R be used [116]. We make use of two histogram methods [117, 118], a Gaussian kernel density estimator with boundary-bias corrections [119], and a logspline-density-estimator. We find the Bayes factors they all estimate to be consistent with those found with the multi-tempered Bayes factor estimators. We more fully describe the methods in Appendix **??**.

5.5 Results

Compared to the standard waveform mode, we find that the *p-g* mode model with priors where $\delta\phi$ is unconstrained gives a Bayes factor of order unity. When we use *p-g* mode priors where $\delta\phi > 0.1$ radians we also find a Bayes factor of order unity. Following the Bayes factor interpretation of [73, 120], these Bayes factors cannot be considered to be statistically significant. A Bayes factor of unity indicates that whatever prior beliefs we had about the plausibility of the *p-g* mode instability prior

to GW170817 is unchanged by the observation of GW170817. For the narrow range of $15 \leq f_0 \leq 100$ Hz where $\delta\phi > 0.1$ rad, we find that the Bayes factors are $\mathcal{B} \sim 0.7$. This is also true of the prior range $10 \leq f_0 \leq 100$ Hz with unconstrained $\delta\phi$. The broader range $15 \leq f_0 \leq 800$ Hz, where $\delta\phi > 0.1$ rad, we find that $\mathcal{B} \sim 0.7$ as well. Our estimated statistical error on Bayes factors due to Monte Carlo error and convergence error is $\sim \pm 0.1$ at the 90% confidence level.

When we consider the way that the nonlinear tides enter the Fourier phase in Eq. (5.3), we see that if $n = 4/3$ then the nonlinear tides enter the Fourier phase of the waveform with the same power law dependence on frequency f as the chirp mass, that is $\Psi(f) \propto f^{-5/3}$. We also note that for the effect of nonlinear tides to be degenerate with chirp mass, they must turn on at a frequency f_0 that is close to the low-frequency limit of the detector's sensitive band. If the effect turns on at higher frequencies, then the phasing will change in the detector's sensitive band and it is more difficult to compensate for the nonlinear tide effect with a change in chirp mass.

The marginalized posterior distributions on parameters shown in Fig. 10 show a strong degeneracy between the source-frame chirp mass $\mathcal{M}^{\rm src}$ and nonlinear tides that creates a tail in the chirp mass posterior skewed towards lower values of chirp mass than the value measured using the standard waveform model, $\mathcal{M}^{\rm src} = 1.1867 \pm 0.0001\, M_\odot$ [3]. We see a peak in the posteriors of n and f_0 at $n \lesssim 4/3$ and $f_0 \lesssim 35$ Hz. This parameter degeneracy is also correlated with large A, where $10^{-8} \lesssim A < 10^{-6}$. The samples with large posterior values of $\delta\phi$ seen in Fig. 10 are strongly correlated with source-frame chirp masses $\mathcal{M}^{\rm src} \lesssim 1.1866$. We have examined the change to the posterior distribution when changing the low-frequency cutoff of the likelihood integration from 20 Hz to 25 Hz, and to 30 Hz. In these analyses, the peak in the posterior of f_0 tracks the low-frequency cutoff of the likelihood integration, confirming that this effect is due to the chirp-mass degeneracy with the low-frequency cutoff. The chirp mass degeneracy is also present in the analysis with the broader range of f_0, however it is not as pronounced in the posterior samples due to the larger prior space being explored. For the prior distributions discussed above, the observation of GW170817 does not provide strong statistical evidence either for or against the presence of nonlinear tides.

Given the observed parameter degeneracies, We now investigate regions of the

parameter space where nonlinear tidal effects are not degenerate with standard waveforms by thresholding the prior distribution of p-g waveforms on their fitting factor with standard waveforms. We combine the results of our analysis on the uniform mass, $\delta\phi$ constrained, narrow f_0 prior distribution model to obtain $22,600$ independent samples. We then examine the fitting factor of every independent sample, from every temperature, with a non-spinning, mass-only template bank of TaylorF2 waveforms with comparable masses to GW170817. For simplicity, we only keep the mass parameters and p-g mode parameters in the overlap calculations, since the correlation between nonlinear tidal dynamics is most apparent in the measured chirp mass. When we examine the fitting factor between nonlinear tidal waveforms and this template bank we observe that there is a very high match between standard templates and nonlinear tidal waveforms when $n = 4/3$. The nonlinear tidal waveforms that least match this template bank tend to be those parameterized by large amplitude and large gravitational-wave phase shift. We then recompute the Bayes factor when discarding samples from the analysis below a particular fitting factor with the template bank. To ensure a robustness of the point-estimate we use a bootstrap method to estimate the Monte Carlo error for this Bayes factor estimate [**?**]. The bootstrap estimated Monte Carlo error tends to be much larger than the convergence error for this analysis and so we neglect inclusion of convergence error in the estimate. A statistically significant Bayes factor of $\sim 30\,(20)$, against nonlinear tides, is found when the waveform has an overlap less than $98.5\,(98.85)\%$ match with the standard waveform, see Fig. 9. While this metric is insufficient to rule out the p-g mode instability, it is a useful metric in understanding why the evidence is nearly identical to the evidence from [3]. We find that portions of the p-g mode parameter space that most contribute towards the evidence come from regions of the parameter space that have a high overlap with standard waveforms. This occurs either through A being too small to induce a large change in the phase of the waveform or through an associated parameter degeneracy with the chirp mass caused by large A, low f_0, and $n \sim 4/3$.

Finally, we examine the leading order estimated energy dissipated through nonlinear tides for the case of a uniform prior on the mass, with $15 \leq f_0 \leq 100$ Hz, with a $\delta\phi > 0.1$ radian constraint. In our analysis, the 95^{th} percentile of the estimated energy dissipated through nonlinear tides from our prior distribution is approximately 2.6×10^{51} ergs at the terminating frequency of the TaylorF2 waveform, f_{ISCO}. The

estimated energy radiated by gravitational waves by neutron stars of the estimated mass range of GW170817 is greater than $\sim 10^{53}$ ergs. Our analysis finds the energy dissipated through nonlinear tides at the 95% posterior credible percentile is 3×10^{50} ergs. We find our 95% posterior credible percentile to be less than the 90% confidence interval constraint of $\lesssim 2.7 \times 10^{51}$ ergs in [2]. Samples from our posterior distribution that have dissipation energies greater than the 90% credible interval tend to come from two modes in the parameter space. The first mode is from parts of the parameter space with large A, for $n \sim 4/3$, low f_0, and $\delta\phi \sim 100$ rad. The second mode is from parts of the parameter space with $A \gtrsim 10^{-8}$, for $1.6 \lesssim n < 3.0$, and $\delta\phi \sim 1 - 10$ rad. The high end of the nonlinear tidal energy constraints are thus dominated by waveforms that are degenerate with the standard signal.

5.6 Discussion

In this paper, we have used the observation of GW170817 and the model of [72] to look for evidence of nonlinear tides from *p*-*g* mode coupling during the inspiral [69, 71, 80]. Over the broad prior space, we find a Bayes factor of unity which gives an inconclusive result on whether nonlinear tides are favored or disfavored in GW170817, consistent with [2]. This Bayes factor can be interpreted as stating that there is insufficient evidence to change our prior beliefs about the credibility of the *p*-*g* mode hypothesis after the observation of GW170817. A closer examination of the posterior distribution lead us to conclude that nonlinear tides are consistent with the signal GW170817 because they either cause very small phase shifts to the waveform, or the nonlinear tides must enter the waveform in a way that is degenerate with the other intrinsic parameters of GW170817. Regions of the nonlinear tide parameter space that have a fitting factor of less than 99% (98.5%) are disfavored by a Bayes factor of 15 (25). we find that waveforms from a *p*-*g* mode instability with overlap > 98.5 %, tend to either induce a very small phase shifts to the waveform or are degenerate with other intrinsic parameters of GW170817. This leads us to conclude that modeling GW170817 with nonlinear tidal parameters may not offer advantages over using a simpler model. We conclude that the consistency of the GW170817 signal with the model of [72] is due to parameter degeneracy and that regions where nonlinear tides produce a measurable effect are strongly disfavored.

In principle, one could improve our analysis by separately parameterizing the amplitude, turn-on frequency, and frequency evolution for each star as in [2]. However, we find our results to be broadly consistent with [2], and so we do not expect these to affect the main conclusion of our paper. Further improvements on the parametric model of p-g mode instability could include a higher order post-Newtonian expansion of the instability, or further understanding of the instability's interaction with neutron star magnetic fields [71]. Nonlinear tides are poorly understood and the contribution from other stellar oscillation modes may yet contribute to a more accurate picture of the interior dynamics of neutron stars [81]. Current models of the gravitational-wave phase shift caused by nonlinear tides from the p-g mode instability suffer from parameter degeneracies with the other intrinsic parameters of a neutron star binary. A measurement of the binary's chirp mass that is independent of gravitational-wave observations would break this degeneracy. However, for a system like GW170817, this would require measurement of the binary's chirp mass to a precision greater than $\sim 0.02\%$ using an electromagnetic counterpart, which is implausible. Absent improved theoretical understanding of nonlinear tides from p-g mode coupling that can excludes degenerate regions of the parameter space *a priori*, we do not expect this situation to improve with future detections.

5.7 Acknowledgments

We thank Reed Essick, and Nevin Weinberg for helpful discussion and pointing out errors in our Bayes factor calculation in an earlier draft of this manuscript [121]. We thank Chaitanya Afle, Nils Andersson, Soumi De, Daniel Finstad, and Pantelis Pnigouras for helpful discussions. We thank Alex Nitz for writing the initial version of the code for nonlinear tides in PyCBC. The authors were supported by the National Science Foundation grant PHY-1707954. Computational work was supported by Syracuse University and National Science Foundation grant OAC-1541396. This research has made use of data obtained from the Gravitational Wave Open Science Center (https://www.gw-openscience.org/about/).

PyCBC Inference [94, 101], emcee [102, 103], LIGO Algorithm Library [93], Matplotlib [122], Scipy [123]

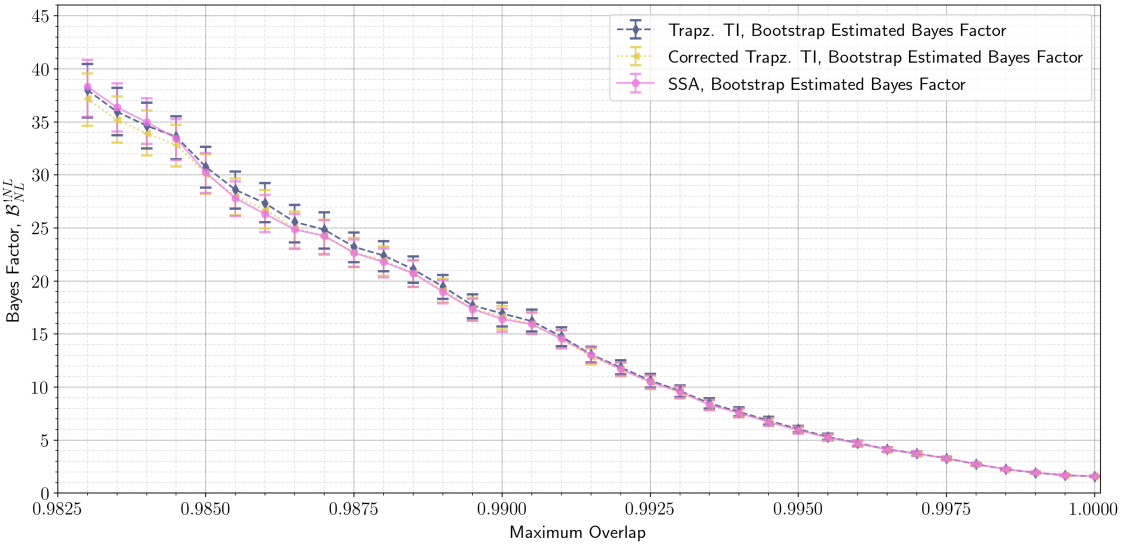

Figure 9: The estimated Bayes factors for nonlinear tidal parameters when the samples are filtered by the fitting factor to a non-spinning, mass-only template bank of TaylorF2 waveforms. The convention in Bayes factor is switched from the main body of the text to represent the Bayes factor for the ratio of evidence for no nonlinear tides, $p\left(\mathbf{d} \,|\, \mathrm{H}_{\mathrm{TaylorF2}}\right)$, to the evidence for nonlinear tides, $p\left(\mathbf{d} \,|\, \mathrm{H}_{\mathrm{TaylorF2+NL}}\right)$. This is abbreviated as $\mathcal{B}^{!\mathrm{NL}}_{\mathrm{NL}}$. The three methods for estimating the Bayes factor are the thermodynamic integration method from trapezoid rule integration (dark grey, dashed line), the thermodynamic integration method from the higher order trapezoid rule (yellow, small-dashed line), and the steppingstone algorithm (dark pink, solid line). A bootstrap method is used to estimate approximate errors on the Bayes Factors. Error bars represent 5^{th} and 95^{th} percentiles. The sampling error becomes large at a fitting factor $\lesssim 99\%$.

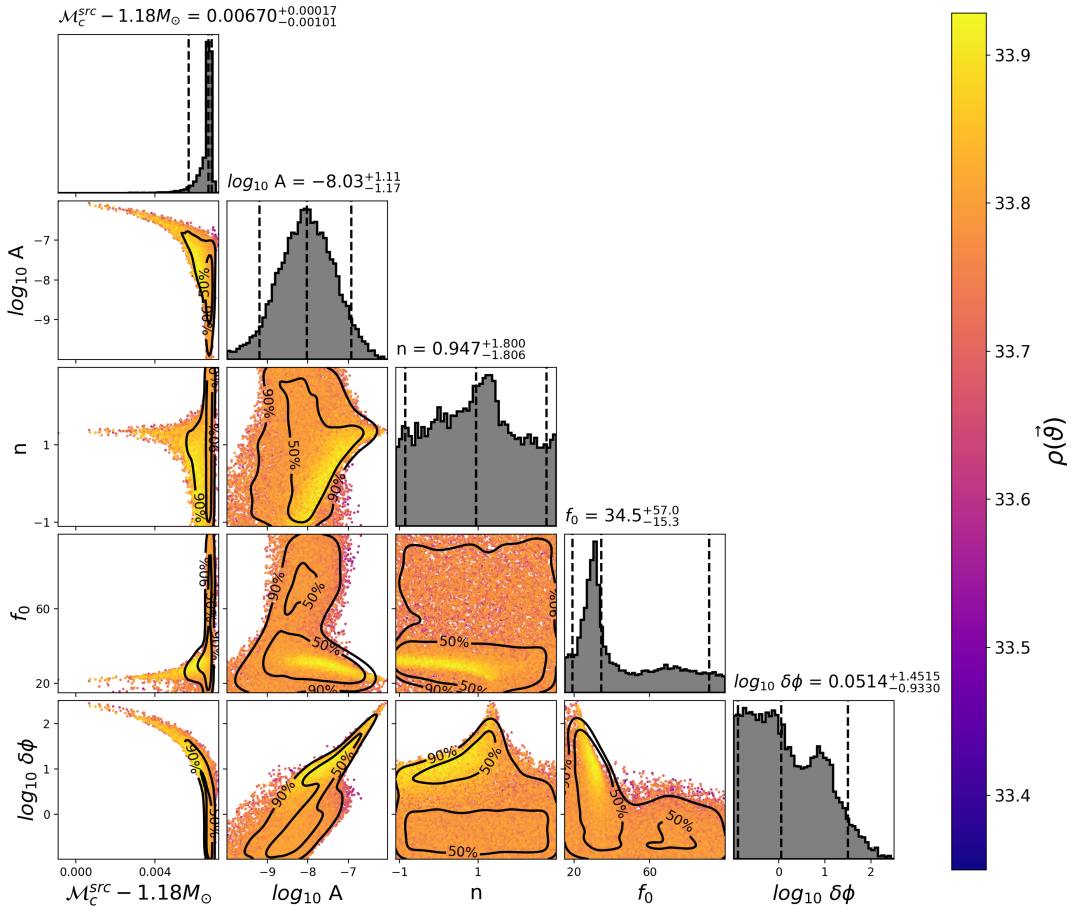

Figure 10: The marginalized posterior distributions for the uniform mass prior and a f_0 restricted to the range 15 and 100 Hz. The vertical lines on the marginalized histograms display the 5th, 50th, and 95th percentiles of the posteriors. The three-detector network signal-to-noise ratio for each sample is given on the color-bar. The posterior scatter plots show 50% and 90% credible interval contours. The posteriors on n is peaked $n \lesssim 4/3$ and for values of f_0 close to the lower end of the detector's low frequency sensitivity. In this region of parameters space, the effect of nonlinear tides is degenerate with chirp mass, causing a secondary peak in the chirp mass posterior. It can be seen from the $\delta\phi$–$\mathcal{M}$ plot (lower left) that large phase shifts due to nonlinear tides are due to points in parameter space where a value of chirp mass can be found that compensates for the phase shift of the nonlinear tides. These are the combined posteriors from 9 runs. It is notable that the the peaks in the f_0 posterior, at $f_0 \approx 30$ Hz and $f_0 \approx 70$ Hz seem to be reversed from those in Fig 2. of [2]. Note that the marginalized posterior for A is diminished for $A < 10^{-8}$ due to the $\delta\phi$ prior constraint.

Chapter 6

Conclusions

The first direct detection of gravitational waves from a binary black hole merger occurred in 2015, beginning the era of gravitational wave astronomy.

In Chapter 3, we introduced the PyCBC offline search analysis pipeline, one of Advanced LIGO's gravitational wave detection pipelines. The PyCBC search pipeline is a matched filter search that searches for gravitational waves from compact binary coalescences by matching the data against gravitational waveform templates. Potential signals that match these waveform templates are designated as triggers and ranked according to a signal to noise ratio ranking statistic in addition to other waveform consistency checks. The statistical significance of these triggers is estimated and statistically significant candidates are considered to be gravitational wave candidates. During advanced LIGO's first observing run, two gravitational wave observations, GW150914 and GW151226 were observed as coming from merging binary black holes.

We also explored the implications of no detections of binary neutron star mergers or neutron-star black hole mergers in Advanced LIGO's first observing run. Using a Bayesian inference framework we modeled the detection of these systems as a Poisson process and estimated the sensitivity of the offline PyCBC search pipeline to mergers of systems with a neutron star through the use of simulated, software injections of these signals. We found that we could constrain the upper limit on the rate of binary neutron star mergers to be less than $12,600$ Gpc^{-3} yr^{-1} at the 90% confidence level. We found that for certain systems of neutron star-black hole systems that the merger rate could be constrained to less than 3600 Gpc^{-3} yr^{-1}. We were also able to constrain

upper limits on gamma-ray burst beaming angles to be greater than ~ 2 degrees.

In Chapter 4, we examined further improvements on the PyCBC offline search analysis and presented 1-OGC, an open gravitational wave catalog. In the work of 1-OGC we analyzed publicly available data from Advanced LIGO's first observing run to search for gravitational waves from compact binary coalescence. Continued work on the gravitational wave ranking statistic have improved PyCBC's ability to discriminate potential gravitational wave signals against other noise transients. We confirmed the observations of gravitational wave signals GW150914, GW151226, and LVT151012. The improvement of the ranking statistic and improvements in statistical significance ranking permitted us to confidently claim GW151012 as an authentic gravitational wave detection with a 97.6% probability of being of astrophysical origin. No other statistically significant gravitational wave candidates were identified. The catalog of events was made open and public for other scientists to investigate.

In Chapter 5, we examined the detection of GW170817, a binary neutron star merger discovered by LIGO and Virgo during their second observing run. Moreover, we conducted a Bayesian parameter estimation and hypothesis testing approach to examine whether nonlinear tides from a nonresonant, nonlinear p-g mode instability were compatible with the observation of GW170817. Our resulting analysis showed that nonlinear tides were broadly compatible with the observation of GW170817, although we found that this occurred because the nonlinear tides did not cause a measurable change to the waveform or the nonlinear tidal parameters entered into the signal through significant degeneracies with the other intrinsic parameters in the signal. As we pursued this problem further we found that we could rule out nonlinear tides from a p-g mode instability that matched standard waveforms with $< 99\%$ match with large statistical significance.

As we collect more gravitational wave events and dig into lower-threshold events GW astronomy will permit us to explore new questions in astrophysics. In this thesis, we presented methods for investigating astrophysical implications for non-detections of gravitational waves from certain binary systems as well as methods for improving the sensitivity of our compact binary coalescence searches towards already detected classes of binary systems. Finally, we developed statistical techniques for testing astrophysical hypotheses on detected signals.

Appendix A

Equations for Pressure-Gravity Mode Instability Analysis

A.1 Derivation of the Fourier phase contribution from the p - g mode instability

We begin our derivation with the energy balance equation presented in [72]

$$\dot{E}_{orbit} = -\dot{E}_{GW} - 2\dot{E}_{NL}, \tag{A.1}$$

for $\dot{E}_{orbit}$ being the rate of energy loss of a quasi-circular orbit, $\dot{E}_{GW}$ being the energy rate loss due to gravitational waves in the point-particle model, and $\dot{E}_{NL}$ being the rate of energy loss from each star's p-g mode instability. We assume that the energy losses from p-g mode instability will be comparable in each star. The $\dot{E}$ notation refers to the derivative of the energy with respect to time. We now give explicit values to these energy rates with respect to gravitational wave frequency, f.

$$\dot{E}_{\rm orbit} = -\frac{G^{2/3}\pi^{2/3}\mathcal{M}^{5/3}\dot{f}}{3f^{1/3}} \tag{A.2}$$

is the orbital energy decay. The gravitational wave energy rate as a function of frequency is given as:

$$\dot{E}_{\rm GW} = \frac{32G^{7/3}\left(\pi\mathcal{M}f\right)^{10/3}}{5c^5} \tag{A.3}$$

Finally, we take from [72] that each star, indexed by i, should have an energy dissipation rate of

$$\dot{E}_{\mathrm{NL},i} = \frac{(2Gm_i)^{2/3} m_1 m_2}{M} (\pi f_{\mathrm{ref}})^{5/3} A \left(\frac{f}{f_{\mathrm{ref}}}\right)^{n+2} \Theta(f - f_0) \tag{A.4}$$

where m_i is the component mass of the neutron star, M is the total mass ($M = m_1 + m_2$). The other parameters are fully described in the Introduction. Assuming that the binaries have equal mass in Eq. A.4 and solving for $\dot{f}$, we arrive at the following expression.

$$\frac{df}{dt} = \pi \left(\frac{f}{f_{\mathrm{ref}}}\right)^{7/3} f_{\mathrm{ref}}^2 \times \left[\frac{96}{5} \left(\frac{G\mathcal{M}\pi f_{\mathrm{ref}}}{c^3}\right)^{5/3} \left(\frac{f}{f_{\mathrm{ref}}}\right)^{4/3} + 6A \left(\frac{f}{f_{\mathrm{ref}}}\right)^{n} \Theta(f - f_0)\right] \tag{A.5}$$

Given this expression, we can now consider a time domain signal of the form, $h(t) = A(t)\ e^{\phi(t)}$, where $h(t)$ is the strain of the gravitational wave at some time t before merger, $A(t)$ is the amplitude of the gravitational wave strain at that same time, and $\phi(t)$ is the orbital phase of the binaries at that same time. The stationary phase approximation lets us approximate the Fourier transform of this time signal according to the following:

$$\tilde{h}(f) = \int_{-\infty}^{\infty} h(t) dt = \int_{-\infty}^{\infty} A(t) e^{-2\pi i f t + \phi(t)} df \approx \tilde{B}(f) e^{-i\Psi(f)} \tag{A.6}$$

where $\tilde{B}(f)$ is the Fourier amplitude of the frequency domain waveform, and $\Psi(f)$ is the Fourier phase of the frequency domain waveform. The full expression for $\Psi(f)$ is

$$\Psi(f) = 2\pi f t(f) - \phi(f). \tag{A.7}$$

One can derive $t(f)$ by solving the differential equation given in Eq. A.5. For convenience we redefine and reorganize this differential equation as:

$$\int_{t_c}^{t} dt = \int_{x_c}^{x} \frac{f_{\mathrm{ref}}}{\kappa} \frac{x^{-7/3} dx}{\alpha x^{4/3} + \Theta(x - x_0)\beta x^n} \tag{A.8}$$

where $x = f/f_{\mathrm{ref}}$, $dx = df/f_{\mathrm{ref}}$, $x_0 = f_0/f_{\mathrm{ref}}$, and $\kappa = \pi f_{\mathrm{ref}}^2$. The integration bounds are the time of coalescence ($t_c = 0$) to some time t prior to merger, and from dimensionless frequency at coalescence ($x_c = f_c/f_{\mathrm{ref}} = \infty$) to some dimensionless frequency x prior to merger. Here α and β are given by the following expressions:

$$\alpha = \frac{96}{5} \left(\frac{G\pi\mathcal{M} f_{\mathrm{ref}}}{c^3}\right)^{5/3} \tag{A.9}$$

$$\beta = 6A \tag{A.10}$$

We can simplify the differential equation given in Eqn. A.8 if we assume that the point particle gravitational wave contribution dominates ($\alpha \gg \beta$), we take a power series expansion assuming large α relative to β. This gives to lowest order in β:

$$\int_{t_c}^{t} dt = \int_{x_c}^{x} \frac{f_{\rm ref}}{\kappa} \left(\frac{1}{\alpha x^{11/3}} + \frac{\Theta(x - x_0)\beta x^{n-5}}{\alpha^2 (n-5)} \right) dx \tag{A.11}$$

The first term in Eq. A.11 corresponds to the zeroth order post-Newtonian result from the point-particle model. Integrating the second term and respecting the $\Theta(x - x_0)$ so as to align the waveform at merger ($t = 0$), we arrive at the leading order contribution of p-g mode instability to $t(f)$:

$$\Delta t(f) = \begin{cases} \frac{-25}{1536} \frac{1}{\pi} \frac{A}{n-4} \left(\frac{G\mathcal{M}\pi f_{\rm ref}}{c^3} \right)^{-10/3} \left(\frac{f_0}{f_{\rm ref}} \right)^{n-4}, & f < f_0 \\ \\ \frac{-25}{1536} \frac{1}{\pi} \frac{A}{n-4} \left(\frac{G\mathcal{M}\pi f_{\rm ref}}{c^3} \right)^{-10/3} \left(\frac{f}{f_{\rm ref}} \right)^{n-4}, & f \geq f_0 \end{cases} \tag{A.12}$$

Following a similar methodology we can calculate $\phi(f)$ via $d\phi = 2\pi f dt = 2\pi (f/\dot{f}) df$. Taking the same power series expansion, integrating so that the waveform coalesces at $t = 0$, and examining the leading order contribution from p-g mode instability we arrive at

$$\int_{\phi_c}^{\phi} d\phi = \int_{x_c}^{x} \frac{f_{\rm ref}}{\kappa} \left(\frac{1}{\alpha x^{8/3}} + \frac{\Theta(x - x_0)\beta x^{n-4}}{\alpha^2 (n-4)} \right) dx \tag{A.13}$$

Integrating this through from ϕ_c, the phase at coalescence, to some earlier ϕ prior to coalescence, and integrating the right hand side of Eq. A.13 we arrive at the zeroth post-Newtonian correction to the phase for the point particle model in integrating the $x^{-8/3}$ term and the lowest order correction due to p-g mode instability in integrating the x^{n-4} term. Thus the correction to the gravitational wave phase due to p-g mode instability can be written as:

$$\Delta\phi(f) = \begin{cases} \frac{-25}{768} \frac{A}{n-3} \left(\frac{G\mathcal{M}\pi f_{\rm ref}}{c^3} \right)^{-10/3} \left(\frac{f_0}{f_{\rm ref}} \right)^{n-3}, & f < f_0 \\ \\ \frac{-25}{768} \frac{A}{n-3} \left(\frac{G\mathcal{M}\pi f_{\rm ref}}{c^3} \right)^{-10/3} \left(\frac{f}{f_{\rm ref}} \right)^{n-3}, & f \geq f_0 \end{cases} \tag{A.14}$$

Finally, we arrive at the expressions for the Fourier phase in terms of Eq. A.12 and A.14 as:

$$\Delta\Psi(f) = \begin{cases} 2\pi f \Delta t(f_0) - \Delta\phi(f_0), & f < f_0 \\ \\ 2\pi f \Delta t(f) - \Delta\phi(f), & f \geq f_0 \end{cases} \tag{A.15}$$

which fully expanded becomes:

$$\Delta\Psi(f) = \begin{cases} -\frac{25}{768} A \left(\frac{f_0}{f_{\rm ref}}\right)^{n-3} \left[\frac{f}{f_0}\frac{1}{n-4} - \frac{1}{n-3}\right], & f < f_0 \\ \\ -\frac{25}{768} A \left(\frac{G\mathcal{M}\pi f_{\rm ref}}{c^3}\right)^{-10/3} \left(\frac{f}{f_{\rm ref}}\right)^{n-3} \left[\frac{1}{n-4} - \frac{1}{n-3}\right], & f \geq f_0 \end{cases} \tag{A.16}$$

We also add the derivation for the first-order energy dissipation from p-g modes in the frequency-domain. This can be solved via:

$$E_{{\rm NL},i}\left(f'\right) = \int_0^{f'} \left(\frac{dE_{{\rm NL},i}}{dt}\right)\left(\frac{dt}{df}\right) df. \tag{A.17}$$

Moving forward we will simply use $E_{{\rm NL},i}$ as a function of f rather than f' as was used here to avoid confusion in the integration bounds. In this derivation we only keep leading order in A and so $\frac{dt}{df}$ can be taken from the point-particle terms in the approximation. Note that this can be taken from the first term in the reciprocal of Eq. (A.5). An A^2 order contribution could be solved for by including the full reciprocal of Eq. (A.5). Some care should be taken however, since the derivation in Eq. (A.5) made use of the simplification that $m_1 = m_2$. We do not take this approach here and so we give $\frac{dt}{df}$ from the point-particle approximation [124]:

$$\frac{dt}{df} = \frac{5}{96}\frac{c^5}{G^{5/3}\pi^{8/3}\mathcal{M}^{5/3}f^{11/3}} \tag{A.18}$$

Plugging Eq. (A.4) and Eq. (A.18) into Eq. (A.17) gives:

$$\frac{dE_{{\rm NL},i}}{df} = \frac{5}{96}\frac{(2m_i)^{2/3}\ m_1\, m_2\, A\, c^5}{G(m_1+m_2)\,\pi\,\mathcal{M}^{5/3}} f_{\rm ref}^{-n-1/3} f^{(n-5/3)}\Theta(f-f_0) \tag{A.19}$$

Integrating this Eq. (A.19) over all frequencies gives us, to linear order, the energy dissipated by the p-g mode instability for a neutron star of mass m_i:

$$E_{{\rm NL},i}(f) = \frac{5}{96}\frac{(2m_i)^{2/3}\ m_1\, m_2\, A\, c^5}{G\,\pi\,(m_1+m_2)\mathcal{M}^{5/3}} f_{\rm ref}^{-n-1/3}\left(f^{n-2/3} - f_0^{n-2/3}\right)\frac{1}{n-2/3}. \tag{A.20}$$

Quick dimensional analysis in SI units confirms that Eq. (A.20) is the form of Joules. In Sec.5.5 we are however only concerned with the energy dissipated by the p-g mode instability at $f_{\rm ISCO}$ when the stars have finally merged. For neutron stars $f_{\rm ISCO}$ is always greater than f_0 in the cases considered in this study, and so this energy dissipation, summing over the contributions from both stars, is:

$$E_{\rm NL}(f_{\rm ISCO}) = \frac{5}{96}\frac{(2m_1+2m_2)^{2/3}\; m_1\, m_2\, A\, c^5}{G\,\pi\,\mathcal{M}^{5/3}} f_{\rm ref}^{-n-1/3}\left(f_{\rm ISCO}^{n-2/3} - f_0^{n-2/3}\right)\frac{1}{n-2/3}. \tag{A.21}$$

A.2 Derivation of the Thermodynamic Integration and Steppingstone methods

A.2.1 The Thermodynamic Integration Method

From the Sec. 5.4 we learned about power-posteriors in Eq. (5.4) and the thermodynamic integral given in Eq. (5.9). We follow resources [112] for the derivation and discussion here. Here we derive the thermodynamic integral by considering the following expression implied by the 2nd Fundamental theorem of Calculus:

$$\ln \mathcal{Z}_{\beta=1}(\mathbf{d}) - \ln \mathcal{Z}_{\beta=0}(\mathbf{d}) = \int_0^1 \left(\frac{d\left(\ln \mathcal{Z}_\beta(\mathbf{d})\right)}{d\beta}\right) d\beta = \int_0^1 \frac{1}{\mathcal{Z}_\beta(\mathbf{d})}\frac{d\,\mathcal{Z}_\beta(\mathbf{d})}{d\beta} d\beta. \tag{A.22}$$

For a properly normalized prior, $\pi(\vec{\theta}$, $\ln \mathcal{Z}_{\beta=0}(\mathbf{d}) = 0$. This leaves the marginal likelihood at $\beta = 1$ that we are interested in, the untempered $\ln \mathcal{Z}(\mathbf{d})$. Now we can expand Eq. A.22 as:

$$\ln \mathcal{Z}(\mathbf{d}) = \int_0^1 \frac{\int \frac{d}{d\beta}\left[\pi\left(\vec{\theta}\right)\,\mathcal{L}\left(\mathbf{d}|\vec{\theta}\right)^\beta\right] d\vec{\theta} d\vec{\theta}}{\int \pi\left(\vec{\theta}\right)\,\mathcal{L}\left(\mathbf{d}|\vec{\theta}\right)^\beta d\vec{\theta}}. \tag{A.23}$$

Suppressing parenthetical arguments on θ and $\mathbf{d}$ for clarity we can arrive at the following expression by taking the derivative in the numerator we then arrive at:

$$\ln \mathcal{Z} = \int_0^1 \frac{\int \pi\;\;(\ln \mathcal{L})\;\;\mathcal{L}^\beta d\theta}{\int \pi \mathcal{L}^\beta d\theta} d\beta. \tag{A.24}$$

Using Bayes theorem we can replace the numerator and denominator with $\mathcal{P}_\beta = \pi\mathcal{L}^\beta/\mathcal{Z}_\beta$ to get:

$$\ln \mathcal{Z} = \int_0^1 \frac{\int \mathcal{P}_\beta\;(\ln \mathcal{L})\, d\theta}{\int \mathcal{P}_\beta d\theta} d\beta = \int_0^1 \langle \ln \mathcal{L}\rangle_{\mathcal{P}_\beta}\, d\beta,. \tag{A.25}$$

Thus, the logarithm of the untempered evidence is given by the one dimensional integral in Eq. (5.9). where $\langle \ln \mathcal{L} \rangle_{\mathcal{P}_\beta}$ represents the average untempered log-likelihood under the measure described by the power-posterior distribution at β. Said in another way, this is the average untempered log-likelihood when drawing samples from the power-posterior distribution at β. We suppress this notation to write $\langle \ln \mathcal{L} \rangle_{\mathcal{P}_\beta} \equiv \langle \ln \mathcal{L} \rangle_\beta$ in the main-body of the text. Thus simulating from power-posterior distributions with β between 0 and 1 provide a means to estimating the logarithm of the untempered evidence for the model and thus present a tractable way towards Bayesian model selection and comparison. This method is an ubiased estimator of the evidence provided that samples of $\langle \ln \mathcal{L} \rangle_\beta$ can be drawn in an unbiased manner from power-posteriors [125].

It is convenient to describe additional derivatives of the thermodynamic integrand as they will be useful as references in the next section. In general, n$^{\text{th}}$ derivatives of the form $\ln \mathcal{Z}$ can be solved as [126][1]:

$$\frac{d^n}{d\beta^n} (\ln \mathcal{Z}) = \sum_{k=1}^{n} \frac{(-1)^{(k+1)} \binom{n}{k}}{k \mathcal{Z}^k} \frac{d^n}{d\beta^n} \left(\mathcal{Z}^k \right). \tag{A.26}$$

The first derivative, $n = 1$, we have already solved as being $\langle \ln \mathcal{L} \rangle_\beta$. The next derivative, $n = 2$, was found in [111] as $\text{Var}(\ln \mathcal{L})_\beta$, which is the variance of the untempered log likelihood samples when drawn from the power-posterior at β. We solve the next derivative, $n = 3$, as:

$$\frac{d^3}{d\beta^3} (\ln \mathcal{Z}) = \langle (\ln \mathcal{L})^3 \rangle_\beta + 2 \langle \ln \mathcal{L} \rangle_\beta^3 - 3 \langle (\ln \mathcal{L})^2 \rangle_\beta \langle \ln \mathcal{L} \rangle_\beta. \tag{A.27}$$

An astute observation would be to recognize that the pattern here follows that the n$^{\text{th}}$ derivative of $\ln \mathcal{Z}$ follow the pattern of the n$^{\text{th}}$ cumulants of the power-posterior distribution at β [111, 125]. The term $\mathcal{Z}$ describes a partition function for the posterior distribution $\mathcal{P}$ [125, 127], and $\ln \mathcal{Z}$ can be thought of as a cumulant generating function [128]. This relationship can make computation of values of high order derivatives more numerically stable since cumulants of order ≥ 2 are shift-invariant [128]. We can make the transformation of variables, $\widetilde{\ln \mathcal{L}} \equiv \ln \mathcal{L} - \ln \mathcal{L}_{\text{max}}$ for every power-posterior before calculating Eq. (A.27). We have tested this on high-order derivatives and found it to be both accurate and numerically stable, although we have found that

[1] Note that the solution in [126] has a minor typo, which we correct here.

the parallel-tempered *emcee* sampler [102, 103] may not be stable enough to permit high-order derivatives to be accurate in all cases.

The thermodynamic integrand with the next two derivatives are shown in Fig. 11 for the permissive $\delta\phi$ prior choice ($\log_{10}A \in U[-10,-5.5]$, $n \in U[-1,3)$, $f_0 \in U[10,100]\ Hz$) with a linear in β scale. We also produce this plot in the logarithmic scale in 12 where inspection of the curvature of the thermodynamic integrand is easier to see. These plots are helpful to inspect for places where the integrand may not be well sampled in β and hence require additional inverse-temperatures [108, 129, 130]. Of particular note is the instability in the second (third) subplot of Fig. 12 where the second (third) derivative is not smooth in β. Even in the first subplot, where we expect the thermodynamic integrand to be smooth and monotonically increasing as β goes from 0 to 1, there is some numerical instability at $\beta \sim 10^{-9}$. This implies, perhaps, the need for a better tempering sampler or bias-corrective terms in the sampling such as those found in [131, 132]. The instability is so slight however that we do not expect it to significantly impact the Bayes factor estimation, but it is a potential source of error in the numerical integration.

Numerical Quadrature

The thermodynamic integral in Eq. (5.9) can be estimated through numerical quadrature rules such as the trapezoidal rule, or Simpson's rule. Because β for thermodynamic integration are typically not uniformly distributed between 0 and 1, it is beneficial to consider integration rules that do not depend on equally spaced abscissa. A polynomial interpolant that does not make use of derivatives of the function or equally spaced abscissa is the Newton's divided difference interpolant, see [133, 134, 135] for examples of how to construct these polynomials. Other interpolants, and thus integration rules, can be constructed, see [135] for examples.

The simplest rule that we consider here is the trapezoidal rule which can be written for thermodynamic integration as:

$$\widehat{\ln \mathcal{Z}}_{\rm Trapz} = \sum_{i=0}^{N_\beta - 1} \frac{1}{2} \left(\beta_{i+1} - \beta_i\right) \left(\langle \ln \mathcal{L} \rangle_{\beta_{i+1}} + \langle \ln \mathcal{L} \rangle_{\beta_i}\right) \tag{A.28}$$

Here N_β represents the number of β being summed over in the integration estimation.

The error corrective term to the trapezoidal rule can be found by integrating the next-to-leading order Taylor polynomial correction [135], yielding:

$$\widehat{\ln \mathcal{Z}}_{\text{Trapz}+} \approx \widehat{\ln \mathcal{Z}}_{\text{Trapz}} + \sum_{i=0}^{N_\beta - 1} -\frac{1}{12} \left(\beta_{i+1} - \beta_i\right)^2 \left(f'(\beta_{i+1}) - f'(\beta_i)\right). \tag{A.29}$$

Here $f'(\beta_i)$ represents the second derivative of $\ln \mathcal{Z}$ with respect to β. It was found in [111] that this corresponds to the variance of the untempered log likelihood as drawn from the power-posterior at β_i.

Simpson's rule for unequally spaced abscissa under Newton's divided difference interpolation [136] is:

$$\widehat{\ln \mathcal{Z}}_{\text{Simps}} = \sum_{\text{i is even}, i=0}^{N_\beta - 2} \frac{h_i + h_{i+1}}{6} \left[A \left\langle \ln \mathcal{L} \right\rangle_{\beta_i} + B \left\langle \ln \mathcal{L} \right\rangle_{\beta_{i+1}} + C \left\langle \ln \mathcal{L} \right\rangle_{\beta_{i+2}}\right], \tag{A.30}$$

for the expressions:

$$\begin{aligned} A &= \tfrac{(2h_i - h_{i+1})}{h_i} \\ B &= \tfrac{(h_i + h_{i+1})^2}{h_i h_{i+1}} \\ C &= \tfrac{(2h_{i+1} - h_i)}{h_{i+1}}. \end{aligned} \tag{A.31}$$

Here $h_i \equiv \beta_{i+1} - \beta_i$, and $h_{i+1} \equiv \beta_{i+2} - \beta_{i+1}$. The error corrective term for Simpson's rule can thus be solved in the same manner as for the trapezoidal rule and we find:

$$\widehat{\ln \mathcal{Z}}_{\text{Simps}+} \approx \widehat{\ln \mathcal{Z}}_{\text{Simps}} + \sum_{\text{i is even},\, i=0}^{N_\beta - 2} \frac{1}{72} \left(\beta_{i+2} - \beta_i\right)^2 \left(\beta_i - 2\beta_{i+1} + \beta_{i+2}\right) \frac{f''(\beta_{i+2}) - f''(\beta_i)}{\beta_{i+2} - \beta_i}. \tag{A.32}$$

Here $f''(\beta_i)$ represents the third derivative of $\ln \mathcal{Z}$ with respect to β, which is in Eq. A.27.

The cubic integration rule for unequally spaced abscissa under Newton's divided difference interpolation can be found in [133, 134, 137] or can be derived through the tools in [135]. We also use the cubic integration rule, and in particular we use the form given in [137]:

$$\widehat{\ln \mathcal{Z}}_{\text{cubic}} = \sum_{\text{i is a multiple of 3}, i=0}^{N_\beta - 3} \frac{h_i + h_{i+1} + h_{i+2}}{12} \left[A \left\langle \ln \mathcal{L} \right\rangle_{\beta_i} + B \left\langle \ln \mathcal{L} \right\rangle_{\beta_{i+1}} + C \left\langle \ln \mathcal{L} \right\rangle_{\beta_{i+2}} + D \left\langle \ln \mathcal{L} \right\rangle_{\beta_{i+3}}\right], \tag{A.33}$$

for expressions:

$$\begin{aligned} A &= \tfrac{3h_i^2-h_{i+1}^2+h_{i+2}^2+2h_ih_{i+1}-2h_ih_{i+2}}{h_i(h_i+h_{i+1})} \\ B &= \tfrac{(h_i+h_{i+1}+h_{i+2})^2(h_i+h_{i+1}-h_{i+2})}{h_ih_{i+1}(h_{i+1}h_{i+2})} \\ C &= \tfrac{(h_i+h_{i+1}+h_{i+2})^2(h_{i+1}+h_{i+2}-h_i)}{h_{i+1}h_{i+2}(h_i+h_{i+1})} \\ D &= \tfrac{h_i^2-h_{i+1}^2+3h_{i+2}^2-2h_ih_{i+2}+2h_{i+1}h_{i+2}}{h_{i+2}(h_{i+1}+h_{i+2})}. \end{aligned} \tag{A.34}$$

Here we have defined $h_i \equiv \beta_{i+1} - \beta_i$, $h_{i+1} \equiv \beta_{i+2} - \beta_{i+1}$, and $h_{i+2} \equiv \beta_{i+3} - \beta_{i+2}$.

We exercise caution in describing the thermodynamic integral through a higher order polynomial quadrature rule as it may not be well described by polynomials. Thus there may be very little incentive for going to higher order polynomial rules as improved accuracy is not always guaranteed by going to higher order polynomial integration rules [138]. It is important to note that we treat the logarithm of the evidence as an unknown quantity which we are trying to infer the value of, and so we must treat the evidence as a random variable. Without prior information we should exercise caution when trusting one of these quadrature rules above the others. Our inference is more confident when these quadrature rules agree on the numerical value of the logarithm of the evidence.

Future studies may make use of Taylor series polynomials for unequally spaced abscissa, ratios of Taylor series polynomials through the Padé approximant for improved accuracy [139], or other interpolant functions. Improvement in numerical integration for thermodynamic integration may also be improved by focusing on increasing the number of inverse-temperatures β and/or by improved placement of β.

Monte Carlo Error

Here we follow the discussion from [112] who provide a heuristic for estimating the Monte Carlo error to estimating the thermodynamic integral as first given in [107]. A variance of the thermodynamic integral estimator, $\widehat{\ln \mathcal{Z}}$, from Monte Carlo error can be found in two steps. First, calculate the thermodynamic integral for each sample of untempered log likelihoods drawn from the power-posterior at β. For N samples drawn from each power-posterior this generates N thermodynamic integral

values. The integration should be done relative to the numerical quadrature technique that one is trying to estimate the Monte-Carlo error for. This represents the sample variance of the thermodynamic integration. The variance of the mean value of the logarithm of the evidence can be calculated via:

$$\sigma^2_{\mathrm{MC}} = \frac{1}{N}\sigma^2_{\mathrm{sample}}. \tag{A.35}$$

Here, σ^2_{MC} represents the Monte-Carlo variance for the thermodynamic integration estimator while $\sigma^2_{\mathrm{sample}}$ is the sample variance and N represents the number of available samples. See Fig. 13 for a visualization of this procedure.

Repeated runs where the random seed for the Markov-Chain Monte Carlo analysis was changed has shown that the variance estimate from presented in [112] is a plausible confidence interval estimate for Monte Carlo error. It has also shown good agreement with the steppingstone Monte Carlo error estimate which uses the same samples as those in thermodynamic integration.

Convergence Error

The procedure of estimating the marginal likelihood from power-posterior simulation requires that the power-posteriors all converge to the proper distribution. To first order, this requires inspection of the thermodynamic integrand over the course of the MCMC analysis. To next order, this would require that sequential cumulants of the power-posterior also stabilizes. In the limit that the MCMC analysis has converged all of the power-posterior distributions will be stationary as a function of MCMC iteration. During the course of the study we did not fully investigate the stationarity of the full power-posterior distribution through inspection of all of the cumulants, but rather focused on the stationarity of the thermodynamic integrand across each temperature. This resulted in investigating the stationarity of the the thermodynamic integral as well.

An accurate depiction of the power-posterior distribution for a particular temperature requires that all of the samples be independent and identically distributed samples (this is sometimes called i.i.d. in the statistics literature) [112]. Gathering independent and identically distributed samples can be done by calculating the autocorrelation length of the MCMC chains from a particular temperature. In practice, PyCBC Inference calculates the autocorrelation length of all of the temperature

chains and uses the largest autocorrelation length as the autocorrelation length for all temperatures. This is a safe and conservative practice for ensuring that samples drawn from the MCMC simulation are not correlated. Thus, to track the thermodynamic integrand at various iterations in the MCMC simulation we divide the MCMC analysis into 12 equally spaced partitions based on the number of MCMC iterations that the analysis has undergone. In practice any number of partitions will do, but it is computationally intensive to sample more partitions. The partitions do not need to be equally spaced in MCMC iterations but we find equally spaced partitions to be useful for visualization of the progression of the thermodynamic integrand. Using this number of partitions, each partition is segmented in half, where the first half is discarded as burn-in samples, and the autocorrelation length is calculated from the remaining samples. Then independent samples are drawn from this segment spaced out by autocorrelation length. This is the generic procedure of the $\mathrm{n_{acl}}$ algorithm implemented in PyCBC for drawing independent samples from the Markov chains. The partitioning is shown in Fig. 14. Having drawn independent samples from 12 segments of the MCMC analysis we can visually inspect the stability of the thermodynamic integrand at 12 iterations in the MCMC analysis. We can also inspect the convergence of the thermodynamic integral. When the logarithm of the evidence has converged to $\mathcal{O}(10^{-2}$ accuracy, we usually consider the power-posteriors to have converged to their final distribution. We base our inference on convergence on the worst quadrature method, the trapezoidal rule for thermodynamic integration, as a means to be conservative. Higher order quadrature rules can converge more rapidly than the trapezoidal rule, however, the more closely aligned the numerical quadrature estimates the more confidence we can attain in the accuracy of the thermodynamic integration. Figure 15 shows the progression of the convergence of the thermodynamic integrand as a function of the MCMC iteration. Figure 16 shows the convergence rate of the thermodynamic integral as a function of the MCMC iteration for a variety of integration techniques.

The absolute value of the difference between the last two thermodynamic integration estimates from this partitioning are then used as the standard deviation of the error for the log evidence due to convergence error, $\sigma_{\mathrm{convergence}}$:

$$\sigma_{\mathrm{convergence}} \sim \mid \ln \mathcal{Z}_{\mathrm{partition}\,N} - \ln \mathcal{Z}_{\mathrm{partition}\,N-1} \mid \tag{A.36}$$

This provides a rough estimate for ensuring that we do not terminate the MCMC analysis too early and thus give ourselves overconfidence about the true value of logarithm of the evidence. Previous analyses such as those in [3] did not make use of this technique and the analysis is terminated based on the number of independent samples collected in the posterior distribution. This is an insufficient metric for inference analyses based on Bayesian model selection through multi-tempered MCMC simulation. We use this estimate for $\sigma_{\text{convergence}}$ in Eq. 5.10.

During the development of this technique a similar technique based on a moving-block bootstrap method was developed in [110] for error analysis of the logarithm of the evidence from the thermodynamic integration method. We have not investigated this technique thoroughly to compare its performance with our own method.

Temperature Placement Bias

The placement of inverse-temperatures β affects the results of the numerical integration for the evidence[106, 109]. This study has a particular bias at the low-end of $\beta \to 0$ because it does not include $\beta = 0$ in the numerical integration. This tail-end bias at small β is much, much lower than the resolution of the Monte Carlo error and the convergence error and so it is irrelevant to the final results. However, future practice should always include $\beta = 0$.

Research into the proper placement of β is ongoing in the field of Statistics [112]. We followed the suggestions in [108] on placing temperatures where the thermodynamic integrand changed rapidly. With 51 inverse-temperatures it is incredibly unlikely that the main results of the Bayes factors being ~ 0.7 are biased by discretization error outside of the statistical uncertainties. This is implied by the agreement of the results between thermodynamic integration, the steppingstone, and the Savage-Dickey density ratio methods. In Fig. 17 there is some slight disagreement on the exact value of the marginal log likelihood, but it seems to cancel out in the Bayes factor since all models share the same temperature ladder (see Fig 18).

One method for improving the temperature ladder would be to use the method of [111] by using the intersection of the slopes of the thermodynamic integrand from two adjacent power-posteriors as a new position for additional temperatures. We could go further by using the intersections of higher-order polynomials using the expressions for the derivatives of the thermodynamic integrand. However, this is not

likely to be fruitful in this study as our error at the moment appears dominated by Monte Carlo error and convergence error. Future studies using multi-tempering techniques may make use of this method in refining temperature placement.

It is our opinion that the placement of inverse-temperatures, if it cannot be solved analytically, should be considered a question of inference. That is to say, given a prior belief on an appropriate distribution on the placement of inverse-temperatures, how should one adjust the placement of inverse-temperatures given the results of the numerical quadrature routine? Bayesian quadrature is a promising area of research meant to address optimal placement of abscissa for numerical integration. It is likely that the thermodynamic integration method would likely greatly benefit from this sort of an approach. See [140] for an initial formulation and [141] for a modern perspective, especially with respect to thermodynamic integration.

A.2.2 The Steppingstone Method

The steppingstone method is very similar in many respects to thermodynamic integration in that it requires multiple inverse-temperatures between 0 and 1 to calculate. The motivation for steppingstone is that it uses importance sampling between adjacent temperatures to estimate the contribution to the marginal likelihood $\mathcal{Z}$ at each interval β_{i-1}-β_i. Before we derive the steppingstone method we provide a brief, but useful derivation of another often used identity, called the harmonic mean estimator for the evidence. For the following section we suppress use of $\vec{\theta}$ and $\mathbf{d}$.

For the derivation of the harmonic mean estimator we follow a simplified version of the derivation presented in [142]. From the definition of the marginal likelihood we can write:

$$\frac{1}{\mathcal{Z}} = \frac{1}{\int \pi \, \mathcal{L} \, d\theta}. \tag{A.37}$$

Since we only deal with proper priors we can substitute the numerator with $\int \pi d\theta = 1$. This gives:

$$\frac{1}{\mathcal{Z}} = \frac{\int \pi \, d\theta}{\int \pi \, \mathcal{L} \, d\theta}. \tag{A.38}$$

Now we multiply both the numerator and denominator by $\mathcal{P}/\mathcal{P}$ to get:

$$\frac{1}{\mathcal{Z}} = \frac{\int \frac{\pi}{\mathcal{P}} \mathcal{P} \, d\theta}{\int \frac{\pi \, \mathcal{L}}{\mathcal{P}} \mathcal{P} \, d\theta}. \tag{A.39}$$

Which we simplify using Bayes theorem to substitute out for $1/\mathcal{P}$ to give:

$$\frac{1}{\mathcal{Z}} = \frac{\int \frac{\pi \mathcal{Z}}{\pi \mathcal{L}} \mathcal{P}\, d\theta}{\int \frac{\pi \mathcal{L} \mathcal{Z}}{\pi \mathcal{L}} \mathcal{P}\, d\theta}. \tag{A.40}$$

Cancelling out terms of π and moving terms of $\mathcal{Z}$ out of the integral to cancel, this gives:

$$\frac{1}{\mathcal{Z}} = \frac{\int \frac{1}{\mathcal{L}} \mathcal{P}\, d\theta}{\int \mathcal{P}\, d\theta} = \int \frac{1}{\mathcal{L}} \mathcal{P}\, d\theta. \tag{A.41}$$

Therefore we can express the inverse of the evidence as:

$$\frac{1}{\mathcal{Z}} = \langle \mathcal{L}^{-1} \rangle_{\mathcal{P}}, \tag{A.42}$$

which is to say that the inverse of the evidence is given as the average value of the inverse of the likelihood when sampled from the measure defined by the posterior distribution. This is the harmonic mean estimator of the evidence, and although it is correct in theory, it typically misbehaves numerically and computationally. It is also worth noting as [109] points out that the harmonic mean estimator is not sensitive to the prior distribution which runs contradictory to the heart of Bayesian model comparison. We will use this identity in the derivation of the steppingstone estimator.

We follow [112] in the derivation of the steppingstone estimator. Recall from Eq. 5.9 that the marginal likelihood can be expressed as:

$$\ln \mathcal{Z} = \ln \mathcal{Z}_{\beta=1} - \ln \mathcal{Z}_{\beta=0}, \tag{A.43}$$

which is equivalent to:

$$\mathcal{Z} = \frac{\mathcal{Z}_{\beta=1}}{\mathcal{Z}_{\beta=0}}. \tag{A.44}$$

For, say, a hundred equally spaced temperatures between 0 and 1 this motivates the following re-expression:

$$\mathcal{Z} = \frac{\mathcal{Z}_{\beta=0.01}}{\mathcal{Z}_{\beta=0}} \times \frac{\mathcal{Z}_{\beta=0.02}}{\mathcal{Z}_{\beta=0.01}} \times \ldots \times \frac{\mathcal{Z}_{\beta=0.99}}{\mathcal{Z}_{\beta=0.98}} \times \frac{\mathcal{Z}_{\beta=1}}{\mathcal{Z}_{\beta=0.99}}. \tag{A.45}$$

The general form for this is:

$$\mathcal{Z} = \prod_{i=1}^{N_\beta} \frac{\mathcal{Z}_{\beta_i}}{\mathcal{Z}_{\beta_{i-1}}}. \tag{A.46}$$

Here we use the ordering on β, as $\beta_0 = 0 < \beta_1 < ... < \beta_{N_\beta - 1} < \beta_{N_\beta} = 1$. Finally, then, consider the evidence for the power-posterior at inverse-temperature β_i given as:

$$\mathcal{Z}_{\beta_i} = \int \pi \mathcal{L}^{\beta_i} \, d\theta. \tag{A.47}$$

We now divide by 1 via $\int \pi \, d\theta$ and multiply by 1 via $\mathcal{P}_{\beta_{i-1}} / \mathcal{P}_{\beta_{i-1}}$ in the numerator and denominator to get:

$$\mathcal{Z}_{\beta_i} = \left(\int \frac{\pi \mathcal{L}^{\beta_i}}{\mathcal{P}_{\beta_{i-1}}} \mathcal{P}_{\beta_{i-1}} \, d\theta \right) \Big/ \left(\int \frac{\pi}{\mathcal{P}_{\beta_{i-1}}} \mathcal{P}_{\beta_{i-1}} \, d\theta \right). \tag{A.48}$$

Using Bayes theorem we substitute $\mathcal{P}_{\beta_{i-1}} = (1/\mathcal{Z}_{\beta_{i-1}}) \, \pi \, \mathcal{L}^{\beta_{i-1}}$ to get:

$$\mathcal{Z}_{\beta_i} = \left(\int \frac{\pi \mathcal{L}^{\beta_i} \mathcal{Z}_{\beta_{i-1}}}{\pi \mathcal{L}^{\beta_{i-1}}} \mathcal{P}_{\beta_{i-1}} \, d\theta \right) \Big/ \left(\int \frac{\pi \mathcal{Z}_{\beta_{i-1}}}{\pi \mathcal{L}^{\beta_{i-1}}} \mathcal{P}_{\beta_{i-1}} \, d\theta \right). \tag{A.49}$$

Terms of $\mathcal{Z}_{\beta_{i-1}}$ are independent of θ and so can be moved out of the integral where they cancel, and we can cancel terms of π to get:

$$\mathcal{Z}_{\beta_i} = \left(\int \frac{\mathcal{L}^{\beta_i}}{\mathcal{L}^{\beta_{i-1}}} \mathcal{P}_{\beta_{i-1}} \, d\theta \right) \Big/ \left(\int \frac{1}{\mathcal{L}^{\beta_{i-1}}} \mathcal{P}_{\beta_{i-1}} \, d\theta \right). \tag{A.50}$$

Finally we recognize that in the denominator we have Eq. A.41 for the inverse of the evidence at the inverse-temperature β_{i-1}, and in the top we can simplify terms so as to get:

$$\mathcal{Z}_{\beta_i} = \mathcal{Z}_{\beta_{i-1}} \int \mathcal{L}^{\beta_i - \beta_{i-1}} \mathcal{P}_{\beta_{i-1}} \, d\theta. \tag{A.51}$$

Thus we arrive at the key ingredient for the steppingstone estimator:

$$\frac{\mathcal{Z}_{\beta_i}}{\mathcal{Z}_{\beta_{i-1}}} = \int \mathcal{L}^{\beta_i - \beta_{i-1}} \mathcal{P}_{\beta_{i-1}} \, d\theta = \langle \mathcal{L}^{\beta_i - \beta_{i-1}} \rangle_{\mathcal{P}_{\beta_{i-1}}}. \tag{A.52}$$

This can be uncomfortably summarized as saying, in the interval of inverse-temperatures between β_{i-1} and β_i, the ratio of the evidences between successive inverse-temperatures is given by the average of the likelihood raised to the difference in the inverse-temperatures when samples for the likelihood are drawn from the power-posterior distribution for the smaller of the inverse-temperatures in the inverse-temperature interval. We suppress some of the notation in Eq. (A.52) such that $\langle \mathcal{L}^{\beta_i - \beta_{i-1}} \rangle_{\mathcal{P}_{\beta_{i-1}}} \equiv \langle \mathcal{L}^{\beta_i - \beta_{i-1}} \rangle_{\beta_{i-1}}$ We finally combine Eq. (A.52) into Eq. (A.46) to achieve the stepping-stone estimator for the evidence:

$$\mathcal{Z} = \prod_{i=1}^{N_\beta} \langle \mathcal{L}^{\beta_i - \beta_{i-1}} \rangle_{\beta_{i-1}}. \tag{A.53}$$

Some care needs to be taken in the implementation of Eq. (A.53) as the form presented is not numerically stable and we often must use the log likelihood and log evidence in place of the likelihood and the evidence. A numerically stable form of the logarithm of Eq. (A.53) is presented in [109]. It is noted to exhibit some level of bias as an estimator of the marginal likelihood due to its logarithmic form. The bias is noted to be small, and it was shown in [109] that the steppingstone estimator typically outperforms the trapezoidal rule for thermodynamic integration in terms of accuracy. This bias can be mitigated by increasing the number of temperatures and/or improving their placement between 0 and 1 [109].

For the steppingstone estimator we can use the same samples as for the thermodynamic integration method. Optimal temperature placement for the steppingstone estimator is an active area of research [112].

Monte Carlo Error

In [109] there is an expression for the estimated variance of the logarithmic steppingstone estimator using an approximation method called the δ method [143]. The expression in [109] for the variance of the logarithm of the evidence is however not presented in a numerically stable version. We use a numerically stabilized version of the variance estimator in our study. We have found the variance estimate from the δ method is typically comparable to the thermodynamic integration method's Monte-Carlo error. Repeated runs where the random seed for the Markov-Chain Monte Carlo analysis was changed has shown that the variance estimate from presented in [109] is a plausible confidence interval estimate for Monte Carlo error.

Convergence Error

The method for calculating the error on the steppingstone estimator due to convergence error is algorithmically identical to the thermodynamic integration method.

Temperature Placement Bias

At the current time optimal placement of inverse-temperatures β remains an active area of research [112]. Due to the high density of inverse-temperatures β we do not

believe that increased number of β nor more optimal placement of β would significantly alter the results of this study. It is regrettably a potential source of bias in our analysis that we cannot fully quantify, although heuristically we believe the bias to be small.

A.3 Derivation of the Savage-Dickey Density Ratio Method

The Savage-Dickey density ratio method for Bayes factor calculation requires consideration of two models, wherein one model is nested in the other model. We derive the method and explain its limitations following [115]. We can consider two models that are parametrized in the following way:

$$\pi\left(\vec{\theta}_{\rm simple}|{\rm H}_{\rm simple}\right) \equiv \pi\left(\left\{\mathcal{M},\eta,\chi_{\rm eff},\tilde{\Lambda},\ldots\right\}|{\rm H}_{\rm simple}\right) \tag{A.54}$$

$$\pi\left(\vec{\theta}_{\rm complex}|{\rm H}_{\rm complex}\right) \equiv \pi\left(\left\{\mathcal{M},\eta,\chi_{\rm eff},\tilde{\Lambda},\ldots,A,f_0,n\right\}|{\rm H}_{\rm complex}\right). \tag{A.55}$$

In the p-g mode instability parametrization setting $A = 0$, effectively reduces the parameter space from the complex parameter space including p-g mode parameters to the simple parameter space denoted as the standard TaylorF2 parameter space in the main text. We abbreviate the notation by writing the prior under the simple hypothesis as $\pi_{\rm !NL}\left(\psi\right)$ and the prior under the more complex hypothesis as $\pi_{\rm NL}\left(\psi, A\right)$. Here the dependence on hypotheses is denoted by the subscript !NL or NL, and ψ denotes all parameters that are not A. The parameters ψ can be considered for the purposes of this derivation to be nuisance parameters. In order for the Savage-Dickey Density Ratio method to hold for the case here we require the following expression be satisfied:

$$\lim_{A\to 0}\pi_{\rm NL}\left(\psi|A\right) = \pi_{\rm !NL}\left(\psi\right). \tag{A.56}$$

In essence, this is stating that setting $A = 0$ reduces the prior parameter space from including p-g mode parameters (and they're potential effect on the likelihood function) down to the TaylorF2 parameter space with point-particle parameters and linear tidal parameters. These conditions are in fact satisfied by setting $A = 0$ and so we can present the Bayes factor as:

$$\mathcal{B}^{\rm NL}_{\rm !NL} = \frac{\mathcal{Z}_{\rm NL}(\mathbf{d})}{\mathcal{Z}_{\rm !NL}(\mathbf{d})}. \tag{A.57}$$

Now, we also know that the denominator can be expressed according to:

$$\mathcal{Z}_{\text{!NL}}(\mathbf{d}) = \int \pi_{\text{!NL}}\left(\psi\right)\, \mathcal{L}_{\text{!NL}}\left(\mathbf{d}|\psi\right) d\psi. \tag{A.58}$$

Since the models are nested, the prior (likelihood) under the NL hypothesis at $A = 0$ is equivalent to the prior (likelihood) under the !NL hypothesis. That is to say:

$$\pi_{NL}\left(\psi, A = 0\right) = \pi_{!NL}\left(\psi\right) \tag{A.59}$$

and

$$\mathcal{L}_{NL}\left(\mathbf{d}|\psi, A = 0\right) = \mathcal{L}_{!NL}\left(\mathbf{d}|\psi\right). \tag{A.60}$$

If we substitute Eqs. A.59 and A.60 into Eq. A.58 we get:

$$\mathcal{Z}_{\text{!NL}}(\mathbf{d}) = \int \pi_{\text{NL}}\left(\psi, A = 0\right)\, \mathcal{L}_{\text{NL}}\left(\mathbf{d}|\psi, A = 0\right) d\psi. \tag{A.61}$$

Integrating this over all ψ, leaves the $A = 0$ unintegrated over leaving us with $\mathcal{Z}_{\text{!NL}} = \mathcal{L}_{\text{NL}}\left(\mathbf{d}|A = 0\right)$. Using Bayes theorem, we can rewrite $\mathcal{L}_{\text{NL}}\left(\mathbf{d}|A = 0\right) = [\mathcal{P}_{\text{NL}}(A = 0|\mathbf{d})\, \mathcal{Z}_{\text{NL}}(\mathbf{d})]/\pi_{\text{NL}}(A = 0)$. This leaves us with:

$$\mathcal{Z}_{\text{!NL}}\left(\mathbf{d}\right) = \frac{\mathcal{P}_{\text{NL}}\left(A = 0|\mathbf{d}\right)\, \mathcal{Z}_{\text{NL}}\left(\mathbf{d}\right)}{\pi_{\text{NL}}\left(A = 0\right)}, \tag{A.62}$$

and thus:

$$\mathcal{B}^{\text{NL}}_{\text{!NL}} = \frac{\pi_{\text{NL}}\left(A = 0\right)}{\mathcal{P}_{\text{NL}}\left(A = 0|\mathbf{d}\right)}. \tag{A.63}$$

The more appropriate manner to write this expression requires the use of limits as expressed in Eq. (5.11). As outlined in the main-body, we can substitute with little risk $A = 0$ with $A = 10^{-10}$. A generalized Savage-Dickey density ratio test that makes fuller use of Eq. A.56 makes an additional correction factor to this Bayes factor [144], but it does not play a role in our analysis.

A.4 Performance of Multi-Tempered Evidence Estimators

We model the logarithm of the evidence from multi-tempered methods for each astrophysical hypothesis as a Gaussian distribution in log-likelihood, with mean $\mu_{\widehat{\ln \mathcal{Z}}}$ given by the particular multi-tempered method's point-estimate, and with standard

deviation $\sigma_{\widehat{\ln \mathcal{Z}}}$ given from Eq.(5.10). Thus we treat the logarithm of the evidence as a random variable described by a Gaussian distribution as follows:

$$p(\widehat{\ln \mathcal{Z}}) = \left(\frac{1}{\sqrt{2\pi\sigma_{\widehat{\ln \mathcal{Z}}}}}\right) \exp\left\{-\frac{\left(\ln \mathcal{Z} - \mu_{\widehat{\ln \mathcal{Z}}}\right)^2}{2\sigma_{\widehat{\ln \mathcal{Z}}}}\right\}. \tag{A.64}$$

We can see the distributions of the logarithm of the evidence for the astrophysical hypothesis on p-g mode instability for the unconstrained $\delta\phi$ prior in Fig. 17.

The logarithm of the Bayes factor is the difference of the logarithm of the evidence for one hypothesis and the logarithm of the evidence for another hypothesis:

$$\ln \mathcal{B}^A_B = \ln \mathcal{Z}_{\rm A} - \ln \mathcal{Z}_{\rm B} \tag{A.65}$$

However, since we treat $\ln \mathcal{Z}_{\rm A}$ as a random variable who's true value is unknown and so we must deal with the uncertainty in $\widehat{\ln \mathcal{Z}_{\rm A}}$. The logarithm of the Bayes factor then becomes the difference between two probability distribution functions. This can be solved via convolution and has been solved for the Gaussian case [145]. Thus the resultant $\widehat{\ln \mathcal{B}^A_B}$ is itself a Gaussian distribution function with mean $\mu_{\widehat{\ln \mathcal{B}^A_B}} = \mu_{\widehat{\ln \mathcal{Z}_A}} - \mu_{\widehat{\ln \mathcal{Z}_B}}$ and standard deviation $\sigma_{\widehat{\ln \mathcal{B}^A_B}} = \sqrt{\sigma^2_{\widehat{\ln \mathcal{Z}_A}} + \sigma^2_{\widehat{\ln \mathcal{Z}_B}}}$ [145]. This thus gives us the following expression for the distribution function that describes our uncertainty on the logarithm of the Bayes factor:

$$p(\widehat{\ln \mathcal{B}^A_B}) = \left(\frac{1}{\sqrt{2\pi\sigma_{\widehat{\ln \mathcal{B}^A_B}}}}\right) \exp\left\{-\frac{\left(\widehat{\ln \mathcal{B}^A_B} - \mu_{\widehat{\ln \mathcal{B}^A_B)}}\right)^2}{2\sigma^2_{\widehat{\ln \mathcal{B}^A_B}}}\right\}. \tag{A.66}$$

The expression in Eq. (A.66) is a Gaussian distribution function in $\widehat{\ln \mathcal{B}^A_B}$, but we often prefer to know the estimate on $\mathcal{B}^A_B$ and so we must transform coordinates. This transformation of coordinates, fortunately, is a well-known distribution called the log-normal distribution and it is able to be described in terms of the coordinates used in Eq. (A.66). Below we write out our log-normal probability distribution function for $\widehat{\mathcal{B}^A_B}$:

$$p(\widehat{\mathcal{B}^A_B}) = \frac{1}{\widehat{\mathcal{B}^A_B}\,\sigma_{\widehat{\ln \mathcal{B}^A_B}}} \frac{1}{2\pi} \exp\left\{-\frac{\left(\ln \widehat{\mathcal{B}^A_B} - \mu_{\widehat{\ln \mathcal{B}^A_B)}}\right)^2}{2\sigma^2_{\widehat{\ln \mathcal{B}^A_B}}}\right\}. \tag{A.67}$$

This is the implementation that we have used to represent that Bayes factor in this study. It is worth noting that for a sufficiently small standard deviation on the

logarithm of the Bayes factor, the probability of the distribution function will look approximately Gaussian in shape. One useful property of the log-normal Bayes factor distribution is that the median of the log-normal Bayes factor distribution is identical to the point-estimate Bayes factor, $\mathcal{B}^A_B = \exp\left[\ln \mathcal{Z}_{\rm A} - \ln \mathcal{Z}_{\rm B}\right]$. Note, that the expectation value (mean) of $\mathcal{B}^A_B$ is always right-skewed of the median, while the mode of the distribution is left-skewed relative to the median. Large standard deviations on the logarithm of the evidence will create very long tails for the distribution of the Bayes factor, which can make decision-making on Bayes factors more risky. Future studies should consider limiting the error on the logarithm of the evidence to mitigate the larger error on the Bayes factor.

Our Bayes factor estimation from 6 multi-tempered estimators on the logarithm of the Bayes factor can be seen in Fig. 18 when comparing the hypothesis on p-g mode instability for the unconstrained $\delta\phi$ prior to the hypothesis presented in [3] for the uniform mass prior with a common equation of state constraint. The different methods appear to give similar probability distributions on the Bayes factor. Those estimators with large standard deviations in $\log \mathcal{B}$ have tails that skew towards a Bayes factor of unity.

Another approach to Bayes factor uncertainty estimation that is less mathematical and easier to gain wrap one's head around is to draw random samples from the Gaussian distribution for the logarithm of the evidence for a particular hypothesis. Subtract the random samples from one another to generate a set of samples for the logarithm of the Bayes factor. Then finally exponentiate the random samples to arrive at a set of samples for the Bayes factor. Finally, create a histogram of the samples to estimate the underlying uncertainty in our Bayes factor estimation. For a sufficiently large number of samples ($\sim \mathcal{O}(10^6)$) the estimate will be very close to the log-normal distribution. We have rigorously checked this through both methods and they agree very well.

The Bayes factors for all hypotheses using all of the multi-tempered methods can be seen in Table 4. The median values of the Bayes factors range between roughly 0.63 and 0.76, with the $5^{\rm th}$ and $95^{\rm th}$ percentile interval being around ± 0.1 with a skew towards Bayes factors of 1. Under a binary choice between the p-g mode instability model and the corresponding model without p-g mode instability we can calculate a posterior probability of one choice over the other choice. Without giving preference

to either model, we can calculate a posterior probability as $p^{\rm NL}_{\rm !NL} = \mathcal{B}^{\rm NL}_{\rm !NL}/(1+\mathcal{B}^{\rm NL}_{\rm !NL})$. Thus, the Bayes factor of 0.63 corresponds to a posterior probability of 39 % and a Bayes factor of 0.76 corresponds to a posterior probability of 43 %. If we consider the $\pm$ 0.1 ends of the Bayes factor estimation, a Bayes factor of 0.53 can be interpreted as a posterior probability of 34% probability, while a Bayes factor of 0.86 can be interpreted as the model having a 46 % posterior probability. If we consider all models collectively, the posterior probability on any one, particular model reduces significantly.

Hypothesis Tested	$\mathcal{B}^{\rm NL}_{\rm !NL}(A)$	$\mathcal{B}^{\rm NL}_{\rm !NL}(B)$	$\mathcal{B}^{\rm NL}_{\rm !NL}(C)$	$\mathcal{B}^{\rm NL}_{\rm !NL}(D)$
H_1 (Uniform Mass, A, n, $f_0 \in$ (15, 100) Hz), $\delta\phi > 0.1$	$0.63^{+0.08}_{-0.07}$	$0.63^{+0.08}_{-0.07}$	$0.64^{+0.07}_{-0.06}$	$0.64^{+0.05}_{-0.05}$
H_2 (Gaussian Mass, A, n, $f_0 \in$ (15, 100) Hz), $\delta\phi > 0.1$	$0.71^{+0.08}_{-0.07}$	$0.71^{+0.08}_{-0.07}$	$0.70^{+0.06}_{-0.06}$	$0.70^{+0.04}_{-0.04}$
H_3 (Uniform Mass, A, n, $f_0 \in$ (15, 800) Hz), $\delta\phi > 0.1$	$0.64^{+0.09}_{-0.08}$	$0.64^{+0.09}_{-0.08}$	$0.64^{+0.08}_{-0.07}$	$0.64^{+0.06}_{-0.06}$
H_4 (Gaussian Mass, A, n, $f_0 \in$ (15, 800) Hz), $\delta\phi > 0.1$	$0.76^{+0.08}_{-0.07}$	$0.76^{+0.08}_{-0.07}$	$0.75^{+0.06}_{-0.06}$	$0.75^{+0.04}_{-0.04}$
H_5 (Uniform Mass, A, n, $f_0 \in$ (10, 100) Hz)	$0.68^{+0.12}_{-0.11}$	$0.68^{+0.13}_{-0.11}$	$0.69^{+0.11}_{-0.1}$	$0.69^{+0.1}_{-0.09}$

Table 4: The various Bayes factors under different multi-tempered integration methods. The column marked with $\mathcal{B}^{\rm NL}_{\rm !NL}(A)$ is the Bayes factor under the thermodynamic integration method using the trapezoid quadrature rule. The (B) column is the Bayes factor from the thermodynamic integration method using the higher-order trapezoid quadrature rule. The (C) column is the Bayes factor from the thermodynamic integration method using Simpson's quadrature rule. The (D) column is the Bayes factor for the thermodynamic integration method using Simpson's higher-order quadrature rule. The (E) column is the Bayes factor for the thermodynamic integration method using a cubic polynomial quadrature rule. And (F) is the Bayes factor from the steppingstone method. The $50^{\rm th}$ percentile with the $5^{\rm th}$ and $95^{\rm th}$ percentiles in the plus and minus superscripts and subscripts, respectively, are shown above.

A.5 Performance of Savage-Dickey Density Tests

Below we will enumerate the methods used to conduct the Savage-Dickey density test for finding the Bayes factor of the on *p-g* mode instability for the unconstrained $\delta\phi$ prior compared to a null-hypothesis, that is a hypothesis where *p-g* mode instability

is not modeled. Formally, this is the model presented in [3] for the uniform mass prior with a common equation of state constraint however the Savage-Dickey density test makes no use of the data from [3]. While this may sound somewhat outrageous, we will find that the results generally agree with those found using multi-tempered model selection techniques. The results of all of the Savage-Dickey density tests are collected and summarized in Table 5.

A.5.1 Histogram Method

An elementary method for estimating the probability density function is to histogram the samples and normalize the histogram to integrate to unity. Under this methodology the only relevant parameter to fitting the histogram to the data is the choice of bin-width, sometimes called bandwidth.

Two methods that we have used to try to maximize the fit of the histogram to the data are through the choice of plug-in bin-width algorithms. The algorithms are designed to minimize the error to the histogram density fit to an underlying density function. The first is Scott's rule [117] and is considered optimal relative to a density function that is normally distributed, and hence the bin-width h is defined as:

$$h = \frac{3.5\,\hat{\sigma}}{N^{1/3}}. \tag{A.68}$$

Here $\hat{\sigma}$ represents the sample standard deviation of the data, and N represents the number of samples in the data.

The second method is described in [118] and makes use of the interquartile range (IQR). The IQR is the difference between the 75^{th} and 25^{th} percentile of the data. The Freedman-Diaconis bin method [118], where the bin-width h is:

$$h = \frac{2\,\text{IQR}}{N^{1/3}} \tag{A.69}$$

Here N represents the number of samples in the data.

Since the marginal prior (posterior) distribution functions on A are distributed logarithmically, it is convenient to do the density estimation in the variable $\tilde{X} \equiv \log_{10} A$. Under this change of variables the marginal prior distribution on $\tilde{X}$ is uniform between $\tilde{X} = -10$ and $\tilde{X} = -5.5$, hence the distribution function is constant over all values of $\tilde{X}$:

$$\pi\left(\tilde{X}\right) = \frac{1}{\tilde{X}_{max} - \tilde{X}_{min}} = 0.22, \qquad -10.0 \leq \tilde{X} \leq -5.5 \tag{A.70}$$

Using the histogram bin-width rules from above we estimate the marginal posterior probability density at $\tilde{X} = -10.0$. To simulate the variability of the algorithm to the data we resample from the the marginal posterior probability density distribution of $\log_{10} A$ via the bootstrap method some $5,000$ times. This gives us a distribution of possible Bayes factor values. We find a $\mathcal{B}^{\rm NL}_{\rm !NL} = 0.66^{+0.08}_{-0.07}$ at the 90% confidence interval for Scott's Rule and a $\mathcal{B}^{\rm NL}_{\rm !NL} = 0.66^{+0.08}_{-0.06}$ at the 90% confidence interval for the Freedman-Diaconis Rule. For a comparison of the density estimates for the marginal posterior probability density on $\log_{10} A$ see Fig. 19. The Bayes factor uncertainty distribution is shown, in comparison to the other density estimators and the thermodynamic integration method uner the higher-order trapezoidal rule, in Fig. 20.

A.5.2 Gaussian Kernel Density Estimator

We use a Gaussian kernel density estimator available in the Python package GetDist [119]. GetDist is a Python package intended to accurately estimate the underlying one-dimensional and two-dimensional posterior probability distribution functions from a Bayesian MCMC analysis. A rough understanding of a Gaussian kernel density estimator is that it uses small truncated-Gaussian distributions centered at samples of the data and combines the sum of the distributions into a smooth probability distribution function.

The advantage that GetDist offers over other Gaussian kernel density estimators is that it comes with robust linear-boundary bias corrections to the Gaussian kernel. Sharp boundaries on the distribution function are known to cause bias to the probability distribution function estimation for Gaussian kernel density estimators [119], and the Savage-Dickey density ratio method in this application requires us to know the density of the posterior distribution function at the boundary. GetDist uses an asymmetric linear addition to the kernel at the boundaries which helps to ameliorate some of the boundary bias. There are additional bias-correction and bandwidth optimization algorithms in the routine that help improve the accuracy of the density estimation. See [119] for the full details.

We follow the same procedure in estimating the posterior probability density of A at 10^{-10} as in the histogram method. We resample the posterior distribution through the bootstrap method to generate $5,000$ estimates of $\mathcal{P}(A = 10^{-10}|\mathbf{d})$. This then yields an estimate of the Bayes factor at the 90% confidence level of $\mathcal{B}^{\rm NL}_{\rm !NL} = 0.66^{+0.13}_{-0.1}$.

For a comparison of the density estimates for the marginal posterior probability density on $\log_{10} A$ see Fig. 19. The Bayes factor uncertainty distribution is shown, in comparison to the other density estimators and the thermodynamic integration method under the higher-order trapezoidal rule, in Fig. 20.

A.5.3 Logspline Density Estimator

In the logspline density estimator of [116] a univariate log-probability density is modeled by a cubic spline. The algorithm places knots of a cubic spline in an algorithmic fashion and uses an internal likelihood function to find a maximum likelihood number of knots (and placement) to use. Internal to the software package is a Bayesian model-selection routine based on the Akaike Information Criterion (AIC) [146] and the Bayesian Information Criterion (BIC) [147] to both ensure goodness of fit and to avoid over-fitting to the data. The details of the procedure are sophisticated and beyond the scope of this study. We utilize the maximum likelihood fit to the probability distribution function from the packages' model selection routine. We note that using the maximum likelihood fit is not very risky as the likelihood for other values of knot and knot placement as given from the package's fit routine provide posterior density estimates that are almost identical to the maximum likelihood fit.

We use the same bootstrap method outlined above to estimate the variability of the logspline algorithm relative to the variability of the data. We find an estimate of the Bayes factor at the 90% confidence level of $\mathcal{B}^{\rm NL}_{\rm !NL} = 0.63^{+0.06}_{-0.05}$. For a comparison of the density estimates for the marginal posterior probability density on $\log_{10} A$ see Fig. 19. The Bayes factor uncertainty distribution is shown, in comparison to the other density estimators and the thermodynamic integration method under the higher-order trapezoidal rule, in Fig. 20.

A.6 Can we improve the chirp mass measurement with an independent EM Observation?

We noted in the introduction that we would require a strong constraint on the chirp mass independent of the gravitational wave data to mitigate the parameter degeneracy from the p-g mode instability. Here we make a rough quantitative analysis of how tight an electromagnetic observation would have to constrain the chirp mass. To do

Hypothesis Tested	$\mathcal{B}^{\rm NL}_{\rm !NL}$(FD)	$\mathcal{B}^{\rm NL}_{\rm !NL}$(Scott)	$\mathcal{B}^{\rm NL}_{\rm !NL}$(Gaussian KDE)	$\mathcal{B}^{\rm N}_{\rm !N}$
H_5 (Uniform Mass, A, n, $f_0 \in (10, 100)\ Hz$)	$0.66^{+0.08}_{-0.07}$	$0.66^{+0.08}_{-0.07}$	$0.66^{+0.13}_{-0.1}$	

Table 5: The various Bayes factors from the Savage-Dickey Density Ratio test under different density estimators. The column marked with $\mathcal{B}^{\rm NL}_{\rm !NL}$(FD) is the Bayes factor from the Freedman Diaconis histogram binning rule. The $\mathcal{B}^{\rm NL}_{\rm !NL}$(Scott) column is the Bayes factor estimate under Scott's histogram binning rule. The $\mathcal{B}^{\rm NL}_{\rm !NL}$(Gaussian KDE) column is the Bayes factor estimate when using the Gaussian kernel density estimator with linear boundary bias corrections as found in the GetDist Python package. The column denoted as $\mathcal{B}^{\rm NL}_{\rm !NL}$(Logspline) is the Bayes factor estimate when using the logspline density estimator. The $50^{\rm th}$ percentile with the $5^{\rm th}$ and $95^{\rm th}$ percentiles in the plus and minus superscripts and subscripts, respectively.

so we must consider the joint posterior distribution, $\mathcal{P}(\mathcal{M} \,|\, \mathbf{d_{GW}}, \mathbf{d_{EM}})$, from two statistically independent data sets, the gravitational wave data $\mathbf{d_{GW}}$, and a mock electromagnetic data set $\mathbf{d_{EM}}$. We must then define a (hyper) prior on the chirp mass that bridges the measured likelihood of the chirp mass for each data set to the posterior distribution. This expression looks as follows:

$$\mathcal{P}(\mathcal{M} \,|\, \mathbf{d_{GW}}, \mathbf{d_{EM}}) = \frac{\pi(\mathcal{M})}{\mathcal{Z}(\mathbf{d_{GW}}, \mathbf{d_{EM}})}\, \mathcal{L}(\mathbf{d_{GW}}, \mathbf{d_{EM}} \,|\, \mathcal{M}) = \frac{\pi(\mathcal{M})}{\mathcal{Z}(\mathbf{d_{GW}}, \mathbf{d_{EM}})}\, \mathcal{L}(\mathbf{d_{GW}} \,|\, \mathcal{M})\, \mathcal{L}(\mathbf{d_{EM}} \,|\, \mathcal{M}). \tag{A.71}$$

The separation of $\mathcal{L}(\mathbf{d_{GW}}, \mathbf{d_{EM}} \,|\, \mathcal{M}) = \mathcal{L}(\mathbf{d_{GW}} \,|\, \mathcal{M})\, \mathcal{L}(\mathbf{d_{EM}} \,|\, \mathcal{M})$ follows from the two being statistically independent measurements of the chirp mass of GW170817. Here, $\mathcal{Z}(\mathbf{d_{GW}}, \mathbf{d_{EM}})$ is the normalizing constant that maintains the equality, which is easy to solve for computationally for a one parameter model. We will call this normalizing constant c from now on. Finding the marginal likelihood of $\mathcal{L}(\mathbf{d_{GW}} \,|\, \mathcal{M})$ given all of the parameters in the analysis would be a prohibitively difficult difficult to construct without the MCMC methods described here for calculating the likelihood marginalized over all parameters. However, an application of Bayes' theorem reduces this problem to one that we can solve from the data in hand. Consider that $\mathcal{L}(\mathbf{d_{GW}} \,|\, \mathcal{M}) = \mathcal{P}(\mathcal{M} \,|\, \mathbf{d_{GW}})/\pi_{\rm GW}(\mathcal{M})$ for the properly normalized marginal posterior and prior distributions on the chirp mass.

$$\mathcal{P}(\mathcal{M}\,|\,\mathbf{d_{GW}},\mathbf{d_{EM}}) = \frac{\pi(\mathcal{M})}{c} \times \frac{\mathcal{P}(\mathcal{M}\,|\,\mathbf{d_{GW}})}{\pi_{\rm GW}(\mathcal{M})} \times \frac{\mathcal{P}(\mathcal{M}\,|\,\mathbf{d_{EM}})}{\pi_{\rm EM}(\mathcal{M})} \tag{A.72}$$

Now, since we don't have a real posterior distribution from electromagnetic data on hand we make a simple assumption, $\mathcal{P}(\mathcal{M}|\mathbf{d_{EM}})$. We let the hyper prior $\pi(\mathcal{M})$ equal to the $\pi_{\rm GW}(\mathcal{M})$ and equal to our mock-analysis $\pi_{\rm EM}(\mathcal{M})$. These priors are uniform in chirp mass in the detector fram between $\mathcal{M} \in (1.1876, 1.2076)$. Now, let $\mathcal{P}(\mathcal{M}|\mathbf{d_{EM}})$ be a Gaussian distribution with mean value μ equivalent to the gravitational wave posterior mode of the chirp mass given a non p-g mode hypothesis. Now, we ask what is the standard deviation σ of $\mathcal{P}(\mathcal{M}|\mathbf{d_{EM}})$ required to regulate the p-g mode inferred marginal posterior distribution on the chirp mass so that it *looks* more like the marginal posterior distribution on the chirp mass from the null hypothesis? We solve this computationally for the unconstrained prior on $\delta\phi$ in the *p-g* mode analysis and for the uniform mass distribution with common equation of state constraint from [3]. The normalizing constant c is solved via a fine-grid trapezoidal rule to normalize the posterior to unity. The net result is in Fig. 21 where we find that an electromagnetic observer would need a constraint on $\sigma_{\mathcal{M}} < 0.0001 M_{\odot}$. This corresponds to a measurement error less than 0.008 %, well outside the realm of current methods.

One might improve on this simple heuristic by using the marginal chirp mass distribution when marginalizing over all *p-g* mode models and then comparing it to the marginal chirp mass distribution when marginalizing over all models in [3]. But the result would be qualitatively identical and so for simplicity we neglect this possible improvement.

A.7 Will there ever be a time when we can rule out the *p-g* mode instability?

In this section we investigate whether we can make a projection about whether future events will ever permit us to rule out the *p-g* mode instability. We will have to make many assumptions in this section regarding the properties of the model, the events that we see, and the rate of mergers for binary neutron stars.

Our goal is to take advantage of the fact that Bayes factors for hypotheses are multiplicative across statistically independent events. That is to say, that with more

binary neutron star events we can accumulate evidence for or against p-g mode instability through continuous testing of these hypotheses on the individual events. The Bayes factors for N events can be combined into one Bayes factor via following expression:

$$\mathcal{B}(\mathbf{d_1}, \mathbf{d_2}, \ldots, \mathbf{d_{N-1}}, \mathbf{d_N} \,|\, \mathrm{H_{NL}}, \mathrm{H_{!NL}}) = \prod_{i=1}^{N} \mathcal{B}(\mathbf{d_i} \,|\, \mathrm{H_{NL}}, \mathrm{H_{!NL}}). \tag{A.73}$$

Here the hypotheses for p-g mode instability is denoted as $\mathrm{H_{NL}}$, while the null hypothesis is denoted $\mathrm{H_{!NL}}$. Note that inference with Bayes factors is equivalent to a Frequentist inference based on the likelihood rather than inference on a posterior probability. We multiply the Bayes factor by a $50-50$ prior odds ratio, $\left(\frac{\pi(\mathrm{H_{NL}})}{\pi(\mathrm{H_{!NL}})}\right)$, effectively stating no preference for either hypothesis, to convert the Bayes factor to a posterior odds ratio $\mathcal{O}$. Doing so permits us to consider the posterior odds ratio as equivalent to the Bayes factor:

$$\mathcal{O}(\mathrm{H_{NL}}, \mathrm{H_{!NL}} \,|\, \mathbf{d_1}, \mathbf{d_2}, \ldots, \mathbf{d_{N-1}}, \mathbf{d_N}) = \frac{\pi(\mathrm{H_{NL}})}{\pi(\mathrm{H_{!NL}})} \times \mathcal{B}(\mathbf{d_1}, \mathbf{d_2}, \ldots, \mathbf{d_{N-1}}, \mathbf{d_N} \,|\, \mathrm{H_{NL}}, \mathrm{H_{!NL}}). \tag{A.74}$$

Here $\mathcal{O}(\mathrm{H_{NL}}, \mathrm{H_{!NL}} \,|\, \mathbf{d_1}, \mathbf{d_2}, \ldots, \mathbf{d_{N-1}}, \mathbf{d_N})$ is the posterior odds ratio, and thus setting $\frac{\pi(\mathrm{H_{NL}})}{\pi(\mathrm{H_{!NL}})}$ equal to unity makes the posterior odds ratio equivalent to the Bayes factor. We will use the Bayes factor instead of the odds ratio for the duration of this section, with the understanding that they are equivalent in this scenario.

It is instructive to examine the behavior of the cumulative logarithm of the Bayes factor for the incredible case that the next several binary neutron stars are identical in signal-to-noise ratio and intrinsic properties as GW170817. Here we consider two estimators for the Bayes factor, the thermodynamic integration method which we found to have a log Bayes factor of $\mu \sim -0.38$, and at worst $\sigma \sim 0.1$, and the logspline estimator with the Savage Dickey density ratio which we found to have a log Bayes factor of $\mu \sim -0.46, \sigma \sim 0.06$. We note that the logspline estimate is not a log-normal distribution, but it this estimate is close enough to the estimates from our analysis for demonstrative purposes. The analysis of [2] found a log Bayes factor of $0.03^{+0.70}_{-0.58}$ at 90% confidence using the Savage-Dickey density ratio. We model this as a Gaussian distribution in the logarithm Bayes factor with $\mu = 0.03, \sigma = 0.4$ so as to have a similar 90% interval width. While for GW170817 these Bayes factor estimates are compatible, if we measure a new GW170817-like binary neutron star and

measure the same Bayes factor for this new event, the cumulative Bayes factor and the uncertainty surrounding this cumulative Bayes factor propagates multiplicatively and quickly begin to exclude each other as more events are aggregated.

To illustrate this consider the case that our MCMC methods estimate the logarithm of Bayes factors for some fixed choice of prior distribution and for cumulative gravitational wave events. Consider the case that the estimator of the logarithm of the Bayes factor is a normal distribution with mean (point-estimate) μ and a standard deviation (uncertainty) σ_i:

$$\widehat{\ln \mathcal{B}^{\rm NL}_{\rm !NL}}(\mathbf{d_i}) = \mathcal{N}(\mu_i, \sigma_i). \tag{A.75}$$

Thus, the cumulative Bayes factor for many neutron star events becomes:

$$\begin{aligned} \widehat{\ln \mathcal{B}^{\rm NL}_{\rm !NL}}(\mathbf{d_1}, \mathbf{d_2}, ..., \mathbf{d_N}) \quad &= \sum_{i=1}^{N} \widehat{\ln \mathcal{B}^{\rm NL}_{\rm !NL}}(\mathbf{d_i}) \\ &= \mathcal{N}\left(\mu = \sum_{i=1}^{N} \mu_i,\ \sigma = \sqrt{\sum_{i=1}^{N} \sigma_i^2}\right). \end{aligned} \tag{A.76}$$

We note that if the Bayes factor point-estimate term μ_i is monotonic across events, and the uncertainty estimate from our MCMC methods σ_i is usually consistent, then the estimator $\widehat{\ln \mathcal{B}^{\rm NL}_{\rm !NL}}(\mathbf{d_1}, ..., \mathbf{d_N})$ will tend towards a log Bayes factor that is statistically significant, for which a decision on whether nonlinear tidal effects from a p-g mode instability is present / measurable in neutron stars. For repetitions of the same event we see that the mean of this Gaussian distribution for the cumulative log Bayes factor grows linearly with mu_i, and the uncertainty grows as $\sqrt{N}\sigma_i$. However, if noise elements in the MCMC or from the data itself dilute the ability for our MCMC estimator to find a $\mu \neq 0$ or a monotonic measurement of μ then the logarithm of the Bayes factor will become overly diluted by a growing uncertainty σ. As seen in Fig. 22, after five events, the cumulative Bayes factor estimations diverge significantly. They may be caused by different methodologies in Bayes factor estimation, by waveform systematics, by power-spectral density estimation differences, by different segments of GPS-analysis times used, the low-frequency cutoff used, or by many other possible variables that are not accounted for in this comparison. As MCMC methods for estimating Bayes factor improve this error in cumulative Bayes factor estimation should also improve.

A more realistic approach is to consider that the distribution of signal-to-noise ratio for binary neutron star events will not all be repeats of GW170817 but rather will follow some other distribution. To do so explore this we need to model the expected signal-to-noise-ratio, ρ, of events that we expect to see with gravitational wave observatories. Fortunately, this work has already been done in [148, 149]. We can expect that for a network of interferometers with a signal to noise ratio detection threshold of $\rho_{\text{threshold}}$ that our distribution will follow the rule:

$$p(\rho) = 3\,\frac{\rho_{\text{threshold}}^3}{\rho^4}. \tag{A.77}$$

This expression is a normalized probability distribution function in so far as we only permit $\rho > \rho_{\text{threshold}}$ and let ρ go to positive infinity. Given this probability distribution we can expect our average ρ to be equal to $\frac{3}{2}\rho_{\text{threshold}}$. If we assume a very conservative $\rho_{\text{threshold}} = 11$, then the probability of attaining gravitational wave neutron star mergers as loud as or louder than GW170817 ($\rho \approx 34$) is slightly higher than 3 %. At a signal-to-noise ratio of ~ 34 we have found that the p-g mode instability hypothesis has a Bayes factor of approximately 1, and we expect that 97% of neutron star detections will be quieter than GW170817 for which the Bayes factor will in all likelihood be unity as well due to parameter degeneracies. Thus we see that in the long-term we may have to wait for hundreds of binary neutron star mergers to be able to make a decision on the validity of the p-g mode instability hypothesis, for-or-against. Biases and difficulties in the MCMC method for hypothesis testing, the noise in the detector, and waveform systematics in all likelihood will make this endeavor all the more difficult.

And so to answer the question of the subsection, "Will there ever be a time when we can rule out p-g mode instability?", depends on a number of factors. Firstly, what do we mean when we say the p-g mode instability, e.g., which priors, which waveform model, with which MCMC method, and under which assumptions? We have ruled out some sorts of p-g mode instability in this paper and the marginal posterior densities from [2] similarly tell us that some p-g mode parameters are not favored by the data. and (b) whether the theory has yielded any new or interesting observations or features of the data that cannot be explained by another model. In all likelihood the theory will dissolve into obscurity, much as it currently stands today. While the p-g mode hypothesis is perhaps not very interesting in its own right, it shows

the ambiguities and difficulties inherent in Bayesian gravitational wave astrophysical analysis. We hope that our exploration and explanation of the problem prompts greater thoughtfulness with respect to statistical analysis and experimental design.

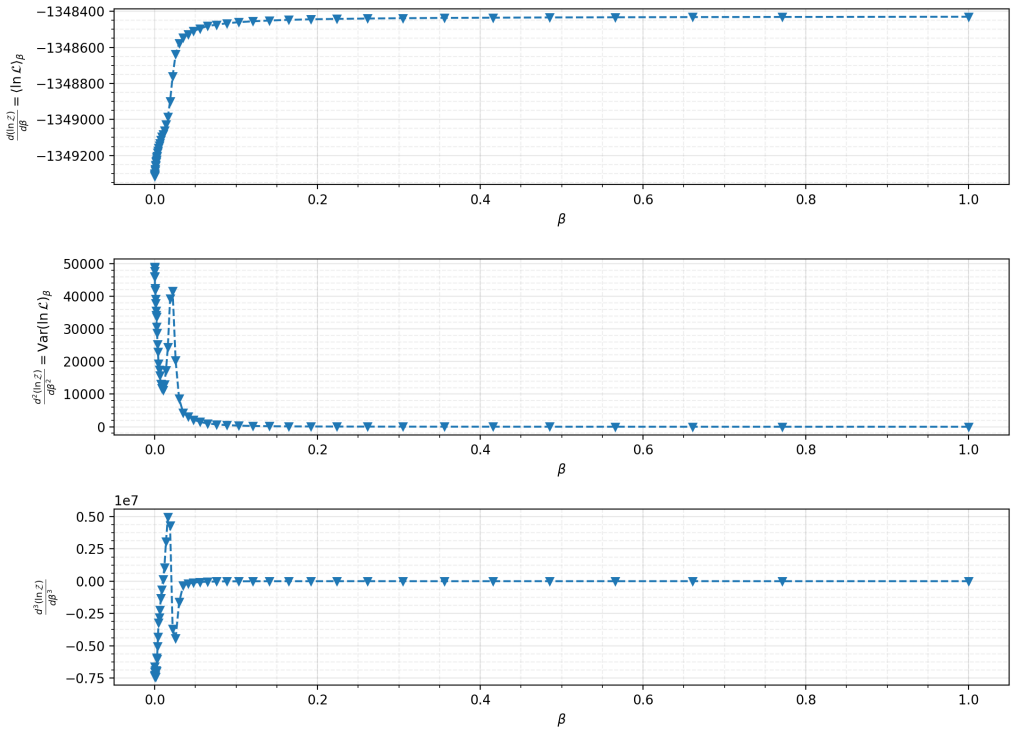

Figure 11: The subplots of the thermodynamic integrand and subsequent derivatives of the thermodynamic integral. (*Top*) The thermodynamic integrand when compared to the inverse-temperature β. The curve should be smooth and montonic, however it is very difficult to inspect the integrand on a linear β scale. (*Middle*) The second derivative of the logarithm of the evidence is the variance of the power-posterior at an inverse temperature β. There is some indication that an inflection point happens in the curvature of the integrand at high temperature. (*Bottom*) The third derivative of the logarithm of the evidence is also the third-order cumulant of the power-posterior distributions at an inverse-temperature β. It is difficult to inspect the behavior of this derivative on the linear β scale.

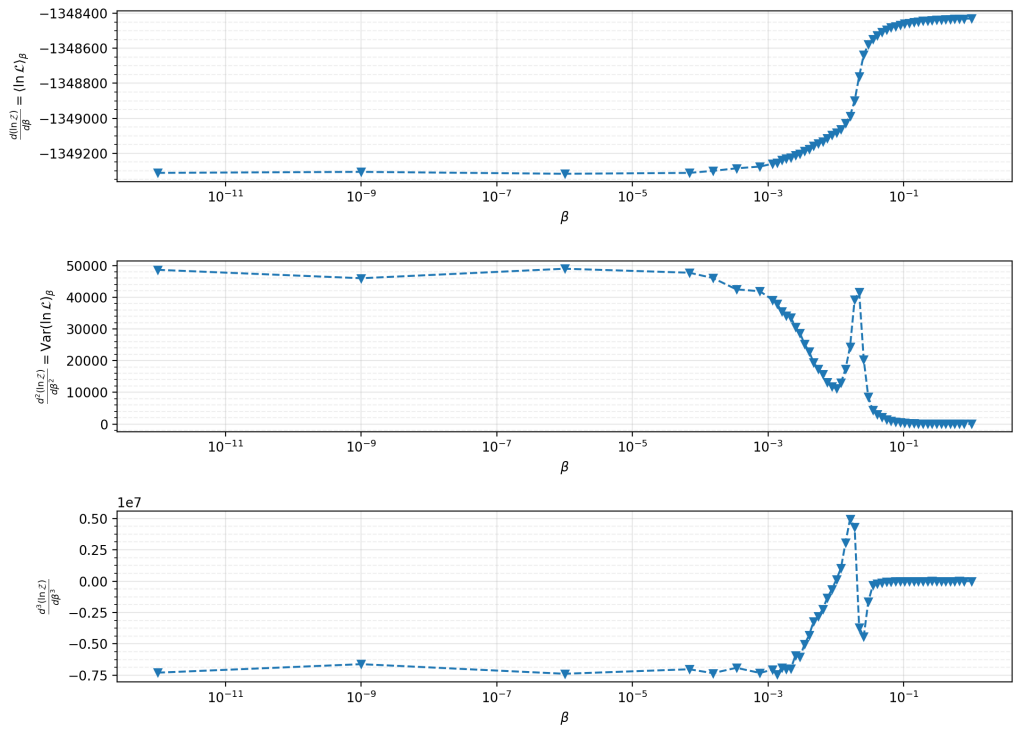

Figure 12: The subplots of the thermodynamic integrand and subsequent derivatives of the thermodynamic integral. (*Top*) The thermodynamic integrand when compared to the inverse-temperature β. The curve should be smooth and montonic, however there is some indication at $\beta = 10^{-9}$ that this condition is not strictly met in the Markov Chain Monte Carlo simulation. (*Middle*) The second derivative of the logarithm of the evidence is the variance of the power-posterior at an inverse temperature β. This function should also be smooth however there is some indication that at high temperature that the derivatives are not stable. (*Bottom*) The third derivative of the logarithm of the evidence is also the third-order cumulant of the power-posterior distributions at an inverse-temperature β. Here we can see that the derivatives are not very sable or smooth. This may motivate moving our analysis to new multi-tempered samplers that are optimized for thermodynamic integration.

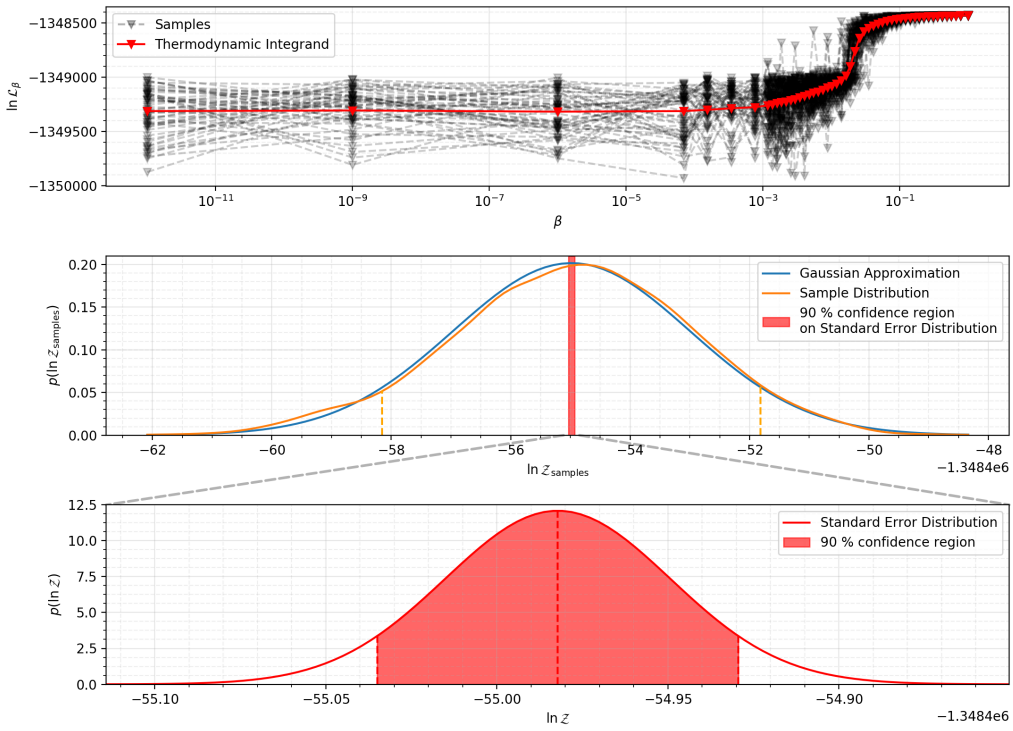

Figure 13: The first subplot denotes the untempered log-likelihood samples when drawn from the power-posteriors at β. The expectation value of the untempered log-likelihood when drawn from these power-posteriors is the thermodynamic integrand and is plotted in red. The thermodynamic integral over all geometric paths given from the samples is drawn in the second subplot. The sample-log-integral distribution is approximately a Gaussian distribution. The standard error of the mean value of the log evidence is given by the sample standard deviation divided by the square root of the number of samples. The 90% confidence interval on the sample distribution in the log-evidence is drawn in dashed orange lines. The 90% confidence region from this standard error is shaded in red. The final subplot is a zoom-in on this 90% confidence region showing the error estimate on the thermodynamic integral due to Monte Carlo sampling. This, when combined with the convergence error, is used in the final error estimate on the log-evidence of the thermodynamic integral.

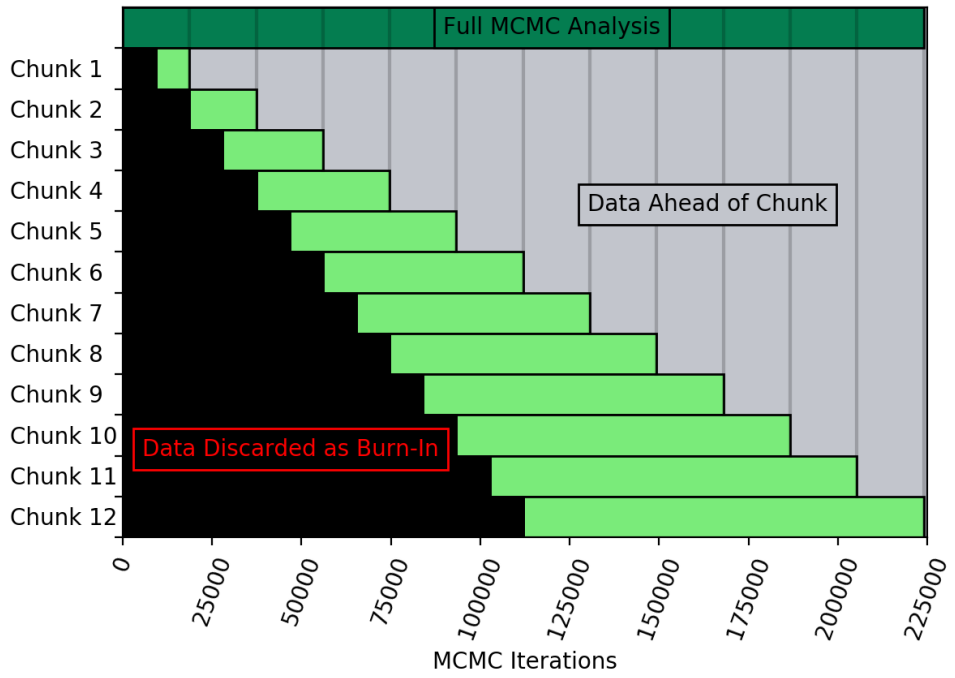

Figure 14: The partitioning of the MCMC analysis to check on the convergence of the thermodynamic integrand and the thermodynamic integration. The dark-green bar at the top represents all of the samples collected by the MCMC analysis. This segment is divided into 12 segments represented by the light gray lines. The light-green segments represent chunks that independent samples can be drawn from. The dark region represents samples discarded as burn-in samples for the MCMC. The dark grey region represents data that is ahead of the chunk and thus not used in drawing independent samples for that chunk. Chunk 12 produces the identical samples as drawing independent samples according to the n_{acl} algorithm from PyCBC at the end of the analysis.

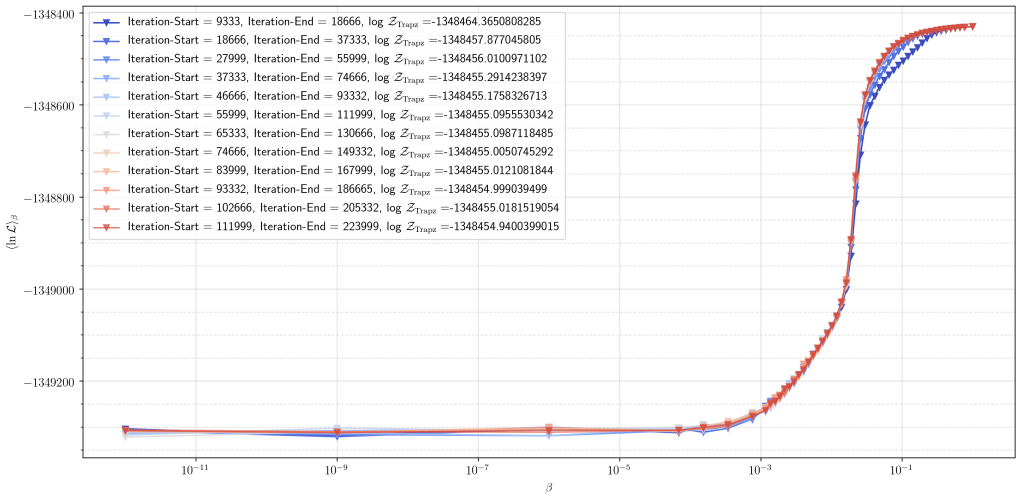

Figure 15: The convergence of the thermodynamic integrand for the unconstrained $\delta\phi$ prior choice on the p-g mode instability model. The Iteration-Start denotes the point is taken from a segment beginning with that MCMC iteration and ending with the MCMC iteration denoted as Iteration-End. These iterations correspond to the segments found in Fig. 14. The logarithm of the evidence is shown also in the figure caption, and it can be noted that as the MCMC analysis progresses the integral converges to a set value. The thermodynamic integrand can be visually seen to converge to the S-like curve seen in the figure. Early in the analysis the curve can be mishaped as the power-posteriors have not all converged. Experience has told us tha the power-posteriors that take the longest to converge tend to be in the region where the average log likelihood changes rapidly. Here this is in the region between $\beta \in (10^{-2} - 1)$.

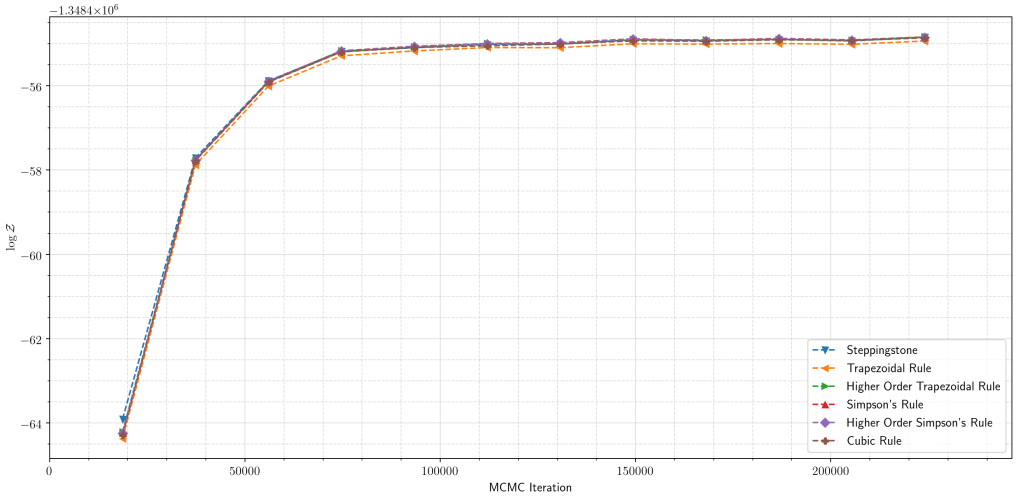

Figure 16: The convergence of the thermodynamic integral for the unconstrained $\delta\phi$ prior choice on the p-g mode instability model as a function of the MCMC iteration. These choice of points of iterations correspond to the segments found in Fig. 14. As the analysis progresses the logarithm of the evidence from all methods tend towards a fixed value.

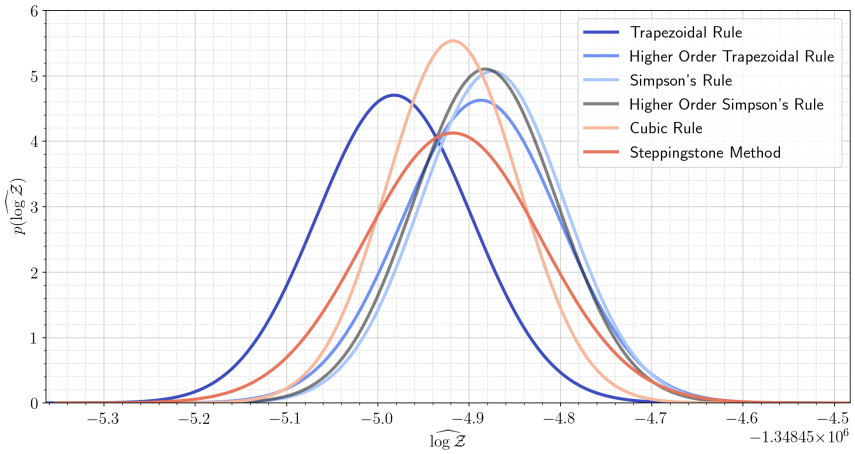

Figure 17: The estimates of the logarithm of the evidence from multi-temper evidence integration methods. We model the logarithm of the evidence as a Gaussian in log-space. These data are for the logarithm of the evidence from the unconstrained $\delta\phi$ prior for the *p*-*g* mode instability model. The trapezoidal rule estimates the lowest log evidence for this model, and the cubic rule has the smallest estimated statistical error uncertainty (the smallest confidence interval). The mean values of the higher order quadrature rules appear to be closer together to one another than they are to the trapezoidal rule.

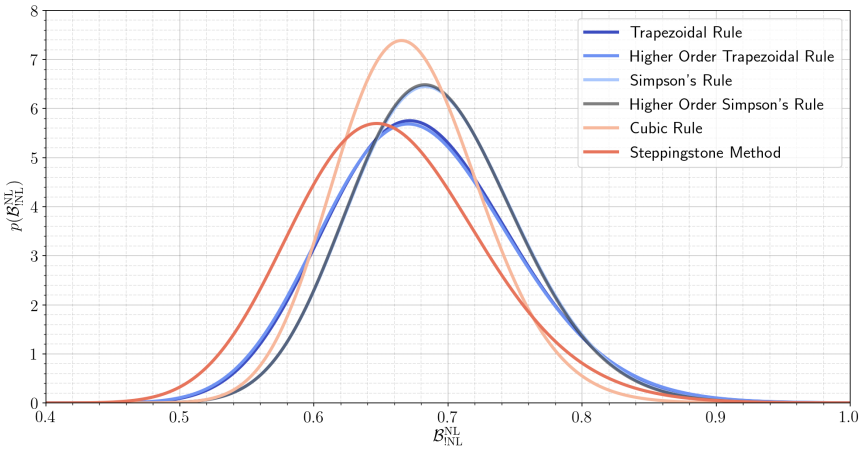

Figure 18: The distribution for the Bayes factor for nonlinear tides from p-g mode instability from the unconstrained $\delta\phi$ prior relative to the uniform mass, common equation of state prior from [3] under the assumption that the logarithm of the evidence for each model is well approximated by a Gaussian distribution. but our method is sufficiently accurate in the high-sample limit. When the uncertainty on the logarithm of the evidences in the Bayes factor estimation are sufficiently small, the Bayes factor distribution is approximately normal in shape, but formally they are log-normal distributions.

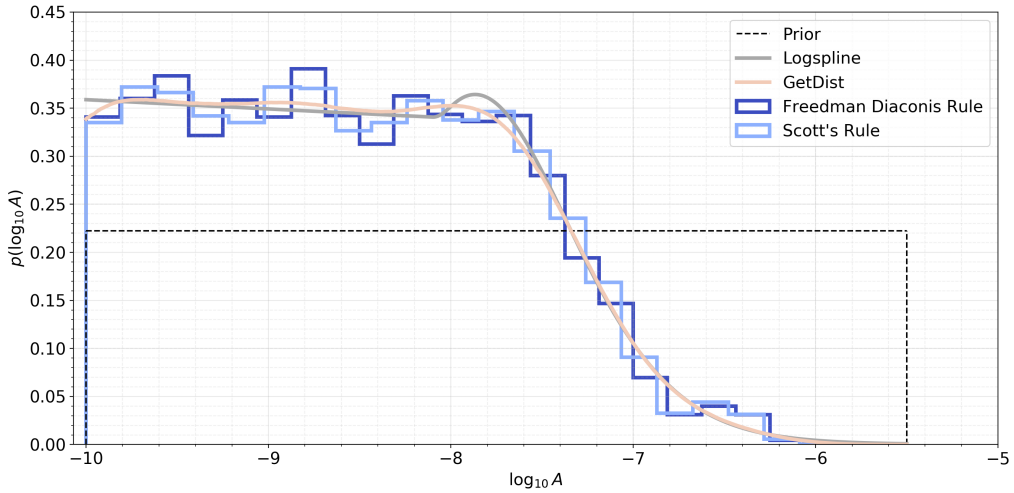

Figure 19: The prior and posterior density estimations from different density estimators for the parameter $\log_{10} A$. The prior density is uniform in $\log_{10}$ and is 0.2. Posterior Logspline is the density estimation under the logspline density estimator, and it is notable that it tends to estimate the posterior density at $\log_{10} A = -10$ more highly than the other estimators. If this is systematically true, then the Bayes factor from the logspline density estimator will be smaller than the other estimators. Posterior GetDist is the Gaussian kernel density estimator described in the Python GetDist package. Posterior FD and Scott are the histogram methods under the Freedman-Diaconis binning rule and Scott's binning rule, respectively. We can see here that there is some *wasted* prior space at large $\log_{10} A$. THe posterior density might more closely match the prior density if we restricted our prior density to smaller $\log_{10} A$.

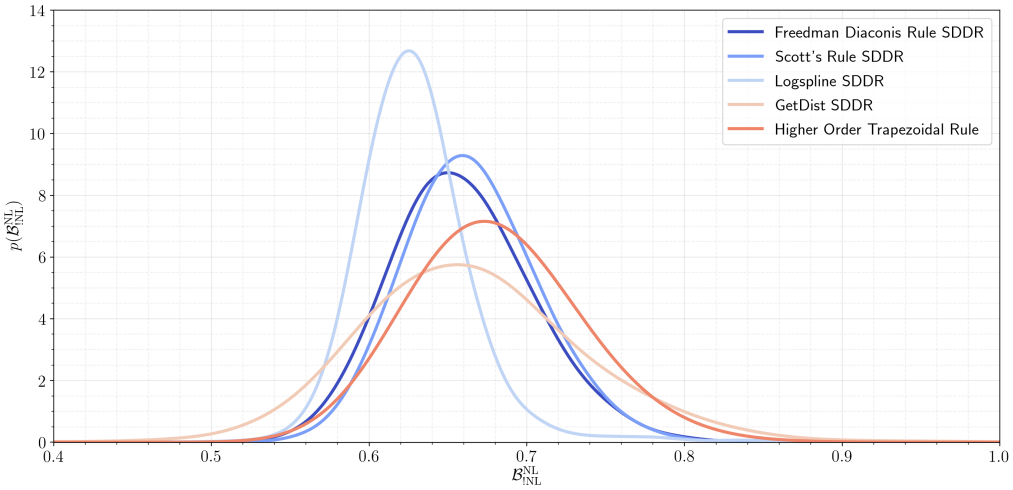

Figure 20: A comparison of the Bayes factor estimates for p-g mode instability with the permissive prior on $\delta\phi$ vs no p-g mode instability from different methods. Here, SDDR refers to the Savage Dickey density ratio test for each corresponding estimator technique. We compare these results to the higher order trapezoidal rule from thermodynamic integration. The other multi-tempered Bayes factors are comparable to the one shown here and so are not displayed. The estimates generally agree and if required we could marginalize over the method for calculating the Bayes factor to report one single value.

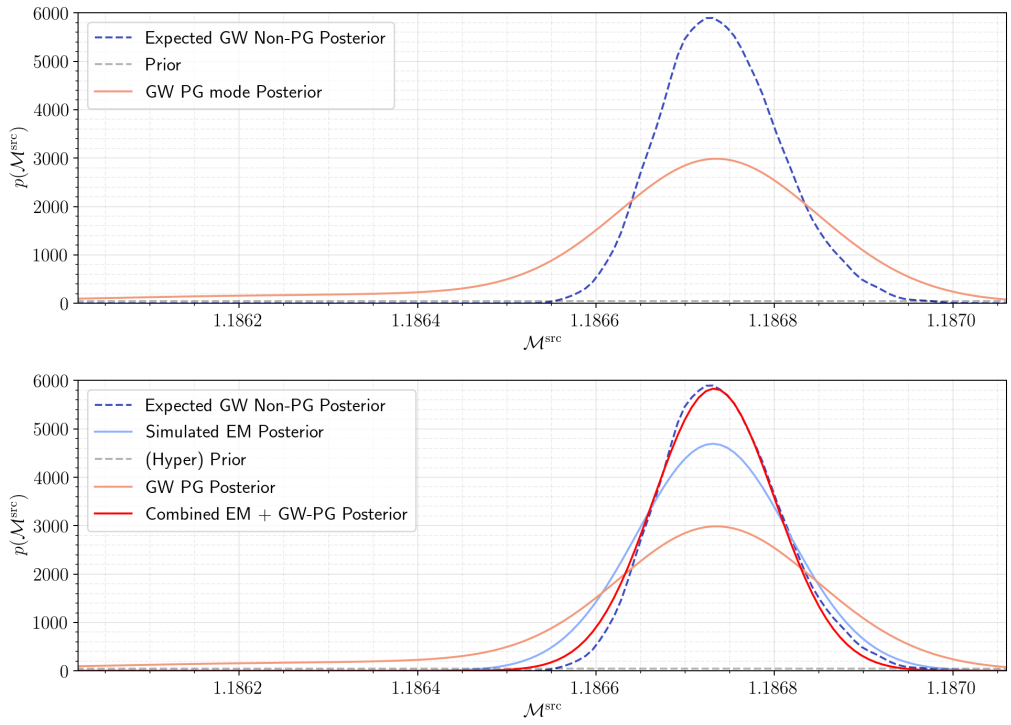

Figure 21: (*Top*) The prior distribution on the chirp mass for two gravitational wave astrophysical hypotheses. The first hypothesis is the uniform mass and constrained equation of state constraint model from [3], while the second model is the *p*-*g* mode instability hypothesis with unconstrained $\delta\phi$. The marginal posterior distributions on the chirp mass are in dashed-blue and solid, light-red, respectively. (*Bottom*) Combining a simulated Gaussian electromagnetic posterior on the chirp mass (light-blue) and a prior on the chirp mass we can combine the posterior distributions from the gravitational wave data with the *p*-*g* mode instability from the unconstrained $\delta\phi$ model with this electromagnetic posterior to construct a joint posterior distribution (solid, red) that closely matches the inferred chirp mass for GW170817 from [3]. The simulated Gaussian electromagnetic posterior has mean centered at the maximum a posteriori value from [3], $\mu = 1.186731\,M_{\odot}$, and standard deviation, $\sigma = 0.000085\,M_{\odot}$.

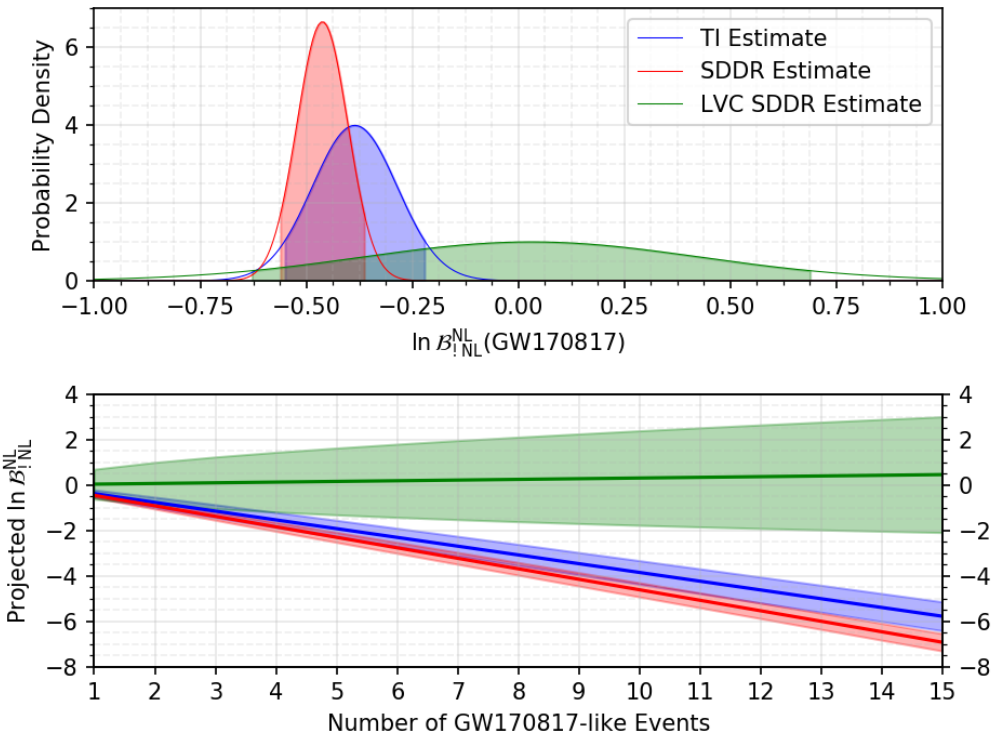

Figure 22: (*Top*) A comparison of Gaussian approximations of the logarithm of the Bayes factor using different estimators or waveform systematics. Note that the LVC estimate here is a rough Gaussian approximation based on the reported bounds in [2]. The 90% confidence regions are shaded in. Positive log Bayes factors are indicative of support for the *p*-*g* mode hypothesis, while negative log Bayes factors are indicative of support for the null hypothesis. (*Bottom*) For repeated GW170817-like binary neutron star mergers the cumulative logarithm of the Bayes factor for the *p*-*g* mode hypothesis vs the null hypothesis begin to diverge in estimation. The solid lines represent the cumulative median estimates, while the shaded regions represent the cumulative 90% confidence intervals. Waveform systematics or uncontrolled variables in the Bayes factor estimation methods may be the main driver of this divergence and future meta-analyses will have to control for these sorts of uncertainty.

Bibliography

[1] B. P. Abbott et al. Binary Black Hole Mergers in the first Advanced LIGO Observing Run. *Phys. Rev.*, X6(4):041015, 2016. [erratum: Phys. Rev.X8,no.3,039903(2018)].

[2] BP Abbott, R Abbott, TD Abbott, F Acernese, K Ackley, C Adams, T Adams, P Addesso, RX Adhikari, VB Adya, et al. Constraining the p-mode–g-mode tidal instability with gw170817. *Physical review letters*, 122(6):061104, 2019.

[3] Soumi De, Daniel Finstad, James M Lattimer, Duncan A Brown, Edo Berger, and Christopher M Biwer. Tidal deformabilities and radii of neutron stars from the observation of gw170817. *Physical review letters*, 121(9):091102, 2018.

[4] Alexander H Nitz, Collin Capano, Alex B Nielsen, Steven Reyes, Rebecca White, Duncan A Brown, and Badri Krishnan. 1-ogc: The first open gravitational-wave catalog of binary mergers from analysis of public advanced ligo data. *The Astrophysical Journal*, 872(2):195, 2019.

[5] Benjamin P. Abbott et al. Sensitivity of the Advanced LIGO detectors at the beginning of gravitational wave astronomy. *Phys. Rev.*, D93(11):112004, 2016. [Addendum: Phys. Rev.D97,no.5,059901(2018)].

[6] B. P. Abbott et al. Observation of Gravitational Waves from a Binary Black Hole Merger. *Phys. Rev. Lett.*, 116(6):061102, 2016.

[7] B. P. Abbott et al. GW151226: Observation of Gravitational Waves from a 22-Solar-Mass Binary Black Hole Coalescence. *Phys. Rev. Lett.*, 116(24):241103, 2016.

[8] Martin A. Green and J. W. Moffat. Extraction of black hole coalescence waveforms from noisy data. *Phys. Lett.*, B784:312–323, 2018.

[9] Javier Roulet and Matias Zaldarriaga. Constraints on Binary Black Hole Populations from LIGO-Virgo Detections. 2018.

[10] Javier M. Antelis and Claudia Moreno. An independent search of gravitational waves in the first observation run of advanced LIGO using cross-correlation. 2018.

[11] Benjamin P. Abbott et al. Upper Limits on the Rates of Binary Neutron Star and Neutron Starblack Hole Mergers From Advanced Ligos First Observing run. *Astrophys. J.*, 832(2):L21, 2016.

[12] B. P. Abbott et al. GW150914: First results from the search for binary black hole coalescence with Advanced LIGO. *Phys. Rev.*, D93(12):122003, 2016.

[13] Alexander H. Nitz, Thomas Dent, Tito Dal Canton, Stephen Fairhurst, and Duncan A. Brown. Detecting binary compact-object mergers with gravitational waves: Understanding and Improving the sensitivity of the PyCBC search. *Astrophys. J.*, 849(2):118, 2017.

[14] Alexander Harvey Nitz. Distinguishing short duration noise transients in LIGO data to improve the PyCBC search for gravitational waves from high mass binary black hole mergers. *Class. Quant. Grav.*, 35(3):035016, 2018.

[15] Tito Dal Canton and Ian W. Harry. Designing a template bank to observe compact binary coalescences in Advanced LIGO's second observing run. 2017.

[16] G. Ashton, E. Burns, T. Dal Canton, T. Dent, H. B Eggenstein, A. B. Nielsen, R. Prix, M. Was, and S. J. Zhu. Coincident detection significance in multimessenger astronomy. *Astrophys. J.*, 860(1):6, 2018.

[17] E. Burns et al. A Fermi Gamma-ray Burst Monitor Search for Electromagnetic Signals Coincident with Gravitational-Wave Candidates in Advanced LIGO's First Observing Run. 2018.

[18] Michele Vallisneri, Jonah Kanner, Roy Williams, Alan Weinstein, and Branson Stephens. The LIGO Open Science Center. *J. Phys. Conf. Ser.*, 610(1):012021, 2015.

[19] Samantha A. Usman et al. The PyCBC search for gravitational waves from compact binary coalescence. *Class. Quant. Grav.*, 33(21):215004, 2016.

[20] Tito Dal Canton et al. Implementing a search for aligned-spin neutron star-black hole systems with advanced ground based gravitational wave detectors. *Phys. Rev.*, D90(8):082004, 2014.

[21] Alexander H. Nitz, Ian W. Harry, Joshua L. Willis, Christopher M. Biwer, Duncan A. Brown, Larne P. Pekowsky, T. Dal Canton, Andrew R. Williamson, Thomas Dent, Collin D. Capano, Thomas J. Massinger, Amber K. Lenon, Alex B. Nielsen, and Miriam Cabero. PyCBC Software. https://github.com/gwastro/pycbc, 2018.

[22] B. P. Abbott et al. Characterization of transient noise in Advanced LIGO relevant to gravitational wave signal GW150914. *Class. Quant. Grav.*, 33(13):134001, 2016.

[23] B P Abbott et al. Effects of data quality vetoes on a search for compact binary coalescences in Advanced LIGOs first observing run. *Class. Quant. Grav.*, 35(6):065010, 2018.

[24] Benjamin P. Abbott et al. Search for intermediate mass black hole binaries in the first observing run of Advanced LIGO. *Phys. Rev.*, D96(2):022001, 2017.

[25] B. P. Abbott et al. Prospects for Observing and Localizing Gravitational-Wave Transients with Advanced LIGO and Advanced Virgo. *Living Rev. Relat.*, 19:1, 2016.

[26] Benjamin P. Abbott et al. GW170104: Observation of a 50-Solar-Mass Binary Black Hole Coalescence at Redshift 0.2. *Phys. Rev. Lett.*, 118(22):221101, 2017. [Erratum: Phys. Rev. Lett.121,no.12,129901(2018)].

[27] B.. P.. Abbott et al. GW170608: Observation of a 19-solar-mass Binary Black Hole Coalescence. *Astrophys. J.*, 851(2):L35, 2017.

[28] B. P. Abbott et al. GW170814: A Three-Detector Observation of Gravitational Waves from a Binary Black Hole Coalescence. *Phys. Rev. Lett.*, 119(14):141101, 2017.

[29] B.P. Abbott et al. GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral. *Phys. Rev. Lett.*, 119(16):161101, 2017.

[30] C. M. Biwer, Collin D. Capano, Soumi De, Miriam Cabero, Duncan A. Brown, Alexander H. Nitz, and V. Raymond. PyCBC Inference: A Python-based parameter estimation toolkit for compact binary coalescence signals. 2018.

[31] Bruce Allen, Warren G. Anderson, Patrick R. Brady, Duncan A. Brown, and Jolien D. E. Creighton. FINDCHIRP: An Algorithm for detection of gravitational waves from inspiraling compact binaries. *Phys. Rev. D*, 85:122006, 2012.

[32] Benjamin J. Owen. Search templates for gravitational waves from inspiraling binaries: Choice of template spacing. *Phys. Rev.*, D53:6749–6761, 1996.

[33] Benjamin J. Owen and B. S. Sathyaprakash. Matched filtering of gravitational waves from inspiraling compact binaries: Computational cost and template placement. *Phys. Rev. D*, 60:022002, 1999.

[34] Duncan A. Brown, Ian Harry, Andrew Lundgren, and Alexander H. Nitz. Detecting binary neutron star systems with spin in advanced gravitational-wave detectors. *Phys. Rev.*, D86:084017, 2012.

[35] Etienne Racine. Analysis of spin precession in binary black hole systems including quadrupole-monopole interaction. *Phys. Rev.*, D78:044021, 2008.

[36] P. Ajith, M. Hannam, S. Husa, Y. Chen, B. Brügmann, N. Dorband, D. Müller, F. Ohme, D. Pollney, C. Reisswig, L. Santamaría, and J. Seiler. Inspiral-merger-ringdown waveforms for black-hole binaries with nonprecessing spins. *Phys. Rev. Lett.*, 106(24):241101, Jun 2011.

[37] A. Taracchini, A. Buonanno, Y. Pan, T. Hinderer, M. Boyle, D. A. Hemberger, L. E. Kidder, G. Lovelace, A. H. Mroué, H. P. Pfeiffer, M. A. Scheel, B. Szilágyi, N. W. Taylor, and A. Zenginoglu. Effective-one-body model for black-hole

binaries with generic mass ratios and spins. *Phys. Rev. D*, 89(6):061502, March 2014.

[38] Alejandro Boh et al. An improved effective-one-body model of spinning, non-precessing binary black holes for the era of gravitational-wave astrophysics with advanced detectors. 2016.

[39] B. S. Sathyaprakash and S. V. Dhurandhar. Choice of filters for the detection of gravitational waves from coalescing binaries. *Phys. Rev. D*, 44:3819–3834, 1991.

[40] Serge Droz, Daniel J. Knapp, Eric Poisson, and Benjamin J. Owen. Gravitational waves from inspiraling compact binaries: Validity of the stationary-phase approximation to the fourier transform. *Phys. Rev. D*, 59:124016, 1999.

[41] Luc Blanchet. Gravitational radiation from post-Newtonian sources and inspiralling compact binaries. *Living Rev. Rel.*, 5:3, 2002.

[42] Guillaume Faye, Sylvain Marsat, Luc Blanchet, and Bala R. Iyer. The third and a half post-Newtonian gravitational wave quadrupole mode for quasi-circular inspiralling compact binaries. *Class. Quant. Grav.*, 29:175004, 2012.

[43] Alexander H. Nitz, Tito Dal Canton, Derek Davis, and Steven Reyes. Rapid detection of gravitational waves from compact binary mergers with PyCBC Live. *Phys. Rev.*, D98(2):024050, 2018.

[44] L. K. Nuttall et al. Improving the Data Quality of Advanced LIGO Based on Early Engineering Run Results. *Class. Quantum Grav.*, 32(24):245005, 2015.

[45] Bruce Allen. A chi**2 time-frequency discriminator for gravitational wave detection. *Phys. Rev. D*, 71:062001, 2005.

[46] S. Babak, R. Biswas, P.R. Brady, D.A. Brown, K. Cannon, et al. Searching for gravitational waves from binary coalescence. *Phys. Rev. D*, 87:024033, 2013.

[47] Benjamin P. Abbott et al. GW170104: Observation of a 50-Solar-Mass Binary Black Hole Coalescence at Redshift 0.2. *Phys. Rev. Lett.*, 118(22):221101, 2017.

[48] Benjamin P. Abbott et al. GW170814: A three-detector observation of gravitational waves from a binary black hole coalescence. *Phys. Rev. Lett.*, 119:141101, 2017.

[49] C. Capano, T. Dent, C. Hanna, M. Hendry, C. Messenger, Y.-M. Hu, and J. Veitch. Systematic errors in estimation of gravitational-wave candidate significance. *Phys. Rev. D*, 96(8):082002, October 2017.

[50] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: A Practical and powerful approach to multiple testing. *J. Roy. Statist. Soc.*, 57:289–300, 1995.

[51] F. Guglielmetti, R. Fischer, and V. Dose. Background-source separation in astronomical images with Bayesian probability theory - I. The method. *Mon. Not. Roy. Astr. Soc.*, 396:165–190, June 2009.

[52] W. M. Farr, J. R. Gair, I. Mandel, and C. Cutler. Counting and confusion: Bayesian rate estimation with multiple populations. *Phys. Rev. D*, 91(2):023005, January 2015.

[53] B. P. Abbott et al. The Rate of Binary Black Hole Mergers Inferred from Advanced LIGO Observations Surrounding GW150914. *Astrophys. J.*, 833(1):L1, 2016.

[54] Krzysztof Belczynski, Alessandra Buonanno, Matteo Cantiello, Chris L. Fryer, Daniel E. Holz, Ilya Mandel, M. Coleman Miller, and Marek Walczak. The Formation and Gravitational-Wave Detection of Massive Stellar Black-Hole Binaries. *Astrophys. J.*, 789(2):120, 2014.

[55] Michal Dominik, Emanuele Berti, Richard O'Shaughnessy, Ilya Mandel, Krzysztof Belczynski, Christopher Fryer, Daniel E. Holz, Tomasz Bulik, and Francesco Pannarale. Double Compact Objects III: Gravitational Wave Detection Rates. *Astrophys. J.*, 806(2):263, 2015.

[56] Feryal Ozel, Dimitrios Psaltis, Ramesh Narayan, and Jeffrey E. McClintock. The Black Hole Mass Distribution in the Galaxy. *Astrophys.J.*, 725:1918–1927, 2010.

[57] Will M. Farr, Niharika Sravan, Andrew Cantrell, Laura Kreidberg, Charles D. Bailyn, et al. The Mass Distribution of Stellar-Mass Black Holes. *Astrophys. J.*, 741:103, 2011.

[58] Laura Kreidberg, Charles D. Bailyn, Will M. Farr, and Vassiliki Kalogera. Mass Measurements of Black Holes in X-Ray Transients: Is There a Mass Gap? *Astrophys. J.*, 757:36, 2012.

[59] Christopher S. Reynolds. Measuring Black Hole Spin using X-ray Reflection Spectroscopy. *Space Sci. Rev.*, 183(1-4):277–294, 2014.

[60] K. Belczynski et al. The origin of low spin of black holes in LIGO/Virgo mergers. 2017.

[61] Daniel Wysocki, Davide Gerosa, Richard O'Shaughnessy, Krzysztof Belczynski, Wojciech Gladysz, Emanuele Berti, Michael Kesden, and Daniel E. Holz. Explaining LIGOs observations via isolated binary evolution with natal kicks. *Phys. Rev.*, D97(4):043014, 2018.

[62] Bernard F. Schutz. Networks of gravitational wave detectors and three figures of merit. *Class. Quantum Grav.*, 28:125023, 2011.

[63] Hsin-Yu Chen and Daniel E. Holz. The Loudest Gravitational Wave Events. 2014.

[64] I. Tews, J. Margueron, and S. Reddy. How well does GW170817 constrain the equation of state of dense matter? 2018.

[65] Elias R. Most, L. Jens Papenfort, Veronica Dexheimer, Matthias Hanauske, Stefan Schramm, Horst Stcker, and Luciano Rezzolla. Signatures of quark-hadron phase transitions in general-relativistic neutron-star mergers. 2018.

[66] Carolyn Raithel, Feryal zel, and Dimitrios Psaltis. Tidal deformability from GW170817 as a direct probe of the neutron star radius. *Astrophys. J.*, 857(2):L23, 2018.

[67] B. P. Abbott et al. GW170817: Measurements of neutron star radii and equation of state. 2018.

[68] B. P. Abbott et al. Properties of the binary neutron star merger GW170817. 2018.

[69] Nevin N. Weinberg, Phil Arras, and Joshua Burkart. An instability due to the nonlinear coupling of p-modes to g-modes: Implications for coalescing neutron star binaries. *Astrophys. J.*, 769:121, 2013.

[70] Tejaswi Venumadhav, Aaron Zimmerman, and Christopher M. Hirata. The stability of tidally deformed neutron stars to three- and four-mode coupling. *Astrophys. J.*, 781:23, 2014.

[71] Nevin N. Weinberg. Growth rate of the tidal p-mode g-mode instability in coalescing binary neutron stars. *Astrophys. J.*, 819(2):109, 2016.

[72] Reed Essick, Salvatore Vitale, and Nevin N. Weinberg. Impact of the tidal p-g instability on the gravitational wave signal from coalescing binary neutron stars. *Phys. Rev.*, D94(10):103012, 2016.

[73] Robert E Kass and Adrian E Raftery. Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795, 1995.

[74] Michael P Hobson, Andrew H Jaffe, Andrew R Liddle, Pia Mukherjee, and David Parkinson. *Bayesian methods in cosmology*. Cambridge University Press, 2010.

[75] T. A. Apostolatos. Search templates for gravitational waves from precessing, inspiraling binaries. *Phys. Rev.*, D52:605–620, 1995.

[76] Luc Blanchet. Gravitational Radiation from Post-Newtonian Sources and Inspiralling Compact Binaries. *Living Rev. Rel.*, 17:2, 2014.

[77] P. C. Peters and J. Mathews. Gravitational radiation from point masses in a Keplerian orbit. *Phys. Rev.*, 131:435–439, 1963.

[78] Walter Del Pozzo, Tjonnie G. F. Li, Michalis Agathos, Chris Van Den Broeck, and Salvatore Vitale. Demonstrating the feasibility of probing the neutron star equation of state with second-generation gravitational wave detectors. *Phys. Rev. Lett.*, 111:071101, 2013.

[79] Lee Lindblom, Benjamin J. Owen, and Duncan A. Brown. Model Waveform Accuracy Standards for Gravitational Wave Data Analysis. *Phys. Rev. D*, 78:124020, 2008.

[80] Yixiao Zhou and Fan Zhang. Equation of State Dependence of Nonlinear Mode-tide Coupling in Coalescing Binary Neutron Stars. 2018. [Astrophys. J.849,114(2017)].

[81] N. Andersson and W. C. G. Ho. Using gravitational-wave data to constrain dynamical tides in neutron star binaries. *Phys. Rev.*, D97(2):023016, 2018.

[82] Lawrence E. Kidder, Clifford M. Will, and Alan G. Wiseman. Spin effects in the inspiral of coalescing compact binaries. *Phys. Rev.*, D47:4183–4187, 1993.

[83] Luc Blanchet, Thibault Damour, Bala R. Iyer, Clifford M. Will, and Alan. G. Wiseman. Gravitational-radiation damping of compact binary systems to second post-Newtonian order. *Phys. Rev. Lett.*, 74:3515–3518, 1995.

[84] Luc Blanchet, Thibault Damour, Gilles Esposito-Farèse, and Bala R. Iyer. Gravitational radiation from inspiralling compact binaries completed at the third post-Newtonian order. *Phys. Rev. Lett.*, 93:091101, 2004.

[85] Alessandra Buonanno, Bala R. Iyer, Evan Ochsner, Yi Pan, and B. S. Sathyaprakash. Comparison of post-Newtonian templates for compact binary inspiral signals in gravitational-wave detectors. *Phys.Rev.*, D80:084043, 2009.

[86] K.G. Arun, Alessandra Buonanno, Guillaume Faye, and Evan Ochsner. 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, 2009.

[87] Sylvain Marsat, Alejandro Boh, Luc Blanchet, and Alessandra Buonanno. Next-to-leading tail-induced spinorbit effects in the gravitational radiation flux of compact binaries. *Class. Quant. Grav.*, 31:025023, 2014.

[88] Alejandro Boh, Sylvain Marsat, and Luc Blanchet. Next-to-next-to-leading order spinorbit effects in the gravitational wave flux and orbital phasing of compact binaries. *Class. Quant. Grav.*, 30:135009, 2013.

[89] Alejandro Boh, Guillaume Faye, Sylvain Marsat, and Edward K. Porter. Quadratic-in-spin effects in the orbital dynamics and gravitational-wave energy flux of compact binaries at the 3PN order. *Class. Quant. Grav.*, 32(19):195010, 2015.

[90] Balazs Mikoczi, Matyas Vasuth, and Laszlo A. Gergely. Self-interaction spin effects in inspiralling compact binaries. *Phys. Rev.*, D71:124043, 2005.

[91] Eanna E. Flanagan and Tanja Hinderer. Constraining neutron star tidal Love numbers with gravitational wave detectors. *Phys. Rev.*, D77:021502, 2008.

[92] Justin Vines, Eanna E. Flanagan, and Tanja Hinderer. Post-1-Newtonian tidal effects in the gravitational waveform from binary inspirals. *Phys. Rev.*, D83:084051, 2011.

[93] LSC Algorithm Library software packages LAL, LALWRAPPER, and LALAPPS.

[94] Alex Nitz, Ian Harry, Duncan Brown, Christopher M. Biwer, Josh Willis, Tito Dal Canton, Larne Pekowsky, Thomas Dent, Collin Capano, Andrew R. Williamson, Soumi De, Miriam Cabero, Bernd Machenschalk, Prayush Kumar, Steven Reyes, Thomas Massinger, Amber Lenon, Stephen Fairhurst, Alex Nielsen, Duncan Macleod, shasvath, Francesco Pannarale, Leo Singer, Stanislav Babak, Hunter Gabbard, dfinstad, John Veitch, CBC Sugar, Sebastian Khan, and Lorena Magaa Zertuche. ligo-cbc/pycbc: Post-o2 release 8, March 2018.

[95] Lee Samuel Finn. Aperture synthesis for gravitational-wave data analysis: Deterministic sources. *Phys. Rev. D*, 63:102001, 2001.

[96] Christian Rover, Renate Meyer, and Nelson Christensen. Coherent Bayesian inference on compact binary inspirals using a network of interferometric gravitational wave detectors. *Phys. Rev.*, D75:062004, 2007.

[97] M. Soares-Santos et al. The Electromagnetic Counterpart of the Binary Neutron Star Merger LIGO/Virgo GW170817. I. Discovery of the Optical Counterpart Using the Dark Energy Camera. *Astrophys. J.*, 848(2):L16, 2017.

[98] Michele Cantiello et al. A Precise Distance to the Host Galaxy of the Binary Neutron Star Merger GW170817 Using Surface Brightness Fluctuations. *Astrophys. J.*, 854(2):L31, 2018.

[99] Feryal Ozel and Paulo Freire. Masses, Radii, and the Equation of State of Neutron Stars. *Ann. Rev. Astron. Astrophys.*, 54:401–440, 2016.

[100] Kent Blackburn et al. *LOSC CLN Data Products for GW170817*, 2017.

[101] CM Biwer, Collin D Capano, Soumi De, Miriam Cabero, Duncan A Brown, Alexander H Nitz, and Vivien Raymond. Pycbc inference: A python-based parameter estimation toolkit for compact binary coalescence signals. *Publications of the Astronomical Society of the Pacific*, 131(996):024503, 2019.

[102] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman. emcee: The MCMC Hammer. *Publ. Astron. Soc. Pac.*, 125:306, March 2013.

[103] W. D. Vousden, W. M. Farr, and I. Mandel. Dynamic temperature selection for parallel tempering in Markov chain Monte Carlo simulations. *Monthly Notices of the Royal Astronomical Society*, 455(2):1919–1937, 2016.

[104] Jian-Sheng Wang and Robert H. Swendsen. Replica monte carlo simulation (revisited). *Progress of Theoretical Physics Supplement*, 157:317–323, 2005.

[105] David J. Earl and Michael W. Deem. Parallel tempering: Theory, applications, and new perspectives. *Phys. Chem. Chem. Phys.*, 7:3910–3916, 2005.

[106] Nicolas Lartillot and Hervé Philippe. Computing bayes factors using thermodynamic integration. *Systematic biology*, 55(2):195–207, 2006.

[107] Nial Friel and Anthony N Pettitt. Marginal likelihood estimation via power posteriors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(3):589–607, 2008.

[108] Peigui Liu, Ahmed S Elshall, Ming Ye, Peter Beerli, Xiankui Zeng, Dan Lu, and Yuezan Tao. Evaluating marginal likelihood with thermodynamic integration method and comparison with several other numerical methods. *Water Resources Research*, 52(2):734–758, 2016.

[109] Wangang Xie, Paul O Lewis, Yu Fan, Lynn Kuo, and Ming-Hui Chen. Improving marginal likelihood estimation for bayesian phylogenetic model selection. *Systematic biology*, 60(2):150–160, 2010.

[110] Patricio Maturana Russel, Renate Meyer, John Veitch, and Nelson Christensen. The stepping-stone sampling algorithm for calculating the evidence of gravitational wave models. 2018.

[111] Nial Friel, Merrilee Hurn, and Jason Wyse. Improving power posterior estimation of statistical evidence. *Statistics and Computing*, 24(5):709–723, 2014.

[112] Jeffrey Annis, Nathan J Evans, Brent J Miller, and Thomas J Palmeri. Thermodynamic integration and steppingstone sampling methods for estimating bayes factors: A tutorial. *Journal of mathematical psychology*, 89:67–86, 2019.

[113] Ward Edwards, Harold Lindman, and Leonard J Savage. Bayesian statistical inference for psychological research. *Psychological review*, 70(3):193, 1963.

[114] James M Dickey. The weighted likelihood ratio, linear hypotheses on normal location parameters. *The Annals of Mathematical Statistics*, pages 204–223, 1971.

[115] Eric-Jan Wagenmakers, Tom Lodewyckx, Himanshu Kuriyal, and Raoul Grasman. Bayesian hypothesis testing for psychologists: A tutorial on the savage–dickey method. *Cognitive psychology*, 60(3):158–189, 2010.

[116] Charles J Stone, Mark H Hansen, Charles Kooperberg, Young K Truong, et al. Polynomial splines and their tensor products in extended linear modeling: 1994 wald memorial lecture. *The Annals of Statistics*, 25(4):1371–1470, 1997.

[117] David W Scott. On optimal and data-based histograms. *Biometrika*, 66(3):605–610, 1979.

[118] David Freedman and Persi Diaconis. On the histogram as a density estimator:l2 theory. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 57(4):453–476, Dec 1981.

[119] Antony Lewis. Getdist: Kernel density estimation. *Available from download at https://cosmologist.info/notes/GetDist.pdf*, 2015.

[120] Harold Jeffreys. *The theory of probability*. OUP Oxford, 1998.

[121] Reed Essick and Nevin N. Weinberg. A Comparison of p-g Tidal Coupling Analyses. 2018.

[122] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing In Science & Engineering*, 9(3):90–95, 2007.

[123] Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for Python, 2001–.

[124] Michele Maggiore. *Gravitational waves: Volume 1: Theory and experiments*, volume 1. Oxford university press, 2008.

[125] David Carlson, Patrick Stinson, Ari Pakman, and Liam Paninski. Partition functions from rao-blackwellized tempered sampling. In *International Conference on Machine Learning*, pages 2896–2905, 2016.

[126] IS Gradshteyn and IM Ryzhik. Table of integrals, series, and products, 8th edn., ed. by d. zwillinger, 2015.

[127] Colin H LaMont and Paul A Wiggins. Correspondence between thermodynamics and inference. *Physical Review E*, 99(5):052140, 2019.

[128] Mehran Kardar. *Statistical physics of particles*. Cambridge University Press, 2007.

[129] Anita de Ruiter and Chris Oostenbrink. Free energy calculations of protein–ligand interactions. *Current opinion in chemical biology*, 15(4):547–552, 2011.

[130] Anita de Ruiter, Stefan Boresch, and Chris Oostenbrink. Comparison of thermodynamic integration and bennett acceptance ratio for calculating relative protein-ligand binding free energies. *Journal of computational chemistry*, 34(12):1024–1034, 2013.

[131] Chris J Oates, Mark Girolami, and Nicolas Chopin. Control functionals for monte carlo integration. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(3):695–718, 2017.

[132] Nathan J Evans and Jeffrey Annis. Thermodynamic integration via differential evolution: A method for estimating marginal likelihoods. *Behavior research methods*, 51(2):930–947, 2019.

[133] Viggo Brun. A generalization of the formula of simpson for non-equidistant ordinates. *Nordisk Matematisk Tidskrift*, 1(1):10–15, 1953.

[134] Ernts S. Selmer. Numerical integration by non-equidistant ordinates. *Nordisk Matematisk Tidskrift*, pages 97–108, 1958.

[135] Milton Abramowitz and Irene A Stegun. *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, volume 55. Courier Corporation, 1965.

[136] Said M Easa. Area of irregular region with unequal intervals. *Journal of Surveying Engineering*, 114(2):50–58, 1988.

[137] Daniel W Chambers. Estimating pit excavation volume using unequal intervals. *Journal of Surveying Engineering*, 115(4):390–401, 1989.

[138] James F Epperson. On the runge example. *The American Mathematical Monthly*, 94(4):329–341, 1987.

[139] William H Press and Saul A Teukolsky. Padé approximants. *Computers in Physics*, 6(1):82–83, 1992.

[140] Persi Diaconis. Bayesian numerical analysis. *Statistical decision theory and related topics IV*, 1:163–175, 1988.

[141] François-Xavier Briol, Chris J Oates, Mark Girolami, Michael A Osborne, and Dino Sejdinovic. Probabilistic integration: A role for statisticians in numerical analysis. *arXiv preprint arXiv:1512.00933*, 1293, 2015.

[142] Michael A Newton and Adrian E Raftery. Approximate bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society: Series B (Methodological)*, 56(1):3–26, 1994.

[143] Gary W Oehlert. A note on the delta method. *The American Statistician*, 46(1):27–29, 1992.

[144] Isabella Verdinelli and Larry Wasserman. Computing bayes factors using a generalization of the savage-dickey density ratio. *Journal of the American Statistical Association*, 90(430):614–618, 1995.

[145] Paul Bromiley. Products and convolutions of gaussian probability density functions. *Tina-Vision Memo*, 3(4):1, 2003.

[146] Hirotugu Akaike. Likelihood of a model and information criteria. *Journal of econometrics*, 16(1):3–14, 1981.

[147] Gideon Schwarz et al. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978.

[148] Bernard F Schutz. Networks of gravitational wave detectors and three figures of merit. *Classical and Quantum Gravity*, 28(12):125023, 2011.

[149] Hsin-Yu Chen and Daniel E Holz. The loudest gravitational wave events. *arXiv preprint arXiv:1409.0522*, 2014.

Curriculum Vitae

Title of Dissertation

Inference and Model Comparison in Gravitational Wave Astronomy

Full Name

Steven Reyes

Place and Date of Birth

Chicago, Illinois May 4, 1992

Colleges and Universities,	Years attended and degrees
The University of Chicago	2010–2014, B.A.
Syracuse University	2014–2019, Ph.D.

Major Department

Physics