# LIGO SURF Interim Report 1: The impact of astrophysical population model choices on post-Newtonian deviation tests of general relativity

Ruby Knudsen, Mentor: Ethan Payne

July 2024

## 1 Background

In the early 20th century, Albert Einstein put forward the theory of general relativity (GR), and summarized his theory mathematically through the Einstein field equations. Part of GR theorized the concept of spacetime, which spatially distorts in the vicinity of massive astrophysical objects. Building off of this work, two years later Einstein posited the existence of gravitational waves (GW) — spatial distortions as transverse waves that originate from non-spherically symmetric quadrupolar disturbances before traveling away at the speed of light [2, 10]. Because GR theorizes the existence of GWs, it is possible to use GWs as a method to test GR. There are several tests of GR, but GWs are the best probe for the theory.

Different tests of GR interrogate the theory to different levels, as summarized by Fig. 1. While all astrophysical objects have gravitational potential and cause spacetime curvature — a key insight of GR — the most extreme tests of GR exist in the strong-field regime. The gravitational potential is proportional to $M/R^3$, and curvature is proportional to $M/R$. The perihelion procession of Mercury probes GR, but not strenuously, since it has potential between $10^{-7}$ and $10^{-8}$, and curvature between $10^{-32}$ and $10^{-33}$ cm$^{-2}$ [18]. Due to the Schwarzschild radius, the event horizon of a black hole yields potential approximately 0.5, and is the most extreme possible potential. While Sagittarius A* and M87 both have potential near 0.5, their curvature is orders of magnitude lower than the spacetime curvature near merging stellar mass black hole binaries [8]. Curvatures between $10^{-14}$ and $10^{-10}$ cm$^{-2}$ are the most extreme for realistic astrophysical signals [8]. Probing GR in this region requires the detection of GWs from black hole and neutron star mergers. Thus, GW tests probe GR at the most extreme curvatures and potentials, and as such are the best tests of GR.

The Laser Interferometer Gravitational-wave Observatory (LIGO) — a joint project between the California Institute of Technology, the Massachusetts Institute of Technology, and the National Science Foundation — was built to detect these GWs and aims to further our understanding of GWs and test GR [1, 5]. LIGO has perpendicular arms that detect compact binary mergers using the interference of light due to the difference in length of the arms that a passing GW causes. This change in length is directly proportional to GW strength [1, 5]. The compact binary mergers, mostly binary black hole mergers, are then analyzed to check for consistency with GR. LIGO is sensitive to GW frequencies between 20 and 1000 Hz, so black holes of stellar mass are easiest to detect [12].

## 2 Introduction to methods for testing GR

### 2.1 Modeling modifications to GR with individual observations

There are many different ways of modeling GR deviations by modeling the GW signal. One of these tests is inferring deviations from the post-Newtonian (PN) approximation. The post-Newtonian approximation at the lowest order is the quadrupole formula, which estimates the emitted radiation from a quadrupolar mass distribution [17]. The PN expansion for gravitational-wave emission involves the dimensionless parameter $v/c$, where $v$ is the orbital velocity of the binary, with the lowest order being the Newtonian-order $v/c$

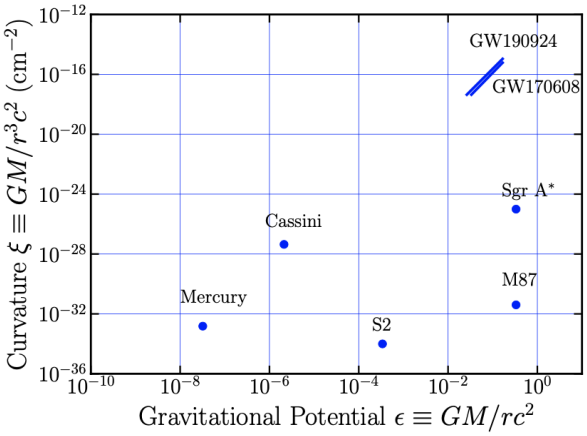


Figure 1: From Ref. [18]. This graphic shows a plot of gravitational potential vs. curvature for various tests of GR. The theoretical limit for the gravitational potential exists between 0.5 and 1.

order, and higher orders being $(v/c)^2$, $(v/c)^3$, and so on [20, 9]. PN tests are run by first constructing a post-Newtonian description of the GW inspiral in the frequency domain, before making modifications to individual parameters in the phase evolution [7, 19]. This is done with the phase, not the amplitude, because LIGO is more sensitive to phase deviations than those of amplitudes; over the course of an inspiral, phase shifts accumulate, while amplitude differences do not.

$$\Phi(f) = 2\pi f t_c - \phi_c - \frac{\pi}{4} + \frac{3}{128} \times \sum_{k=0}^{7} \frac{1}{\eta^{\frac{k}{5}}} (\varphi_k + \varphi_{k,l} \ln \tilde{f}) \tilde{f}^{\frac{k-5}{3}} \tag{1}$$

In this equation, $\Phi(f)$ is the frequency domain GW phase under the stationary phase approximation $\tilde{f}$, the red-shifted chirp-mass multiplied by $f$ (the frequency), divided by $c^3$. Additionally, $t_c$ and $\phi_c$ are the coalescence time and phase respectively, $\eta$ is the symmetric mass ratio, and $\varphi_k$, $\varphi_{k,l}$ are PN coefficients for the $k/2$ PN order [7, 19]. After running these PN tests, a posterior is obtained by running Bayesian inference.

Bayesian inference makes statements about the Universe from data by inferring the distribution probable for parameters of each event analyzed, known as the posterior distribution [23]. Calculating this posterior distribution requires a likelihood, to describe the measurement, and a prior, to encode prior beliefs about each event's parameters $\theta$. Bayesian inference uses Bayes' Theorem,

$$p(\theta|d) = \frac{\mathcal{L}(d|\theta)\pi(\theta)}{\mathcal{Z}} \tag{2}$$

In this equation $\mathcal{L}(d|\theta)$ is the likelihood and $\pi(\theta)$ is the prior, and the evidence, which normalizes the distribution, is denoted as $\mathcal{Z}$ [23]. This same formalism can be used to study multiple events, with a modified version of the likelihood function.

## 2.2 Testing GR with many observations

Hierarchical inference is a method that uses posterior distributions from individual events to model the overall population of the astrophysical distribution from which the events originated. The analysis returns a distribution for the hyper-parameters of the population model, which is called the hyper-posterior [6, 23]. In this hyper-posterior, the astrophysical population has been inferred under the assumption of the choice of population model. A population model is important because otherwise an incorrect astrophysical population may be implicitly assumed, which in turn leads to bias in supposed measured deviations from GR [16].

To look at the population properties of a collection of events, the prior for the parameters $\theta$ is made conditional on hyper-parameters $\Lambda$, which encodes the astrophysical distribution from which the $\theta$s are

drawn [23]. These hyper-parameters are very important because they parameterize the shape of the inferred astrophysical population, and one goal of population inference is estimating the posterior of these hyper-parameters. One key piece of estimating the hyper-posterior is the population likelihood, which is calculated using Eq. 3.

A population likelihood is necessary for hierarchical inference. The population likelihood is a formula that allows simultaneous astrophysical population inference and GR testing, decreasing this bias [16, 14]. The equation for the population likelihood is

$$p(\{d\}|\Lambda) = \frac{1}{\xi(\Lambda)^N} \prod_{i=1}^{N} \int d\theta_i p(d_i|\theta_i)\pi(\theta_i|\Lambda) \tag{3}$$

Here, $\{d\}$ is the collection of $N$ observations, $\xi(\Lambda)$ accounts for selection biases and is the detectable fraction of observations given the population hyper-parameters, $p(d_i|\theta_i)$ is the likelihood for each individual event, and $\pi(\theta_i|\Lambda)$ are the hyper-priors for each event [23]. Thus, $\pi(\theta_i|\Lambda)$ is where the population distribution is encoded.

This equation shows how individual observations are put together for hierarchical inference. As stated earlier, this hierarchical approach requires a population model, so we therefore must choose one population model to use. Currently, we are not sure in what manner the different choices of population model impact GR tests. This leads to my project.

# 3 Project motivation

Population distributions depend in part on the choice of population model, thus the choice of population model must be carefully considered. So far, this approach of joint inference has, at the population level, yielded GR deviations more consistent with GR by about $0.4\sigma$, when using a PowerLaw+Peak population distribution for the mass of a black hole [16, 3]. However, it is possible that deviation from general relativity could be absorbed or hidden by an incorrect, assumed astrophysical population, so it is important to study the impact of different astrophysical population models on inferred GR deviation constraints. It is true that at some point that the incorrect choice will also lead to biases, the question is when, and how badly. Testing this using different population models is what I am working on.

# 4 Current progress

To work on population inference calculations, I began by recreating the analysis in Ref. [13] using a toy model to "test" general relativity. This required generating a simulated population of 100 events, each with a mean and a standard deviation, shown in Fig. 2. Each standard deviation was drawn randomly between 0 and 1, while the means were each drawn from a Gaussian with a mean of 0 and the associated standard deviation. In the beginning, I went straight to the likelihood of the population by using an analytical approximation for the integral of the product of two Gaussian distributions. I ended up with a result consistent with my original distribution, which implied that my toy model agreed with general relativity, as seen in Fig. 3.

After creating this simple version of the problem, I made the test more consistent with current analyses of LIGO data by drawing posterior distributions and running the population likelihood from these using Markov-Chain Monte Carlo sampling methods. From each event I draw 10,000 samples, which made up the posterior distribution for each event. A histogram of one posterior distribution is shown in Fig. 4. When I run my injected signals with the population models I choose for black holes, this will be the method I use. To make the computation easier, I used the natural logarithm of the likelihood instead of the likelihood — a simplified version of the likelihood is shown in Eq. 4. After some adjustments to correct errors, I ended up with results consistent with general relativity in my toy model again, as is shown in Fig. 5, where (0,0) for the mean and standard deviation is present in the plot. One issue is that the minimum effective sample size is quite small, so I am working on testing whether this is an issue that is biasing my results, and if so how to fix it.

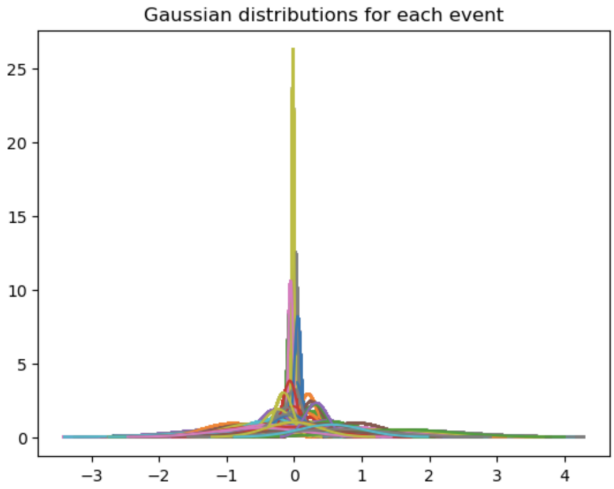


Figure 2: This figure shows the Gaussian distributions for each of the 100 events. The Gaussians were created from the means and standard deviations of each event.

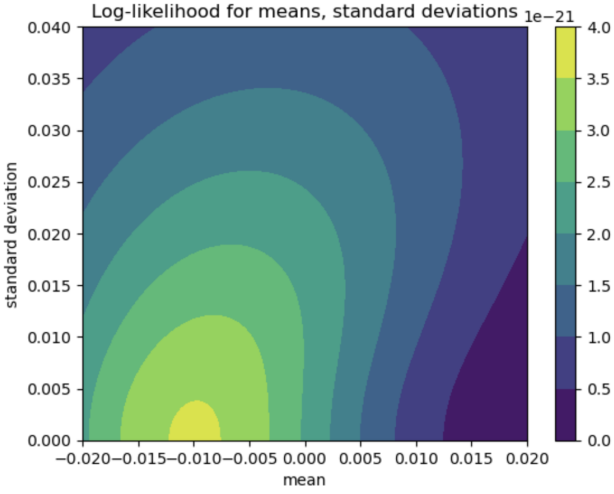


Figure 3: This figure shows the population likelihood over a grid of means and standard deviations, with means on the x-axis and standard deviations on the y-axis. The point (0,0) is contained in the tail, which implied consistency with general relativity in this toy model, and thus confirms that my code works.

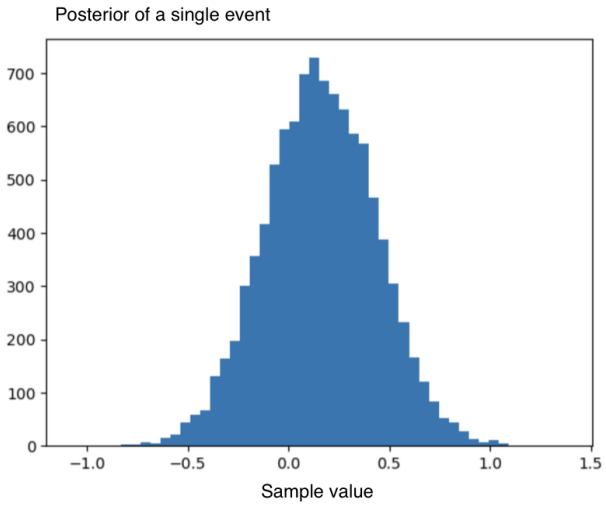


Figure 4: This figure shows a histogram of the 10,000 samples drawn from the Gaussian of a single event.

$$\mathcal{L}(\{d\}|\Lambda) \propto \prod_{i}^{N} \sum_{k}^{n_s} \frac{\pi(\theta_i^k|\Lambda)}{\pi(\theta_i^k|\varnothing)} \tag{4}$$

# 5 Challenges and Looking Forward

Over the last few weeks, I have been: getting set up with NumPyro and Jax in JupyterNotebook, writing code to run hierarchical population inference calculations, reading literature to identify population models to use in my analysis, and actually visiting LIGO Hanford, which was an amazing experience.

The first challenges that I encountered were primarily computer challenges; getting setup with the appropriate software, accounts, permissions, and more. Luckily, those challenges have been resolved, and I anticipate that going forward I will only experience minor technical difficulties that I will be able to work through easily.

My second main challenge was creating runnable code. The population likelihood includes a product of an integral of the product of two Gaussians; this is computationally taxing, so my mentor steered me in the correct direction for an analytical solution for this integral.

Once I had that up and running, it involved three nested for loops, which requires a lot of computing power and is quite inefficient. Thus, I had to rewrite my code using meshgrid from NumPy, and move to a map function. This cut runtime down drastically. After solving this challenge, another change to improve on efficiency was moving to a Markov-Chain Monte Carlo simulation with posterior distributions.

Outside of my coding work, I have been reading up on population models and brainstorming which to include in my future simulations. My mentor and I have discussed having one population model that is obviously misspecified, probably a simple PowerLaw. The first three observing runs of LIGO have found that black hole mass does not follow this model, and has substructure and other features [3]. Thus I am currently considering population models such as PowerLaw+Peak, and possibly Broken PowerLaw to more accurately capture the complex structure [22].

Choosing population models has been a challenge. Because there have not been a high number of observations so far, it is difficult to determine the distribution of black hole masses, spins, etc. Therefore, my quest to find a "most accurate" population model is difficult. For example, with two of the models for spin, Default and Gaussian, both agree with each other within the margin of error [4, 24, 15]. I plan to use the Default model for spin, and Power-law Evolution model for redshift, along with my three mass population models [11, 21].

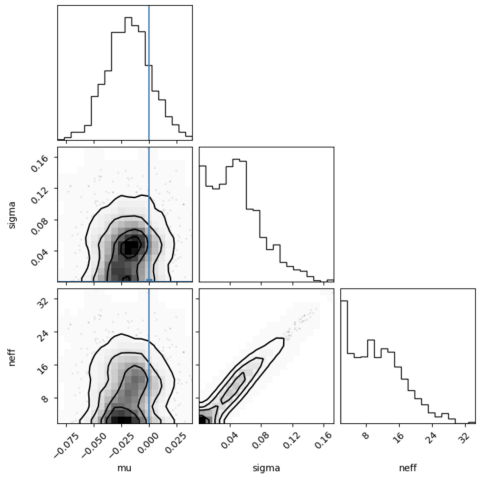


Figure 5: This figure shows the distribution of means and standard deviations for the population, as well as the effective sample size. The blue line shows where a mean of 0 is represented. For consistency with GR, the point (0,0) needs to be inside the distribution, and it is. The graph of $n_{eff}$ shows the effective sample size. Each of the 100 events had 10,000 samples drawn to create the posterior, therefore an effective sample size below 32 represents an extremely small percentage of the samples. I am working on investigating this issue right now.

# 6 Acknowledgments

This work was supported by the National Science Foundation Research Experience for Undergraduates (NSF REU) program, the LIGO Laboratory Summer Undergraduate Research Fellowship program (NSF LIGO), and the California Institute of Technology Student-Faculty Programs. I would like to thank my mentor, Ethan Payne, my fellow LIGO SURF-ers, and all of the faculty and mentors at the LIGO SURF program, especially Alan Weinstein and Jonah Kanner.

# References

[1] Junaid Aasi, BP Abbott, Richard Abbott, Thomas Abbott, MR Abernathy, Kendall Ackley, Carl Adams, Thomas Adams, Paolo Addesso, RX Adhikari, et al. Advanced ligo. *Classical and quantum gravity*, 32(7):074001, 2015.

[2] Benjamin P Abbott, Richard Abbott, TDe Abbott, MR Abernathy, Fausto Acernese, Kendall Ackley, Carl Adams, Thomas Adams, Paolo Addesso, Rana X Adhikari, et al. Observation of gravitational waves from a binary black hole merger. *Physical review letters*, 116(6):061102, 2016.

[3] R Abbott, TD Abbott, F Acernese, K Ackley, C Adams, N Adhikari, RX Adhikari, VB Adya, C Affeldt, D Agarwal, et al. Population of merging compact binaries inferred using gravitational waves through gwtc-3. *Physical Review X*, 13(1):011048, 2023.

[4] Rich Abbott, TD Abbott, S Abraham, Fausto Acernese, K Ackley, A Adams, C Adams, RX Adhikari, VB Adya, Christoph Affeldt, et al. Population properties of compact objects from the second ligo–virgo gravitational-wave transient catalog. *The Astrophysical journal letters*, 913(1):L7, 2021.

[5] Fet al Acernese, M Agathos, K Agatsuma, Damiano Aisa, N Allemandou, Aea Allocca, J Amarni, Pia Astone, G Balestri, G Ballardin, et al. Advanced virgo: a second-generation interferometric gravitational wave detector. *Classical and Quantum Gravity*, 32(2):024001, 2014.

[6] Matthew R Adams, Neil J Cornish, and Tyson B Littenberg. Astrophysical model selection in gravitational wave astronomy. *Physical Review D*, 86(12):124032, 2012.

[7] KG Arun, Bala R Iyer, Bangalore Suryanarayana Sathyaprakash, and Pranesh A Sundararajan. Parameter estimation of inspiralling compact binaries using 3.5 post-newtonian gravitational wave phasing: The nonspinning case. *Physical Review D—Particles, Fields, Gravitation, and Cosmology*, 71(8):084008, 2005.

[8] Tessa Baker, Dimitrios Psaltis, and Constantinos Skordis. Linking tests of gravity on all scales: from the strong-field regime to cosmology. *The Astrophysical Journal*, 802(1):63, 2015.

[9] Luc Blanchet and Thibault Damour. Post-newtonian generation of gravitational waves. In *Annales de l'IHP Physique théorique*, volume 50, pages 377–408, 1989.

[10] Albert Einstein and Emil Warburg. *Die Relativitätstheorie*. Springer, 1911.

[11] Maya Fishbach, Daniel E Holz, and Will M Farr. Does the black hole merger rate evolve with redshift? *The Astrophysical Journal Letters*, 863(2):L41, 2018.

[12] Davide Gerosa, Geraint Pratten, and Alberto Vecchio. Gravitational-wave selection effects using neural-network classifiers. *Physical Review D*, 102(10):103020, 2020.

[13] Maximiliano Isi, Katerina Chatziioannou, and Will M Farr. Hierarchical test of general relativity with gravitational waves. *Physical Review Letters*, 123(12):121101, 2019.

[14] Ryan Magee, Maximiliano Isi, Ethan Payne, Katerina Chatziioannou, Will M. Farr, Geraint Pratten, and Salvatore Vitale. Impact of selection biases on tests of general relativity with gravitational-wave inspirals. *Phys. Rev. D*, 109(2):023014, 2024.

[15] Simona Miller, Thomas A Callister, and Will M Farr. The low effective spin of binary black holes and implications for individual gravitational-wave events. *The Astrophysical Journal*, 895(2):128, 2020.

[16] Ethan Payne, Maximiliano Isi, Katerina Chatziioannou, and Will M. Farr. Fortifying gravitational-wave tests of general relativity against astrophysical assumptions. *Phys. Rev. D*, 108(12):124060, 2023.

[17] William H Press and Kip S Thorne. Gravitational-wave astronomy. *Annual Review of Astronomy and Astrophysics*, 10(1):335–374, 1972.

[18] Dimitrios Psaltis, Colm Talbot, Ethan Payne, and Ilya Mandel. Probing the black hole metric: Black hole shadows and binary black-hole inspirals. *Physical Review D*, 103(10):104036, 2021.

[19] Bangalore Suryanarayana Sathyaprakash and SV Dhurandhar. Choice of filters for the detection of gravitational waves from coalescing binaries. *Physical Review D*, 44(12):3819, 1991.

[20] Hideyuki Tagoshi, Masaru Shibata, Takahiro Tanaka, and Misao Sasaki. Post-newtonian expansion of gravitational waves from a particle in circular orbit around a rotating black hole: Up to o (v 8) beyond the quadrupole formula. *Physical Review D*, 54(2):1439, 1996.

[21] Colm Talbot and Eric Thrane. Determining the population properties of spinning black holes. *Physical Review D*, 96(2):023012, 2017.

[22] Colm Talbot and Eric Thrane. Measuring the binary black hole mass spectrum with an astrophysically motivated parameterization. *The Astrophysical Journal*, 856(2):173, 2018.

[23] Eric Thrane and Colm Talbot. An introduction to bayesian inference in gravitational-wave astronomy: parameter estimation, model selection, and hierarchical models. *Publications of the Astronomical Society of Australia*, 36:e010, 2019.

[24] Daniel Wysocki, Jacob Lange, and Richard O'Shaughnessy. Reconstructing phenomenological distributions of compact binaries via gravitational wave observations. *Physical Review D*, 100(4):043012, 2019.