

© 2022 by Scott Ellis Perkins. All rights reserved.

TESTS OF FUNDAMENTAL PHYSICS WITH GRAVITATIONAL WAVES:
THEORY AND APPLICATION

BY

SCOTT ELLIS PERKINS

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Physics
in the Graduate College of the
University of Illinois Urbana-Champaign, 2022

Urbana, Illinois

Doctoral Committee:

Assistant Professor Helvi Witek, Chair
Professor Nicolás Yunes, Director of Research
Professor Gilbert Holder
Assistant Professor Gautham Narayan

Abstract

The era of gravitational wave astronomy began in 2015 with the first direct detection of gravitational radiation. Those ripples in spacetime ushered in a new age in astronomy, astrophysics, fundamental physics and a host of other fields in science. In this dissertation, we examine how this method of observation can be used to constrain alternative theories of gravity using inspiral tests of general relativity. We begin by using a variety of simulation techniques to evaluate the efficacy of different methods and future detectors for investigating the gravitational interaction. Specifically, we start by examining the existence of the graviton mass, where we explore the impact of screening effects on the measureability of this modification with future detectors with a Fisher analysis. We then widen the scope of our Fisher analysis, considering a variety of different gravitational theories as well as theory-agnostic models. By considering astrophysically realistic populations of black holes and the most accurate gravitational wave detector timelines, we make predictions about the next three decades of fundamental science through gravitational wave observation. Next, we investigate potential sources of systematic error in theory-agnostic inspiral tests of general relativity. To accomplish this, we propose a new waveform model and prior inspired by post-Newtonian calculations. We then use it to analyze synthetic data, through which we show that current constraints are robust to these systematic effects. We then transition to working with current data released by the LIGO-Virgo-Kagra collaboration in order to place constraints on viable alternatives to general relativity. Specifically, we focus on two theories of gravity, Einstein-dilaton-Gauss-Bonnet and dynamical Chern Simons, and successfully place a constraint on the size of the relevant lengthscales of Einstein-dilaton-Gauss-Bonnet. In order to obtain these constraints, we use observations from the first, second and third gravitational wave catalogs. To ensure the robustness of our conclusions, we verified that two key, potential sources of systematic error, possible mis-modeling both within and without general relativity, do not impact our result.

To my mentors and collaborators who showed me the inherent beauty of Science through their enthusiasm and insight. And to my friends and family who continually remind me through their support and encouragement that Science is not the only thing that is beautiful.

Acknowledgments

This dissertation was the result of support, both professional and personal, from a number of institutions and people. The physics and astronomy department at Texas A&M University fostered my early love of science. This enthusiasm was subsequently encouraged and promoted at the physics department at Montana State University and at the physics department at the University of Illinois Urbana-Champaign. The faculty and staff who helped me to understand physics and to navigate the hurdles of undergraduate and graduate school have my deep gratitude.

In particular, I would like to thank Nicolás Yunes for his mentorship throughout my graduate career. Your guidance and support has helped shape me into the physicist I am today, and I am very grateful for your investment in me. You have started me out in my professional career in science with a solid foundation in research, allowing me to be a responsible and effective researcher for many years to come. I have also been very fortunate to work with several collaborators on various chapters of this dissertation: Remya Nair, Hector Okada da Silva, and Emanuele Berti. It was wonderful to learn about science and physics through your insights and knowledge. I would like to thank Emanuele Berti in particular, as our conversations about physics and careers in science were very encouraging and enlightening. Your investment in me was above and beyond scientific collaboration, and I am very grateful to you for that. I would also like to thank the many collaborators and scientists who I have had the fortune to interact with during my time in graduate school and who have taught me much through our discussions about research and physics: Neil Cornish, Kent Yagi, Daniel Holz, Gilbert Holder, Gautham Narayan, Helvi Witek, Deep Chatterjee, and many, many others. In particular, I would like to thank Helvi Witek, Gilbert Holder, and Gautham Narayan for agreeing to be on my defense committee. Your feedback, both as my committee and through our general discussions, has been very useful.

I would also like to thank the many collaborators, fellow scientists, and friends that have made my time in graduate school much better: Alejandro Cárdenas-Avendaño, Pratik Wagle, Hung Tan, Caroline Owen, Kristen Schumacher, Yiqi Xie, Rohit Chandramouli, Alex Deich, Jameson Dong, Nijaid Arredondo and Simone Mezzasoma, among many, many others. Working with these people has made my graduate

school career much more successful and much more enjoyable. In particular, I would like to thank Alejandro Cárdenas-Avendaño, Pratik Wagle, and Hung Tan for being there since the beginning of graduate school. In the early years of my graduate career, our work together and our friendships were particularly important. Furthermore, I would like to thank those people whose friendship made my time in Illinois much more meaningful: Alex Coleman, Lee Gaines, Anna Casey, Mary Casey, and Raundi Quevedo. Our friendships were a real source of joy these past few years.

My family has also been a wonderful source of support. My parents (Brad and Judi Perkins) and my brothers (Cory and Casey Perkins) have been there for me my entire life. Your love and support has given me the confidence and ability to pursue my passions in life. Finally, I would like to thank Dana Cronin for being a constant source of encouragement and support in the last stage of writing my dissertation. Our time spent together has been a source of peace and calm for me that made the lows of graduate school so much more bearable and the highs of graduate school so much more sweet.

Many parts of this work made use of the Illinois Campus Cluster, a computing resource that is operated by the Illinois Campus Cluster Program (ICCP) in conjunction with the National Center for Supercomputing Applications (NCSA) and which is supported by funds from the University of Illinois Urbana-Champaign. The following software libraries were used at various stages in the analysis for this work: `GSL` [1], `numpy` [2], `scipy` [3], `filltex` [4], `matplotlib` [5], `corner` [6], `Astropy` [7, 8], `Eigen` [9], `ADOL-C` [10], `OpenMP` [11], `FFTW` [12] and `Mathematica` [13].

This work was supported by many funding organizations, including support through NSF grants No. PHY-1759615 and No. PHY-1949838, NASA grants No. NNX16AB98G, No. 80NSSC17M0041 and No. 80NSSC18K1352, NASA ATP Grant No. 17-ATP17-0225, Simons Foundation through Award number 896696, and the Center for AstroPhysical Surveys (CAPS) at the National Center for Supercomputing Applications (NCSA), University of Illinois Urbana Champaign through the ICASU/CAPS Graduate Fellowship.

Table of Contents

List of Tables	ix
List of Figures	xii
Chapter 1 Introduction	1
1.1 Executive Summary	8
Chapter 2 Probing Screening and the Graviton Mass with Gravitational Waves	12
2.1 Introduction	12
2.2 Gravitational Wave Propagation in Screened Theories	15
2.2.1 Vainshtein Screening	15
2.2.2 Modifications to GW Propagation Speed	18
2.2.3 Phase Modifications from Finite λ_g	21
2.3 Fisher Analysis and Computational Framework	23
2.3.1 Fisher Information Matrix	23
2.3.2 Waveform Model	24
2.4 Projected Constraints on Screened Massive Gravity	26
2.4.1 Theory Agnostic	26
2.4.2 dRGT and bigravity	29
2.5 Conclusions and Future Work	29
Chapter 3 Probing Fundamental Physics with Gravitational Waves: The Next Generation 32	
3.1 Introduction	32
3.2 Detector Networks	41
3.2.1 Estimated Timeline	41
3.2.2 Estimated Sensitivity	44
3.2.3 Estimated Location	46
3.3 Statistical Methods for Population Simulations	46
3.3.1 Terrestrial Detection Probability	47
3.3.2 Space Detection Probability	50
3.3.3 Waveform Model for Population Estimates	51
3.4 Population Simulations	52
3.4.1 Population Models	52
3.4.2 Detection Rate Calculations	55
3.4.3 Synthetic Catalog Creation	56
3.5 Parameter Estimation	60
3.5.1 Basics of Fisher Analysis	60
3.5.2 Waveform Model for the Fisher Analysis	62
3.5.3 Numerical Implementation	63
3.6 Tests of General Relativity	64
3.6.1 Constraints on Generic Modifications	64
3.6.2 Specific Theories	74

3.6.3	Effect of Precession on the Constraints	94
3.7	Conclusions	98
Chapter 4 Are Parametrized Tests of General Relativity with Gravitational Waves Robust to Unknown Higher Post-Newtonian Order Effects?		
4.1	Introduction	101
4.2	Testing GR with GWs	102
4.2.1	Bayesian Analysis of GW Data	107
4.2.2	Current Parametric Tests of GR	109
4.3	Improved Parametric Tests of GR	111
4.3.1	Restrictions Based on Mathematical Structure of Modified Theories	112
4.3.2	Implementation of Improvements in Parametric Inspiral Tests	114
4.4	Experimental Design	117
4.5	Bayesian Results	120
4.6	Alternative Parametrizations	125
4.7	A Concrete Example: ssGB theory	129
4.8	Conclusions	135
Chapter 5 Fundamental Physics Implications on Higher-Curvature Theories from the Binary Black Hole Signals in the LIGO-Virgo Catalog GWTC-1		
5.1	Introduction	138
5.2	Quadratic gravity	140
5.3	Order of magnitude constraints	141
5.4	Fisher-estimated constraints on LIGO-Virgo data	143
5.5	Bayesian-estimated constraints on LIGO-Virgo data	143
5.6	Fundamental physics implications	145
Chapter 6 Improved gravitational-wave constraints on higher-order curvature theories of gravity		
6.1	Introduction	147
6.2	Basics of Quadratic Gravity	151
6.3	Constraints on Quadratic Gravity	153
6.3.1	Bayesian inference in GW science	153
6.3.2	Selection of Sources	156
6.3.3	Fisher analysis predictions	157
6.3.4	Bayesian analysis: single events	160
6.3.5	Bayesian analysis: stacked events	161
6.4	Robustness of Constraints	164
6.4.1	Changing the GR base model	164
6.4.2	Adding higher-order PN order GR modifications	166
6.5	Conclusions and Future Directions	171
Chapter 7 Conclusions		
Appendix A Bayesian Theory and Fisher Analysis Details		
Appendix B Mapping to Specific Theories		
B.1	Dipole Radiation	181
B.2	Black Hole Evaporation	182
B.3	Local Position Invariance Violation	183
B.4	Parity Violation	183
B.5	Lorentz Violation	184
B.6	Modified Dispersion	184

Appendix C	Inspiral/merger/ringdown vs. inspiral waveforms	186
Appendix D	Concerns over source classification	188
Appendix E	Methods for combining posteriors	190
Bibliography		192

List of Tables

2.1	Properties of the binary systems considered in Fig. ?? to estimate projected bounds on the graviton mass and the screening radius. The parameter A is a factor used to re-scale the y-axis of the figure.	15
2.2	Parameters of the models used in this study and the subsequent predicted bounds on β for each model. All quantities are in the source frame. The bounds on β are a 1σ bound. The sources for aLIGO were picked to emulate previous detections by aLIGO/VIRGO [139, 188–192]	27
2.3	Projected 1σ bounds for m_g for dRGT- and bigravity-type screening radii. The second column shows the percent change in the constraint on m_g between the screened and unscreened constraints, while the screening radius r_V is measured in Mpc, and the graviton mass in eV. .	30
3.1	Summary of the constraints we predict on the theory-agnostic ppE modification parameter β as a function of the PN order parameter b , as defined in Eqs. (3.25) and (3.26) below. We compare these constraints against current constraints from pulsar tests [49] and GW observations from the LVC [57], denoted by (*). The LVC analysis used a slightly different formalism, so we mapped their results to the ppE framework for 4 specific sources (GW150914, GW170104, GW170608, and GW170814), we computed the standard deviation of the Markov Chain Monte Carlo (MCMC) samples, and then combined the posteriors assuming a normal distribution to obtain a rough order-of-magnitude estimate of current ppE bounds from the LVC results. The columns list, from left to right: the PN order of each particular modification, the current constraint (if one exists), the best and worst constraints from our simulations, and the class of astrophysical sources those constraints come from. All the constraints are 1σ bounds, and we only show worst-case constraints that still improve on existing bounds. The source class acronyms are as follows: MB stands for multiband observations of SOBHs, T stands for terrestrial-only observations of SOBHs, and MBH stands for space-based detection of MBHs.	37
3.2	Summary of forecasted constraints on specific modifications of GR. The source class acronyms are the same as in Table 3.1. A (*) symbol denotes constraints coming from previous BBH observations, as opposed to other experimental evidence. When necessary, we have mapped all existing constraints to 1σ constraints by assuming the posterior to be normally distributed. We only show worst-case constraints that improve on existing GW bounds. For consistency with previous work, $\dot{M}$ is given in units of $M_\odot/\text{yr}$, while we use geometrical units (so that $\delta\dot{E}$ is dimensionless) for the generic dipole radiation bound. Note that the necessary factor for transforming between the two is $c^3/G = 6.41 \times 10^{12} M_\odot/\text{yr}$. The time derivative of the gravitational constant, $\dot{G}$, is normalized to the current value of G , and it does indeed have units of yr^{-1} in geometrical units (where $G = c = 1$).	39
3.3	The above timeline tabulates the exact terrestrial detector evolution utilized by this study. There is a single timeline of detectors until 2035, when we model three separate scenarios that could play out in the next three decades: Scenario 1, 2, and 3. A graphical representation is shown in Fig. 3.1. The various sensitivity curves in column 3 are shown in Fig. 3.2.	42
3.4	Detector locations used in this paper.	45

3.5	Configurations used at each stage of our analysis to calculate the probability of detection for a given binary for the terrestrial detector network. Note that networks involving multiple detectors are labelled by the network nodes and not just their number, because the relative position of the detectors impacts the calculation of the detection probability. Our calculation depends on the assumption that all the detectors have approximately the same sensitivity curve, and so the curve used at each stage is given in the last column. Because of this assumption, and the extreme disparity in sensitivity between second- and third-generation detectors, we only use the CE detector to calculate rates when CE is part of the network.	48
3.6	Detection rates for the detector networks and population models examined in this study. For SOBH populations, the first number in the parentheses is the detection rate for the terrestrial-only network (neglecting LISA), while the second number is the detection rate for multiband events seen in both the terrestrial network and LISA. For MBH populations, we show the detection rate for LISA for the PopIII, light-seeding scenario, as well as for the Q3, heavy-seeding scenario. In the case of Q3, the first number in parentheses corresponds to delayed mergers (Q3delays) and the second number to the nondelayed version (Q3nodelays).	59
3.7	A summary of the theories examined in this work (adapted and updated from [161, 210]). The columns (in order) list the theory in question (unless a generic deviation is being examined), the physical interpretation of the modification, the way the modification is introduced into the waveform, the PN order at which the modification is introduced, the equation specifying the ppE-theory mapping, and the b parameter in the ppE framework. The practical ramifications between “generation” vs “propagation” effects relates to how the modification is introduced into the waveform, as explained in Appendix C.	61
4.1	Choices of GW sources for injection campaigns. The source parameters are the following: m_1 and m_2 are the masses of the larger and smaller black holes, respectively, D_L is the luminosity distance from Earth to the source, χ_i is the aligned, dimensionless spin of the i -th black hole, SNR_{2g} and SNR_{3g} are the SNRs of the source as measured by a 2g and 3g detector network respectively. The “heavy” source and the light “source” both have an SNR of 20 (for the 2g detector network), but have different total masses and different spin configurations. All source parameters were kept the same between the analyses involving the 2g and the 3g detector networks.	120
4.2	Marginalized 1σ constraints on the leading PN order deformations for the “heavy” source (left) and the “light” source (right) using the 2g network (see Sec. 4.4 for details). The columns represent the PN order (relative to the Newtonian term in GR) at which the GR deformation is first introduced. The rows corresponds to the number of PN corrections that are added on top of the leading PN order one in ascending PN order. The functional form of the phase is given in Eq. (4.13), where the priors are presented in Sec. 4.3. Observe the (order of magnitude) consistency of the constraints as one includes more and more sub-leading PN order deviations (i.e. as one moves down the column for a fixed leading-order term). This indicates that higher PN order corrections to the current modified gravity ppE waveforms will not invalidate current bounds placed on modified theories with only leading PN order deformations.	122
4.3	Same as Table 4.2, but for a 3g network, using the “heavy” source (top) and the “light” source (bottom). Once more, the trends found with the 2g detector networks continue when considering a 3g detector network.	126
5.1	Current constraints on EdGB and dCS gravity from low-mass x-ray binary and Solar System observations respectively, with the Fisher-estimated constraints, and Bayesian constraints using LVC (testing GR) posteriors for GW151226 and GW170608 [57, 324]	143
6.1	Current state of quadratic theories of gravity and the bounds resulting from this work. The columns are, in order, the theory, the topological invariant that couples non-minimally to the scalar field, the constraints as the result of this work, and the previous strongest constraints on these theories.	148

6.2	Results from a Fisher projection, calculated using the median values for the source parameters from an MCMC analysis within GR. The results show the 90% confidence interval for the root of the coupling parameter in km for each source. The values in parenthesis are the upper limits on the validity of the small coupling approximation (6.6). The last row of the table shows the combined constraint assuming each posterior is normally distributed, consistent with the assumptions of a Fisher analysis.	160
-----	---	-----

List of Figures

2.1	(Color Online) Projected constraints on the graviton mass as a function of the screening radius, assuming the detection of three expected GW sources with aLIGO at design sensitivity, ET, and LISA. As the sensitivity to these screening effects varies widely for each detector, the y-axis is re-scaled appropriately through a factor A presented in Table 2.1. The effectiveness of screening becomes very apparent when the screening radius is about $D_L/2$, since then the constraint on m_g rapidly falls to zero.	14
2.2	Schematic of the binary/detector system with various distances labeled for clarity. z is the redshift to the source, and z_s and z_{MW} are the redshifts to the edges of the screening effects for the host galaxy and the Milky Way, respectively. $r_{V,h}$ ($r_{V,MW}$) is the Vainshtein radius for the host galaxy (Milky Way). $D_{V,MW}$ is the distance from the source to the edge of the Milky Way Vainshtein regime, equal to the luminosity distance to the source minus the Vainshtein radius of the Milky Way ($D_L - r_{V,MW}$)	21
2.3	(Color Online) Projected constraint on the mass of the graviton as a function of the Vainshtein radius for a variety of systems detected with aLIGO (top), ET (middle) and LISA (bottom). The shaded regions would be disallowed given injections consistent with GR.	28
3.1	Graphical representation of Table 3.3. The shaded regions in the figure represent periods of active observation, and the colors/hatching corresponds to the noise curve being implemented, as shown in Fig. 3.2.	40
3.2	Noise curves for the various detector configurations studied in this work. The shaded bands observed for the Virgo+ phase 2 and KAGRA sensitivities reflect uncertainties in estimates of their anticipated power spectral densities.	41
3.3	Detection probability p_{det} for the four networks examined in this paper. The black curve is for a single detector (where global position no longer matters, so this is valid for any single right-angle Michelson interferometer). The blue curve is specifically for the Hanford, Livingston, and Virgo (HLV) network. The red curve is for the Hanford, Livingston, Virgo, and KAGRA (HLVK) network. Finally, the green curve represents a network comprised of CE and ET (which includes all three of the ET detectors as well as the 60° angle between each set of arms).	47
3.4	Distributions of the different source properties detected by each network. For each detector network, labeled across the y-axis, we plot the distribution of the total detector-frame mass $M_z = M(1+z)$, mass ratio $q = m_2/m_1 < 1$, redshift z , and SNR ρ in log-space (base 10). Each plot is split, with the upper (grey) half coming from the $\sigma = 265$ km/s SPOPS simulations, and the lower (green) half coming from the $\sigma = 0$ km/s simulations.	53
3.5	Distributions of the different MBH binary source properties detected by LISA. For each MBH binary simulations, labeled across the y-axis, we plot the distribution of the total detector-frame mass $M_z = M(1+z)$, mass ratio $q = m_2/m_1 < 1$, redshift z , and SNR ρ in log-space (base 10). Each plot is split in two, with the upper (grey) half corresponding to a “nominal” four-year LISA mission, and the lower (green) half corresponding to an extended ten-year mission. . .	54

- 3.6 Properties of detected merger events for various detector networks and population models. The left panels refer to terrestrial-only sources, while MBHs and multiband sources are shown on the right. The points and thick lines show the mean values, while the shaded regions and error bars encompass the optimistic and pessimistic scenarios. The assumed detector network is shown in the top x-axis (using the notation of Table 3.3), while the corresponding years are shown on the bottom x-axis. The top panels show the rates of detected mergers for each class of sources; circles refer to the PopIII MBH population. The middle panels show the cumulative number of observed sources: here the three different multiband scenarios are identical, as the choice of terrestrial network has little impact on the number of multiband sources we can detect [204]. The bottom panels show the average $\log_{10}$ SNR. Here the lower (upper) bounds correspond to subtracting (adding) the standard deviation to the mean value of the most pessimistic (optimistic) scenario. 58
- 3.7 Constraints on modifications to GR at various PN orders as a function of time. The colors represent different classes of populations (including SOBH terrestrial-only sources, SOBH multiband sources, MBH sources from the Q3 heavy-seed scenario, and MBH sources from the light-seed PopIII scenario). The bands in all of these scenarios – except for PopIII – correspond to astrophysical uncertainties: kick velocities $\sigma = 265$ km/s and $\sigma = 0$ km/s give the upper and lower bounds for SOBHs, while the inclusion of delays affects Q3 scenarios. Greyscale patches at the top of each panel correspond to the observation period for each network, labeled across the top. Multiband sources and MBHs yield strong constraints at negative PN orders. Terrestrial-only SOBH sources begin to contribute substantially at positive PN orders for all detector networks, with the optimistic scenario S1 yielding the best constraints. We overlay as horizontal lines the most stringent current bounds, where available and competitive, from pulsars [49] and LVC observations of GWs [57]. 67
- 3.8 Scaling relations discussed in Sec. 3.6.1. The ratio $\sigma_{\beta\beta}^{\text{TERR}}/\sigma_{\beta\beta}^{\text{MBH}}$, calculated from the full Fisher simulations including the realistic noise curves shown in Fig. 3.2 and the IMRPhenomPv2 waveform, is shown in blue. The empirically measured trend is derived from averaging the constraints from each terrestrial network and each population model, then calculating the ratios of every combination of terrestrial network and SOBH model against each MBH heavy-seeding model. The blue line shows the mean ratio, and the blue shaded region is the area bounded by the maximum and minimum ratios. The red line and the red shaded region refer instead to the ratio between the terrestrial-only constraints and the multiband constraints, i.e. $\sigma_{\beta\beta}^{\text{TERR}}/\sigma_{\beta\beta}^{\text{MB}}$. For this class of sources, we calculate the ratio for each population model and detector network, one at a time. That is, the terrestrial-only constraints from the S1 network derived from the SPOPS 265 model are compared against the multiband constraints from the S1 network and the SPOPS 265 model. The trends predicted analytically in the text are shown in black and grey for MBH and multiband sources, respectively. The trend lines we show for our predictions have been shifted along the y-axis to better compare the with the data. 69
- 3.9 Empirically determined values of N_{eff} for the CEK (Scenario 1) network and the SPOPS 0 catalog, derived from our full Fisher analysis, including the noise curves shown in Fig. 3.2 and the IMRPhenomPv2 waveform. The parameter N_{eff} is defined as the number of sources needed from the full catalog in order to achieve a threshold constraint $\sigma_{\beta,\text{thr}}$, using the most constraining sources first. Here we choose $\log_{10} \sigma_{\beta,\text{thr}} = 0.95 \log_{10} \sigma_{\beta}$, where σ_{β} is the cumulative bound from the full Fisher analysis for the entire catalog. The values of the threshold constraint (blue + signs) are shown alongside the full constraint (red × signs) in the lower panel. The number of required sources grows exponentially as a function of PN order: large catalogs benefit positive PN orders, but they are not as important for highly negative PN orders. . . . 70

3.10	Three different distributions in the $\log_{10} \mathcal{M} - \rho$ plane for the CEK network and the SPOPS 0 population model. The blue heat map shows the distribution of the sources directly in the $\log_{10} \mathcal{M} - \rho$ plane, and it is the same for all PN orders. The black contours show the constraints from individual sources. The red scatter plots show the sources needed to obtain a threshold cumulative constraint $\log_{10} \sigma_{\beta, \text{thr}} = 0.95 \log_{10} \sigma_{\beta}$, where the shade of red indicates the strength of the individual bounds (in log base 10). We utilized a 2σ gaussian filter over the data to smooth out the noise and create more easily interpretable contour plots. In conjunction with Fig. 3.9, the growing number of scatter points as a function of PN order illustrates the increasing dependence of the cumulative constraint on the size of the source catalog. Furthermore, the relation between chirp mass, SNR, and individual bound can be seen to shift significantly between positive and negative PN orders, agreeing with the commonly held intuition that lower-mass sources are better for constraining negative PN effects. In more detail, the negative PN orders benefit highly from low-mass systems, with slight dependence on SNR, while positive PN order effects depend much more strongly on the SNR and have more minimal dependence on the chirp mass. Finally, the range of individual bounds (~ 4 orders of magnitude at negative PN orders and ~ 2 orders of magnitude at positive PN orders) helps to explain the different scaling relations between the cumulative bounds and the total number of sources.	71
3.11	Projected cumulative constraints on generic dipolar radiation for the detector networks and population models examined in this paper. The multiband sources outperform all other source classes by at least ~ 2 orders of magnitude, with MBH sources and the most optimistic terrestrial scenario performing comparably.	75
3.12	Result of the scaling analysis outlined in Sec. 3.6.2 performed on the data synthesized with the HLVKIO8 network and the SPOPS 0 population. The left panel shows a heat map of the constraint on $\delta\dot{E}$ versus the SNR of the source. The solid blue lines correspond to the strongest and weakest single-source constraint, and the area between these two bounds is shown in hatching. The cumulative bound from the entire catalog is shown as the solid green line. The power-law fit to the data in the left panel is shown as the solid black curve, and our prediction for the scaling is shown as the solid red curve. The right panel shows three distinct slices of the catalog, with ranges in SNR from 10 to 11 (blue), 20 to 22 (green), and 50 to 55 (red). These ranges are highlighted in the left panel. The right panel shows the density of the constraint versus the chirp mass, with empirical trends shown in black and predicted trends shown in red. There is a noticeable transition point in the distribution, so low-mass and high-mass systems were analyzed separately. The powers used in all trend lines are shown in the legend. For trend lines, the (logarithmic) offset for the predicted scaling relations has been adjusted to coincide with the empirically fit offset, to better compare the slopes of the trends. Of particular interest is the strong trend relating the SNR and the bound, as well as the tight correlation between chirp mass and constraint for low-mass systems, which seems to taper off for high-mass systems.	76
3.13	Projected cumulative constraints on the time derivative of the gravitational constant $\dot{G}$ for the detector networks and population models examined in this paper. Multiband sources outperform all other source classes by $\sim 1 - 2$ orders of magnitude, with MBH sources performing the next best. SOBHS observed by the terrestrial network alone perform the worst, but with Scenario 1 outperforming Scenarios 2 and 3 due to the high SNR of the observations in the former network.	78
3.14	Result of the scaling analysis outlined in Sec. 3.6.2 performed on the data synthesized with the HLVKIO8 network and the SPOPS 0 population. The plotting style is the same as in Fig. 3.12. The left panel shows a heat map of the constraint on $\dot{G}$ versus the SNR of the source. The right panel shows the density of the constraint versus $\mathcal{M}$, with empirical trends shown in black and predicted trends shown in red. Again, the strong trend relating the SNR and the bound agrees well with the prediction, and there seems to be a tight correlation between $\mathcal{M}$ and constraint, well approximated by our analysis in Sec. 3.6.2.	79

3.15	Projected cumulative constraints on $\sqrt{\Lambda}$ for the detector networks and population models examined in this paper. Terrestrial-only catalogs, with their populations of millions of sources, seem to dominate any future constraint on this particular deviation, with an improvement by 1–2 orders of magnitude over any other source classification. This conclusion seems independent of the particular terrestrial scenario we pick, with comparable performance from all three.	81
3.16	Result of the scaling analysis outlined in Sec. 3.6.2 performed on the data synthesized with the HLVKIO8 network and the SPOPS 0 population. The plotting style is the same as in Fig. 3.12. The left panel shows a heat map of the constraint on $\sqrt{\Lambda}$ versus the SNR of the source. The right panel shows the density of the constraint versus the chirp mass, with empirical trends shown in black and predicted trends shown in red. The small range of constraints from the catalog lead to considerable enhancements of the cumulative bound when stacking observations, and the weak scaling with chirp mass and moderate scaling with SNR further benefit SOBH sources over other source classes.	82
3.17	Projected cumulative constraints on $\sqrt{\alpha_{\text{dCS}}}$ for the detector networks and population models examined in this paper. Terrestrial-only catalogs, with their populations of millions of sources, dominate any future constraint on this particular deviation, with an improvement of 2-5 orders of magnitude over other source classification. This conclusion is independent of the terrestrial scenario we pick, with comparable performance from all three. Multiband sources, with their low chirp masses, seem to perform the next best.	83
3.18	Result of the scaling analysis outlined in Sec. 3.6.2 performed on the data synthesized with the HLVKIO8 network and the SPOPS 0 population. The plotting style is the same as in Fig. 3.12. The left panel shows a heat map of the constraint on $\sqrt{\alpha_{\text{dCS}}}$ versus the SNR of the source. The right panel shows the density of the constraint versus the chirp mass, with empirical trends shown in black and predicted trends shown in red. Our prediction for the SNR scaling is considerably less accurate than for previous theories, presumably from covariances with other source parameters and competing scaling trends with the chirp mass. The tight range of constraints and large improvement of the cumulative bound over all other single source constraints, seen in the left panel, indicate strong dependence on the total number of sources in the catalog.	84
3.19	Histogram of spin-related terms contributing to the relevant Fisher element for dCS and EdGB. The sources were taken from the catalog derived from the HLVKIO8 network and SPOPS 0 population model. For dCS, this only includes the term to first order in spin. The wide range of magnitudes that this term can take (5–6 orders of magnitude) helps to explain the breakdown of our ability to predict trends concerning the constraints on these theories. From Fig. 3.18 we see that the SNR and chirp mass only span a range of 1 or 2 orders of magnitude, and as such, the trends we would expect to see for these parameters could be completely washed out by this additional spin-dependent term, which we have neglected in our simple analysis.	86
3.20	Projected cumulative constraints on $\sqrt{\alpha_{\text{EdGB}}}$ for the detector networks and population models examined in this paper. Terrestrial-only catalogs, with their populations of millions of sources, seem to most efficiently constrain EdGB, but multiband sources are not far behind. The modified scaling of the constraint with SNR and chirp mass work in favor of terrestrial networks, but the fact that EdGB produces a negative PN modification to leading order benefits multiband sources. MBHs are not effective at constraining EdGB, and will not contribute much to future bounds on this theory.	87

3.21	Result of the scaling analysis outlined in Sec. 3.6.2 performed on the data synthesized with the HLVKIO8 network and the SPOPS 0 population. The plotting style is the same as in Fig. 3.12. The left panel shows a heat map of the constraint on $\sqrt{\alpha_{\text{EdGB}}}$ versus the SNR of the source. The right panel shows the density of the constraint versus the chirp mass, with empirical trends shown in black and predicted trends shown in red. Because of the small range in single-observation constraints (about 1–2 orders of magnitude), the cumulative bound greatly benefits from large numbers of observations, despite this being a negative PN effect that would typically be dominated by a small cadre of favorable sources. The predicted trend for the constraint-SNR relationship fails, presumably due to covariances introduced through the Jacobian. The predicted trend for the constraint- $\mathcal{M}$ relationships performs fairly well, as the correlation is enhanced through the Jacobian.	88
3.22	Projected cumulative constraints on the rate of black hole evaporation $\dot{M}$, for the detector networks and population models examined in this paper. Our models predict multiband sources to perform the best from the three classes of sources examined in this paper, followed next by MBH observations by LISA. Terrestrial-only observations from the most optimistic scenario are competitive with LISA’s MBH sources, but the other two scenarios considered in this work trail behind by 2-3 orders of magnitude.	90
3.23	Result of the scaling analysis outlined in Sec. 3.6.2 performed on the data synthesized with the HLVKIO8 network and the SPOPS 0 population. The plotting style is the same as in Fig. 3.12. The left panel shows a heat map of the constraint on $\dot{M}$ versus the SNR of the source. The right panel shows the density of the constraint versus the chirp mass, with empirical trends shown in black and predicted trends shown in red. The wide distribution of constraints in this catalog indicate that the benefit of large catalogs is minimal, and the total bound is dominated by a select few, highly favorable observations. The distribution of the sources in the $\Delta\dot{M}$ - $\mathcal{M}$ plane is to a very good approximation linear, showing a tight correlation between the two quantities. The $\Delta\dot{M}$ -SNR relationship also agrees fairly well with our predictions.	91
3.24	Projected cumulative constraints on the mass of the graviton, m_g , for the detector networks and population models examined in this paper. Our models show that MBH sources observed by LISA will perform the best at constraining this modification, but only slightly better than the terrestrially-observed only sources. Multiband sources perform the worst, as they received no benefits from the Jacobian and already perform only moderately well for positive PN order effects.	92
3.25	Result of the scaling analysis outlined in Sec. 3.6.2 performed on the data synthesized with the HLVKIO8 network and the SPOPS 0 population. The plotting style is the same as in Fig. 3.12. The left panel shows a heat map of the constraint on m_g versus the SNR of the source. The right panel shows the density of the constraint versus the redshift z , with empirical trends shown in black and predicted trends shown in red. Because of the narrow range of constraints in the catalog and the large enhancement of the cumulative bound over the strongest single observation, stacking observations is quite efficient for this modification. The right panel shows that there is indeed a trend in the Δm_g - z relation (although the distributions are moderately wide) which would favor sources far from Earth, and would primarily benefit MBH sources.	93
3.26	Distributions of single-source constraints on the GR-modifying parameters $\sqrt{\alpha_{\text{dCS}}}$ (top) and $\sqrt{\alpha_{\text{EdGB}}}$ (bottom) from the two population models SPOPS 0 (blue) and SPOPS 265 (orange) as detected by the CEK network. The histograms are normalized to provide a comparison of the shapes of the distributions, as opposed to the raw numbers of sources. We see that the distributions only diverge slightly, towards the larger-constraint side of the spectrum. This suggests that the larger precessional effects seen in the SPOPS 265 catalog do not significantly modify the typical constraints attainable by individual sources, or that any effect we may have seen was washed out by the differences in the distributions of other source parameters, such as the total mass and mass ratio. This lack of difference could also be an artifact of our waveform model (IMRPhenomPv2), which is not the most up-to-date waveform available, or of the Fisher approximation, which could be improved upon by a full MCMC analysis.	96

3.27	To create the data involved in this figure, we have created a set grid in parameter space with total mass ranging from $5 M_{\odot}$ to $20 M_{\odot}$, mass ratio in the range 0.05 to 1, and aligned-spin components for each binary ranging from -0.8 to 0.8 . The rest of the parameters were populated with random numbers in the usual ranges, and the luminosity distance was set such that the typical SNRs ranged from ~ 20 – 150 . We computed Fisher matrices for each set of parameters, with the in-plane component of the spin set to 0, and then we recomputed them setting the in-plane component of the spin $\chi_p = \sqrt{1 - \chi_1^2}$, so that the binary is approximately “maximally” precessing. The top panel shows the distribution of the bounds for the two binary subsets – precessing (blue) and non-precessing (green) – as a function of PN order. The solid line denotes the average of the synthetic catalog, while the shaded region denotes the 1σ interval. The lower panel shows the ratio $\sigma_{\beta, p}/\sigma_{\beta, \text{nonp}}$. Each ratio is calculated for a single parameter set, and the mean of these ratios is shown as a solid black line, with the 1σ spread shown by the shading. Even in this more extreme comparison, the improvement in constraint as the result of larger precession effects only amounts to a factor of ~ 2 . However, more drastic difference may be possible if we performed a full MCMC analysis, or if we used different waveform models.	97
4.1	Ratio of the PN coefficients of the Fourier GW phase (Eq. (4.5)) for a non-spinning binary as functions of the symmetric mass ratio η . The ratios are evaluated at an orbital separation of $r_{\text{eval}} = 100m$, corresponding to $v_{\text{eval}} = 0.1$. Each line corresponds to a different pair of coefficients, and the dotted black line simply identifies the threshold we use for convergence ($ p_i/(p_{i+1}u_{\text{eval}}) > 1$). The spikes shown here correspond to when the coefficient $p_6 = 0$ at $\eta \sim 0.053$. Observe that all ratios are above unity for symmetric mass ratios $\eta > 0.09$, indicating that our choice of v_{eval} has led to a convergent structure of PN coefficients.	116
4.2	The one dimensional, marginalized prior on a set of deformations. The leading order deformation enters at Newtonian order relative to GR, and each of the subsequent deformations are labeled as the $i/2$ PN order relative to this leading order deformation. The priors for this parametrization are non-trivial, and can be misleading. They appear to disallow the GR limit $\delta\bar{\psi}_i = 0$ for several of the different PN orders. This is resolved when looking at the transformed priors of the right panel, rewritten in terms of the absolute modifications, $\bar{\Delta}_i$	118
4.3	Marginalized 1σ constraints on the leading PN order deformations for the “heavy” source (left) and the “light” source (right) using the 2g network (see Sec. 4.4 for details) as a function of the number of sub-leading PN order terms added in the modified sector (in ascending order). Each line corresponds to modifications that start at different leading PN order. The functional form of GW phase is given in Eq. (4.13), and the priors used are those presented in Sec. 4.3. The top panels show the 1σ constraint on the leading PN order deformation, while the lower panel shows the strengthening factor, as defined in Eq. (4.21). Observe how the lines in the upper panels trend downward, which means that constraints with just a single parameter deviation are <i>conservative</i> (i.e. adding higher PN order corrections strengthens the constraint obtained with a single parametric deviation). Moreover, observe how the slope of the lines are small, which means that the strengthening of the constraint is mild, with improvements of at most roughly one order of magnitude.	121
4.4	Marginalized 1σ constraint on $\bar{\gamma}$ for a modification that first enters at 0PN order as a function of number of PN terms kept in the modified sector, using 3 different choices of $r_{12, \text{eval}}$ for the PN prior. In all cases, we here focus on the light source and the 2g detector network. Observe that in all cases the constraint on $\bar{\gamma}$ becomes stronger the more PN order terms are added. Observe also that the larger we choose $r_{12, \text{eval}}$ to be, the stronger the “pressure” on $\bar{\gamma}$ and thus the stronger the constraint.	124
4.5	Same as Fig. 4.3, but for a 3g network, using a “heavy” source (left) and a “light” source (right). Observe that the trends found with the 2g detector networks continue when considering 3g detector networks.	125

4.6	The one dimensional, marginalized prior on a set of deformations in the three models considered in Sec. 4.6. The leading-order deformation enters at Newtonian order relative to GR, and each of the subsequent deformations are labeled as the $i/2$ PN order relative to this leading order deformation. Included in this figure is the original model described in Sec. 4.3, shown as a solid blue line. The first alternative model uses the $\delta\psi_i$ parametrization described in Eq. (4.13) but with simple, uniform ranges with fixed (increasingly larger with higher PN order) boundaries, shown as a dotted orange line. The second alternative model is described by a series of $\bar{\Delta}_i$ parameters with uniform distributions with fixed (increasingly larger with higher PN order) boundaries, defined in Eq. (4.19), shown as a dashed green line. Note that the range of the prior on the $\delta\psi_i$ parameters in the second alternative parametrization extends beyond the frame of the figure, but the range was restricted for visual purposes.	127
4.7	The final, marginalized posterior distributions on $\bar{\gamma}$ coming from the three models discussed in Sec. 4.6 using the synthetic data from the “light” source as observed by the 2g detector network. The top panel shows a larger range of $\bar{\gamma}$ and the lower panel zooms-in the range of $\bar{\gamma}$ to present details in the two narrower distributions. The probabilities on the y-axis are normalized to the shown range of $\bar{\gamma}$ in both panels. The distribution coming from the model using a single deformation parameter at Newtonian order is shown as a solid blue line. The distribution coming from the model described by the first alternative parametrization with six deformations (with a prior uniform in $\delta\psi_i$) is shown as the dotted orange line. The distribution coming from the model described by the second alternative parametrization with six deformations (with a prior uniform in $\bar{\Delta}_i$) is shown as the dashed green line. The distribution coming from the model described by the original parametrization with six deformations (with a prior that strictly enforces our notion of convergence) is shown as the dotted-dashed red line. Note that the first alternative parametrization and the original parametrization are almost totally overlapping in the top panel.	130
4.8	Absolute value of the terms in the phase deformation at different PN order (left) and ratio of terms (right) in ssGB as a function of symmetric mass ratio, evaluated at $v_{\text{eval}} = 0.1$. Observe that, for this choice of v_{eval} , the leading PN order term is larger than the next-to-leading order one, which in turn is larger than the next-to-next-to-leading order, except at $\eta \sim 0.19$, where the 1PN term vanishes identically. Observe also that the ratios are all larger than unity, except again for a specific value of η at which the 1PN term vanishes.	132
4.9	For reference, we show the two distributions on $\bar{\gamma}$ coming from the two generic deformation models (models 3 and 4). This figure shows that the distributions on the generic parameter $\bar{\gamma}$ for these two models are comparable, with a slight enhancement when using the 6 deformation variation. This same information is shown graphically in Fig. 4.3 and tabulated in Table 4.2. When compared with Fig. 4.10, we see that this relation persists. Even once the posteriors are transformed to theory-specific constants, constraints are not meaningfully changed when using these two models. Furthermore, note that the posterior on $\bar{\gamma}$ is fully consistent with GR (the model that truly describes the injected data), which is further evidence that the deviation away from GR in Fig 4.10 for the generic models is a artifact of the transformation.	133
4.10	Marginalized posterior distributions on $\sqrt{\alpha}$ for a GR injection extracted with the four models described in the text. We here considered the “light” source, defined by Table 4.1, and a 2g detector network. When carrying out the Bayesian studies with the multi-parameter ppE model, we employed the PN prior with $r_{12,\text{eval}} = 100$, as done in the rest of this paper. The constraints on $\bar{\gamma}$ obtained with models 3 and 4 were mapped to constraints on $\sqrt{\alpha}$ to enable comparisons with the results obtained with models 1 and 2. Observe that clearly the posterior distributions are all consistent with each other (modulo singularities in the transformation at $\alpha = 0$, discussed in detail in [123, 124]). This shows clearly that in ssGB theory, leading PN order constraints are conservative and sufficient to place bounds on the theory.	134

- 4.11 Marginalized posterior distributions on $\sqrt{\alpha}$ for a GR injection extracted with the two generic models described in the text. One of the models uses a single deformation while the second model uses a series of six deformations. We here considered the “light” source, defined by Table 4.1, and a 2g detector network. These two data sets are models 3 and 4, described in the text and shown in Fig. 4.10. For comparison, we also show the derived prior on $\sqrt{\alpha}$, calculated by taking a uniform prior on $\bar{\gamma}$ and the usual priors on source parameters, then mapping it to $\sqrt{\alpha}$ with the same prescription as was used for transforming the two generic model constraints. This figure illustrates the interesting behavior around $\sqrt{\alpha} \rightarrow 0$ (the GR limit). The fact that the derived prior on $\sqrt{\alpha}$, as transformed from a uniform prior on $\bar{\gamma}$, disallows $\sqrt{\alpha} = 0$ explains the bias in the posteriors for models 3 and 4. The issue is related to the Jacobian of the transformation, discussed in the text, and is not of serious concern in analysis such as these. 136
- 5.1 Posterior distributions of $\alpha_{\text{dCS}}^{1/2}$ (left panel) and $\alpha_{\text{EdGB}}^{1/2}$ (right panel) obtained using GW151226 and GW170608. For the GW events shown in both panels, $m_2/M_\odot = 7.7^{+2.2}_{-2.6}$ (GW151226) and $m_2/M_\odot = 7.6^{+1.3}_{-2.1}$ (GW170104) at 90% credibility. This implies that the small-coupling approximation is valid only when $\alpha_{\text{dCS,EdGB}}^{1/2} \lesssim 5.6$, shown as vertical lines in the plots. For dCS gravity (left-panel) we see that most of the support of the posterior distributions of these two events lays *passed* the bounds set by the small-coupling approximation. Consequently, one cannot place constraints on $\alpha_{\text{dCS}}^{1/2}$ with these two events. For EdGB gravity (right-panel) most ($> 90\%$) of the posteriors’ support lays *within* the bound, therefore allowing us to constrain the theory with these two events. For the other three events, which contain a large m_2 ($\gtrsim 13 M_\odot$) BH [18], the vertical lines are pushed towards the left, leaving most of the posterior’s support outside the small-coupling approximation bound. We stress that the location of the peaks in the posteriors *are not an indication of a deviation from GR*. Instead, as detailed in the main text, the lack of support at zero is an artifact of the choice of the sampling variable $\delta\phi_i$. . . 144
- 6.1 The detector frame total mass and the effective inspiral spin, χ_{eff} , of the events analyzed in this work. These values were inferred by LIGO assuming GR correctly describes the gravitational sector [18, 351]. 158
- 6.2 The individual constraints on the coupling constant of EdGB, $\sqrt{\alpha_{\text{EdGB}}}$. Each panel shows the marginalized posterior distribution of the square root of the coupling constant in EdGB. Overlaid is the 90% confidence value of the coupling constant, shown as the vertical solid line, with the upper limit of validity for the small coupling approximation shown as a vertical dashed line. On top of the discrete histogram, the KDE approximation used to determine the joint distribution from all the sources is also shown as a solid curve. All six sources shown satisfy the small coupling approximation at 90% confidence, resulting in a robust bound on $\sqrt{\alpha_{\text{EdGB}}}$ 162
- 6.3 The individual constraints on the coupling constant of dCS, $\sqrt{\alpha_{\text{dCS}}}$. Each panel shows the marginalized posterior distribution of the square root of the coupling constant in dCS. Overlaid is the 90% confidence value of the coupling constant, shown as the vertical solid line, with the upper limit of validity for the small coupling approximation shown as a vertical dashed line. On top of the discrete histogram, the KDE approximation used to determine the joint distribution from all the sources is also shown as a solid curve. None of the sources shown satisfy the small coupling approximation, and as such, we can still not place a meaningful constraint on the dCS coupling parameter purely through GW observations. 162
- 6.4 In the bottom panel, we show the histogram representation of the probability density of the value of the EdGB coupling constant, $\sqrt{\alpha_{\text{EdGB}}}$, for the six events we have chosen for this analysis. The top panel shows the 90% confidence constraint on the magnitude of the coupling parameter and the maximum value for validity of the small-coupling approximation for each of the individual events, with the cumulative constraint shown at the top. After combining the information of all the sources, we can achieve a constraint on $\sqrt{\alpha_{\text{EdGB}}}$ of less than 1.7 km at 90% confidence. 163

6.5	In the bottom panel, we show the histogram representation of the probability density of the value of the dCS coupling constant, $\sqrt{\alpha_{\text{dCS}}}$, for the six events we have chosen for this analysis. The top panel shows the 90% confidence constraint on the magnitude of the coupling parameter and the maximum value for validity of the small-coupling approximation for each of the individual events, with the cumulative constraint shown at the top. As no single constraint satisfies the small-coupling approximation, our cumulative bound is also untrustworthy. We only show the combined posterior distribution for illustrative purpose.	163
6.6	Comparison of the constraints obtained through two separate analyses of the event GW151226. In the first iteration, we used a more robust template, IMRPhenomPv2 , which encodes information about precession. We then repeated the analysis using the base template IMRPhenomD , which only models spin-aligned and spin-anti-aligned binaries. Our results indicate that the constraints we have placed on the coupling constant of EdGB are robust to changes in the GR-template used to recover the parameters of the signal.	165
6.7	Comparison of priors for the different parametrization schemes outlined in Sec. 6.4.2. The left panel shows the priors of each parametrization transformed to the parametrization $\delta\phi_{\text{v1}}$, shown in Eq. (6.29). The middle panel shows each parametrization transformed to that shown in Eq. (6.30). Finally, the right panel shows the priors transformed to the parametrization shown in Eq. (6.32). From these figures, we can see that the first parametrization differs substantially from parametrizations 2 and 3, while parametrizations 2 and 3 are very similar.	167
6.8	The results of analyzing the GW151226 event with EdGB and three different versions of EdGB with some generic, higher PN order modification. The histograms in the lower panel show the posterior probability on the value of the EdGB coupling constant, $\sqrt{\alpha_{\text{EdGB}}}$ in km. All three iterations include the -1PN effect, and the three versions labeled GH0 (generic higher order) incorporate some generic modification at 0PN (relative to the Newtonian term in GR), outlined in the text. The upper panel shows the 90% confidence constraint on $\sqrt{\alpha_{\text{EdGB}}}$, and the maximum value the coupling constant can take and still satisfy the small coupling approximation. We see good agreement between these four methods, giving us confidence that our results in this work will be robust to any additional modifications that are derived in the future.	171
6.9	The characteristic strain compared against an example sensitivity curve representative of the LIGO detectors in O2. The contribution to the SNR from the merger is negligible for almost all the events, ensuring that our analysis is robust to our lack of knowledge of the highly nonlinear dynamics of the merger itself.	172
C.1	Comparison between the constraints on β at 1.5PN predicted by using the generation modification (INS), as opposed to the propagation modification (IMR). We used the same catalogs and networks (SOBH Base and SOBH S1) in both cases. The difference is negligible when considering the order-of-magnitude constraints of interest in this work.	187
D.1	The posterior distributions on $\sqrt{\alpha_{\text{EdGB}}}$ for the GW190814 event. The top panel was derived from an analysis that assumed GW190814 was a BBH system, while the lower panel shows results that were derived assuming the source is a NSBH system. The small coupling criteria for this source, when considering it as a NSBH source, is approximately 18 km, and lies beyond the edge of the plot. As is immediately apparent, the upper limit on $\sqrt{\alpha_{\text{EdGB}}}$ based on each of these assumptions differs by an order of magnitude.	189
E.1	Comparison of the two different methods for combining posterior distributions from independent experiments into a single, joint distribution on $\sqrt{\alpha_{\text{dCS}}}$. The two methods in question involve the fitting of functions to the individual distributions (KDE method) and the direct product of the histograms (Direct Hist. Product), as well as a smoothed version of the latter used for the calculation of the confidence intervals. While there seems to be a bias towards higher values of α_{dCS} , any conclusion from this distribution should be tempered, as this analysis still doesn't provide reliable results, due to violations to the small coupling approximation.	191

E.2	Comparison of the two different methods for combining posterior distributions from independent experiments into a single, joint distribution on $\sqrt{\alpha_{\text{EdGB}}}$. The two methods in question involve the fitting of functions to the individual distributions (KDE method) and the direct product of the histograms (Direct Hist. Product), as well as a smoothed version of the latter used for the calculation of the confidence intervals.	191
-----	--	-----

Chapter 1

Introduction

In the past decade, the astronomy and physics scientific communities have seen an explosion of new discoveries in a variety of different topics and areas of research. The diverse new science undertaken in recent years is partly due to the novel detection of gravitational waves (GWs), a form of radiation predicted by Einstein over one hundred years ago. This phenomenon was only observed directly for the first time in 2015 [14] by the Laser Interferometer Gravitational wave Observatory (LIGO) [15], igniting a new revolution in many areas of science. This new window into the universe has drastically expanded the global scientific community’s ability to detect some of the most energetic processes in the universe: the coalescence of compact binaries. Compact binaries are a class of systems defined by their high densities and strong gravitational fields, including objects like black holes (BHs) and neutron stars (NS) orbiting one another until they eventually merge in a cataclysmic collision. Binary black holes (BBHs) reach orbital speeds exceeding $\sim 0.3c$ and achieve gravitational fields (roughly defined by some characteristic mass $\mathcal{M}_C$ over some characteristic length $\mathcal{R}_C$ or $\sim G\mathcal{M}_C/(c^2\mathcal{R}_C)$) of the order 0.1 for sources currently targeted by LIGO, Virgo [16], and Kagra [17] (LVK). These highly relativistic compact binaries are remarkably effective at probing many areas of physics that are completely unreachable with more conventional experiments.

Furthermore, the catalog of events that LVK has detected is quickly growing. Starting with the first catalog released in 2019 (GWTC-1) [18] and through the latest catalog released by the collaboration (GWTC-3) [19], the global network of detectors has increased the number of confident detections eight fold in three years, from 11 sources to 90. The statistical weight of these catalogs can be powerful tools now, but even more so in the near future. Studies of the compact object populations are making good use of these new sets of observations especially (c.f. [20–22]), but these catalogs can help tremendously in a variety of topics, as will be evidenced later in this thesis. Furthermore, the growing number of unexpected sources [23–29] is pushing the bounds in several areas of physics (c.f. [30–36]). This population of “surprising” sources is an exciting frontier for scientists to explore.

With an entirely new way of observing astrophysical systems, which areas of physics stand to benefit the most? The list can be lengthy, as these sources can be used to explore everything from nuclear physics (c.f. [31,

37–39]) to cosmology (c.f. [40–42]). In this thesis, we will focus specifically on a particularly effective use of these new observations: putting general relativity (GR) to the test. For the reasons outlined above, these observations can shed light on the gravitational interaction like nothing else in the solar system; even among other astrophysical tests, GWs are a comparatively powerful tool. Specifically in the case of astrophysically realistic BBHs, the dynamics of the inspiral, merger and post-merger signals are effectively described by the gravitational force exclusively. This leads to very low systematics as compared with other astrophysical tests, for example with tests based on accretion disks [43, 44]. With many astrophysical systems, the uncertainties surrounding the dynamics of the system *even in GR* can wash out any inferences one might hope to make concerning the gravitational interaction [44]. Similar to GW signals, binary pulsar observations (measured through electromagnetic (EM) observations) have low systematics and are very effective at constraining alternatives to GR [45–48], but they can be less effective at probing certain types of modifications [49]. Despite these hurdles, impressive progress has been made within the topic of gravity using astrophysical systems observed through EM observations, but GW observations can complement these works in new and exciting ways [49, 50].

Einstein’s famous theory of gravity has been rigorously tested for over a century (c.f. [51, 52]), but the challenges it faces are only increasing as our measurements become more precise and reach new energy scales. The first major hurdle that GR surpassed was the prediction of the orbital precession of Mercury, which had been a major open question in science for more than fifty years [51]. From there, GR has continued to best every test that has been thrown at it, including the accurate prediction of the bending of light around the sun in the 1920’s, the gravitational redshift of light [53–55] and the orbital decay of the Hulse-Taylor pulsar [56], just to name a few historical mile-markers. In recent years, this trend has continued even in the face of the measurement of GWs (c.f. [57–61]) and the imaging of the shadow of a black hole [62, 63].

If GR has been completely successful to date, why continue testing it? The answer generally lies in two very broad categories: (i) observational irregularities and (ii) theoretical curiosities. Category (i) does not include failings of GR *per se*, but includes open questions in physics that could potentially be resolved by amending our description of the gravitational interaction. For example, the anomalous rotation curves of galaxies (commonly attributed to dark matter [64, 65]) could potentially be resolved by modifying the gravitational sector [66–76]. The late-time acceleration of our universe [77, 78] could also be explained by modifying GR [79–83], instead of using vacuum energy or some unnaturally small cosmological constant. Finally, the disagreement between quantum mechanics and GR is a well known issue that needs a resolution [84], and the solution might be to add higher energy corrections to GR [85–87].

Within category (ii), there are several fundamental assumptions that GR relies upon. These can be

succinctly summarized by Lovelock’s theorem [52, 88]. Included in these assumptions are: the gravitational interaction is mediated by the metric tensor alone, the metric tensor is a massless tensor field, spacetime is four-dimensional, the theory of gravity is position-invariant and Lorentz-invariant and the gravitational action is parity invariant. These tenets of GR are *assumptions* and do not necessarily need to be true in our universe. Therefore, these tenets can help provide a framework to systematically search for deviations from GR. By looking for theories that might potentially solve some of the open questions in physics falling in category (i), we can construct and investigate theories by systematically questioning each of the fundamental assumptions falling in category (ii).

With these general motivations in mind, how should we use GW observations to determine what theory of gravity our universe actually obeys? There are a multitude of methods to go about this, which is a testament to the wealth of information in these observations [89]. Here, we will outline some of the more standard tests of GR using GWs, but this is only a sampling of the ways scientists are using this data to look for new physics.

The most straightforward way of testing whether GR actually describes the signals measured by our detectors begins by taking a model of the waveform within GR and fitting it to the data (giving the maximum likelihood parameters of the system, for example). The fitting process is a nuanced business, which we will discuss below. Once the fitted waveform model is in hand, the maximum likelihood prediction from GR can simply be subtracted from the actual data stream. At this point, we would like to know if what is left over (the residual) is consistent with stationary Gaussian noise (which is the assumed noise model for our terrestrially based GW detectors at current sensitivity). A simple way to do this is to calculate the signal to noise ratio (SNR) of the residual, which is a measure of how much power is in the residual. A high SNR for the residual signal would indicate there is structure in the signal that is not being captured by the GR waveform model (suggesting the waveform model is incomplete).

In practice, LVK uses a slightly more complicated version of this method to take advantage of the fact that there are multiple GW detectors in the global network [57]. Deviations between the signal and the GR waveform model should be consistently present across detectors, and that extra information can improve the test. To consistently test the residuals for a network of detectors, the fitting process is repeated on the *residual* data, using a totally physics-agnostic waveform model called **BayesWave** [90, 91]. This agnostic waveform model uses a set of orthogonal wavelets to search for coherent signals between detectors, minimally assuming the signal to come from a well localized astrophysical source and that the wave is elliptically polarized. The residual *network* SNR is then calculated with this agnostic reconstruction (repeating the SNR calculation with the coherent, residual signal for each detector and summing in quadrature).

This type of test has a lot of benefits. For example, as a type of null-test of GR, no alternative waveform model has to be specified. If there is a lacking in GR to describe the data in any way, there will be a residual SNR from mis-modelling. On the other hand, this test is sensitive to systematic errors *within GR*. That is, GR could still be right, but our deficiency to model the waveform of a system perfectly within GR would leave power in the residual. If a residual SNR is detected, there would be no way to determine if that residual structure came from systematics within GR or because GR is lacking as a theory in some way.

A similar test to the residual test is the IMR consistency test [92, 93]. In this test, the signal is separated into two distinct portions of data: the inspiral portion (characterized by relatively slow velocities and wide separations) and the merger-ringdown portion (characterized by highly relativistic dynamics). The separation between the two is fixed at the innermost stable circular orbit (ISCO). The gravitational experimentalist can fit a waveform model in GR to the inspiral portion, then separately fit a merger-ringdown waveform model to the later portion of the data. Once these two sections of data have been independently fit, estimates for the final BH mass and spin are calculated for the inspiral waveform model using fitting functions calibrated to numerical relativity data[94]. The waveform model for the merger-ringdown portion of the signal already provides an estimate for the final BH mass and spin. If the data is truly described by GR, the estimates for these two quantities coming from the inspiral signal and the merger-ringdown signal should be consistent. The methods used to model the dynamics of these two different parts of the waveform are very different, where post-Newtonian [95, 96] and effective-one-body (EOB) [97–99] calculations are used for the inspiral, while the merger relies on numerical relativity and black hole perturbation theory. Consistency between the predictions of these various methods, all still within GR, would be an impressive consistency check for the theory.

Another test of GR focused on the ringdown portion of the signal is black hole spectroscopy [100, 101]. This test is a search for what is known as BH hair. Within GR, a BH in vacuum is fully described by three parameters: the mass, angular momentum (spin) and the electric charge [102]. In astrophysical systems, it is generally expected that the charge would be neutralized on short time scales, leaving the remaining two degrees of freedom only. According to BH perturbation theory, a perturbed BH in isolation will radiate energy out through gravitational waves as exponentially damped sinusoids called quasinormal modes (QNMs) [103, 104]. Within GR and for a known perturbation, the real and imaginary components of the frequency of the QNMs are fully determined for all the modes if the mass and angular momentum of the BH are known. For BBHs, there will be a single, perturbed BH for a short time after the binary coalesces, as the remnant settles down to a stationary BH. If one can measure the full complex frequency for several of these modes independently, the consistency between these parameters could prove (or disprove) the no-hair theorem

(possibly calling GR into question).

The final category of tests we will outline are the tests upon which we focus for a majority of this dissertation. In the early inspiral of a binary, the dynamics can be calculated by looking at the equations of GR and expanding them around Minkowski spacetime and assuming the velocities are slow. This is called a post-Newtonian (PN) expansion [95, 96], and is very effective at describing the trajectories of the binary components to surprisingly late times in the inspiral [105]. In this formalism, the equations are then solved, order by order, to obtain the gravitational waveform as a series in the orbital velocity, v . In waveforms currently used in GW analysis, the series has been calculated up to v^7 , and the coefficients can be found in a variety of works, e.g. [106].

This waveform model gives a well defined, precise prediction for the GW strain within GR. One can now ask, what would change in this series when considering alternative theories of gravity? Assuming the theory admits a series solution, *à la* PN expansions, the only thing to change would be shifts in the coefficients (possibly including new terms in the series) [107]. This calculation has been done in a variety of different alternative theories of gravity (c.f. [107–114]). With these calculations done in specific theories, one can simply replace the GR inspiral waveform model with one of these alternative waveform models. The waveform model can then be fit to the data, and various statements can subsequently be made about which waveform model best fits the data.

This method of re-analyzing the data with each contending waveform model has its drawbacks, however. Repeating the analysis for every theory is a computationally expensive endeavor and would require a significant amount of compute time. Furthermore, our list of viable alternative theories is hardly exhaustive. In principle, there are an infinite number of theories one could create. Finally, calculations of this sort in theories beyond GR are still being worked out. There are only a handful of alternative theories where the PN calculation has been completed to sufficient order to be useful in this context. With this in mind, one can take the calculations in these alternative theories as inspiration for a theory-agnostic variation of this analysis. This method was first proposed as the *parametrized post-Einsteinian* (*ppE*) framework [107, 115–117], a variation of which was later incorporated into the standard analysis by LVK [57–61, 118–122]. Using this technique, one would simply add a generic parameter to each coefficient in the series (typically one coefficient at a time), using this generic shift in the coefficient to capture *any* physics that might not already be in GR. By construction, GR is a subset of the new waveform model’s parameter space. Specifically, GR is recovered when this coefficient is taken to 0 (in typical parametrizations). One can now fit this new waveform model to the data to determine if the data is better described by introducing this new coefficient. If the coefficient does not significantly improve the waveform model’s ability to describe the data, then GR (the simpler waveform

model, as it has one fewer parameter) is preferred.

This generic method of testing each of the coefficients of the PN series is efficient, in that the analysis does not need to be repeated for every single theory. Furthermore, it has the nice quality that a theory does not need to be specified at all, allowing us to test theories that have not yet been considered at all. However, at the end of the day, tying back these generic tests of GR to fundamental physics is necessary to motivate and direct theoretical efforts. Fortunately, we still retain this connection for any theories where the PN solutions have already been derived. If there exists a PN calculation in a specific theory, the results for the generic parameter can simply be mapped to any relevant parameters without rerunning the entire analysis (with the caveat that changes in the prior deserve some attention [123, 124]). This class of tests (focused on the inspiral portion of the signal, where analytic calculations can be done to find the binary's dynamics) will be the focus of a majority of this dissertation.

Now that we have a waveform model to predict the GW signal in GR, some alternative theories of gravity and in a theory agnostic way, how exactly do we use these waveform models to constrain physics beyond GR? This leads us into some details of GW data analysis, elaborated in much more detail in several chapters of this dissertation. As a brief summary, however, the types of GW data analysis that we will focus on is often carried out in the framework of Bayesian statistics. Within Bayesian statistics there are several quantities that can provide this type of information: the posterior distribution and the model evidence. The model evidence gives a measure of how well the model as a whole describes the data and can be compared against the evidence of other models to determine which model is preferred. This way of quantitatively comparing models can be very useful, but it should be noted that this involves the comparison of the *entire* model, of which the *waveform* model is only a part. The full model also involves how one encodes any prior held beliefs about the data and how the detectors operate, as well as the waveform model which predicts the generation and propagation of GWs.

The work in this dissertation, however, will focus on the other quantity of interest in Bayesian statistics: the posterior distribution. The posterior distribution is a probability distribution on the parameters of the model, conditional on the data set and the exact model. These parameters typically include the masses, angular momentum, sky location and orientation of the binary, as well as any parameters controlling additional physics one might like to test. For example, one may include additional parameters to control tidal effects of the bodies, or, in our case, additional parameters to control the magnitude of deformations from GR. In the case of generic deviations in inspiral tests of GR (mentioned above), the deformation to a PN coefficient would be made a part of this parameter list. For specific theories of gravity, this additional parameter might be a coupling constant in the action (defining the relevant scales of the theory). Ultimately,

this distribution determines the probability the parameter vector takes certain values considering the data has been observed. Scientists can then take this distribution and calculate quantities of interest, such as confidence intervals on parameters of the model. Calculating these confidence intervals on parameters that control deviations from GR leads to the constraints on beyond-GR physics.

But how does one calculate the posterior distribution or the model evidence? The full problem (involving at least fifteen dimensions in the BBH case) cannot be analytically solved, being too difficult for even basic root-finding or gradient descent solvers to succeed. With few other options, the typical solution is to use sampling algorithms (like nested sampling, ensemble sampling or Markov chain Monte Carlo (MCMC) sampling), which can be very computationally expensive to perform. Samples drawn directly from the posterior can be histogrammed into a discretized version of the posterior, where the accuracy of this approximated distribution roughly scales inversely with the square root of the number of points being drawn. In practice, this is how one derives constraints on physics beyond GR using data from GW detector networks.

While analyzing current data is the ultimate goal for physicists seeking to actually constrain theories of modified gravity, there should always be an eye set on the future. Funding, developing and preparing for future GW detectors is a big part of conducting fundamental science with GW data. Part of that work is evaluating proposed detector designs and upgrades in the context of achievable science goals, allowing for funding agencies and scientific collaborations to make informed decisions. At the end of the day, the usual criterion that proposed detectors and upgrades are critiqued against is the following: what is the scientific return on investment? Another facet of this topic includes the evaluation of what analysis and modelling techniques are worth the research community's time and resources. Does this new technique, which might take a large amount of time and computer resources to develop, allow for significantly better science to be accomplished? This type of work requires significant modelling and numerical investigation to develop the necessary answers.

These numerical investigations, which typically revolve around simulating the process of observing and analyzing GW signals, vary in scope and cost. A variety of different types of simulation techniques are used in this dissertation. Many situations call for highly accurate, fully realistic simulations of the data analysis framework applied to current data. We leverage this style of simulation in this thesis when accuracy is required. In essence, this method requires the creation of synthetic data (which can be a full numerical relativity simulation of the orbital dynamics of compact binaries or using one of the simpler GW waveform models already used for data analysis), which is then sent through the exact pipeline used on real data. This yields results that are maximally robust, shedding light on nuanced aspects of the analysis that might be important for future research directions.

On the other side of the cost spectrum, there are ways to approximate various steps in the GW analysis, allowing for fast calculations at a slight cost to accuracy. A common choice (and one used to great effect in this dissertation) is the Fisher Information Matrix [125] (or Fisher, for short). The Fisher is an approximation method to simplify the likelihood function, which is the most costly component of most GW data analysis pipelines. The Fisher approximation, in its most simple form, amounts to expanding the probability distribution of the model parameters about the most likely values. By making this approximation, one can (semi-)analytically write down the posterior distribution, speeding up the problem of calculating the posterior by many orders of magnitude. This calculation opens many doors to numerical experiments that would be utterly intractable using the full pipeline used for real data. For example, questions that revolve around the combination of multiple detections (where the number and type of sources used are critical) now become accessible, and questions of these type are essential for evaluating future methodology and detector development.

1.1 Executive Summary

This dissertation is organized as follows. We begin in Chapter 2 by examining one particular modification to general relativity that has been of much interest since GR’s original inception, but we also include more nuanced effects that had been neglected thus far: giving the graviton a mass in the presence of screening mechanisms. One of the tenets of GR is that the metric field is massless. When considering quantizations of this field (the graviton) in this case, the particle would also be massless. An interesting way to modify GR that has implications for cosmological topics [81, 82, 126–130] is to give the graviton a small mass, which has been considered in past literature (c.f. [131, 132]). However, there are screening mechanisms, first discovered in 1972 [133], that could nullify the observable effect of a massive graviton in certain environments (depending on the specifics of the theory). This effect has been considered in static systems and in the generation of GWs (c.f. [134]), but had not been considered in the case of graviton propagation. We considered a class of scalar-tensor theories that produced massive graviton-like effects in the propagation of GWs, but that would also be susceptible to these screening effects. We used Fisher matrices to make predictions about the measurability of the graviton mass by a variety of GW detectors while considering screening mechanisms during the GW propagation.

Taking our calculations from Chapter 2 as a starting point, we widened our scope to try and predict the future for tests of GR with GW observations for approximately the next three decades in Chapter 3. We increased the number of modifications we were considering to seven, as well as included theory agnostic ppE

deformations. A critical component of this analysis was getting the right distributions of observed sources. For predictions on constraints of GR modifications, both the number and the type of GW sources being observed, as well as the capabilities of the detectors, significantly impact the conclusions. As there is still much uncertainty about the exact underlying population of binary BHs in the universe, we used several different models for BBH populations spanning a range in masses from stellar-origin BBHs to supermassive BBHs. To account for the uncertainty in the capability of the global GW detector network in the coming decades, we also considered a variety of different detector funding/development scenarios. We used all the combinations of the above models to bracket the uncertainties on the observed population of GW sources. With these synthetic, observed catalogs of GW sources, we computed Fisher matrices for each of the sources for each of the modifications we were interested in. This not only gave us insight into the expected constraints on each of the modifications in our list, but also allowed us to quantify the impact of various components of our analysis (like the number and distribution of sources) on our conclusions for each individual modification.

From here, we move on to Chapter 4 where we use the full MCMC, injection/recovery method to simulate a GW detection. In this method, we create synthetic data with a known model and set of parameters, then sample the posterior distribution with an MCMC algorithm to maintain maximum accuracy. This type of analysis yields predictions that are closest to what one might actually expect if the injected signal were actually detected by the synthetic detector, as it uses an identical analysis as one might use on real data. In this work, we use this method to address a variety of concerns in the GW science community about the proper way to test for deviations from GR with generic parametrizations (like the ppE parametrization). A popular use of generic, theory-agnostic parametrizations to test GR entails adding a single deformation parameter to a single coefficient in the GW phase at a time. However, if the purpose of these deformations is to generically represent modifications that would appear in the phase due to a modification to the action (as was the original inspiration [107, 115–117]), then one would expect an *infinite series* of modifications to the phase. Ideally, searches for physics beyond GR would allow deformations to appear at many terms in the PN series, but that ends up making the problem intractable. As has been known for a long time, methods using too many of these parameters suffer from overfitting the data, and the constraints widen significantly [119, 135, 136]. In this chapter, we outline why this pessimistic outlook is probably unwarranted by creating a physically motivated parametrization and prior inspired by PN calculations. We then take this model (comprised of the GW waveform, the particular parametrization, the prior, and the likelihood) and use it to analyze synthetic data. With these simulations of GW detections, we are able to show that the constraints actually improve when adding additional deformations, but only if a *physics-inspired* model is used. This leads us to conclude that the constraints using single deformations are most likely overly conservative, if

anything.

From here, we will move into placing constraints on alternative theories of gravity using current data. This marks an important shift, as the previous chapters were focused on making predictions and evaluating future networks, measurement techniques, and waveform models. In Chapter 5, we focus on a specific class of theories called quadratic theories of gravity. These theories are classified as such because they introduce a (pseudo) scalar field and couple this auxiliary field to the gravitational sector through a coupling that involves terms that are quadratic in curvature invariants (like the Ricci scalar, Ricci tensor, and Riemann tensor). We look at two theories in particular, dynamical Chern Simons (dCS) [137] and Einstein-dilaton-Gauss-Bonnet (EdGB) [138]. By transforming previous generic results from the LVK [18] into constraints on parameters in these specific theories, we are able to place a constraint on the lengthscale of EdGB.

Next, we improve upon this work in Chapter 6. In Chapter 5, we used two GW sources and utilized previously derived results, leaving certain undesired artifacts in the posterior distributions. In Chapter 6, we analyze six different GW observations and combine the information from these different observations into one constraint. Each of the individual analyses was performed directly within the theory in question, removing the previously-observed artifacts from the posterior distribution. This new analysis resulted in a constraint more than three times stronger for EdGB. Finally, we summarize our findings and discuss future avenues of work in Chapter 7.

Chapters 2, 3, 5, and 6 are based on works already published in peer reviewed journals and Chapter 4 is submitted for review, as of the time of writing. These works were written in collaboration with different coauthors. The information for each manuscript is shown below.

1. *Probing Screening and the Graviton Mass with Gravitational Waves*

Scott Perkins and Nicolás Yunes

Class. Quant. Grav. 36, 055013 (2019)

Chapter 2

2. *Probing Fundamental Physics with Gravitational Waves: The Next Generation*

Scott E. Perkins, Nicolás Yunes, and Emanuele Berti

Phys. Rev. D 103, 044024 (2021)

Chapter 3

3. *Are Parametrized Tests of General Relativity with Gravitational Waves Robust to Unknown Higher Post-Newtonian Order Effects?*

Scott Perkins and Nicolás Yunes

Submitted to Phys. Rev. D., arXiv:2201.02542

Chapter 4

4. *Fundamental Physics Implications on Higher-Curvature Theories from the Binary Black Hole Signals in the LIGO-Virgo Catalog GWTC-1*

Remya Nair, Scott Perkins, Hector O. Silva, and Nicolás Yunes

Phys. Rev. Lett. 123, 191101 (2019)

Chapter 5

5. *Improved gravitational-wave constraints on higher-order curvature theories of gravity*

Scott E. Perkins, Remya Nair, Hector O. Silva, and Nicolás Yunes

Phys. Rev. D 104, 024060 (2021)

Chapter 6

As the first author of publications 1, 2, 3, and 5, I personally carried out the relevant calculations and had a substantial part in writing the manuscripts, under the advisement and direction of my coauthors. For publication 4, I also played a significant role in carrying out the details of the calculation and the writing of the manuscript, again under the direction and in collaboration with my coauthors.

Chapter 2

Probing Screening and the Graviton Mass with Gravitational Waves

This chapter is based on: *Perkins, S.*¹ and *Yunes, N.*² (2019). *Probing Screening and the Graviton Mass with Gravitational Waves. Class. Quant. Grav.* 36, 055013

Abstract Gravitational waves can probe fundamental physics, leading to new constraints on the mass of the graviton. Previous tests, however, have neglected the effect of screening, which is typically present in modified theories that predict a non-zero graviton mass. We here study whether future gravitational wave observations can constrain the graviton mass when screening effects are taken into account. We first consider model-independent corrections to the propagation of gravitational waves due to screened massive graviton effects. We find that future observation can place constraints on the screening radius and the graviton mass of $\mathcal{O}(10^2)$ – $\mathcal{O}(10^4)$ Mpc and $\mathcal{O}(10^{-22})$ – $\mathcal{O}(10^{-26})$ eV respectively. We also consider screening effects in two specific theories, ghost-free massive gravity and bigravity, that might not realize these types of propagation modifications, but that do provide analytic expressions for screening parameters relevant to our analysis allowing for more concrete results. However, the constraints we are able to place are small. The reason for this is that second- and third-generation detectors are sensitive to graviton masses that lead to very small screening radii in these particular models. The effect of screening, however, can become important as constraints on the graviton mass are improved through the stacking of multiple observations in the near future.

2.1 Introduction

The onset of gravitational wave (GW) astrophysics has allowed new constraints of modifications to General Relativity (GR), including new bounds on the graviton mass [139]. The latter is a particularly important quantity in a variety of fields, from quantum extensions of GR [140–143] to cosmological modified theories of gravity [81, 82, 126–130]. Because of this widespread interest, the graviton mass has been constrained through

¹Carried out the calculations in the paper; wrote software related to waveform modelling; wrote statistical analysis software; conducted the numerical experimentation; wrote the first draft of the manuscript

²Conceived of the project; supervised and advised on the calculations and experimentation; edited the manuscript

a host of different experiments [144–151], from weak field Solar System bounds resulting in $m_g < 10^{-24}\text{eV}$ [152], to strong field, stacked GW bounds resulting in $m_g < 10^{-23}\text{eV}$ [139]. While Solar System constraints are currently more competitive than their GW counterparts, the former probe fifth-force type forces, while the latter constrains the propagating sector of such theories [153, 154].

In the context of GW observations, estimated bounds on the graviton mass have used the work of Will [131] and extensions [132]. Working in the post-Newtonian (PN) framework [95], one first postulates a modified dispersion relation during the propagation of GWs that is based on the special-relativistic dispersion relation of massive particles. One then studies how these modifications trickle down into the observed GW phase through the stationary phase approximation [125]. The end result is a correction that enters at 1PN order in the GW phase relative to the leading-order GR term and that scales with the distance traveled.

Previous GW work on the graviton mass, however, has neglected a mechanism that is typically present in cosmological modified gravity theories: screening [133, 155–157]. This mechanism suppresses modifications to GR inside of some screening radius through nonlinear self-interactions of an auxiliary field, thus allowing these theories to pass Solar System tests with ease. GWs, however, typically exit this radius during their travel from the source to Earth, thus possibly acquiring modifications in their propagation *only* outside the screening radius. Intuitively, one would therefore expect screening to soften constraints on the graviton mass, as its modifications to the GW phase need not always be active. Because of this softening, one should consider the effect of screening in future bounds, like those estimated in [132, 158–161], to ensure modified theories are not prematurely ruled out.

Whether screening is present in modifications to the dispersion relation of GWs depends strongly on the particular theory considered. For example, in quartic and quintic Galileon theories [162], screening is present in the dynamics of the Galileon scalar field, but this does not percolate into the GW dispersion relation [163]. As such, it is not expected that this effect be present in theories like ghost-free massive gravity (dRGT) [83] or bigravity [164], which are generalizations of the Galileon family of theories [153, 165]. In other theories, however, these propagation effects may be present, provided the propagation equations for the helicity-2 mode couple directly to the scalar field.

Given this, we here take on an agnostic approach to study whether GWs can probe the graviton mass if screening is present. We first derive the correction to the GW Fourier phase when screening is present within the PN framework and using the stationary phase approximation. We find that the correction to the phase takes the same exact functional form as in the unscreened case, except that the distance parameter that enters the correction is not the usual luminosity distance, but rather a new effective quantity. This new effective distance is smaller than the luminosity distance, yielding a measure of the distance through which

massive graviton modifications are active.

We then study how well such a screened massive graviton modification could be constrained with second- and third-generation GW detectors, like aLIGO [166], the Einstein Telescope (ET) [167], and LISA [168]. We first modify the quasi-circular IMRPhenomD model [106, 169] to construct screened massive graviton waveforms that span the whole frequency range, from inspiral to merger and ringdown. We then assume second- and third-generation detectors have detected a signal consistent with GR and predict the constraints one could place on screened modified gravity effects through a Fisher analysis [125], which should be accurate at the large signal-to-noise ratios we consider.

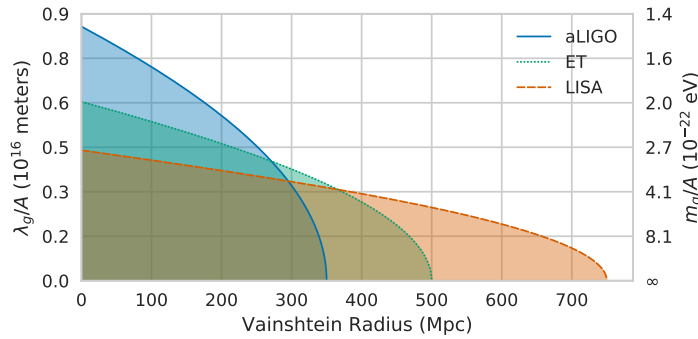


Figure 2.1: (Color Online) Projected constraints on the graviton mass as a function of the screening radius, assuming the detection of three expected GW sources with aLIGO at design sensitivity, ET, and LISA. As the sensitivity to these screening effects varies widely for each detector, the y-axis is re-scaled appropriately through a factor A presented in Table 2.1. The effectiveness of screening becomes very apparent when the screening radius is about $D_L/2$, since then the constraint on m_g rapidly falls to zero.

Constraints on the modifications to the GW phase lead to degenerate constraints on the graviton mass and the screening radius, when these parameters are treated as independent. Figure ?? shows these degenerate constraints, where the shaded regions would be disallowed, given GW observations consistent with GR and produced by the systems in Table 2.1. When the screening radius is small, then constraints on the graviton mass reduce to previous estimates [132, 158–161]. When the screening radius is large, the projected constraints deteriorate, until for sufficiently large radii no constraints on the graviton mass are possible any longer. The latter occurs when the screening radius is roughly half the luminosity distance to the source, as then the graviton mass effects are completely screened.

We conclude with a study of projected constraints on screening and the graviton mass in specific cosmological modified gravity theories. In particular, we consider two massive gravity theories, dRGT and bigravity, in which the screening radius is a function of the graviton mass [155, 165]. The effect of screening on the scalar field in these theories does not appear to affect the propagation of the tensor modes [153, 165], but these models give explicit functional forms for the screening radius as a function of the graviton mass.

Detector	m_1 ($M_\odot$)	m_2 ($M_\odot$)	D_L (Mpc)	A
aLIGO	20	15	700	1
ET	35	10	10^3	10
LISA	10^6	10^5	1.5×10^3	10^4

Table 2.1: Properties of the binary systems considered in Fig. ?? to estimate projected bounds on the graviton mass and the screening radius. The parameter A is a factor used to re-scale the y-axis of the figure.

This allows for a more tangible result that might be indicative of what one may expect in other more general theories that do include screening of the tensor modes. In these theories, however, the scaling of the screening radius is such that the graviton mass that second- and third-generation detectors are sensitive to lead to tiny screening effects. Only once constraints on the graviton mass become of $\mathcal{O}(H_0)$ do screening effects begin to become important in these theories. Such graviton mass constraints, however, could be achievable through the stacking of multiple observations with third-generation detectors [92].

The remainder of this paper deals with the details of the calculations that led to the results summarized above. Section 2.2 presents the basics of screening and its impact on the GW response function. Section 2.3 describes the Fisher analysis methodology that we employ to estimate constraints. Section 2.4 presents the projected constraints we obtained. Section 2.5 concludes and points to future research. Henceforth, we follow mostly the conventions of [102], where Greek letters represent spacetime indices, Latin letters represent parameter indices, and for the most part we employ geometric units in which $G = 1 = c$.

2.2 Gravitational Wave Propagation in Screened Theories

In this section, we calculate how the propagation of GWs is affected by a screening radius in theories with a massive graviton. We begin by describing the basics of Vainshtein screening, therefore setting notation. We then continue with the calculation of the modifications to the propagation speed of GWs, and how this percolates to the Fourier transform of the response function in the stationary phase approximation.

2.2.1 Vainshtein Screening

Vainshtein screening is the process by which nonlinear interaction in the field equations suppress modifications to GR in a certain regime of spacetime. Let us here outline the basic foundation behind Vainshtein screening; a more in depth description can be found in several reviews on the topic, e.g. [155, 165, 170, 171] or in works like [172–176].

First discovered in 1972 by Vainshtein [133], the idea of screening has seen waves of interest, with a renewed

popularity recently due to recent work on theories like dRGT and bigravity. Originally, the mechanism was introduced as the solution to the van Dam-Veltman-Zakharov discontinuity [177], a problem that had plagued massive gravity in the 1970's. This discontinuity arises as an inability of massive gravity to match (at linear order) GR solutions in the limit of a vanishing graviton mass, resulting in easily measured discrepancies between predictions and observations in the Solar System and gravitational lensing experiments. However, Vainshtein noticed that these deviations came at linear order in the metric perturbation, and that by including higher order terms, these modifications could be suppressed around massive sources, reviving the viability of massive gravity theories [133, 178].

Later on, a different issue, the so-called Boulware-Deser ghost [179], began to haunt massive gravity theories. This unbounded-from-below sixth degree of freedom again muted interest for a number of years until de Rham, Gabadaze, and Tolley formulated dRGT to specifically eradicate this unphysical degree of freedom [83]. This gave rise to renewed interest in massive gravity theories and spawned bigravity as an extension, which also exhibits Vainshtein screening effects [156]. Since then, the Vainshtein mechanism has been crucial in keeping theories viable in the wake of highly accurate Solar System tests, as it has been shown to be effective at suppressing “fifth force” modifications to the Newtonian potential [165].

Screening, however, is usually analyzed in the static, spherically symmetric case. It remains a bit of an open topic of research whether the screening persists within dynamical systems for massive gravity, and how effective the screening actually is. Recently, the mechanism has been examined in the context of the cubic Galileon theories (a simpler, but related theory to dRGT) numerically and it has been shown that the energy loss of binary systems through scalar radiation is indeed suppressed (albeit not to the full extent of the suppression of the fifth force in static scenarios) [134].

To illustrate the effect of screening in a simple scenario, let us focus on a static, spherically-symmetric scenario in the context of the cubic Galileon theories. Following [165], the cubic Galileon action in geometric units is

$$S_{\text{CG}} = \int d^4x \sqrt{-g} \left[\frac{R}{16\pi} - \frac{1}{2} \partial_\mu \phi \partial^\mu \phi - \frac{1}{\Lambda^3} \partial_\mu \phi \partial^\mu \phi \square \phi + \sqrt{8\pi} \phi T \right] + S_m[g], \quad (2.1)$$

where $\Lambda = [m_g^2/(\hbar^2 \sqrt{8\pi})]^{1/3}$, the Ricci scalar R is associated with the metric tensor $g_{\mu\nu}$, the scalar field ϕ couples to the trace of the matter stress-energy tensor T and the metric tensor couples to matter through S_m .

Varying Eq. (2.1) in a static and spherically symmetric system with a stress energy tensor trace $T =$

$-M\delta(r)/4\pi r^2$ results in the equation of motion of the scalar field

$$\frac{1}{r^2} \frac{d}{dr} \left[r^2 \frac{d\phi}{dr} + \frac{r}{\Lambda^3} \left(\frac{d\phi}{dr} \right)^2 \right] = \frac{M}{\sqrt{2\pi}} \frac{\delta(r)}{r^2}, \quad (2.2)$$

which, after direct integration of the first derivative, gives

$$\frac{1}{r} \frac{d\phi}{dr} + \frac{1}{r^2 \Lambda^3} \left(\frac{d\phi}{dr} \right)^2 = \frac{M}{\sqrt{2\pi}} \frac{1}{r^3}. \quad (2.3)$$

The importance of Vainshtein screening is illustrated in Eq. (2.3). With the definition

$$r_v = \frac{1}{\Lambda} \left(\frac{M}{\sqrt{2\pi}} \right)^{1/3}, \quad (2.4)$$

the separation of scales is apparent, and allows us to approximate the gradient of the scalar field from Eq. (2.3) via

$$\frac{d\phi}{dr} \approx \begin{cases} \frac{M}{\sqrt{2\pi}} \frac{1}{r^2} & r \gg r_v, \\ \frac{M}{\sqrt{2\pi}} \frac{1}{r^2} \left(\frac{r}{r_v} \right)^{3/2} & r \ll r_v. \end{cases} \quad (2.5)$$

For regimes of spacetime $r \ll r_v$, the evolution of the scalar field is suppressed by the $(r/r_v)^{3/2}$ factor, which then suppresses the “fifth force” modifications this scalar field would introduce.

For dynamical systems, the situation is slightly more complicated. Let us leave cubic Galileon theories and consider instead the generic action

$$S = \int d^4x \frac{\sqrt{-g}}{8\pi} \left[\frac{R}{2} - \frac{1}{2} \partial_\mu \phi \partial^\mu \phi - \mathcal{U}(g, \phi, \partial\phi) + \mathcal{L}_{\text{matter}}(g, \phi) \right], \quad (2.6)$$

where $\mathcal{U}$ is a kinetic self-interaction term that depends on the metric, as well as derivatives of the auxiliary field. The field equations associated with this action are schematically

$$G_{\mu\nu} \propto T_{\mu\nu} - \partial_\mu \phi \partial_\nu \phi - \frac{\delta \mathcal{U}}{\delta g_{\mu\nu}}, \quad (2.7a)$$

$$\square \phi + \frac{\delta \mathcal{U}}{\delta \phi} \propto \frac{\delta \mathcal{L}_{\text{matter}}}{\delta \phi}. \quad (2.7b)$$

The equation of motion for the scalar field resembles that in Eq. (2.2), with the identification of the variation of $\mathcal{U}$ in Eq. (2.7b) with the non-linear term in ϕ in Eq. (2.2).

The analysis of the evolution of the scalar field now follows closely the analysis of Eq. (2.2). If the kinetic, self-interaction term of Eq. (2.7b) dominates through the Vainshtein mechanism, (a) the scalar mode is not excited because of a suppression of its coupling to matter (verified numerically in [134]), and (b) the impact of the scalar field on the metric perturbation in the radiation regime, but still in the Vainshtein region, is also suppressed. This implies that GR is recovered within the Vainshtein radius, even in terms of propagation effects. Outside the Vainshtein radius, the self-interaction term in Eq. (2.7b) becomes much smaller than $\delta\mathcal{L}_{\text{matter}}/\delta\phi - \square\phi$, and ϕ is allowed to grow, in turn allowing $\delta\mathcal{U}/\delta g_{\mu\nu}$ to affect the propagation of the GW tensor modes.

Before proceeding, let us define the screening radius for massive objects that we will use later. This quantity differs from theory to theory, but as an example, the screening radius for a galaxy with Schwarzschild radius r_s in dRGT and bigravity takes the form [165]

$$r_v = r_s^{1/3} \lambda_g^{2/3}, \quad (2.8)$$

whereas the screening radius in non-linear Fierz-Pauli gravity with a general potential is [155]

$$r_v = r_s^{1/5} \lambda_g^{4/5}, \quad (2.9)$$

with the reduced Compton wavelength $\lambda_g = \hbar/m_g$.

To remain agnostic throughout most of the rest of the paper, the Vainshtein radius will be kept as an independent variable r_v , only studying specific models close to the end.

2.2.2 Modifications to GW Propagation Speed

Let us focus on the modifications to the GW observable introduced by a massive dispersion relation with screening effects. The modifications without screening were originally derived by Will [131], and then extended in [132] to consider more generic dispersion relations. We begin by reviewing Will's original derivation and then incorporating screening effects.

Let us then postulate that a graviton with a non-zero mass obeys the generic dispersion relation of a massive particle

$$E^2 = p^2 + m_g^2. \quad (2.10)$$

Recasting the energy and momentum of the graviton as $E = \hbar\omega$ and $p = \hbar k$, the *classical* or *particle* speed of

propagation is the same as the group velocity for the graviton, namely

$$\left(\frac{d\omega}{dk}\right)^2 = v_g^2 = 1 - \frac{m_g^2}{E^2}. \quad (2.11)$$

Equation (2.11) is where we incorporate additional modifications induced by screening. More specifically, let us consider a modified theory of gravity that excites extra degrees of freedom beyond the two of GR. When these extra degrees of freedom are suppressed, the modified theory reduces essentially to GR, while when they are excited modifications are present. Moreover, let us assume that the extra degrees of freedom are non-linearly coupled, such that they are suppressed inside some screening radius, but not suppressed at larger radii in intergalactic space. The speed of propagation of gravitons is then modified from Eq. (2.11) into

$$v_g^2 = 1 - \frac{m_g^2}{E^2} \Theta(r - r_{V,h}) \Theta(D_{V,MW} - r), \quad (2.12a)$$

$$D_{V,MW} = D_L - r_{V,MW}, \quad (2.12b)$$

where $r_{V,h}$ and $r_{V,MW}$ are the screening radii of the host galaxy and the Milky Way respectively, D_L is the luminosity distance, and the coordinate r is measured from the center of mass of the source. Although Eq. (2.12) is purely phenomenological, it is still interesting to consider what such screening effects would do to constraints on the graviton mass.

A bound for the Compton wavelength of the graviton purely from the difference in travel time between a photon and graviton can quickly be derived. Noting that the Compton wavelength $\lambda_g = h/m_g$, we have that

$$\begin{aligned} \Delta t &= \Delta t_a - (1+z)\Delta t_e \approx D(1 - v_g), \\ \Rightarrow \lambda_g &= \sqrt{\frac{D}{2f^2\Delta t}}, \end{aligned} \quad (2.13)$$

with D some distance measure of *unscreened* space between the source and the detector and f being the graviton's frequency.

A more rigorous calculation is required to figure out exactly what this unscreened distance is. Let us then compute the coordinate distance traveled by a massive graviton in a FLRW universe with metric

$$ds^2 = -dt^2 + a(t)^2 [d\chi^2 + \Sigma(\chi)^2 d\Omega^2]. \quad (2.14)$$

The radial momentum $p_\chi = a(t_e)(E_e^2 - m_g^2)^{1/2}$ is conserved because it is normalized via $p_\mu p^\mu = -m_g^2$, where t_e is the emission time. Since the classical or particle velocity of the graviton $d\chi/dt = p^\chi/p^t$, we then have

that

$$\frac{d\chi}{dt} = \frac{1}{a} \left(1 + \frac{a^2 m_g(r)^2}{p_\chi^2} \right)^{-1/2}, \quad (2.15)$$

where $m_g(r)^2 = m_g^2 \Theta(r - r_{v,h}) \Theta(D_{v,MW} - r)$. This differential equation can be expanded in the small parameter $m_g(r)/E_e \ll 1$, and then solved to find the coordinate distance at emission χ_e , namely

$$\begin{aligned} \chi_e &\approx \int_{t_e}^{t_a} \frac{dt}{a(t)} - \frac{1}{2} \frac{m_g^2}{a^2(t_e) E_e^2} \\ &\times \int_{t_e}^{t_a} a(t) \Theta(r - r_{v,h}) \Theta(D_{v,MW} - r) dt \\ &= t_a - (1+z)t_e - \frac{1+z}{2} \frac{m_g^2}{E_e^2} \mathcal{D}, \end{aligned} \quad (2.16)$$

where we have used that $\Delta t_e \ll H^{-1}$, $1+z = a_0/a_e$ and $\mathcal{D}$ is a new cosmological distance measure defined by

$$\begin{aligned} \frac{\mathcal{D}}{1+z} &= \int_{t_e}^{t_a} a(t) \Theta(r - r_{v,h}) \Theta(D_{v,MW} - r) dt \\ &= \int_0^{z_s} \frac{1}{H(z')} \frac{dz'}{(1+z')^2} - \int_0^{z_{MW}} \frac{1}{H(z')} \frac{dz'}{(1+z')^2}, \end{aligned} \quad (2.17)$$

where z_s and z_{MW} represent the redshift of the edge of the screening effects for the source galaxy and the Milky Way respectively, while H is the Hubble parameter, which for a Λ CDM universe at late times is $H \approx H_0 \sqrt{\Omega_m(1+z)^3 + \Omega_\Lambda}$, with $H_0 = 67.31$ km/s/Mpc the Hubble constant, $\Omega_m = 0.315$ the matter density and $\Omega_\Lambda = 0.685$ the dark energy density [180]. Figure 2.2 shows a schematic view of the system with the quantities labeled for clarity. When screening is absent, $z_s \rightarrow z$ and $z_{MW} \rightarrow 0$, rendering Eq. (2.17) identical to Eq. (2.5) in [131]. The effect of the screening corrections is to “soften” the effect of a massive graviton, hiding the mass for stretches of the propagation distance.

We can now compare the coordinate distances traveled by a massive graviton and a massless photon between Earth and the source. The coordinates describe the same location, so the difference gives

$$\chi_g - \chi_{\text{null}} = 0 = \Delta t - \frac{(1+z)}{2} \frac{m_g^2}{E_e^2} \mathcal{D}. \quad (2.18)$$

Given a observed time delay Δt between an impinging gravitational wave and an impinging electromagnetic

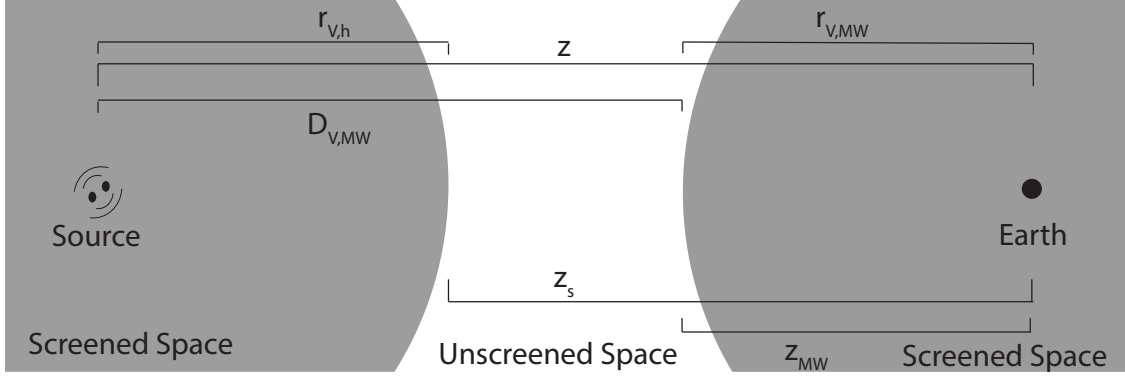


Figure 2.2: Schematic of the binary/detector system with various distances labeled for clarity. z is the redshift to the source, and z_s and z_{MW} are the redshifts to the edges of the screening effects for the host galaxy and the Milky Way, respectively. $r_{V,h}$ ($r_{V,MW}$) is the Vainshtein radius for the host galaxy (Milky Way). $D_{V,MW}$ is the distance from the source to the edge of the Milky Way Vainshtein regime, equal to the luminosity distance to the source minus the Vainshtein radius of the Milky Way ($D_L - r_{V,MW}$)

wave, one could then place the constraint

$$\lambda_g \geq \sqrt{\frac{(1+z)\mathcal{D}}{2\Delta t f^2}}, \quad (2.19)$$

where $\mathcal{D}$ is defined in Eq. (2.17). This expression for the constraint on λ_g is equivalent to the one obtained using Eq. (2.13), where we now see that the *unscreened* distance D appearing in that equation is really $(1+z)\mathcal{D}$.

2.2.3 Phase Modifications from Finite λ_g

While the above calculations illustrate the effectiveness of constraining the mass of the graviton through direct comparison of GW and photon travel times, more stringent constraints can be derived by examining the phase modifications of a gravitational signal coming from coalescing binaries. This phase modification has already been investigated using the GW150914 signal, which has led to the constraint $\lambda_g \gtrsim 10^{13}$ km at 90% confidence through matched filtering [181, 182] and later a bound of $\lambda_g > 1.6 \times 10^{13}$ km by combining several observations [139] (assuming no screening effects are present). For different constraints on the graviton mass using the GW150914 observation, see [183]. In that work, the authors were concerned with constraints on the screening of the *Newtonian potential*, as opposed to the screening of modifications to GR analyzed in this work.

Let us then derive the modifications to the phase when screening is present. Emulating the derivation of Will [131], the coordinate difference calculation of Sec. 2.2.2 can be applied to two gravitons with different energies E and E' , instead of a photon and a graviton. The difference in travel time between two gravitons is

then useful for deriving the phase modifications to a gravitational wave signal. The resulting difference in travel time is then given by

$$\Delta t_a = (1+z) \left[\Delta t_e + \frac{\mathcal{D}}{2\lambda_g^2} \left(\frac{1}{f_e^2} - \frac{1}{f_e'^2} \right) \right]. \quad (2.20)$$

We see that this equation is identical to that derived in [131] except that here the quantity $\mathcal{D}$ contains screening modifications.

The later observation then allows us to immediately write down the modification to the GW Fourier phase. Using the restricted PN approximation, the Fourier transform of the response function for a GW emitted by a compact binary is simply

$$\tilde{h}(f) = A(f) e^{i\Psi(f)}, \quad (2.21a)$$

where the Fourier amplitude is

$$A(f) = \sqrt{\frac{\pi}{30}} \frac{\mathcal{M}_z^2}{D_L} u^{-7/6}, \quad (2.21b)$$

and the Fourier phase is

$$\begin{aligned} \Psi(f) = & 2\pi f t_c - \phi_c - \pi/4 + \frac{3}{128} u^{-5/3} - \beta u^{-1} \\ & + \frac{5}{96} \left(\frac{743}{336} + \frac{11}{4} \eta \right) \eta^{-2/5} u^{-1} - \frac{3\pi}{8} \eta^{-3/5} u^{-2/3}, \end{aligned} \quad (2.21c)$$

with $u = \pi \mathcal{M}_z \tilde{f}$ and $\mathcal{M}_z = (1+z)\eta^{3/5}m$ is the redshifted chirp mass. The Fourier phase is presented here to 1.5PN order, but it is straightforward to include the higher PN order terms that have already been calculated (see e.g. the references in [106, 169]).

The Fourier phase clearly depends on quantities related to the binary itself, like the redshifted chirp mass $\mathcal{M}_z$, with $\eta = m_1 m_2 / m^2$ the symmetric mass ratio and $m = m_1 + m_2$ the total mass, as well as the time and phase of coalescence t_c and ϕ_c . But this phase also depends on massive graviton modifications, encoded in the parameter β , which is given by

$$\beta = \frac{\pi^2 \mathcal{D} \mathcal{M}_z}{\lambda_g^2 (1+z)}, \quad (2.22)$$

where $\mathcal{D}$ is defined as in Eq. (2.17). Clearly then, the modifications to the Fourier phase are identical to those of a massive graviton, except that the quantity $\mathcal{D}$ is an effective luminosity distance that accounts for screening effects.

2.3 Fisher Analysis and Computational Framework

In the next section, we outline the details behind our model creation and statistical analysis. We briefly introduce the basic foundations of a Fisher analysis and define the calculations used to find the estimated bounds on our model parameters. We then go on to describe the algorithm chosen to model our sample waveforms, and finally, we present the parameters of the models we used in this study.

2.3.1 Fisher Information Matrix

The accuracy to which parameters λ^a in a model (the waveform template in this case) can be estimated from a data set (the GW signal in this case) can be approximated through the Fisher information matrix [154] (referred to as the Fisher matrix) in the high signal-to-noise ratio (SNR) limit [184, 185]. The Fisher matrix is defined via

$$\Gamma_{ab} = (\partial_a h | \partial_b h), \quad (2.23)$$

where the inner product is

$$(h_1 | h_2) = 2 \int \frac{h_1 h_2^* + h_2 h_1^*}{S(f)} df, \quad (2.24)$$

with $S(f)$ the sensitivity noise density of the detector and the star representing complex conjugation. The inner product also leads to the definition of the SNR via

$$\rho^2 = (h | h). \quad (2.25)$$

For computing the Fisher elements, we use the projected noise density curves of two ground-based detectors, advanced LIGO at design sensitivity [166] and the ET (or specifically ET-D) [167], a proposed third-generation detector. ET is projected to improve drastically upon aLIGO (by a factor of 10 to 100, especially at low frequencies) and is expected to come online in the 2030s. For space-based detectors, we will use the final LISA design submitted by the ESA [168]. This design will consist of six-links with 2.5 Gm arms and low acceleration noise demonstrated possible with LISA Pathfinder [186].

The calculation of the Fisher matrix requires the evaluation of derivatives of the waveform model and then the integration of these derivatives normalized by the spectral noise density [see e.g. Eq. (2.24)]. The integration for each element of the Fisher matrix is here truncated at the frequency for which the signal is about a tenth of the noise spectrum. If the model was based on a system that contained a neutron star, the cut off frequency was instead set to an approximation of the frequency at contact $f_{\text{contact}} = (Mc^3/\pi^2(24 \text{ km})^3)^{1/2}$. This cutoff was chosen to ensure the accuracy of the model, as this frequency is a conservative estimate of

the point at which the faithfulness of our waveforms breakdown.

Given the Fisher matrix, the variance of any estimated model parameter $\hat{\lambda}^a$ can then be approximated from the Cramer-Rao bound

$$\sigma_{\lambda^a} \geq \sqrt{\Sigma^{aa}}, \quad (2.26)$$

where $\Sigma^{ab} = (\Gamma_{ab})^{-1}$ is the variance-covariance matrix, and no sum over the index a is here implied. In this paper, we will assume a GW signal consistent with GR at a sufficiently high SNR has been detected; we will synthesize this injection through a waveform model evaluated at $\beta = 0$. We will then estimate the accuracy to which the parameters of a waveform model, including β , can be estimated using a Fisher matrix approach.

Given an estimated bound on β , we can then calculate a projected bound on the physical constants that β depends on. Simple propagation of error gives

$$\begin{aligned} \sigma_{\bar{m}_g^2}^2 = & \left(\frac{\partial \bar{m}_g^2}{\partial \beta} \right)^2 \sigma_\beta^2 + \left(\frac{\partial \bar{m}_g^2}{\partial \mathcal{M}} \right)^2 \sigma_{\mathcal{M}}^2 + \left(\frac{\partial \bar{m}_g^2}{\partial D_L} \right)^2 \sigma_{D_L}^2 \\ & + \sum_{i \neq j \in \{\beta, \mathcal{M}_z, D_L\}} 2 \frac{\partial \bar{m}_g^2}{\partial \lambda_i} \frac{\partial \bar{m}_g^2}{\partial \lambda_j} \sigma_i \sigma_j \end{aligned} \quad (2.27)$$

where we have defined $\bar{m}_g \equiv m_g \mathcal{D}^{1/2}/h$. The dependence on the luminosity distance D_L comes in implicitly as $z = z(A_0) = z(\mathcal{M}_z, D_L)$. Since the derivatives with respect to the model parameters are proportional to β and evaluated at the injected parameters ($\beta = 0$), Eq. (2.27) reduces to

$$\sigma_{\bar{m}_g^2} = \frac{1+z}{\pi^2 \mathcal{M}_z} \sigma_\beta, \quad (2.28)$$

where we have used Eq. (2.22) to evaluate the derivative. From this expression, we can derive similar expressions for the variance of the Compton wavelength of the graviton, divided by the effective distance. Clearly, $\mathcal{D}$ (or equivalently r_v) and λ_g (or equivalently m_g) cannot be independently constrained due to the way they enter the β parameter in Eq. (2.22).

2.3.2 Waveform Model

To produce the waveforms needed to compute the Fisher elements, we use the IMRPhenomD model by Khan, et. al. [106, 169]. The model combines analytic, PN inspiral waveforms with phenomenological functions calibrated with numerical relativity simulations for the merger-ringdown phase. The fitting data spanned $\chi \in [-.95, .95]$ and $q = m_1/m_2 \leq 18$ and is designed to emulate spin-aligned or anti-aligned systems. The

functional form of the waveform is

$$\tilde{h}_{\text{GR}}(f) = A_{\text{GR}}(f)e^{i\varphi_{\text{GR}}(f)}, \quad (2.29)$$

where $A(f)$ and $\varphi(f)$ are piece-wise functions defined as

$$\varphi_{\text{GR}}(f) = \begin{cases} \phi_{\text{ins}} & f < 0.018/M, \\ \phi_{\text{int}} & 0.018/M < f < 0.5f_{\text{RD}}, \\ \phi_{\text{mr}} & 0.5f_{\text{RD}} < f, \end{cases} \quad (2.30)$$

$$A_{\text{GR}}(f) = \begin{cases} A_0 A_{\text{ins}} & f < 0.014/M, \\ A_0 A_{\text{int}} & 0.014/M < f < f_{\text{peak}}, \\ A_0 A_{\text{mr}} & f_{\text{peak}} < f, \end{cases} \quad (2.31)$$

with

$$A_0 = \sqrt{\frac{\pi}{30}} \frac{\mathcal{M}_z^2}{D_L} (\pi \mathcal{M}_z f)^{-7/6}. \quad (2.32)$$

where f_{RD} and f_{peak} are the ringdown and peak frequencies, respectively. The specific functional forms for each piece of the waveform can be found in [106, 169].

This model only reproduces GR waveforms, and to extend it to non-GR modifications, specifically to include β [Eq. (2.22)] in the phase, we modify the model via

$$\tilde{h}(f) = \tilde{h}_{\text{GR}}(f)e^{i\beta u^{-1}} = A_{\text{GR}}(f)e^{i(\varphi_{\text{GR}}(f) + \beta u^{-1})}. \quad (2.33)$$

Note that this modification enters in all phases of coalescence, because it is sourced by a correction to the dispersion relation, which is a *propagation* effect and not a modification in the generation of the GWs. The modified waveform model depends on the eight parameters $\lambda^a = [\ln A, \Phi_c, t_c, \ln \mathcal{M}_z, \ln \eta, \chi_s, \chi_a, \beta]$.

As modifications to GR can come in through both generation and propagation effects generally, the veracity of our model might be questionable. However, it was shown in [112] that the effects of modifications to the propagation of GWs overwhelm the effects due to the generation of GWs in general massive theories of gravity. This difference comes about because the propagation effect accumulates over the entire distance traveled, while the generation effects only develop while the system is producing GWs in band. Furthermore, confining our study to theories that exhibit Vainshtein screening should generally suppress any modification to the generation of gravitational waves. This is due to the fact that the scalar field is suppressed around the source, as was numerically verified in cubic Galileon by [134, 187]. These reasons lead us to believe that GW

events in the modified theories we are examining are sufficiently similar to corresponding events in GR that calibration of our waveforms using GR numerical simulations is justified.

2.4 Projected Constraints on Screened Massive Gravity

In this section, we describe the projected constraints we find using a Fisher analysis for a variety of astrophysical parameters that are expected sources for the various ground-based and space-based detectors, as well as for the previously announced LIGO/VIRGO observations. The parameters of the chosen injections that we will study are shown in Table 2.2. We begin by presenting theory-agnostic, projected constraints, and then specialize to constraints on particular theories.

2.4.1 Theory Agnostic

The constraints on β for each of the injections we considered are listed in the last column of Table 2.2. We can then map these constraints to bounds on r_v and λ_g (or m_g) through Eq. (2.28), leaving r_v as an independent parameter. This allows for the results to be interpreted in the context of any theory that exhibits Vainshtein screening with a graviton mass. The Vainshtein radius, however, generally depends on the Schwarzschild radius of the Milky Way and that of the host galaxy, in principle creating an asymmetric screening scenario if the host galaxy mass differs substantially from the mass of the Milky Way. To simplify the analysis below, we assume the host galaxy mass is comparable to that of the Milky Way, therefore creating a symmetric system.

The projected constraints in the m_g - r_v (or λ_g - r_v) plane are shown in Fig. 2.3 for aLIGO, LISA, and ET. For aLIGO sources, the projected constraints follow a uniform pattern. For each source, the effect of screening increasingly grows more pronounced until reaching the midpoint between the source and the Milky Way, corresponding to a fully screened graviton mass. Systems with a higher mass and at closer distances allow for stronger constraints on β (and therefore on m_g and r_v), but systems at much higher redshifts allow for a deeper probing of the screening mechanism at much larger distances from screening sources. The low SNR, low chirp mass of NSNS binaries conspire to inhibit our ability to constrain m_g , and as such, are poor sources to constrain these effects.

The prospects of probing screened massive gravity improve dramatically when incorporating potential LISA and ET detections, as these detectors will be able to observe sources much farther out and with a higher SNR. Due to the nonlinear relation between the cosmological distance in Eq. (2.17) and the luminosity distance, not all distances are equal for this type of test. Sources farther away are not just rescaled versions of closer sources, but actually probe more complex domains of the m_g - r_v relation. The abrupt change in our

Table 2.2: Parameters of the models used in this study and the subsequent predicted bounds on β for each model. All quantities are in the source frame. The bounds on β are a 1σ bound. The sources for aLIGO were picked to emulate previous detections by aLIGO/VIRGO [139, 188–192]

Model	$M_1 (M_\odot)$	$M_2 (M_\odot)$	χ_1	χ_2	D_L (Mpc)	SNR	$\Delta\beta$
aLIGO							
GW150914	36.0	29.0	0.32	0.44	410.0	4.57×10^1	6.61×10^{-2}
GW151226	14.2	7.5	0.2	0.01	440.0	1.64×10^1	7.75×10^{-2}
GW170104	31.2	19.4	0.45	0.47	880.0	1.83×10^1	1.65×10^{-1}
GW170814	30.5	25.3	0.01	0.01	540.0	2.88×10^1	7.20×10^{-2}
GW170817	1.48	1.26	0.02	0.02	44.7	3.08×10^1	1.23×10^{-1}
GW170608	12.0	7.0	0.0	0.0	340.0	1.89×10^1	7.77×10^{-2}
Einstein Telescope							
1	1.4	1.0	0.03	0.01	50.0	3.55×10^2	2.93×10^{-3}
2	500.0	100.0	0.1	0.3	1500.0	5.47×10^2	3.84×10^{-3}
3	70.0	50.0	0.7	0.9	400.0	1.22×10^3	7.19×10^{-4}
4	50.0	3.0	0.7	0.7	400.0	2.63×10^2	6.80×10^{-4}
5	80.0	40.0	0.9	0.2	600.0	7.77×10^2	1.35×10^{-3}
6	100.0	40.0	0.4	0.3	2000.0	2.63×10^2	1.02×10^{-2}
LISA							
7	6×10^6	5×10^6	0.32	0.44	1.00×10^4	3.31×10^3	1.23×10^{-3}
8	6×10^6	5×10^6	0.7	0.8	3.00×10^4	9.21×10^2	7.38×10^{-3}
9	5×10^6	4×10^5	0.45	0.47	8.80×10^3	1.26×10^3	1.16×10^{-3}
10	5×10^5	4×10^4	0.01	0.01	5.40×10^3	1.57×10^3	3.72×10^{-4}
11	5×10^4	4×10^3	0.7	0.9	1.60×10^4	1.99×10^2	9.80×10^{-4}
12	5×10^6	4×10^6	0.7	0.9	4.80×10^4	5.32×10^2	1.39×10^{-2}

ability to constrain the graviton mass gives way to a much softer effect at farther screening radii where large changes in screening radius impact our constraints much less. The development of third generation detectors, like ET and LISA, will also allow for a large population of observations, which could also be leveraged to

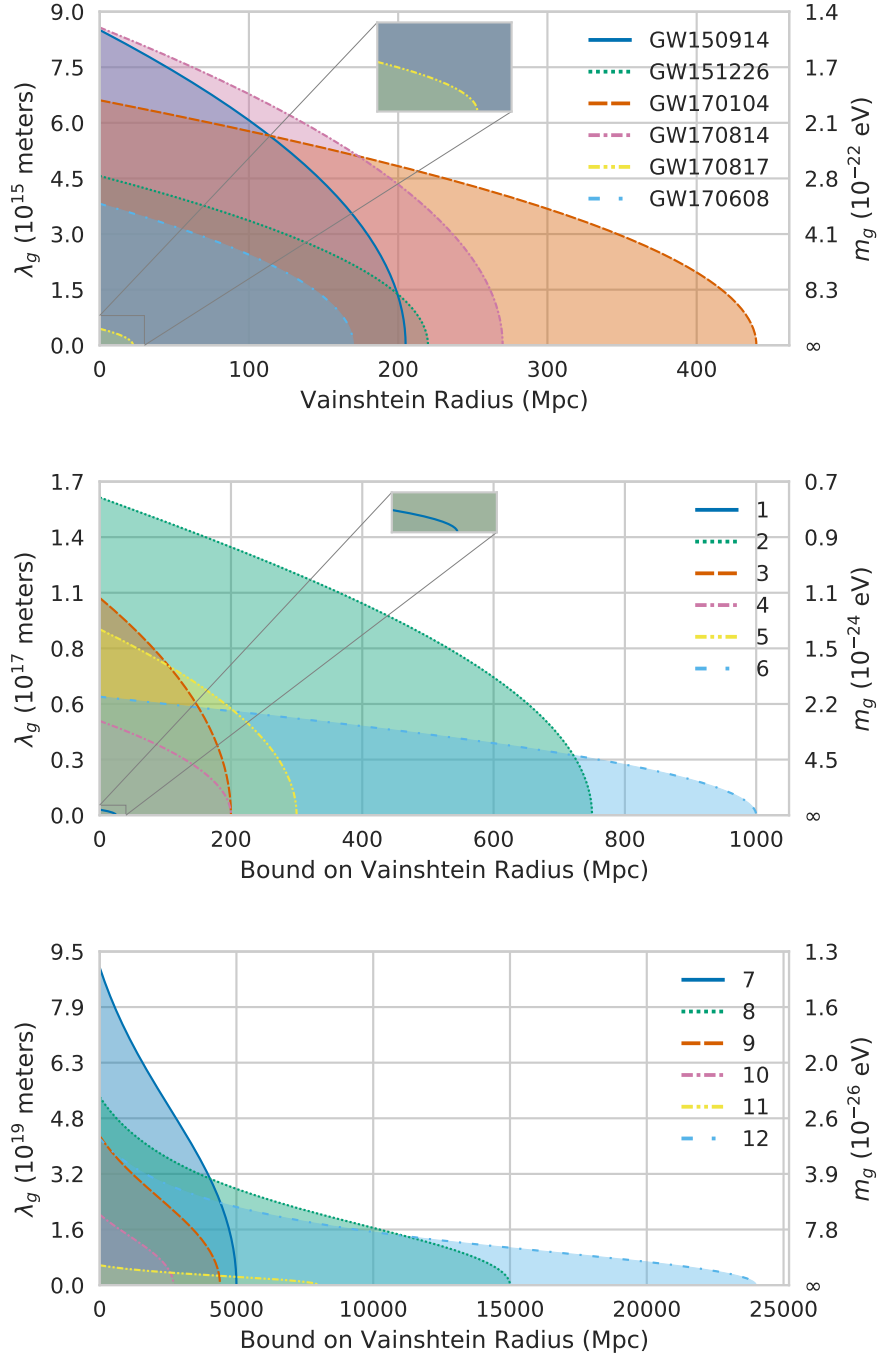


Figure 2.3: (Color Online) Projected constraint on the mass of the graviton as a function of the Vainshtein radius for a variety of systems detected with aLIGO (top), ET (middle) and LISA (bottom). The shaded regions would be disallowed given injections consistent with GR.

improve constraints via stacking. As the redshift grows, these constraints become increasingly sensitive to the cosmological model employed because of the dependence of $\mathcal{D}$ on the Hubble function $H(z)$.

2.4.2 dRGT and bigravity

Let us now consider projected constraints on specific modified gravity models that predict a massive graviton and screening. In the context of specific theories like dRGT and bigravity, the screening radius becomes a function of the graviton mass. Having analytic expressions for the screening radius is useful for exploring the possible bounds on these parameters, but it should be noted that the theories themselves do not necessarily realize these types of modifications [153, 165]. Replacing the parameter r_v with the relation shown in Eq. (2.8) analytically breaks the degeneracy between r_v and m_g . To calculate the graviton mass constraint, as the relationship between $\mathcal{D}$ and r_v is highly nonlinear, we numerically solve Eq. (2.22). The results are shown in Table 2.3, along with the corresponding screening radius for a Milky Way sized galaxy and the percent change in the graviton mass bound if screening were not incorporated.

In the context of dRGT- and bigravity-type screening radii, screening has a minimal effect given the projected sensitivity of second- and third-generation detectors. This is because saturating the bounds on m_g gives screening radii of $\mathcal{O}(1\text{pc})$, since r_v is proportional to $1/m_g^2$. Once detectors become sensitive enough to begin probing the graviton mass at the $\mathcal{O}(H_0)$ level, the screening radius for a Milky Way sized galaxy will grow to $\mathcal{O}(\text{Mpc})$. With a screening radius of this magnitude, the effects might warrant further investigation.

Examining Eq. (2.22), we see that the ideal candidate source to place constraints will have a large chirp mass and a low redshift, as screening effects encompassed by $\mathcal{D}$ are enhanced and degraded by high chirp mass and high redshift, respectively. As space-based detectors will target this higher range of chirp masses, sources being targeted by future LISA-type missions, like the Big Bang Observer [193], would be the ideal subjects of future studies on this mechanism. Even though the average redshift of these millihertz systems will be higher, the average chirp mass will be many orders of magnitude larger than targeted sources for current and future ground-based detectors, overwhelming any loss due to higher distance.

2.5 Conclusions and Future Work

We studied the effect of screening on projected, future constraints on the mass of the graviton from a model-agnostic viewpoint and then within a model-specific approach. When studying model-independent constraints, we found that the effect of screening is to generically deteriorate bounds on the graviton mass. The reason for this is that screening eliminates the correction to the dispersion relation during part of the graviton travel, thus decreasing the overall correction to the GW phase. When studying model-dependent constraints, we found that the effect of screening on graviton mass bounds is actually quite small for the models considered. The reason for this is that the relation between the screening radius and the mass of

Model	Δm_g (Screened)	Percent Change (%)	r_v
aLIGO			
GW150914	1.46×10^{-22}	0.03	3.76×10^{-8}
GW151226	2.72×10^{-22}	0.03	2.49×10^{-8}
GW170104	1.88×10^{-22}	0.06	3.18×10^{-8}
GW170814	1.45×10^{-22}	0.04	3.78×10^{-8}
GW170817	2.82×10^{-21}	0.0	5.22×10^{-9}
GW170608	3.25×10^{-22}	0.02	2.21×10^{-8}
ET			
1	4.44×10^{-22}	0.0	1.79×10^{-8}
2	7.79×10^{-24}	0.09	2.65×10^{-7}
3	1.14×10^{-23}	0.03	2.06×10^{-7}
4	2.63×10^{-23}	0.03	1.18×10^{-7}
5	1.33×10^{-23}	0.04	1.86×10^{-7}
6	2.09×10^{-23}	0.1	1.37×10^{-7}
LISA			
7	1.37×10^{-26}	0.24	1.82×10^{-5}
8	2.31×10^{-26}	0.29	1.29×10^{-5}
9	2.92×10^{-26}	0.23	1.10×10^{-5}
10	6.17×10^{-26}	0.19	6.68×10^{-6}
11	2.19×10^{-25}	0.27	2.87×10^{-6}
12	2.97×10^{-26}	0.3	1.09×10^{-5}

Table 2.3: Projected 1σ bounds for m_g for dRGT- and bigravity-type screening radii. The second column shows the percent change in the constraint on m_g between the screened and unscreened constraints, while the screening radius r_v is measured in Mpc, and the graviton mass in eV.

the graviton is such that, for the range of graviton masses that second- and third-generation detectors are sensitive to, the screening radius is very small.

Future work could focus on a variety of topics. As we enter the era of third-generation detectors, or even second-generation detectors at design sensitivity, the number of GW observations will greatly increase. Constraints on the mass of the graviton, therefore, can be enhanced by combining multiple observations. Doing so can have interesting effects on model-agnostic constraints on the screening radius. In the model-dependent case, stacked constraints could become enough that the effect of screening is not negligible and ultimately ought to be taken into account.

Another possible avenue for future work could involve the more careful study of screening within known models. As we described in this paper, Vainshtein screening has really only been studied in detail within some greatly simplified physical models, such as in stationarity and spherical symmetry. The compact binaries that generate the GWs that we detect, however, are neither stationary, nor spherically symmetric. The degree of

effectiveness of screening in such cases should be investigated further.

A final possibility for future work could consist on the study of other cosmological modified theories to determine whether they also predict screening. Theories such as bigravity, or Horndeski and beyond Horndeski theories could be studied further to determine formally whether the dispersion of GWs is partially screened in regions of high density or high curvature. It is possible that screening is not present for GWs in some of these theories, as is the case in quartic and quintic Galileon theories, but this ought to be demonstrated with careful calculation.

Acknowledgements The authors would like to thank Claudia de Rham, Rachel Rosen, Hector Okada da Silva, and Mark Trodden for useful conversations and input. We would also like to thank Kent Yagi for comparisons with his Fisher code. The authors acknowledge support from NSF grant PHY-1759615 and NASA grants NNX16AB98G and 80NSSC17M0041.

Chapter 3

Probing Fundamental Physics with Gravitational Waves: The Next Generation

This chapter is based on: *Perkins, S.¹, Yunes, N.² and Berti, E.³ (2021). Probing Fundamental Physics with Gravitational Waves: The Next Generation. Phys. Rev. D 103, 044024*

Abstract Gravitational wave observations of compact binary mergers are already providing stringent tests of general relativity and constraints on modified gravity. Ground-based interferometric detectors will soon reach design sensitivity and they will be followed by third-generation upgrades, possibly operating in conjunction with space-based detectors. How will these improvements affect our ability to investigate fundamental physics with gravitational waves? The answer depends on the timeline for the sensitivity upgrades of the instruments, but also on astrophysical compact binary population uncertainties, which determine the number and signal-to-noise ratio of the observed sources. We consider several scenarios for the proposed timeline of detector upgrades and various astrophysical population models. Using a stacked Fisher matrix analysis of binary black hole merger observations, we thoroughly investigate future theory-agnostic bounds on modifications of general relativity, as well as bounds on specific theories. For theory-agnostic bounds, we find that ground-based observations of stellar-mass black holes and LISA observations of massive black holes can each lead to improvements of 2–4 orders of magnitude with respect to present gravitational wave constraints, while multiband observations can yield improvements of 1–6 orders of magnitude. We also clarify how the relation between theory-agnostic and theory-specific bounds depends on the source properties.

3.1 Introduction

Einstein’s general relativity (GR) has been wildly successful. The agreement with the observed perihelion precession of Mercury and the 1919 eclipse expedition to verify the prediction of relativistic light-bending

¹Carried out the calculations in the paper; wrote the software related to waveform modelling; wrote the statistical analysis software; wrote the software used to generate synthetic catalogs; conducted the numerical experimentation; wrote the first draft of the manuscript

²Contributed to the conception of the project; supervised and advised on the calculations and experimentation (particularly with the analytic calculations and those calculations related to inspiral tests of GR); edited the manuscript

³Contributed to the conception of the project; supervised and advised on the calculations and experimentation (particularly with the BBH population-related calculations and the creation of synthetic catalogs); edited the manuscript

around the Sun were the beginning of a century of thorough vetting [51]. The theory has passed every experimental test so far, and it was recently validated in the strong-field regime, most notably through the imaging of a black-hole (BH) shadow in the electromagnetic spectrum by the Event Horizon Telescope [194] and through the observation of coalescing binary black holes (BBHs) by the LIGO/Virgo Collaboration [57, 195].

One century of experimental triumphs did not deter theoretical work on observationally viable extensions of GR for mainly two sets of reasons [52]. The first is observational: some of the most outstanding open questions in physics might be explained by modifying the gravitational sector. For example, one could introduce an additional scalar field to the gravitational action [79, 80] or allow the graviton to be massive [81–83] to explain the late-time acceleration of the Universe [196, 197] without invoking the cosmological constant or dark energy. The second set of reasons is theoretical: string theory and other ultraviolet completions of the Standard Model usually add higher-order curvature corrections to the Einstein-Hilbert action, implying deviations from GR at high energies and large curvatures [85–87]. Therefore it is important to systematically test the assumptions underlying GR, which are often summarized in terms of Lovelock’s theorem [52, 88]. More specifically, GR assumes that the gravitational interaction is mediated by the metric tensor alone; the metric tensor is massless; spacetime is four-dimensional; the theory of gravity is position-invariant and Lorentz-invariant; and the gravitational action is parity-invariant. There is no *a priori* reason why these assumptions should be true, and therefore it is reasonable to explore alternatives to GR by systematically questioning each of them [52, 154]. Our study is motivated by a combination of these two reasons: we will focus on theories that may address long-standing problems in physics, while questioning the validity of the main assumptions behind GR.

The LIGO-Virgo-KAGRA network of Earth-based detectors just completed their third observing run (O3). A fourth observing run (O4) is planned in 2022, and future observations will combine data from LIGO Hanford [198], LIGO Livingston [198], Virgo [16], KAGRA [17], LIGO India [199], and third-generation (3g) detectors such as Cosmic Explorer (CE) [200] and the Einstein Telescope [201]. The space-based observatory LISA [168], scheduled for launch in 2034, will extend these observations to the low-frequency window. As existing ground-based detectors are improved, new ones are built and space-based detectors are deployed, our ability to test GR will be greatly enhanced, but to what level?

The main goal of this study is to combine the anticipated timeline of technological development for Earth- and space-based gravitational-wave (GW) detectors with astrophysical models of binary merger populations to determine what theories will be potentially ruled out (or validated) over the next three decades. We estimated parameters by running $\sim 10^8$ Fisher matrix calculations using waveform models including the

effects of precession [106, 169, 202]. Our null hypothesis is that GR correctly describes our Universe, and that all modifications must reduce to GR in some limit for the coupling constants of the modified theory [154]. Under this assumption, we employ the parameterized post-Einsteinian (ppE) framework [107, 115–117] to place upper limits on the magnitudes of any modification, assuming future GW observations to be consistent with GR. As our GW observatories are most sensitive to changes in the GW phase, we ignore modifications to the GW amplitude, an approximation that has been shown to be very good [203].

Executive Summary

For the reader’s convenience, here we provide an executive summary of the main results of this lengthy study.

(i) We use public catalogs of BBH populations observable by LISA and by different combinations of terrestrial networks over the next thirty years, and extract merger rates and detection-weighted source parameter distributions.

While this was not the main goal of this work, we did require astrophysical population models to realistically model GW science over the next three decades. In the pursuit of constructing forecasts of constraints on GR, we developed useful statistics concerning the distribution of intrinsic parameters for detectable merging BBHs for a variety of population models and detectors.

Useful quantities calculated here and related to BBH mergers are the expected detection rates for a large selection of population models and detector networks. These rates are listed in Table 3.6, and discussed in Secs. 3.3.1 and 3.3.2. Detection rates depend not only on the population model, but also on the detector network. For LISA, we follow the method outlined in Ref. [204] to compute detection rates for multiband and massive black hole (MBH) sources.

We constructed synthetic catalogs by filtering the datasets coming from the full population models based on their signal-to-noise ratio (SNR). This yields a detection-weighted distribution of source parameters (discussed in Sec. 3.4.3) which is useful to understand detection bias and to understand the typical sources accessible by different networks over the next three decades. In Figs. 3.4 and 3.5 we show these distributions for a large selection of detection network/population model combinations, considering both stellar-origin black holes (SOBHs) and MBHs.

The main conclusions of this analysis are summarized in Fig. 3.6, which shows the typical detection rates and SNR distributions for different source models and networks. This plot contains key information on the relative constraining performance of different population model/detector network combinations, which will be important for the following discussion of tests of GR.

(ii) We find that improvements over existing GW constraints on theory-agnostic modifications

to GR range from 2 to 4 orders of magnitude for ground-based observations, from 2 to 4 orders of magnitude for LISA observations of MBHs, and from 1 to 6 orders of magnitude for multiband observations, depending on what terrestrial network upgrades will be possible, on LISA’s mission lifetime, and on the astrophysical distribution of merging BBHs in the Universe.

The main issue addressed in this work is the scientific return on investment of future detector upgrades in terms of future explorations of strong gravity theories beyond GR. What future detectors and network upgrades are most efficient at constraining beyond-GR physics? Our models use astrophysical populations of SOBHs and MBHs and three reasonable development scenarios for ground-based detectors (ranging from optimistic to pessimistic) to try and answer this question. We first consider generic (theory-agnostic) modifications of GR, and then focus on specific classes of theories that test key assumptions underlying Einstein’s theory.

Our primary conclusions for generic modifications to GR are summarized in Fig. 3.7 and in Sec. 3.6.1, where we show bounds on generic deviations from GR at a variety of post-Newtonian (PN) orders, separated by the class of source and marginalized over the detector configurations and population models. A term in the GW phase that is proportional to $(\pi\mathcal{M}f)^{b/3}$, where $\mathcal{M}$ is the chirp mass of the binary and f is the GW frequency, is said to be of $(b + 5)/2$ PN order. While the range in constraints between the different models and scenarios is large, we have plotted constraints from current pulsar and GW tests of GR for comparison, where available and competitive. There are several trends present in this figure, most notably:

- 1) SOBH multiband sources observed by both LISA and terrestrial networks are the most effective at setting bounds on negative PN effects, outperforming all other classes of sources by at least an order of magnitude. This observation must be tempered, however, because no multiband sources are observed at all in some of the scenarios we have analyzed. The detection rate of multiband sources is an open question [204, 205]. We hope that their importance for tests of GR, outlined here and elsewhere [206–212], will stimulate further work on this class of sources.
- 2) The MBH mergers observed by LISA outperform SOBH sources observed only in the terrestrial band for negative PN orders in the more pessimistic ground-based detector scenarios. For most negative PN orders, LISA MBH observations perform at least comparably to the most optimistic terrestrial network scenario, and greatly outperform the other two terrestrial scenarios analyzed in this work.
- 3) Terrestrially observed SOBH sources are most effective at constraining positive PN effects, outperforming MBHs and multiband sources. Furthermore, for positive PN effects, the difference between the different

terrestrial network scenarios closes dramatically. The constraining power between the different terrestrial networks shrinks, spanning a range of 4 orders of magnitude at negative PN orders but showing significant overlap for positive PN orders. This suggests that highly sensitive detectors are less important for constraining deviations that first enter at positive PN order, as opposed to negative PN order.

In terms of what detectors would have the highest return on investment, LISA’s contribution to constraints on negative PN effects is quite high. Multiband sources are, by far, the most effective testbeds for fundamental physics in the early inspiral of GW signals, but even in the absence of multiband sources (a realistic concern), MBH sources perform as well or better than even the most optimistic terrestrial network scenario we examined. The difference in terrestrial network scenarios is fairly drastic for negative PN effects, and so ground-based detector upgrades would play an important role if LISA were not available. The strongest improvement occurs in our most optimistic scenario (including CE and ET), but there is also a clear separation between the “pessimistic” and “realistic” scenarios.

Terrestrial networks perform the best for positive PN effects, but not by orders of magnitude. Even at positive PN orders, LISA MBH sources are still as effective as the more pessimistic terrestrial network scenarios. Furthermore, while constraining positive PN effects, no single terrestrial network scenario drastically outperforms the others: there is a clear hierarchy between the three scenarios, but with significant overlap.

These conclusions are also summarized in Table 3.1, where we show a concise overview of current constraints on generic ppE parameters coming from observations of pulsars [49] and GWs [57], and we compare them against forecasts from our simulations.

(iii) LISA and future terrestrial network constraints on theory-agnostic modifications to GR follow trends which depend on the PN order, the underlying population of sources, and the detector network.

Using suitable approximations, we derive analytical expressions that help to elucidate the reason for the hierarchy of constraining power observed in our simulations. We first examine single observations, and show how different source properties influence the constraints. We then attempt to quantify the importance of stacking multiple observations to develop a cumulative constraint from an entire catalog of observations.

In Sec. 3.6.1 [Eqs. (3.32) and (3.33)] we show that, to leading order, the relative constraining power of one class of sources over another depends on the binary masses and on the initial frequency of observation, raised to a power which depends on the PN order in question. As this power changes sign going from negative to positive PN orders, this scaling explains why multiband and MBH sources are more competitive at negative PN orders, while terrestrial networks are more effective at positive PN orders. This trend is succinctly summarized in Fig. 3.8.

PN order (ppE b)	Current Constraint	Best (Worst) Constraint	Best (Worst) Source Class
-4 (-13)	—	10^{-25} (10^{-14})	MB (T)
-3.5 (-12)	—	10^{-23} (10^{-14})	MB (T)
-3 (-11)	—	10^{-21} (10^{-12})	MB (T)
-2.5 (-10)	—	10^{-19} (10^{-11})	MB (T)
-2 (-9)	—	10^{-17} (10^{-10})	MB (T)
-1.5 (-8)	—	10^{-15} (10^{-9})	MB (T)
-1 (-7)	2×10^{-11}	10^{-13} (10^{-11})	MB (MBH)
-0.5 (-6)	1.4×10^{-8}	10^{-11} (10^{-8})	MB (T)
0 (-5)	1.0×10^{-5}	10^{-7} (10^{-5})	MBH (T)
.5 (-4)	$4.4 \times 10^{-3*}$	10^{-7} (10^{-5})	MB (T)
1 (-3)	$2.5 \times 10^{-2*}$	10^{-6} (10^{-4})	MB/T (T)
1.5 (-2)	0.15^*	10^{-5} (10^{-3})	T (MB)
2 (-1)	0.041^*	10^{-4} (10^{-2})	T (MB)

Table 3.1: Summary of the constraints we predict on the theory-agnostic ppE modification parameter β as a function of the PN order parameter b , as defined in Eqs. (3.25) and (3.26) below. We compare these constraints against current constraints from pulsar tests [49] and GW observations from the LVC [57], denoted by (*). The LVC analysis used a slightly different formalism, so we mapped their results to the ppE framework for 4 specific sources (GW150914, GW170104, GW170608, and GW170814), we computed the standard deviation of the Markov Chain Monte Carlo (MCMC) samples, and then combined the posteriors assuming a normal distribution to obtain a rough order-of-magnitude estimate of current ppE bounds from the LVC results. The columns list, from left to right: the PN order of each particular modification, the current constraint (if one exists), the best and worst constraints from our simulations, and the class of astrophysical sources those constraints come from. All the constraints are 1σ bounds, and we only show worst-case constraints that still improve on existing bounds. The source class acronyms are as follows: MB stands for multiband observations of SOBHs, T stands for terrestrial-only observations of SOBHs, and MBH stands for space-based detection of MBHs.

Besides single-source trends, in Sec. 3.6.1 we quantify the effect of stacking observations and the benefit of large catalogs. In Fig. 3.9 we show that, as the PN order of the modification goes from negative to positive, the number of single observations meaningfully contributing to the cumulative bound from a catalog rises exponentially. This helps to further explain the improvement of terrestrial-only catalogs over LISA catalogs for higher PN orders: the very large catalogs coming from third-generation detectors are effectively leveraged to produce much stronger bounds, but only for positive PN orders. As shown in Fig. 3.10, this depends on the relation between the three parameters of primary concern (the SNR, the chirp mass, and the constraint), and on how their relation evolves as a function of the PN order.

These considerations help us understand the behavior observed in our simulations. The single-source scaling implies that MBHs and multiband sources should be more efficient at negative PN orders, because of the typical masses and initial frequencies of the observations. At positive PN orders the balance shifts in favor of terrestrial-only catalogs, further enhanced by the fact that large catalogs bear much more weight for positive PN effects.

The considerations made above also explain the significant overlap of different terrestrial detection scenarios at positive PN orders, and their separation at negative PN orders: negative PN effects are well constrained by single, loud events (favoring the most optimistic detector scenarios), while positive PN effects benefit from large catalogs. As detection rates are comparable for all three terrestrial scenarios, they perform comparably for positive PN effects.

(iv) We quantify the expected improvement over current constraints on theory-specific coupling parameters. We derive trends for theory-specific scalings and find that some conclusions following from generic modifications must be *reversed*.

The analysis of generic deviations from GR is a good theory-agnostic diagnostic tool for estimating the efficacy of future efforts to constrain fundamental physics. This is useful to perform null tests of GR, but at the end of the day, tests of GR focused on specific contending candidates provide the most meaningful physical insights [213]. Many of the trends observed for generic modifications remain valid when considering specific theories, but the scaling relations we observe in our simulations can change significantly for some of our target theories.

A bird’s eye summary of our conclusions can be found in Table 3.2. There we identify the current bound on theory-specific parameters, our predicted bounds after thirty years, and the class of sources which is most effective at improving the bounds. In this table we only include constraints obtained from actual data with a robust statistical analysis, in an effort to limit our comparisons to reliable experimental limits (as opposed to forecasts, simulations, etcetera). In-depth results by source class and trend derivations are presented in

Theory	Parameter	Current bound	Most (Least) Stringent Forecasted Bound	Most (Least) Constraining Class
Generic Dipole	$\delta\dot{E}$	1.1×10^{-3} [112, 161]*	10^{-11} (10^{-6})	MB (T)
Einstein-dilaton -Gauss-Bonnet	$\sqrt{\alpha_{\text{EdGB}}}$	1 km [214] 3.4 km [124]*	10^{-3} (1) km	T (MBH)
Black Hole Evaporation	$\dot{M}$	–	10^{-8} (10^2) $M_\odot/\text{yr}$	MB (T)
Time Varying G	$\dot{G}$	$10^{-13} - 10^{-12} \text{ yr}^{-1}$ [215–219]	10^{-9} (10) yr^{-1}	MB (T)
Massive Graviton	m_g	10^{-29} eV [149, 220–222] 10^{-23} eV [57, 145]*	10^{-26} (10^{-24}) eV	MBH (MB)
dynamic Chern Simons	$\sqrt{\alpha_{\text{dCS}}}$	5.2 km [223]	10^{-2} (10) km	T (MB)
Non-commutative Gravity	$\sqrt{\Lambda}$	$2.1 l_p$ [224]*	10^{-3} (10^{-1}) l_p	T (MB)

Table 3.2: Summary of forecasted constraints on specific modifications of GR. The source class acronyms are the same as in Table 3.1. A (*) symbol denotes constraints coming from previous BBH observations, as opposed to other experimental evidence. When necessary, we have mapped all existing constraints to 1σ constraints by assuming the posterior to be normally distributed. We only show worst-case constraints that improve on existing GW bounds. For consistency with previous work, $\dot{M}$ is given in units of $M_\odot/\text{yr}$, while we use geometrical units (so that $\delta\dot{E}$ is dimensionless) for the generic dipole radiation bound. Note that the necessary factor for transforming between the two is $c^3/G = 6.41 \times 10^{12} M_\odot/\text{yr}$. The time derivative of the gravitational constant, $\dot{G}$, is normalized to the current value of G , and it does indeed have units of yr^{-1} in geometrical units (where $G = c = 1$).

Sec. 3.6.2. We refer the reader to that section for a detailed discussion of individual theories. In broad terms, the process of mapping generic constraints to theory-specific parameters can impose significant modifications to the trends observed in the analysis of generic constraints. These modifications can be significant enough to completely reverse the conclusions derived from generic deviations. This should temper any interpretation of our conclusions from general modifications. We also remark that our analysis for specific theories is far from comprehensive: there is, in principle, a very large number of GR modifications that have different mappings to ppE parameters, and therefore different trends in connection with source distributions.

Our conclusions on the best return of investment from GW detector development from the generic modification analysis *generally* hold also for specific theories. EdGB gravity (Sec. 3.6.2) and massive graviton theories (Sec. 3.6.2) are two notable exceptions: in these cases, the dependence of the theory-agnostic parameters on source mass, spin and distance implies that the generic modifications predictions (at -1PN and 1PN orders, respectively) must be reversed.

The remainder of the paper presents the calculations summarized above in much more detail. The plan of the paper is as follows. In Sec. 3.2 we give details on the detector networks implemented in this work. This section includes information about the proposed timelines of detector development, as well as the specific

sensitivity curves we have implemented at each stage. In Sec. 3.3 we discuss the statistics with which this network is used to filter astrophysical populations, including the calculation of detection probabilities for both terrestrial and space-based detectors. In Sec. 3.4 we describe the population models, then discuss the calculation of detection rates and the creation of our synthetic catalog. In Sec. 3.5 we outline the statistics of parameter estimation procedures and waveform models, including a brief overview of Fisher analysis and the modified-GR waveforms implemented in this study. In Sec. 3.6 we present the results of our numerical investigation, as well as an analytical analysis to break down certain trends that have appeared in our findings. Finally, in Sec. 3.7 we discuss limitations of this study and directions for future work. To improve readability, some technicalities about Bayesian inference and Fisher matrix calculations, the mapping of the ppE formalism to specific theories and our waveform models are relegated to Appendices A, B and C, respectively. Throughout this paper we will use geometrical units ($G = c = 1$) and we assume a flat Universe with the cosmological parameters inferred by the Planck Collaboration [180].

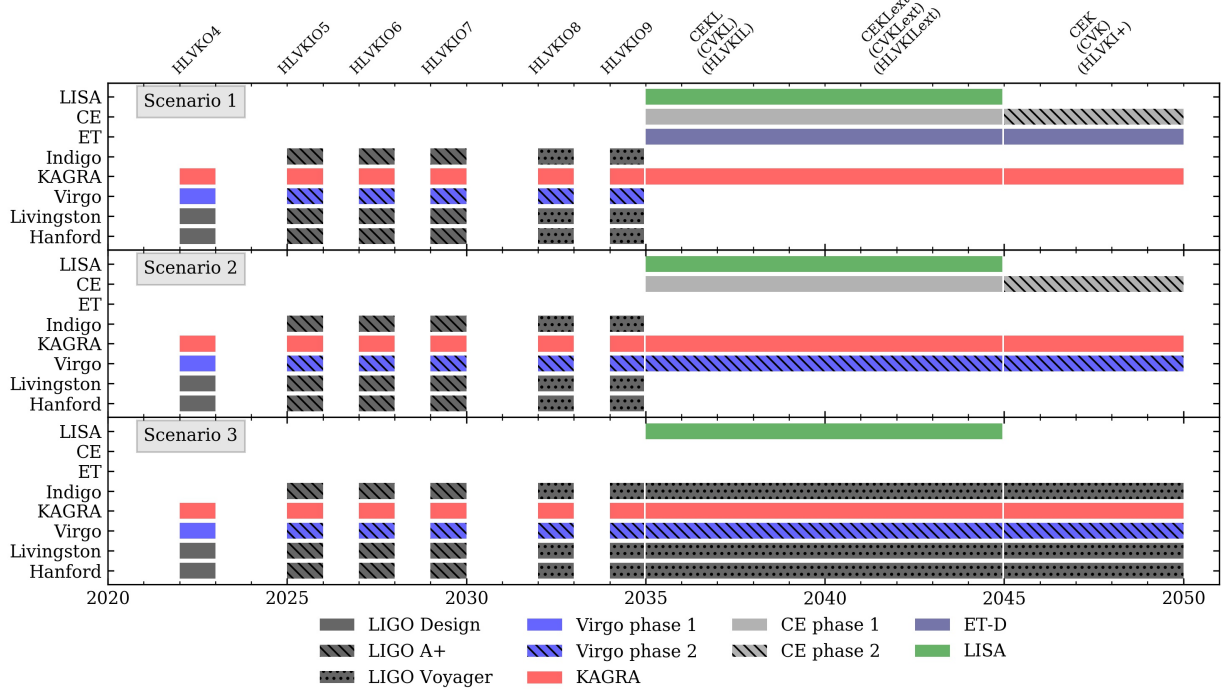


Figure 3.1: Graphical representation of Table 3.3. The shaded regions in the figure represent periods of active observation, and the colors/hatching corresponds to the noise curve being implemented, as shown in Fig. 3.2.

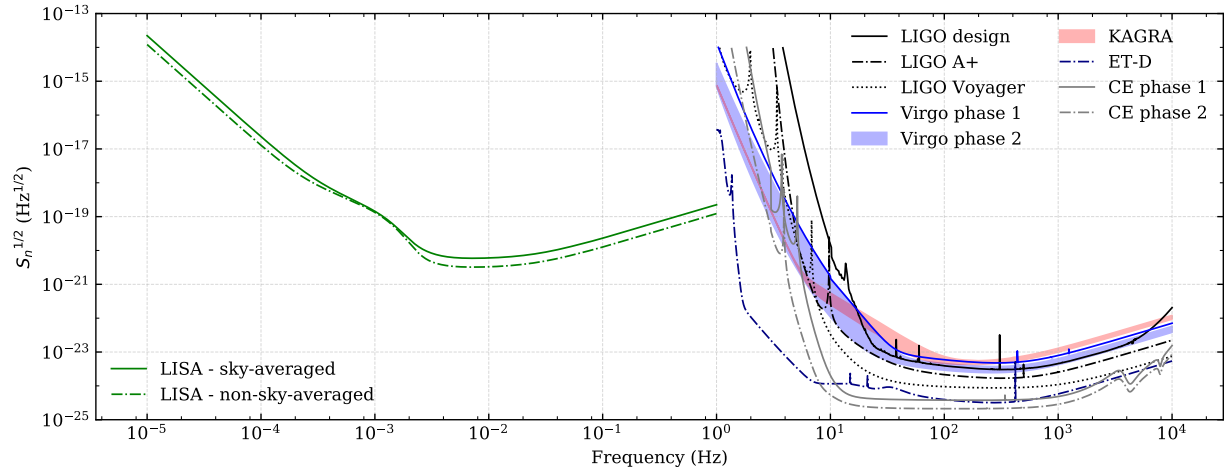


Figure 3.2: Noise curves for the various detector configurations studied in this work. The shaded bands observed for the Virgo+ phase 2 and KAGRA sensitivities reflect uncertainties in estimates of their anticipated power spectral densities.

3.2 Detector Networks

The construction and enhancement of GW detectors across the world and in space is expected to proceed steadily over the next thirty years. Tests of GR using GW observations are fundamentally tied to this global timeline of detector development, so it is important to have a realistic range of models for detector networks that spans the inevitable uncertainties intrinsic in planning experiments over such a long time. In this section we describe potential timelines for upgrades and deployment of new detectors, our assumptions on the location of the detectors, and their expected sensitivities.

3.2.1 Estimated Timeline

Three plausible scenarios for the GW detector roadmap as of the writing of this paper are schematically presented in Fig. 3.1, with more details in Table 3.3. The timeline starts with the fourth observing run (O4) of the LIGO-Virgo-KAGRA detectors, which are scheduled to take data at their design sensitivities for one year starting in 2022. After this run, the instruments would be taken offline to be upgraded to higher sensitivity, with the next set of one-year-long observing runs starting in 2025. At this point, the network would also be joined by LIGO-India. Subsequent upgrades for the LIGO detectors to LIGO Voyager are planned for the early 2030's. The plans for 3g detectors are understandably more uncertain, with CE and ET potentially joining the network in 2035. After a 5–10 year observing run, CE is expected to be taken offline for upgrades, with a second set of runs expected in 2045. Meanwhile, LISA is scheduled to fly in 2034, with a minimum mission lifetime of 4 years and a possible extension by 6 additional years, for a total of 10 years of

observation [225].

Given the timeline described above, one can identify several distinct periods of observations in which a different combination of detectors would be simultaneously online. During the O4 run, LIGO Hanford (H), LIGO Livingston (L), Virgo (V) and KAGRA (K) are expected to collect data simultaneously, creating the HLVKO4 network. LIGO India is expected to join the data collection effort in the late 2020’s for the O5, O6 and O7 observation campaigns, creating the HLVKIO5/O6/O7 networks. In the early 2030’s, the LIGO detectors (Hanford, Livingston, and Indigo) will be upgraded to the Voyager design, reflected in the HLVKIO8/09 networks.

The timeline beyond 2035 is quite uncertain, and we cannot model every possible scenario. Therefore, we chose to model three different timelines:

- 1) After 2035, an optimistic detector schedule would see the Virgo and LIGO detectors replaced by the Einstein Telescope (E) and CE (C) detectors, respectively. Furthermore, LISA (L) is targeting around 2035 as the beginning of its data collection, with a nominal 4-year mission and an additional 6-year extension. These assumptions correspond to the CEKL and CEKLext networks, respectively. We follow up the multiband observation campaigns with a final terrestrial-only observation period from 2045-2050 for the CEK network. This timeline is shown as “Scenario 1” in Table 3.3.
- 2) A less optimistic scenario might see one terrestrial 3g detector receive full funding and come online in the 2030’s. We chose to use CE as our one 3g terrestrial detector to create the CVKL, CVKLext, and CVK networks. This is “Scenario 2” in Table 3.3.
- 3) We also consider a pessimistic scenario where no terrestrial 3g detectors will be observing before the 2050’s. The network will remain at its O9 sensitivity, but it will still be joined by LISA in the 2030’s. This scenario includes the HLVKIL, HLVKILext, and HLVKI+ networks, and is denoted as “Scenario 3” in Table 3.3.

Because these last three observation periods for all three scenarios are less defined and span a wide time range, we assume an 80% duty cycle when estimating terrestrial-only detection rates, but we use the full observation period for calculating multiband rates.

Table 3.3: The above timeline tabulates the exact terrestrial detector evolution utilized by this study. There is a single timeline of detectors until 2035, when we model three separate scenarios that could play out in the next three decades: Scenario 1, 2, and 3. A graphical representation is shown in Fig. 3.1. The various sensitivity curves in column 3 are shown in Fig. 3.2.

Year	Detectors	Noise curves	Moniker(s)
------	-----------	--------------	------------

Table 3.3 (cont.)

2022-2023 [226]	LIGO Hanford	Advanced LIGO design [227]	HLVKO4
	LIGO Livingston	Advanced LIGO design	
	Virgo	Advanced Virgo+ phase 1 [227]	
	KAGRA	KAGRA 80Mpc or 128Mpc [227]	
2025-2030 [226] (one year observations in alternating years)	LIGO Hanford	Advanced LIGO A+ [227]	HLVKIO5 HLVKIO6 HLVKIO7
	LIGO Livingston	Advanced LIGO A+	
	Virgo	Advanced Virgo+ phase 2 high or low [227]	
	KAGRA	KAGRA 80Mpc or 128Mpc	
	LIGO India	Advanced LIGO A+	
2032-2035 (one year observations in alternating years)	LIGO Hanford	Advanced LIGO Voyager [228]	HLVKIO8 HLVKIO9
	LIGO Livingston	Advanced LIGO Voyager	
	Virgo	Advanced Virgo+ phase 2 high or low	
	KAGRA	KAGRA 80Mpc or 128Mpc	
	LIGO India	Advanced LIGO Voyager	
Scenario 1			
2035-2039 [225, 229]	Cosmic Explorer	CE phase 1 [230]	CEKL
	Einstein Telescope	ET-D [231]	
	KAGRA	KAGRA 128Mpc	
	LISA	LISA [232, 233]	
2039-2045 [225, 229]	Cosmic Explorer	CE phase 1	CEKLext
	Einstein Telescope	ET-D	
	KAGRA	KAGRA 128Mpc	
	LISA	LISA	
2045-2050 [225, 229]	Cosmic Explorer	CE phase 2 [230]	CEK
	Einstein Telescope	ET-D	
	KAGRA	KAGRA 128Mpc	
Scenario 2			
2035-2039	Cosmic Explorer	CE phase 1	CVKL
	Virgo	Advanced Virgo+ phase 2 high	
	KAGRA	KAGRA 128Mpc	

Table 3.3 (cont.)

	LISA	LISA	
	Cosmic Explorer	CE phase 1	
2039-2045	Virgo	Advanced Virgo+ phase 2 high	CVKLext
	KAGRA	KAGRA 128Mpc	
	LISA	LISA	
	Cosmic Explorer	CE phase 2	
2045-2050	Virgo	Advanced Virgo+ phase 2 high	CVK
	KAGRA	KAGRA 128Mpc	
Scenario 3			
	LIGO Hanford	Advanced LIGO Voyager	
	LIGO Livingston	Advanced LIGO Voyager	
2035-2039	Virgo	Advanced Virgo+ phase 2 high or low	HLVKIL
	KAGRA	KAGRA 80Mpc or 128Mpc	
	LIGO India	Advanced LIGO Voyager	
	LISA	LISA	
	LIGO Hanford	Advanced LIGO Voyager	
	LIGO Livingston	Advanced LIGO Voyager	
2039-2045	Virgo	Advanced Virgo+ phase 2 high or low	HLVKILext
	KAGRA	KAGRA 80Mpc or 128Mpc	
	LIGO India	Advanced LIGO Voyager	
	LISA	LISA	
	LIGO Hanford	Advanced LIGO Voyager	
	LIGO Livingston	Advanced LIGO Voyager	
2045-2050	Virgo	Advanced Virgo+ phase 2 high or low	HLVKI+
	KAGRA	KAGRA 80Mpc or 128Mpc	
	LIGO India	Advanced LIGO Voyager	

3.2.2 Estimated Sensitivity

The detector sensitivities can be characterized in terms of their power spectral density S_n , which we present in Fig. 3.2.

Detector	Latitude ($^{\circ}$)	Longitude ($^{\circ}$)
LIGO Hanford	46.45	-119.407
LIGO Livingston	30.56	-90.77
Virgo	43.63	10.50
KAGRA	36.41	137.31
LIGO India	14.23	76.43
Cosmic Explorer	40.48	-114.52
Einstein Telescope	43.63	10.50

Table 3.4: Detector locations used in this paper.

We assume that the LIGO detectors will start operating at design sensitivity (“LIGO design” [227] in Fig. 3.2) in O4, but will be upgraded to the A+ configuration (“LIGO A+” [227] in Fig. 3.2) in time for the O5 observing run. In the early 2030’s, the LIGO detectors will be upgraded to the Voyager sensitivity (“LIGO Voyager” [228] in Fig. 3.2). Virgo observations begin with the Advanced Virgo+ phase 1 noise curve (“Virgo phase 1” [227] in Fig. 3.2) in O4, and they will subsequently be upgraded to Advanced Virgo+ phase 2 (“Virgo phase 2” [227] in Fig. 3.2) beginning in O5. To bracket uncertainties, we consider both an optimistic (“high”) configuration and a pessimistic (“low”) configuration for Virgo+ [227]. We model the KAGRA detector using the “128Mpc” and “80Mpc” configurations from Ref. [227] for optimistic and pessimistic outlooks, respectively (“KAGRA” in Fig. 3.2). LIGO India is planned to join the network in O5 with sensitivity well approximated by the A+ noise curve, mirroring the Hanford and Livingston detectors. LIGO India will follow the same development path as its American counterparts, and be upgraded to Voyager sensitivity in the early 2030’s.

The US-led 3g detector, CE, may replace the LIGO detectors in 2035 at phase 1 sensitivity (“CE phase 1” in Fig. 3.2). After upgrades are completed in the early 2040’s, the detector may come back online with phase 2 noise sensitivity (“CE phase 2” in Fig. 3.2) [229].

The European-led 3g counterpart ET could replace the Virgo detector in 2035. ET will be modeled with the ET-D sensitivity in this study (“ET-D” in Fig. 3.2). In reality, ET is comprised of 3 individual detectors arranged in an equilateral triangle, and a fully consistent treatment of ET would incorporate the three detectors separately. However, after testing on subsets of our populations, we concluded that modeling ET as three identical copies of one of the constituent detectors minimally impacts our estimates on constraints of modified gravity, because of the small correlations between modified gravity modifications to the phase and the extrinsic parameters of the source, like sky location and orientation. This approximation significantly reduces the computational resources required to perform this study, so we opted to use it when constructing the Fisher matrices themselves (as discussed in Sec. 3.5). When calculating the detection probability, however, we do account for the three detectors separately (cf. Sec. 3.4.2). This is because the different orientations

and positions of the detectors affect the rates more than they affect parameter estimation.

For networks that include a mixture of 3g and 2g detectors, we will only model the 2g detectors with the most optimistic sensitivity curve, i.e. the “high” configuration for Virgo and the “128Mpc” configuration for KAGRA. The impact of the different 2g sensitivities is small when implemented alongside a 3g detector, and the shrinking of the parameter space for our models significantly reduces the computational cost of the problem.

For LISA, we model the noise curve using the approximations in Ref. [232]. At different points in this work, we required both sky-averaged and non-sky-averaged response functions to various detectors. For LISA this can be more complicated than terrestrial interferometers, so we plot the sky-averaged noise curve directly from Ref. [232] (“LISA – sky-averaged” in Fig. 3.2) and the full (non-sky-averaged) sensitivity produced in Ref. [233] (“LISA – non-sky-averaged” in Fig. 3.2). However, in contrast to Ref. [233], we do include the factor of 2 to account for the second channel, mirroring the approximation we made for ET.

3.2.3 Estimated Location

The relative locations of the various detectors affects the global response function, and thus it impacts the analysis performed in this paper. For terrestrial detectors, the various geographical locations of each site are shown in Table 3.4. The sites of detectors currently built or under construction were taken from data contained in LALSuite [234]. Since a site has yet to be decided upon for CE, we chose a reasonable location near the Great Basin desert, in Nevada. For LISA, the detector’s position and orientation as a function of time must be taken into account, so we use the time-dependent response function derived in Refs. [235, 236]. Unlike those papers we use the polarization angle defined by the total angular momentum $\mathbf{J}$, instead of the orbital angular momentum $\mathbf{L}$, because the latter precesses in time, while $\mathbf{J}$ remains (approximately) constant.

3.3 Statistical Methods for Population Simulations

Both terrestrial and space-borne GW detectors have nonuniform sensitivity over the sky. This effect is important when attempting to estimate the expected detection rate and the resulting population catalog.

Terrestrial detector networks can mitigate this selection bias by incorporating more detectors into the network, which can “fill in” low-sensitivity regions in the sky. The incorporation of the most accurate combination of detectors and their locations can be important. This is why in Sec. 3.2.3 we specified the locations used in this study.

For space-borne detectors, some signals may be detectable for much longer than the observation period,

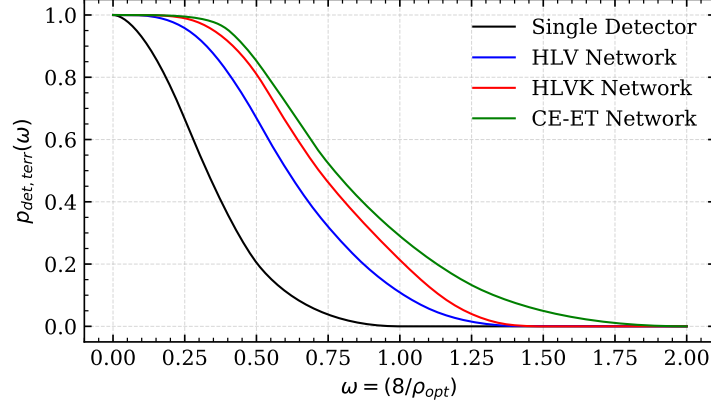


Figure 3.3: Detection probability p_{det} for the four networks examined in this paper. The black curve is for a single detector (where global position no longer matters, so this is valid for any single right-angle Michelson interferometer). The blue curve is specifically for the Hanford, Livingston, and Virgo (HLV) network. The red curve is for the Hanford, Livingston, Virgo, and KAGRA (HLVK) network. Finally, the green curve represents a network comprised of CE and ET (which includes all three of the ET detectors as well as the 60° angle between each set of arms).

so random sky locations map to random spacetime locations, and the effect of only seeing a portion of the signal must be accounted for.

These issues with terrestrial networks and space detectors, and their associated detection probabilities, are discussed in Secs. 3.3.1 and Sec. 3.3.2, respectively.

We wish to calculate the probability that the GWs emitted by some source will be detected by a terrestrial network of instruments, which we will refer to as the detection probability. We will focus primarily on two classes of sources: SOBH binaries [237] and MBH binaries [238]. We will use publicly available SOBH population synthesis models to produce synthetic catalogs which are mainly of interest for the terrestrial network, but can also be observed as “multiband” events by both the terrestrial network and LISA. We will also use MBH binary simulations to create synthetic catalogs for LISA (these sources are typically well outside the frequency band accessible to terrestrial networks). Intermediate-mass BH binaries could also be of interest [239], but we do not consider them here, mainly because their astrophysical formation models and rates have large uncertainties [207, 240, 241].

3.3.1 Terrestrial Detection Probability

An accurate calculation of the detection probability for each source requires injections into search pipelines. A simplifying, while still satisfactorily accurate, assumption used in most of the astrophysical literature (see

Detection network	Detector locations	Detector sensitivity curve
HLVKO4	Hanford site Livingston site Virgo site	Ad. LIGO design [227]
HLVKIO5-O7	Hanford site Livingston site Virgo site KAGRA site	Ad. LIGO A+ [227]
HLVKIO8-O9	Hanford site Livingston site Virgo site KAGRA site	Ad. LIGO Voyager [228]
CEKL(ext)	Cosmic Explorer site All ET sites	CE phase 1 [230]
CVKL(ext)	Cosmic Explorer site	CE phase 1
HLVKIL(ext)	Hanford site Livingston site Virgo site KAGRA site	Ad. LIGO Voyager
CEK	Cosmic Explorer site All ET sites	CE phase 2 [230]
CVK	Cosmic Explorer site	CE phase 2
HLVKI+	Hanford site Livingston site Virgo site KAGRA site	Ad. LIGO Voyager

Table 3.5: Configurations used at each stage of our analysis to calculate the probability of detection for a given binary for the terrestrial detector network. Note that networks involving multiple detectors are labelled by the network nodes and not just their number, because the relative position of the detectors impacts the calculation of the detection probability. Our calculation depends on the assumption that all the detectors have approximately the same sensitivity curve, and so the curve used at each stage is given in the last column. Because of this assumption, and the extreme disparity in sensitivity between second- and third-generation detectors, we only use the CE detector to calculate rates when CE is part of the network.

e.g. [125, 242, 243]) involves computing the SNR ρ , defined by

$$\rho^2 = 4 \operatorname{Re} \left[\int \frac{\tilde{h} \tilde{h}^*}{S_n(f)} df \right], \quad (3.1)$$

where we recall that $S_n(f)$ is the noise power spectral density of the detector, while $\tilde{h} = \tilde{h}(f)$ is the Fourier transform of the contraction between the GW strain and the detector response function. We can factor out all the detector-dependent quantities from the SNR in the form of the “projection parameter” ω defined as [242, 243]

$$\omega^2 = \frac{(1 + \cos^2 \iota)^2}{4} F_+^2(\theta, \phi, \psi) + \cos^2 \iota F_\times^2(\theta, \phi, \psi), \quad (3.2)$$

where ι is the inclination of the binary relative to the line of sight, θ and ϕ are the spherical angles of the source relative to the vector perpendicular to the plane of the detector, and ψ is the polarization angle. The single-detector antenna pattern functions F_+ and $F_\times$ are given by

$$\begin{aligned} F_+ &= \frac{1}{2} (1 + \cos^2 \theta) \cos 2\phi \cos 2\psi - \cos \theta \sin 2\phi \sin 2\psi, \\ F_\times &= \frac{1}{2} (1 + \cos^2 \theta) \cos 2\phi \sin 2\psi + \cos \theta \sin 2\phi \cos 2\psi. \end{aligned} \quad (3.3)$$

With the projection-parameter approximation, we can approximate the SNR as

$$\rho^2 \approx \omega^2 \rho_{\text{opt}}^2, \quad (3.4)$$

where ρ_{opt} is the SNR for an optimally oriented binary with $\theta = 0$, $\iota = 0$, and $\psi = 0$. This relation is approximate if the binary is precessing, so that ι is a function of time, but it is exact otherwise.

The calculation of the detection probability can then be rephrased as a search for the extrinsic source parameters that satisfy $\omega \approx \rho/\rho_{\text{opt}} \geq \rho_{\text{thr}}/\rho_{\text{opt}} \equiv \omega_{\text{thr}}$ for some ρ_{thr} . The probability that ω satisfies the above criteria translates into finding the cumulative probability distribution [242]

$$p_{\text{det,terr}}(\vec{\lambda}) = \int \Theta(\omega'(\theta, \phi, \psi, \iota) - \omega_{\text{thr}}) \frac{\sin \theta d\theta d\phi}{4\pi} \frac{d\psi}{\pi} \frac{d \cos \iota}{2}, \quad (3.5)$$

where $\Theta(\cdot)$ is the Heaviside function, which ultimately describes the selection effects of our terrestrial networks. This cumulative probability clearly depends on the source parameter vector $\vec{\lambda}$, inherited from $\omega_{\text{thr}} = \omega_{\text{thr}}(\vec{\lambda})$.

Equation (3.5) can be extended to multiple-detector networks by expanding our definition of ω to

$$\omega_{\text{network}}^2 = \sum_i \omega_i^2, \quad (3.6)$$

where ω_i is the projection parameter for a single detector in the network, and $\omega_{\text{network}} = \rho_{\text{network-thr}}/\rho_{\text{opt}}$ with some threshold network SNR, $\rho_{\text{network-thr}}$, and single-detector optimal SNR, ρ_{opt} . In the case of a multiple-detector network, the locally defined position coordinates θ and ϕ are replaced with the globally defined position coordinates α (the right ascension angle) and δ (the declination angle). The polarization angle ψ is changed to the globally defined polarization angle $\bar{\psi}$, which is defined with respect to an Earth-centered coordinate axis instead of the coordinate system tied to a single detector.

Evaluating Eq. (3.5) for each network, with the network projection operator defined as Eq. (3.6), provides a good estimation of the probability we are seeking: a weighting factor for a given binary that incorporates

the sensitivity and global geometry of a given detector network, as well as the impact that the intrinsic properties of the source have on its detectability. Importantly, the intrinsic source parameters themselves only enter into Eq. (3.5) through the calculation of ρ_{opt} in ω_{thr} . Once a threshold SNR ρ_{thr} is set, the detection probability function can be seen as a function of only one number ω_{thr} (for a given network), through its dependence on ρ_{opt} . As Eq. (3.5) is a four-dimensional integral and must be calculated numerically, this detail can significantly save on computational cost if we can approximate the full function $p_{\text{det,terr}}(\omega_{\text{thr}})$ once for each network. To do this, we form a grid in ω_{thr} with approximately 100 grid points, and evaluate Eq. (3.5) for each grid point with 10^9 samples uniformly distributed in $\bar{\psi}$, $\cos \iota$, α , and $\sin \delta$. Interpolating across the grid in ω_{thr} produces an approximation for $p_{\text{det,terr}}(\omega_{\text{thr}})$. This approximation must be calculated for each specific network, as the quantity ω' in Eq. (3.5) depends on the number and relative location of the detectors, but it only needs to be evaluated once per network, rather than once per source.

The resulting probability functions for the four terrestrial networks examined in this paper are shown in Fig. 3.3. Note that the relative location of each detector in a network impacts the form of $p_{\text{det,terr}}$, so we label the curves by the detector nodes and not just their number (i.e. the form of $p_{\text{det,terr}}$ will be slightly different for a Hanford, Livingston, and Virgo network when compared to a Hanford, Livingston, and KAGRA network). Furthermore, an important assumption in this calculation is that the sensitivity of each detector is identical. This is not a good approximation when jointly considering second- and third-generation detectors, so in these cases we neglect all the 2g detectors in the network. The configurations used at each stage are summarized in Table 3.5.

3.3.2 Space Detection Probability

For space-based detectors, which operate at much lower frequencies, the picture changes quite drastically. The terrestrial detection probability of Sec. 3.3.1 addresses the issue of random sky location and orientation of the sources, but an important effect for detectors like LISA is the time spent in band. Because signals observable by LISA can be detected for much longer than the observation time T_{obs} of the LISA mission, the time spent in the frequency range accessible to LISA will characterize the detectability of the binary. We characterize this effect as outlined below (we refer the reader to Ref. [204] for a more thorough derivation and further details).

To determine the time the binary spends in the observational frequency band of LISA, we look for the roots of

$$\rho(t_{\text{merger}}) - \rho_{\text{thr}} = 0, \quad (3.7)$$

where t_{merger} is the time before merger at which the signal starts, ρ_{thr} is some threshold SNR, and the SNR

$\rho(t_{\text{merger}})$ is defined as

$$\rho(t_{\text{merger}}) = 4 \text{Re} \left[\int_{f(t_{\text{merger}})}^{\min(f(t_{\text{merger}} - T_{\text{obs}}), 1 \text{ Hz})} \frac{\tilde{h} \tilde{h}^*}{S_n(f)} df \right]. \quad (3.8)$$

Note that, at variance with Ref. [204], we use 1 Hz as the upper cutoff for the LISA noise curve.

Once the roots of Eq. (3.7) (say T_1 and T_2) have been found, we can obtain the probability of mergers for LISA via

$$p_{\text{det,space}}^{\text{SOBH}}(\vec{\lambda}) = p_{\text{det,terr}}(\vec{\lambda}) \times \min \left[\frac{T_1 - T_2}{T_{\text{obs}}}, \frac{T_{\text{wait}} - T_2}{T_{\text{obs}}} \right] \quad (3.9)$$

for SOBH binaries, and

$$p_{\text{det,space}}^{\text{MBH}}(\vec{\lambda}) = \min \left[\frac{T_1 - T_2}{T_{\text{obs}}}, \frac{T_{\text{wait}} - T_2}{T_{\text{obs}}} \right] \quad (3.10)$$

for MBH binaries. The probability $p_{\text{det,space}}^{\text{SOBH}}$ is weighted by $p_{\text{det,terr}}$ because all SOBH binaries we consider for LISA are also candidate multiband events, which must be observed both by LISA and by a terrestrial network to be considered “true” multiband sources. In these expressions, T_{wait} is some maximum waiting time for the binary to merge, which (following Ref. [204]) we choose to be $5 \times T_{\text{obs}}$ for each detector network iteration.

3.3.3 Waveform Model for Population Estimates

When computing the detection probability of a given source, we need a model for the Fourier transform of the time-domain response function $h = F_+ h_+ + F_\times h_\times$. In the terrestrial case, we implement the full precessing inspiral/merger/ringdown model **IMRPhenomPv2** [106, 169, 202] with an inclination angle of $\iota = 0^\circ$ to calculate the optimal SNR, ω_{opt} . For the space-based estimates in the next section, we will use the spinning (but nonprecessing) sky-averaged **IMRPhenomD** waveform model [106, 169], with a small modification: since we are interested in LISA rather than terrestrial, right-angle interferometers, we replace the usual factor of 2/5 (that arises from sky-averaging) in favor of the sky-averaged LISA sensitivity curve from [232], which accounts for the second LISA data channel, sky-averaging, and the 60° angle between the detector arms. This waveform model depends on parameters $\vec{\lambda}_D = [\alpha, \delta, \theta_L, \phi_L, \phi_{\text{ref}}, t_{c,\text{ref}}, D_L, \mathcal{M}, \eta, \chi_1, \chi_2]$, where α is the right ascension, δ is the declination, θ_L and ϕ_L are the polar and azimuthal angles of the binary’s orbital angular momentum $\mathbf{L}$ in equatorial coordinates at the reference frequency, ϕ_{ref} and $t_{c,\text{ref}}$ are the orbital phase and the time of coalescence at the reference frequency, D_L is the luminosity distance, $\mathcal{M}$ and η are the redshifted chirp mass and the symmetric mass ratio, and $\chi_i = \hat{\mathbf{L}} \cdot \mathbf{S}_i / m_i^2$ are the dimensionless spin components along $\hat{\mathbf{L}} = \mathbf{L} / |\mathbf{L}|$ with spin angular momentum $\mathbf{S}_i$.

For space-based detectors we must also choose a way to map between time and frequency. The limits of the SNR integral (3.1) and the antenna patterns (which for LISA are functions of time) depend on this mapping. For multiband SOBH binaries we use the leading-order PN relation [204, 236, 244]

$$f(t_{\text{merger}}) = \frac{5^{3/8}}{8\pi} (\mathcal{M})^{-5/8} t_{\text{merger}}^{-3/8}, \quad (3.11)$$

where again t_{merger} is the time before merger. For massive black hole (MBH) binaries, observed by LISA only through merger, this PN approximation is insufficient, so we use instead [245, 246]

$$t_{\text{merger}} = \frac{1}{2\pi} \frac{d\phi}{df}, \quad (3.12)$$

where ϕ is the GW Fourier phase. When calculating detection rates, we will invert these relations numerically as needed.

3.4 Population Simulations

A key ingredient of our work is the use of astrophysically motivated BBH population models (Sec. 3.4.1). Our methodology for computing detection rates and for creating synthetic catalogs from the models is explained in Sec. 3.4.2 and in Sec. 3.4.3, respectively.

3.4.1 Population Models

For ease of comparison with previous work, we use the SPOPS catalogs [237] for SOBH binaries (Sec. 3.4.1) and the MBH binary merger catalogs used in Ref. [238] (Sec. 3.4.1).

Stellar Mass Simulations

We use the public SPOPS catalog of population synthesis simulations [237] in an effort to accurately capture the full spin orientations of the binaries at merger. The SPOPS catalog uses multiscale solutions of the precessional dynamics [247, 248] computed through the public code PRECESSION [249] to quickly evolve the binary’s spin orientations in time until the binary is about to merge.

The catalog is parameterized by three different variables: the strength of the BH natal kicks, the BH spin magnitudes at formation, and the efficiency of tidal alignment [237]. In this model, natal kicks are caused by asymmetric mass ejection during core collapse, imparting a torque on one of the constituents of the binary, while the tidal alignment reflects spin-orbital angular momentum coupling through tidal interactions that

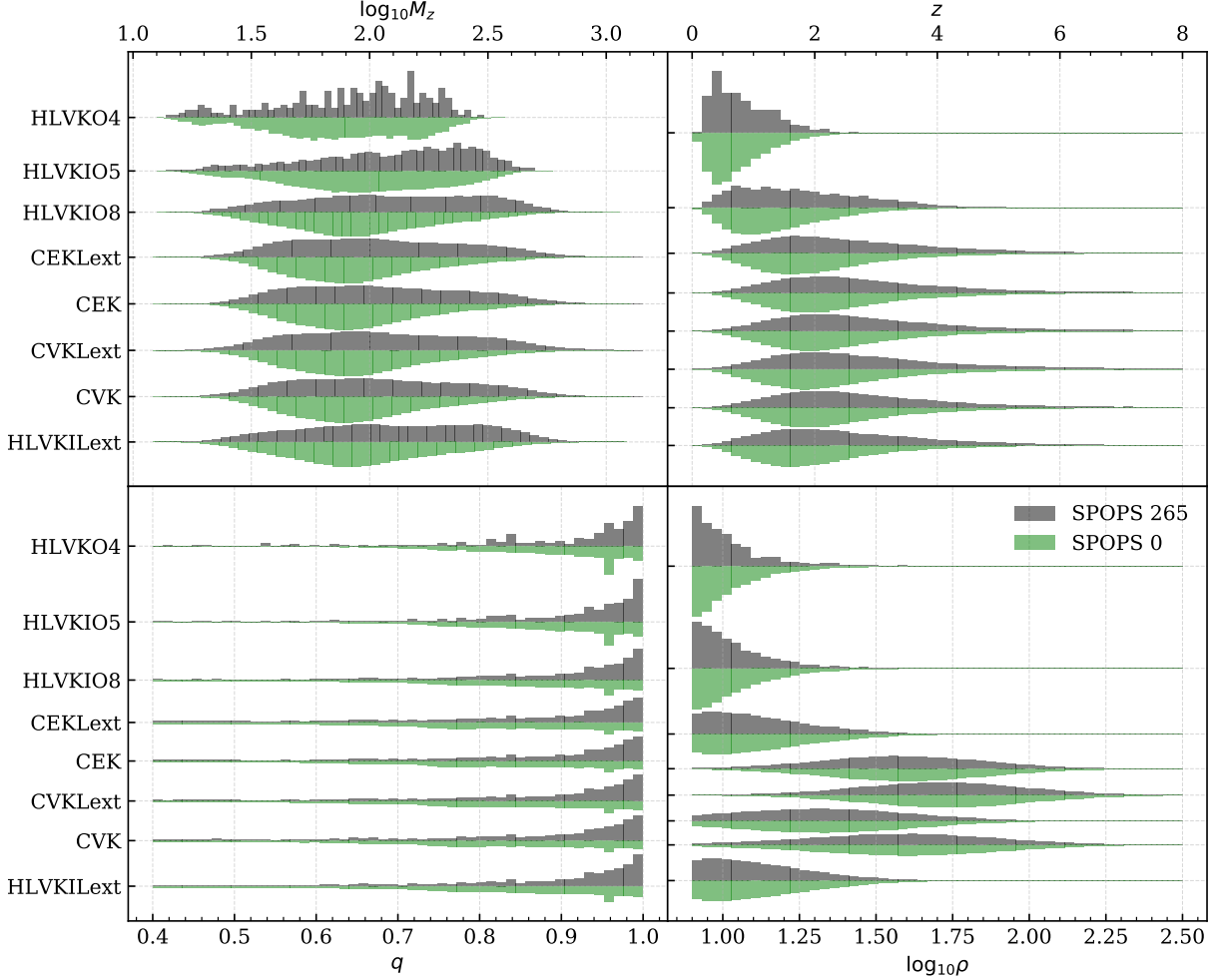


Figure 3.4: Distributions of the different source properties detected by each network. For each detector network, labeled across the y-axis, we plot the distribution of the total detector-frame mass $M_z = M(1+z)$, mass ratio $q = m_2/m_1 < 1$, redshift z , and SNR ρ in log-space (base 10). Each plot is split, with the upper (grey) half coming from the $\sigma = 265$ km/s SPOPS simulations, and the lower (green) half coming from the $\sigma = 0$ km/s simulations.

can realign the spin vectors with the orbital angular momentum vector (see Ref. [237] for further details).

Following Ref. [204], we choose to vary only one parameter of these models while keeping the others fixed. More specifically, we consider a uniform distribution in spin magnitude and the most realistic (“time”) prescription for tidal alignment of Ref. [204], while varying the natal kick. To estimate lower and upper constraints on the rates given uncertainties in our population modelling, we use the two most extreme natal kick models, corresponding to $\sigma = 0$ km/s and $\sigma = 265$ km/s, where σ is the one-dimensional dispersion of the Maxwellian distribution the kicks are drawn from. The zero-kick scenario results in a lack of precessional effects and the highest detection rates for all detectors, while the $\sigma = 265$ km/s choice corresponds to a soft

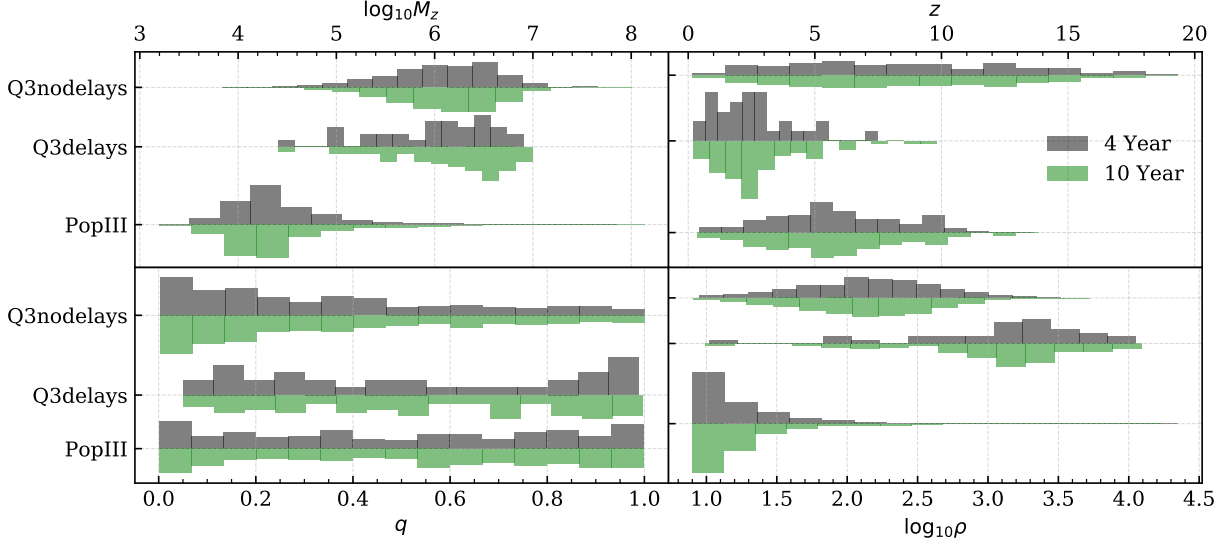


Figure 3.5: Distributions of the different MBH binary source properties detected by LISA. For each MBH binary simulations, labeled across the y-axis, we plot the distribution of the total detector-frame mass $M_z = M(1+z)$, mass ratio $q = m_2/m_1 < 1$, redshift z , and SNR ρ in log-space (base 10). Each plot is split in two, with the upper (grey) half corresponding to a “nominal” four-year LISA mission, and the lower (green) half corresponding to an extended ten-year mission.

upper bound on the size of the kicks, which imparts the largest spin tilts and results in the lowest detection rate. The two chosen values of σ result in optimistic and pessimistic bounds on our projected constraints, and at the same time they provide a useful comparison between highly precessing systems and nonprecessing systems.

Massive Black Hole Simulations

To model MBH binary populations, we adopt the semianalytical models of early Universe BH formation [250–252] used in the LISA parameter estimation survey of Ref. [238]. As in that work, we focus on three populations models, characterized by different BH seeding mechanisms and different assumptions on the time delay between BH mergers and the mergers of their host galaxies. These population models are denoted as

1. PopIII – seeds are produced from the collapse of population III stars in the early Universe (a light-seed scenario);
2. Q3delays – seeds are produced from the collapse of a protogalactic disk (heavy-seed scenario), and there are delays between galaxy mergers and BH mergers;
3. Q3nodelays – seeds are produced from the collapse of a protogalactic disk (heavy-seed scenario), and there are no delays between galaxy mergers and BH mergers.

These three models embody two seed formation mechanisms, with two models representing optimistic and pessimistic heavy-seed scenarios. The difference between PopIII simulations with and without delays is less than a factor of two, so, following Ref. [238], we consider only the more conservative estimate, in which delays are incorporated.

3.4.2 Detection Rate Calculations

With population synthesis simulations at our disposal, we can now estimate expected detection rates for a given detector network. This involves taking a model for our Universe that predicts a certain rate of merging BBHs per comoving volume, and filtering the model through the lens of a particular detector configuration and sensitivity. The detection rate r for a given network follows from the following relation [204, 253]:

$$r = \iint dz d\vec{\lambda} \mathcal{R}(z) p(\vec{\lambda}) \frac{dV_c(z)}{dz} \frac{1}{1+z} p_{\text{det}}(\vec{\lambda}, z), \quad (3.13)$$

where z is the cosmological redshift, $\mathcal{R}$ is the intrinsic merger rate (a function of the redshift), p is the probability of a binary forming and merging given a set of intrinsic source parameters $\vec{\lambda} = \vec{\lambda}_D$ (discussed in Sec. 3.3.3), and dV_c/dz is a shell of comoving volume V_c at redshift z .

The quantity p_{det} is the probability of a binary being detected by a given detector network with some threshold SNR, as discussed in Sec. 3.3. The type of detector network affects the quantity p_{det} only, while the other terms in the integral above depend only on information contained in the population simulation. For this study, we have used a threshold SNR of 8 for terrestrial and space detections, while for multiband detections we require the terrestrial SNR and the LISA SNR to both be above 8 independently. Because of the intrinsic difference in the duration of signals observed by space detectors and terrestrial networks, we treat the calculation of p_{det} slightly differently between the two cases, as discussed in Sec. 3.3.1 for terrestrial detectors, and in Sec. 3.3.2 for space-based detectors.

For all binaries, we evaluate the integral in Eq. (3.13) through a large population of binary systems that are evolved to the point of becoming BBHs, and are weighted according to the probability that a binary of this type would actually be found in the Universe given some population model. This probability is comprised of factors like the star formation rate (SFR), cosmological evolution of the metallicity, the distribution of masses for these stellar populations, etc.; the continuous equation in Eq. (3.13) then becomes a discrete sum

$$r = \sum_i r_i p_{\text{det}}(\vec{\lambda}_i), \quad (3.14)$$

where the index i refers to samples in the simulation, r_i is the intrinsic merger rate, which depends on

parameters like the SFR and the mass distribution, and $p_{\text{det}}(\vec{\lambda}_i)$ is the detection probability evaluated for the source parameters of the particular sample. This detection probability is $p_{\text{det,terr}}$ when considering a terrestrial network only, $p_{\text{det,space}}^{\text{SOBH}}$ when considering multiband events, or $p_{\text{det,space}}^{\text{MBH}}$ when considering MBH binaries detectable only by LISA.

The intrinsic merger rate r_i varies depending on the catalog used. For the case of the SPOPS simulations, we utilized the original **StarTrack** data at the foundation of each SPOPS catalog (cf. Ref. [253] for details) to construct the intrinsic merger rate in Eq. (3.13). For MBH catalogs, the intrinsic merger rate r_i becomes [238]

$$r_i = 4\pi W_{\text{PS},i} \left(\frac{D_L(z_i)}{1+z_i} \right)^2, \quad (3.15)$$

as outlined in the data release [238, 254]. The parameter $W_{\text{PS},i}$ is the weight on the Press-Schechter mass function divided by the number of realizations [250].

3.4.3 Synthetic Catalog Creation

Calculating the BBH detection rate only gets us half-way to our end goal. Once we have the number of mergers we expect to detect for each network and simulated population, we still need to synthesize BBH catalogs to use for the later Fisher analysis in this paper.

To create these synthetic catalogs, we sample directly from the population simulations, using Monte Carlo rejection sampling. The probability of accepting a sample is based on the intrinsic merger rate r_i in Eq. (3.14), evaluated for a single simulation entry, which comes directly from the simulation data itself. This gives a distribution of sources that reflects the expected BBH distributions for each evolution prescription. With a distribution of “intrinsic” mergers in this realization of the Universe, we assign any remaining parameters according to reasonable distributions. For sky-location and orientation, this distribution is uniform in α , $\sin \delta$, $\cos \theta_L$, and ϕ_L .

For the binary’s merger time, we use a uniform distribution in GMST for the terrestrial networks, which impacts the orientation of the terrestrial network at the time of merger. This effect is completely degenerate with the right ascension of the binary, which is also randomly uniform in α . We use a similar prescription for MBH binaries, where the signal duration is typically shorter than the observation period. We employ a uniform distribution in time from 0 to T_{obs} , which again translates to a uniform distribution in detector orientation (random position of LISA in its orbit).

Candidates for multiband detection are more nuanced. The signal is typically detectable for much longer than the observation period, and the frequency-time relation is nonlinear because of the familiar chirping

behavior of GW signals. For this class of sources, we randomly assign a signal starting time, which has a power-law relation with the starting frequency: cf. Eq. (3.11). In this case, the position of the binary in time not only affects the orientation of LISA, but also the initial and final frequencies of the signal. This assignment of time is important, as assigning a uniformly random initial frequency would create a bias towards seeing sources close to merger.

Once the full parameter vector has been specified, we proceed to calculate the SNR for the source in question. Sources meeting the SNR threshold requirements are retained in the final catalog. This process is repeated as necessary until we have a catalog of sources that matches the number of BBHs predicted by our rate calculations in Sec. 3.4.2.

There are some drawbacks to this scheme. If this process is repeated enough times, sources in the simulation will begin to be reused, as there are a fixed number of possible sources to draw from. For this study, however, these effects are negligible, as the number of the sources in the simulations is larger than any single catalog we construct. Furthermore, the effects will be further mitigated by randomly assigning the rest of the parameter vector not coming from the simulation, which will imbue at least slightly different properties to each source, even if one were reused.

To recap, our process can be broken down into the following steps:

1. Perform rejection sampling on the simulation entries according to the probability of merging, neglecting detector selection effects.
2. Keep the “successful” events, and randomly draw the rest of the requisite parameters according to their individual distributions.
3. Calculate the SNR for the given detector network. If the binary meets the threshold requirements, keep the source in the final catalog.

The source properties of the various *detected* catalogs are shown in Fig. 3.4 for the SOBH populations, and in Fig. 3.5 for the MBH populations targeted by LISA. Both figures show the distributions of the redshifted total mass M_z , the mass ratio $q = m_2/m_1 < 1$, the redshift z , and the SNR ρ of the detected populations of sources for different detector configurations and population models. For the SOBH sources shown in Fig. 3.4, the y-axis labels correspond to different detector combinations, while the upper (grey) and lower (green) histograms correspond to the two different kick magnitudes ($\sigma = 265$ km/s and $\sigma = 0$ km/s) chosen to bracket SOBH population models.

In the LISA SMBH case of Fig. 3.5, the same properties are plotted for the three populations models and for a four-year and ten-year LISA mission. Note that the y-axis label now corresponds to different

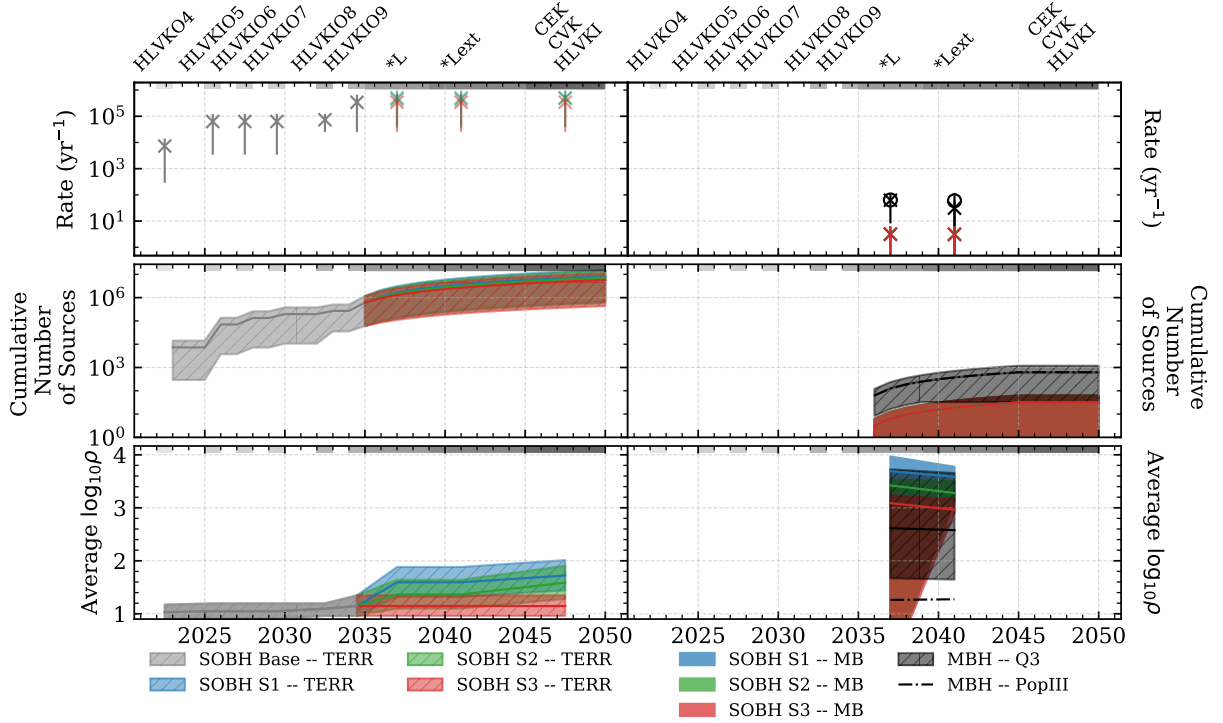


Figure 3.6: Properties of detected merger events for various detector networks and population models. The left panels refer to terrestrial-only sources, while MBHs and multiband sources are shown on the right. The points and thick lines show the mean values, while the shaded regions and error bars encompass the optimistic and pessimistic scenarios. The assumed detector network is shown in the top x-axis (using the notation of Table 3.3), while the corresponding years are shown on the bottom x-axis. The top panels show the rates of detected mergers for each class of sources; circles refer to the PopIII MBH population. The middle panels show the cumulative number of observed sources: here the three different multiband scenarios are identical, as the choice of terrestrial network has little impact on the number of multiband sources we can detect [204]. The bottom panels show the average $\log_{10}\rho$. Here the lower (upper) bounds correspond to subtracting (adding) the standard deviation to the mean value of the most pessimistic (optimistic) scenario.

population models, and each half of the violin plot corresponds to different mission durations: the upper (grey) half corresponds to the “nominal” four-year LISA mission, and the lower (green) half corresponding to an extended ten-year mission.

The detection rates, cumulative detected sources, and average SNR for each class of sources are shown in Fig. 3.6, where sources are broken down into 4 distinct categories:

- (i) “SOBH - TERR”: SOBH candidates detected only by a terrestrial network;
- (ii) “SOBH - MB”: SOBH candidates detected by both a terrestrial network and LISA (multiband);
- (iii) “MBH - PopIII”: MBH sources from the PopIII model (light seeds);
- (iv) “MBH - Q3”: MBH sources from both Q3 (heavy seeds) models, with shaded bands indicating the

SOBH Rates (yr^{-1})		
Network	SPOPS 0 (T, MB)	SPOPS 265 (T, MB)
HLVKO4	$(1.43 \times 10^4, 0)$	$(2.90 \times 10^2, 0)$
HLVKIO5-O7	$(1.22 \times 10^5, 0)$	$(3.43 \times 10^3, 0)$
HLVKIO8-O9	$(6.60 \times 10^5, 0)$	$(2.48 \times 10^4, 0)$
Scenario 1		
CEKL	$(9.70 \times 10^5, 2.58)$	$(3.96 \times 10^4, 0.0854)$
CEKLext	$(9.70 \times 10^5, 6.24)$	$(3.96 \times 10^4, 0.210)$
CEK	$(9.72 \times 10^5, 0)$	$(3.97 \times 10^4, 0)$
Scenario 2		
CVKL	$(8.36 \times 10^5, 2.58)$	$(3.36 \times 10^4, 0.0854)$
CVKLext	$(8.36 \times 10^5, 6.24)$	$(3.36 \times 10^4, 0.210)$
CVK	$(9.26 \times 10^5, 0)$	$(3.77 \times 10^4, 0)$
Scenario 3		
HLVKIL	$(6.60 \times 10^5, 2.58)$	$(2.48 \times 10^4, 0.0854)$
HLVKILext	$(6.60 \times 10^5, 6.24)$	$(2.48 \times 10^4, 0.210)$
HLVKI+	$(6.60 \times 10^5, 0)$	$(2.48 \times 10^4, 0)$
MBH Rates (yr^{-1})		
Network	PopIII	Q3 (delay,nodelay)
LISA	62.5	(8.11,119.1)

Table 3.6: Detection rates for the detector networks and population models examined in this study. For SOBH populations, the first number in the parentheses is the detection rate for the terrestrial-only network (neglecting LISA), while the second number is the detection rate for multiband events seen in both the terrestrial network and LISA. For MBH populations, we show the detection rate for LISA for the PopIII, light-seeding scenario, as well as for the Q3, heavy-seeding scenario. In the case of Q3, the first number in parentheses corresponds to delayed mergers (Q3delays) and the second number to the nondelayed version (Q3nodelays).

range of uncertainty on delays between galaxy mergers and BH mergers.

The year is shown across the bottom x-axis, while the detector network timeline is shown across the top x-axis using the acronyms defined in Table 3.3. The solid lines and markers represent the mean values of the different quantities when considering each population model and optimistic/pessimistic detector configurations. The error bars and shaded regions represent the most optimistic and most pessimistic scenarios, except in the case of the SNR in the third panel, where the upper and lower bounds are the optimistic (pessimistic) average plus (minus) the standard deviation of the optimistic (pessimistic) distribution. There is no error for the PopIII model, as we only have one iteration of this model and only one noise curve for LISA. The detection rates for SOBHs and MBHs in the different scenarios are also listed in Table 3.6.

Roughly speaking, the power of a detector network to reveal new physics comes from a combination of (i) the number of sources the network can detect, and (ii) the typical quality of each signal (as measured by the SNR). Figure 3.6 attempts to capture the zeroth-order difference between each detector configuration and

population model in these two aspects. The punchline is that although LISA will be able, on average, to see events with much larger SNR, these are just a few compared to the abundant number of sources that ground-based detectors will observe (albeit at typically lower SNR). The precision of GR tests scales as ρ^{-1} and it is approximately proportional to $\sqrt{N}$ for N events [255], therefore it is not immediately obvious which set of observations will be best at testing GR. With our catalogs this question can be answered quantitatively. As we discuss below, ground-based and space-based detectors are complementary to each other.

3.5 Parameter Estimation

In this section we describe the statistical methods we will use to carry out projections on the strength of tests of GR in the future, as well as our waveform model and the numerical implementation.

3.5.1 Basics of Fisher Analysis

The backbone of this work is built on the estimation of the posterior distributions that might be inferred based on our synthetic signals. Given a loud signal with a large enough SNR, the likelihood of the data, i.e., the probability that one would see a data set d given a model with parameters $\vec{\theta}$, can be expanded about the maximum likelihood (ML) parameters $\vec{\theta}_{\text{ML}}$. This expansion taken out to second order results in the following approximate likelihood function (where we focus on a single detector for the moment) [236, 256]:

$$\mathcal{L} \propto \exp \left[-\frac{1}{2} \Gamma_{ij} \Delta\theta^i \Delta\theta^j \right], \quad (3.16)$$

where $\Delta\theta^i = \theta_{\text{ML}}^i - \theta^i$ are deviations from the ML values, and Γ_{ij} is the Fisher information matrix

$$\Gamma_{ij} = (\partial_i h | \partial_j h) |_{\text{ML}}. \quad (3.17)$$

As before, h is the template response function, and the noise-weighted inner product is given by

$$(A|B) = 4 \text{Re} \left[\int \frac{\tilde{A} \tilde{B}^*}{S_n(f)} df \right], \quad (3.18)$$

with $S_n(f)$ the noise power spectral density. By truncating the expansion at second order, we have effectively represented our posterior probability distribution as a multidimensional Gaussian with a covariance matrix given by $\Sigma^{ij} = (\Gamma^{-1})^{ij}$. The variances of individual parameters can then be read off to be $\sigma^i = \sqrt{\Sigma^{ii}}$, where index summation is not implied.

Theory or physical process	Physical modification	G/P	PN order	β	Theory parameter	b
Generic dipole radiation	Dipole radiation	G	-1	(B.2)	$\delta\dot{E}$	-7
Einstein-dilaton Gauss-Bonnet	Dipole radiation	G	-1	(B.3)	$\sqrt{\alpha_{\text{EdGB}}}$	-7
Black Hole Evaporation	Extra dimensions	G	-4	(B.6)	$\dot{M}$	-13
Time varying G	LPI	G	-4	(B.7)	$\dot{G}$	-13
Massive Graviton	Nonzero graviton mass	P	1	(B.11)	m_g	-3
dynamical Chern-Simons	Parity violation	G	2	(B.8)	$\sqrt{\alpha_{\text{dcs}}}$	-1
Noncommutative gravity	Lorentz violation	G	2	(B.10)	$\sqrt{\Lambda}$	-1

Table 3.7: A summary of the theories examined in this work (adapted and updated from [161, 210]). The columns (in order) list the theory in question (unless a generic deviation is being examined), the physical interpretation of the modification, the way the modification is introduced into the waveform, the PN order at which the modification is introduced, the equation specifying the ppE-theory mapping, and the b parameter in the ppE framework. The practical ramifications between “generation” vs “propagation” effects relates to how the modification is introduced into the waveform, as explained in Appendix C.

In an attempt to capture the hard boundaries on the spin components (the dimensionless spin magnitudes $|\chi_i|$ and in-plane spin component χ_p in GR should not exceed 1), we incorporate a Gaussian prior on these two parameters with a width of 1. We do so by adding to the Fisher matrix diagonal terms of the form [236, 256, 257]

$$\Gamma_{ij} \rightarrow \Gamma_{ij} + \Gamma_{ij}^0, \quad (3.19)$$

where Γ_{ii}^0 represents our prior distribution and is given by

$$\Gamma_{ij}^0 = \delta_{\chi_1, \chi_1} + \delta_{\chi_2, \chi_2} + \delta_{\chi_p, \chi_p}. \quad (3.20)$$

In the case of multiple observations for a single source, we simply generalize the above results through sums. For example, the likelihood for a single event observed with N detectors can be expanded quadratically via

$$\mathcal{L} \propto \exp \left[-\frac{1}{2} \Delta\theta^i \Delta\theta^j \sum_k^N \Gamma_{ij,k} \right], \quad (3.21)$$

where the subscript k labels the k -th detector, and we have assumed that the parameters $\vec{\theta}$ are globally

defined. This gives the final covariance matrix

$$\Sigma^{ij} = \left(\left(\sum_k^N \Gamma_k + \Gamma^0 \right)^{-1} \right)^{ij}. \quad (3.22)$$

To improve readability, additional details on the calculation of the Fisher matrix are given in Appendix A.

3.5.2 Waveform Model for the Fisher Analysis

For the Fisher studies carried out in this paper, we model binary merger waveforms using the phenomenological waveform model `IMRPhenomPv2` [106, 169, 202], which allows us to capture certain spin precessional effects from inspiral until merger. The software used in this work was predominantly written from scratch, but the software library `LALSuite` [234] was used for comparison and to verify our implementation. For the actual parameter estimation calculation with LISA, we rescale the sensitivity curve to remove the sky-averaging numerical factor, and we account for the geometric factor of $\sqrt{3}/2$ manually in the LISA response function (“LISA – non-sky-averaged” in Fig. 3.2), following Ref. [233].

To fully specify the waveform produced by the `IMRPhenomPv2` template in GR, we need a 13-dimensional vector of parameters:

$$\vec{\lambda}_{\text{Pv2,GR}} = [\alpha, \delta, \theta_L, \phi_L, \phi_{\text{ref}}, t_{c,\text{ref}}, D_L, \mathcal{M}, \eta, \chi_1, \chi_2, \chi_p, \phi_p]. \quad (3.23)$$

The first 11 parameters are the same as those introduced for the `IMRPhenomD` model in Sec. 3.3.3. The parameters χ_p and ϕ_p define the magnitude and direction of the in-plane component of the spin, defined as [18]

$$\chi_p = \frac{1}{B_1 m_1^2} \max(B_1 S_{1\perp}, B_2 S_{2\perp}), \quad (3.24)$$

where $B_1 = 2 + 3q/2$, $B_2 = 2 + 3/(2q)$, $q = m_2/m_1 < 1$ is the mass ratio, and $S_{i\perp}$ is the projection of the spin of BH i on the plane orthogonal to the orbital angular momentum $\mathbf{L}$.

This `IMRPhenomPv2` is then deformed through parameterized post-Einsteinian corrections to model generic, theory-independent modifications to GR [107, 115–117]. We worked with deformations of two types:

$$\tilde{h}_{\text{gen}}(\vec{\lambda}_{\text{Pv2}}, \beta) = \begin{cases} \tilde{h}_{\text{GR}} e^{i\beta(\mathcal{M}\pi f)^{b/3}} & f < 0.018m \\ \tilde{h}_{\text{GR}} & 0.018m < f, \end{cases} \quad (3.25)$$

$$\tilde{h}_{\text{prop}}(\vec{\lambda}_{\text{Pv2}}, \beta) = \tilde{h}_{\text{GR}} e^{i\beta(\mathcal{M}\pi f)^{b/3}}, \quad (3.26)$$

where the first waveform h_{gen} represents deviations from GR caused by modified generation mechanisms, and h_{prop} represents deviations from GR caused by modified propagation mechanisms. Details (including the motivation for these implementations, and the disparity of the results between the two types of deviations) are discussed in Appendix C. As outlined there, differences are minor, and therefore from now on we will focus on the propagation mechanism, unless otherwise specified. The parameter β controls the magnitude of the deformation, and b controls the type of deformation considered. The ppE version of the `IMRPhenomPv2` model is then controlled by the parameters

$$\vec{\lambda}_{\text{Pv2,ppE}} = \vec{\lambda}_{\text{Pv2,GR}} \cup \{\beta\}. \quad (3.27)$$

Recall that, in PN language [95], a term in the phase that is proportional to $(\pi \mathcal{M} f)^{b/3}$ is said to be of $(b+5)/2$ PN order. The waveform model above is identical to the `gIMR` model coded up in LAL, and used by the LVC when performing parameterized PN tests of GR on GW data.

The main power of the ppE approach is its ability to map the ppE deformations to known theories of gravity. Table 3.7 presents the mapping between (β, b) and the coupling constants in various theories of gravity (see Appendix B for a more detailed review of these mappings).

This table makes it clear then that ppE deformations are not false degrees of freedom, in the language of [213]. Once a constraint is placed on β , one can easily map it to a constraint on the coupling constants of a given theory through Table 3.7. This reparameterization is typically computationally trivial, and therefore it saves significant resources by reusing generic results, instead of repeating the analysis for every individual theory.

3.5.3 Numerical Implementation

Common methods for calculating the requisite derivatives for the Fisher matrices typically involve either symbolic manipulation software, such as `Mathematica` [13], or the use of numerical differentiation based on a finite difference scheme. The calculation of the derivatives is always followed by some sort of numerical integration, which can be based on a fairly simple method such as Simpson’s rule, or some more advanced integration algorithm that might appear prepackaged in `Mathematica`.

All of these methods have their respective benefits: symbolic manipulation and complex integration algorithms provide the most accuracy, while numerical differentiation and simpler integration schemes are typically much faster. All methods also come with their respective drawbacks. The maximally accurate method of adaptive integration and symbolic differentiation in `Mathematica` can be computationally taxing,

while the fully numerical approach can be prone to large errors if the stepsizes are not tuned correctly, both for the differentiation with respect to the source parameters $\vec{\theta}$, as well as for the frequency spacing in the Fisher matrix integrals. On top of these aspects, using a program like `Mathematica` can be cumbersome at times, as interfacing with lower-level (or even scripting) languages adds an extra layer of complexity.

A combination of the two extremes implemented in one low-level language would be ideal, and it is the route chosen for this work. While symbolic manipulation is not available in the language that we chose (C++), we instead implemented an automatic differentiation (AD) software package natively written in C/C++: `ADOL-C` [10]. The basic premise of AD (as implemented in `ADOL-C`) is to use operator-overloading to perform the chain-rule directly on the program itself. By hard-coding a select number of derivatives on basic mathematical functions and operations (such as trigonometric functions, exponentials, addition, multiplication, etc.) and tracing out all the operations performed on an input parameter as it is transformed into an output parameter, `ADOL-C` can stitch together the derivative of the original function. This results in derivatives that are exact to numerical precision. As no final, mathematical expression is output, this does not exactly constitute symbolic differentiation, but perfectly fulfills our requirements.

To complete the Fisher calculation, we take our exact derivatives (to floating-point error) and integrate them with a Gaussian quadrature scheme based on Gauss-Legendre polynomials, as in Ref. [236]. To calculate the weighting factors and the evaluation points, we have implemented a modified version of the algorithm found in Ref. [258]. While this typically incurs a high computational cost to calculate the weights and abscissas, we mitigate this fact by doing the calculation only once, and reusing the results for each Fisher matrix. This results in integration errors orders of magnitude lower than a typical ‘‘Simpson’s rule’’ scheme, with the same computational speed per data point.

3.6 Tests of General Relativity

In this section we summarize the main results of the analysis described above. We begin with the constraints on generic modifications as a function of time for each population and network (Sec. 3.6.1). Next, we translate these into constraints on specific theories (Sec. 3.6.2, and in particular Table 3.7).

3.6.1 Constraints on Generic Modifications

Let us begin by showing in Fig. 3.7 the projected strength of constraints on modifications at various PN orders (shown in different panels) as a function of time. Detector scenarios are labeled at the top, and the various astrophysical population classes are separated to facilitate visual comparisons. Recall from Sec. 3.2

that we consider three detector scenarios (S1, S2, and S3) bracketing funding uncertainties in the development of the future detector network. The source classes include the following:

- (i) SOBH - TERR: SOBH populations as seen by only terrestrial networks;
- (ii) SOBH - MB: SOBH events observed by both terrestrial networks and LISA;
- (iii) MBHs: heavy-seed (Q3) and light-seed (PopIII) scenarios as seen by LISA.

When relevant, the error estimates shown in the figures below come from the different versions of the population model (i.e. SPOPS 265 vs SPOPS 0 and Q3delays vs Q3nodelays), as well as marginalization over the different estimates of the noise curves (i.e. the “high” and “low” sensitivity curve for Virgo and the “128Mpc” and “80Mpc” curves for KAGRA). The uncertainties correspond to the minimum and maximum bounds from all the combinations we studied at that point in the timeline.

Figure 3.7 is one of the main results of this paper. It allows us to draw many conclusions, itemized below for ease of reading⁴:

- (i) **Multiband sources yield the best constraints at negative PN orders.** This is expected from previous work [161, 206]: the long, early (almost monochromatic) inspiral signals coming from LISA observations stringently constrain deviations at low frequencies.
- (ii) **LISA MBH observations do better than terrestrial SOBH observations at negative PN orders.** Constraints coming from the large-SNR MBH populations outperform the terrestrial networks at negative PN order, despite the large number of expected SOBH sources in the terrestrial network.
- (iii) **Terrestrial SOBH observations can do slightly better than LISA MBH observations at positive PN orders.** Positive PN order effects can be constrained better when the merger is in band. The terrestrial networks begins to benefit from the millions of sources in the SOBH catalogs, but the extremely high-SNR sources in the MBH catalogs mean that LISA constraints are still competitive with terrestrial constraints.
- (iv) **Terrestrial network improvements make a big difference at negative PN orders.** The different terrestrial network scenarios are widely separated for the negative PN effects, with the most optimistic S1 scenario vastly outperforming the S2 and S3 scenarios. This conclusions is robust with respect to astrophysical uncertainties in the population models.

⁴Throughout this analysis, the 0th PN order in the GW phase refers to the first (often called “Newtonian”) term in the GR series, which is proportional to $v^{-5} \propto f^{-5/3}$. Consistently, negative (positive) PN orders identify modifications entering in at lower (higher) powers of v , relative to this leading-order term.

(v) **Network improvements are less relevant at higher PN order.** In this case the three different scenarios overlap considerably (but the S1 scenario maintains a clear edge over the other two).

To understand some of these features, it can be illuminating to model the scaling behavior of bounds at different PN orders with respect to various source parameters. Below we consider an analytical approximation that can reproduce most of the observed features. We first model constraints on individual sources, and then fold in the enhancement achieved by stacking multiple events.

Analytical scaling: individual sources

A good first approximation is to ignore any covariances between parameters by treating the Fisher matrix as approximately diagonal, so that the bounds on the generic ppE parameter β is roughly

$$\sigma_{\beta\beta} \approx \left(\frac{1}{\Gamma_{\beta\beta}} \right)^{1/2} = \left[4 \operatorname{Re} \int_{f_{\text{low}}}^{f_{\text{high}}} \frac{(\pi \mathcal{M} f)^{2b/3} |\tilde{h}|^2}{S_n(f)} df \right]^{-1/2}, \quad (3.28)$$

where f_{low} and f_{high} are the lower and upper bounds of integration. This expression can be simplified further by assuming white noise, so that $S_n(f) = S_0$ is constant, and by ignoring PN corrections to the amplitude, i.e. $|\tilde{h}| = A f^{-7/6}$, where $A \propto \mathcal{M}^{5/6}/D_L$ is an overall amplitude (see e.g. [259]). This leads to

$$\sigma_{\beta\beta} \approx \left[\frac{6A^2}{S_0} \frac{(f_{\text{low}}^{2(b-2)/3} - f_{\text{high}}^{2(b-2)/3}) (\pi \mathcal{M})^{2b/3}}{2-b} \right]^{-1/2}, \quad (3.29)$$

as long as $b \neq 2$. We can further simplify the expression for $\sigma_{\beta\beta}$ by using the fact that, within the same approximations, the SNR scales like

$$\rho^2 = 4 \operatorname{Re} \left[\int_{f_{\text{low}}}^{f_{\text{high}}} \frac{h h^*}{S_n(f)} df \right] \approx \frac{3A^2}{S_0} (f_{\text{low}}^{-4/3} - f_{\text{high}}^{-4/3}), \quad (3.30)$$

which then leads to

$$\sigma_{\beta\beta} \approx \frac{(\pi \mathcal{M})^{-b/3}}{\rho} \left[\left(1 - \frac{b}{2} \right) \frac{f_{\text{low}}^{-4/3} - f_{\text{high}}^{-4/3}}{f_{\text{low}}^{2(b-2)/3} - f_{\text{high}}^{2(b-2)/3}} \right]^{1/2} \quad (3.31)$$

Assuming the higher frequency cutoff to be at the Schwarzschild ISCO, so that $f_{\text{high}} = f_{\text{ISCO}} = 6^{-3/2} \eta^{3/5} / (\pi \mathcal{M})$, and expanding to leading order in the small quantity $\pi \mathcal{M} f_{\text{low}} \ll 1$, we finally obtain the approximate scaling

$$\sigma_{\beta\beta} \approx \left[6^{b-2} \left(\frac{b}{2} - 1 \right) \right]^{1/2} \frac{(\pi \mathcal{M} f_{\text{low}})^{-2/3}}{\eta^{(b-2)/5} \rho}, \quad b > 2, \quad (3.32)$$

$$\sigma_{\beta\beta} \approx \left(1 - \frac{b}{2} \right)^{1/2} \frac{(\pi \mathcal{M} f_{\text{low}})^{-b/3}}{\rho}, \quad b < 2. \quad (3.33)$$

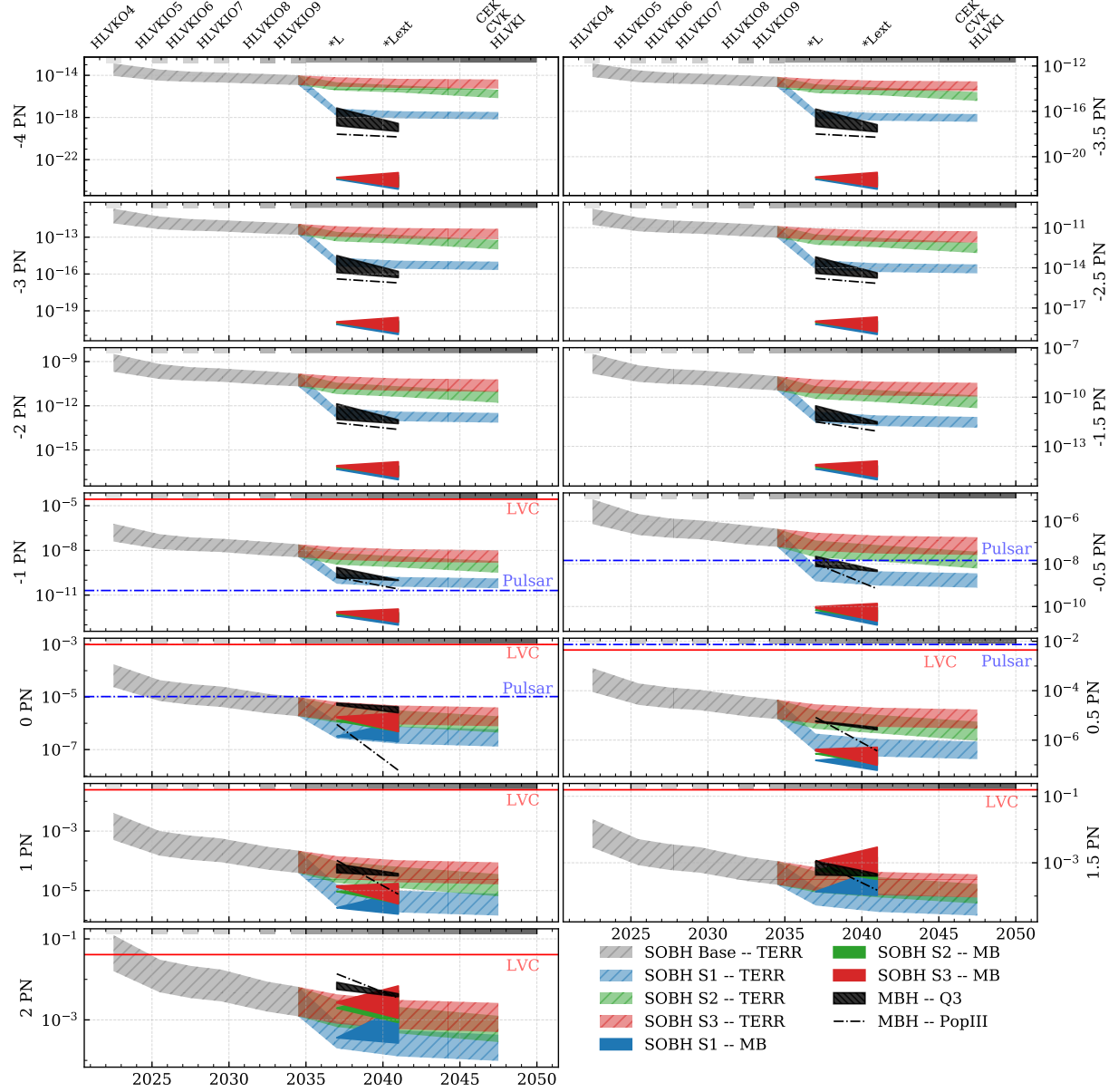


Figure 3.7: Constraints on modifications to GR at various PN orders as a function of time. The colors represent different classes of populations (including SOBH terrestrial-only sources, SOBH multiband sources, MBH sources from the Q3 heavy-seed scenario, and MBH sources from the light-seed PopIII scenario). The bands in all of these scenarios – except for PopIII – correspond to astrophysical uncertainties: kick velocities $\sigma = 265$ km/s and $\sigma = 0$ km/s give the upper and lower bounds for SOBHs, while the inclusion of delays affects Q3 scenarios. Greyscale patches at the top of each panel correspond to the observation period for each network, labeled across the top. Multiband sources and MBHs yield strong constraints at negative PN orders. Terrestrial-only SOBH sources begin to contribute substantially at positive PN orders for all detector networks, with the optimistic scenario S1 yielding the best constraints. We overlay as horizontal lines the most stringent current bounds, where available and competitive, from pulsars [49] and LVC observations of GWs [57].

The expressions above do not apply to the case $b = 2$, as the integration would lead to a logarithmic scaling. Recall that $b > 2$ corresponds to PN orders higher than 3.5.

As expected, all bounds on generic ppE parameters approximately scale as the inverse of the SNR, regardless of the PN order at which they enter. What is more interesting is that they also scale with the chirp mass as $\mathcal{M}^{-b/3}$ when $b < 2$, or as $\mathcal{M}^{-2/3}$ when $b > 2$. For a single event, we then have the ratio

$$\frac{\sigma_{\beta\beta}^{\text{TERR}}}{\sigma_{\beta\beta}^{\text{MBH}}} \approx \frac{\rho^{\text{MBH}}}{\rho^{\text{TERR}}} \left(\frac{\mathcal{M}^{\text{TERR}}}{\mathcal{M}^{\text{MBH}}} \right)^{-b/3} \left(\frac{f_{\text{low}}^{\text{TERR}}}{f_{\text{low}}^{\text{MBH}}} \right)^{-b/3}, \quad (3.34)$$

for $b < 2$. Since $\rho^{\text{MBH}}/\rho^{\text{TERR}} \sim 10^2$, $\mathcal{M}^{\text{TERR}}/\mathcal{M}^{\text{MBH}} \sim 10^{-4}$ and $f_{\text{low}}^{\text{TERR}}/f_{\text{low}}^{\text{MBH}} \sim 10^5$, we conclude that the ratio $\sigma_{\beta\beta}^{\text{TERR}}/\sigma_{\beta\beta}^{\text{MBH}} \approx 10^{3-b/3}$. This ratio is large (favoring MBH sources) when b is negative and large, i.e. at highly negative PN orders, and slowly transitions to favor terrestrial, SOBH sources at positive PN orders, explaining the observations in items (ii) and (iii) above. The ratio degrades by approximately four orders of magnitude between -4 PN and 2 PN, in favor of the terrestrial network, and in agreement with Fig. 3.7. This scaling with b holds true regardless of the typical SNRs of the sources, as the ratio of SNRs depends on the ratio of the chirp masses of the sources, but not on the PN order.

Let us now consider the scaling of the bounds with PN order in more detail. Figure 3.8 shows an averaged ratio $\sigma_{\beta\beta}^{\text{TERR}}/\sigma_{\beta\beta}^{\text{MBH}}$ computed from the full numerical simulations of Fig. 3.7 (solid blue line), together with the prediction in Eq. (3.34) that the ratio should scale as $\propto 10^{-b/3}$ (solid black line). The numerical results (blue line, with an “uncertainty” quantified by the shaded blue region) were computed as follows. We first averaged the constraints for each population model at each PN order and for each detector network that concurrently observes with LISA; this allowed us to isolate the effect of the combination of source class and detector, neglecting the sometimes significant contribution from stacking. Ratios of the averaged quantities were then calculated for each combination of SOBH model (SPOPS 0 and SPOPS 265) and heavy-seeding MBH model (Q3delays and Q3nodelays) and for each detector network – the CEKLext, CVKLext, and HLVKILext (optimistic and pessimistic) configurations – resulting in 16 combinations in all at each PN order, assuming an extended ten-year LISA mission duration. The average of these combinations is shown as the solid blue line in Fig. 3.8, and the region bounded by the minimum and maximum ratios is shown shaded in blue. Observe that the scaling of Eq. (3.34) is consistent with the averaged ratio in the entire domain; the small dip at $b = -5$ (or 0PN order) is due to degeneracies with the chirp mass, which the scaling relation does not account for.

The relation $\sigma_{\beta\beta}^{\text{TERR}}/\sigma_{\beta\beta}^{\text{MBH}}$ can be pushed further by comparing multiband sources against the rest of the SOBH sources detected *only* by the terrestrial network. For these two classes of sources, the masses would

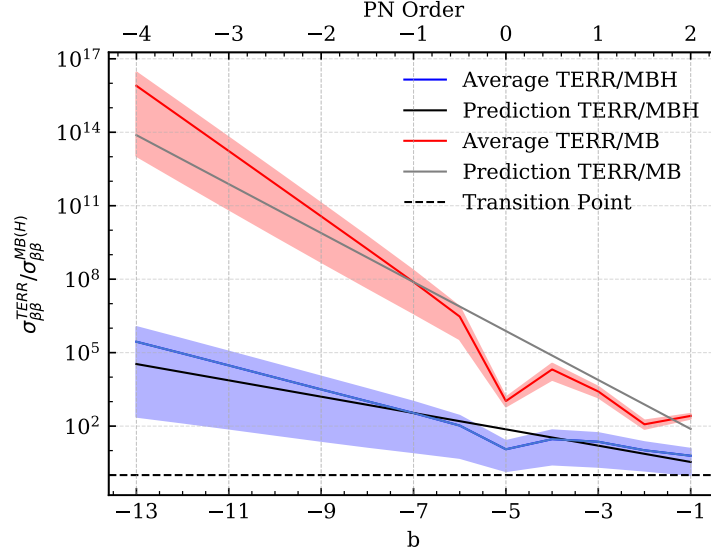


Figure 3.8: Scaling relations discussed in Sec. 3.6.1. The ratio $\sigma_{\beta\beta}^{\text{TERR}}/\sigma_{\beta\beta}^{\text{MBH}}$, calculated from the full Fisher simulations including the realistic noise curves shown in Fig. 3.2 and the IMRPhenomPv2 waveform, is shown in blue. The empirically measured trend is derived from averaging the constraints from each terrestrial network and each population model, then calculating the ratios of every combination of terrestrial network and SOBH model against each MBH heavy-seeding model. The blue line shows the mean ratio, and the blue shaded region is the area bounded by the maximum and minimum ratios. The red line and the red shaded region refer instead to the ratio between the terrestrial-only constraints and the multiband constraints, i.e. $\sigma_{\beta\beta}^{\text{TERR}}/\sigma_{\beta\beta}^{\text{MB}}$. For this class of sources, we calculate the ratio for each population model and detector network, one at a time. That is, the terrestrial-only constraints from the S1 network derived from the SPOPS 265 model are compared against the multiband constraints from the S1 network and the SPOPS 265 model. The trends predicted analytically in the text are shown in black and grey for MBH and multiband sources, respectively. The trend lines we show for our predictions have been shifted along the y-axis to better compare the with the data.

be comparable. Let us focus on the impact of the early inspiral observation. The ratio of the SNRs in the LISA band is of $\mathcal{O}(1)$ for typical sources, so we will neglect it for now. Typical initial frequencies, however, are quite different, with multiband sources having initial frequencies of about 10^{-2}Hz for SOBH sources that merge within several decades in the terrestrial band. This makes the ratio $f_{\text{low}}^{\text{TERR}}/f_{\text{low}}^{\text{MB}} \sim 10^3$, and thus, the constraining power of multiband sources relative to that of terrestrial-only sources is approximately $\sigma_{\beta\beta}^{\text{TERR}}/\sigma_{\beta\beta}^{\text{MB}} \sim 10^{-b}$, which explains the scaling observed in item (i) above. In Fig. 3.8 we show the averaged ratio measured from our full simulations including the noise curves shown in Fig. 3.2 and the IMRPhenomPv2 waveform (solid red line) as well as the 10^{-b} scaling derived from Eq. (3.34) (solid gray line). Again, we average the constraints from each population model at each PN order, assuming a ten-year LISA mission duration. However we do not consider every combination of population models and detector networks, but instead compare the multiband constraints from each network and SOBH model against the terrestrial-only constraints from the same combination of terrestrial network and SOBH model. That is, we compare S1

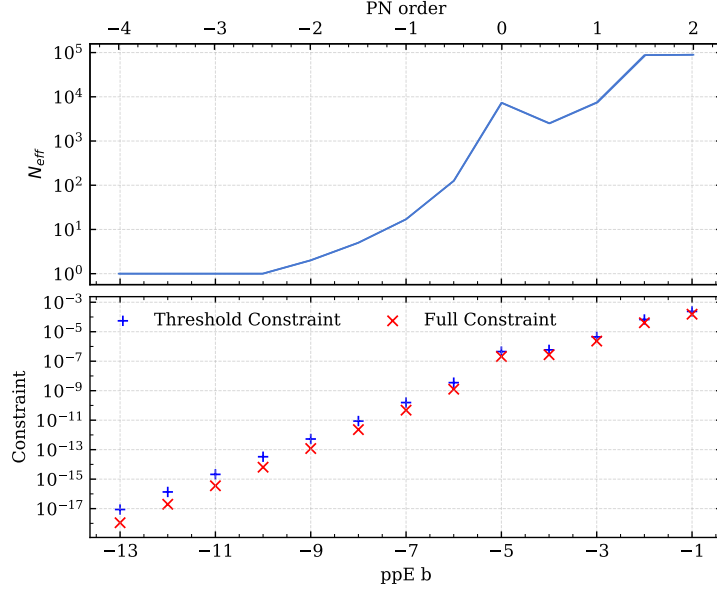


Figure 3.9: Empirically determined values of N_{eff} for the CEK (Scenario 1) network and the SPOPS 0 catalog, derived from our full Fisher analysis, including the noise curves shown in Fig. 3.2 and the IMRPhenomPv2 waveform. The parameter N_{eff} is defined as the number of sources needed from the full catalog in order to achieve a threshold constraint $\sigma_{\beta, \text{thr}}$, using the most constraining sources first. Here we choose $\log_{10} \sigma_{\beta, \text{thr}} = 0.95 \log_{10} \sigma_{\beta}$, where σ_{β} is the cumulative bound from the full Fisher analysis for the entire catalog. The values of the threshold constraint (blue + signs) are shown alongside the full constraint (red x signs) in the lower panel. The number of required sources grows exponentially as a function of PN order: large catalogs benefit positive PN orders, but they are not as important for highly negative PN orders.

terrestrial-only constraints derived from the SPOPS 265 model against the multiband constraints with the S1 network and from the SPOPS 265 model, repeating the procedure for each terrestrial network and population model. This yields 8 different combinations of population models and networks. The red line shows the average ratio for all the combinations considered, and the red-shaded region shows the area bounded by the maximum and minimum ratios. The simple analytical scaling reproduces the numerics quite well at negative PN orders, where the contribution to the constraint on the ppE parameter primarily comes from LISA observations. At positive PN orders the scaling relation breaks down for two main reasons: (i) our scaling relation neglects covariances, and (ii) the dominant source of information is no longer LISA’s observation of the early inspiral, but the signal from the merger-ringdown seen by the terrestrial network.

Analytical scaling: multiple sources

Our analysis above helps to elucidate some of the trends observed in our numerical simulations by examining individual sources, but it fails to capture the power of combining observations to enhance constraints on modified theories of gravity. Especially when considering terrestrial networks, this element is critical in

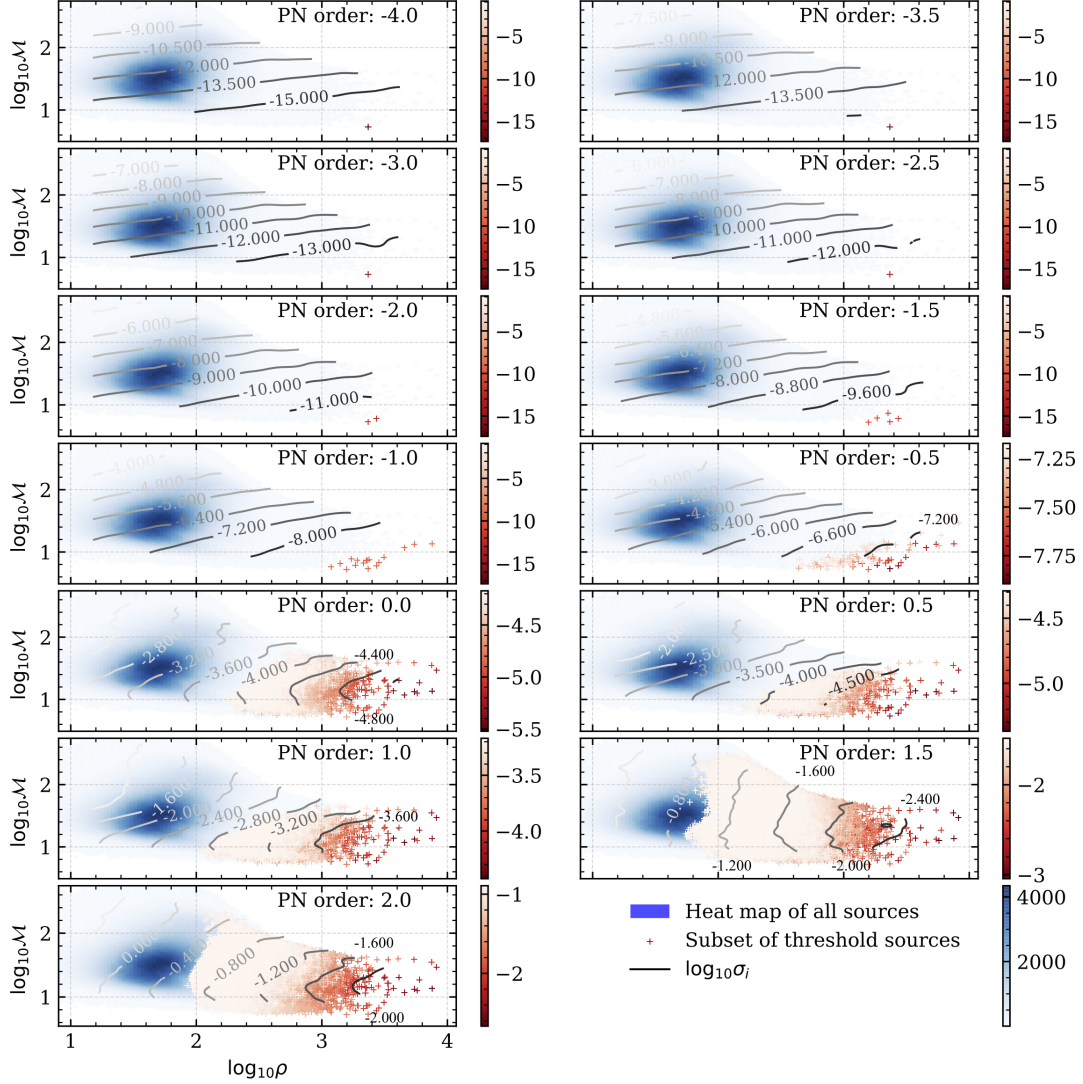


Figure 3.10: Three different distributions in the $\log_{10} \mathcal{M} - \rho$ plane for the CEK network and the SPOPS 0 population model. The blue heat map shows the distribution of the sources directly in the $\log_{10} \mathcal{M} - \rho$ plane, and it is the same for all PN orders. The black contours show the constraints from individual sources. The red scatter plots show the sources needed to obtain a threshold cumulative constraint $\log_{10} \sigma_{\beta, \text{thr}} = 0.95 \log_{10} \sigma_{\beta}$, where the shade of red indicates the strength of the individual bounds (in log base 10). We utilized a 2σ gaussian filter over the data to smooth out the noise and create more easily interpretable contour plots. In conjunction with Fig. 3.9, the growing number of scatter points as a function of PN order illustrates the increasing dependence of the cumulative constraint on the size of the source catalog. Furthermore, the relation between chirp mass, SNR, and individual bound can be seen to shift significantly between positive and negative PN orders, agreeing with the commonly held intuition that lower-mass sources are better for constraining negative PN effects. In more detail, the negative PN orders benefit highly from low-mass systems, with slight dependence on SNR, while positive PN order effects depend much more strongly on the SNR and have more minimal dependence on the chirp mass. Finally, the range of individual bounds (~ 4 orders of magnitude at negative PN orders and ~ 2 orders of magnitude at positive PN orders) helps to explain the different scaling relations between the cumulative bounds and the total number of sources.

predicting future constraints, and it is connected with our observations (iv) and (v) in the previous list.

To fully explore this facet of our predictions, we try to isolate the impact of the total number of sources on the final, cumulative constraint for a given network. As shown in Eq. (A.11) of Appendix A, the combined constraint from an ensemble of simulated detections is

$$\sigma_\beta^2 = \left(\sum_i^N \frac{1}{\sigma_{\beta,i}^2} \right)^{-1}, \quad (3.35)$$

where $\sigma_{\beta,i}$ is the variance on β of the i -th source marginalized over the source-specific parameters, including all detectors and priors, and N is the total number of sources in the ensemble. The effect of the population on all the different combinations of detector networks and PN orders can be summarized by the distribution in $\sigma_{\beta,i}$, and we find empirically that they all lie somewhere in the spectrum bounded by the following extreme scenarios:

- (a) all the constraints contribute more or less equally,
- (b) the total constraint is dominated by a single (or a few) observations.

When the covariances are all approximately equal, the sum above reduces to $\sigma_\beta \approx \sigma_{\beta,i}/\sqrt{N}$, but when one constraint (say $\sigma_{\beta,\text{strongest}}$) dominates the ensemble, the sum reduces to $\sigma_\beta \approx \sigma_{\beta,\text{strongest}}$. Naturally, in the case where all sources are more or less equally important, the power of large catalogs is maximized, and one would expect terrestrial networks observing hundreds of thousands to millions of sources to outperform networks with smaller populations, such as MBHs and multiband sources (everything else being equal). When one observation dominates the cumulative bound because of loud SNR or source parameters that maximize the constraint, then large catalogs are not as important.

In an attempt to quantify this effect, we can ask the following question: what is the minimum number of sources we can retain and still achieve a similar constraint on β ? To answer this question, we take all the variances calculated with our Fisher analysis for a given population model and detector network, and order them according to the strength of the constraint from each individual source. With some threshold constraint set, we can work our way down the list, calculating the cumulative bound for the “best” N' sources at a time. We define N_{eff} as the value of N' such that our threshold constraint is achieved. Comparing the values of N_{eff} at each PN order for a single population model and network provides useful insights into how generic constraints benefit from the catalog size.

The upper panel of Fig. 3.9 shows the values of N_{eff} calculated using the results from our full Fisher analysis, including the noise curves shown in Fig. 3.2 and the IMRPhenomPv2 waveform, for the CEK network with the SPOPS 0 population model and a threshold constraint of $\log_{10} \sigma_{\beta,\text{thr}} = 0.95 \log_{10} \sigma_\beta$. A pronounced

trend is evident: positive PN orders require up to $\sim 10^5$ sources to retain a constraint equal to our threshold value, while the most negative PN effects only require a single, highly favorable source to reach the threshold value. The lower panel of Fig. 3.9 merely shows the value of the full numerical constraint (red $\times$ signs) compared with our value of the threshold constraint (blue $+$ signs): by our own definition, the threshold constraint captures most (i.e. 95%) of the full constraint.

Figure 3.10 shows several different facets of the data relevant to the analysis of Fig. 3.9. For each PN order, we have plotted three different quantities: (i) a heat map of all the sources in the catalog in the $\log_{10} \mathcal{M} - \rho$ plane (shown in blue), which is the same for all PN orders, (ii) the contours showing the strength of the individual constraints from each source for the entire catalog (in black), and (iii) the subset of sources required to meet the threshold constraint $\sigma_{\beta, \text{thr}}$ (in red), where the shade corresponds to the strength of the individual bounds.

Several interesting conclusions can be drawn from this figure. First, the relation between the constraint, the SNR, and the chirp mass changes as a function of PN order. The highly positive PN orders benefit highly from loud sources, with only a slight preference for the lower mass systems (if at all), while highly negative PN effects benefit greatly from low-mass systems, with a slight preference for louder sources. This agrees with our intuition about low-mass systems being most important for negative PN effects: in Eq. (3.32) the chirp mass is raised to the $-b/3$ power, significantly enhancing the impact of low-mass systems for negative PN effects, while minimizing their impact for positive PN effects (assuming $b < 2$). As these figures are constructed from our fully numerical data, these trends take into account the nonlinear relation between SNR and chirp mass, as these are not independent parameters when considering realistic population models. Reasonably accurate population models are important in studies of this type, as bounds can be significantly altered by changing the distributions of source properties.

A second observation one can draw from Fig. 3.10 relates to the change in the relation between SNR and individual constraints, which explains why the constraining-power gap between the different terrestrial network scenarios closes at positive PN orders (items (iv) and (v) from above). The relaxation in the SNR-constraint correlation at high positive PN orders means that the huge boost in SNR from utilizing 3g detectors, as compared to a 2g only network, has only a moderate impact on the cumulative bound, *if* the 2g network is sensitive enough to observe a comparable number of sources to the 3g network. In the case of the Voyager network (HLVKI+), the much lower average SNR (shown in Fig. 3.4 and Fig. 3.6) hinders the network’s capability greatly at negative PN orders, but only minimally at positive PN orders, as compared with the CEK or CVK networks shown in Fig. 3.7. This is because the total number of sources observed in each scenario is comparable with Scenario 3, only differing by $\sim 30\%$, and allowing HLVKI+ to maintain

competitive constraining power through comparably sized catalogs.

A third observation that we can make about Fig. 3.10 is that the range in individual bounds is also clearly PN-order dependent. The most negative PN corrections change by ~ 4 orders of magnitude, while the most positive PN corrections only change by ~ 2 orders of magnitude. This change in constraint range lends credence to the interpretation outlined above. When constraints are clustered closer together and contribute equally, the cumulative constraint scales strongly with the number of sources. The opposite is true when the clustering is weaker and one constraint dominates over the whole ensemble. The analysis performed here, coupled with that done in Sec. 3.6.1, further clarifies the trend observed in items (ii) and (iii). The combination of the individual source scaling favoring LISA at negative PN orders is enhanced by the significant benefit from large catalogs for terrestrial networks for positive PN orders.

3.6.2 Specific Theories

We can now recast the constraints on generic ppE parameters from Sec. 3.6.1 into constraints on relevant quantities in a variety of specific modified gravity theories. We list and categorize these theories in Table 3.7.

We will utilize the scaling analysis outlined in the previous section, with the additional step

$$\Gamma_{\text{theory}} = \mathcal{J}^T \cdot \Gamma_{\text{ppE}} \cdot \mathcal{J}, \quad (3.36)$$

where $\mathcal{J}$ is the Jacobian $\partial \vec{\theta}_{\text{ppE}} / \partial \vec{\theta}_{\text{theory}}$ of the transformation, and $(\cdot)^T$ is the transpose operation. In our case, the Jacobian is diagonal. This is because the off-diagonal components are all proportional to the theory-specific modifying parameter; as we inject with GR models, these are always set to zero for any specific beyond-GR theory. We can then write

$$\Gamma_{\alpha_{\text{theory}} \alpha_{\text{theory}}} = \left(\frac{\partial \beta}{\partial \alpha_{\text{theory}}} \right)^2 \Gamma_{\beta \beta}, \quad (3.37)$$

where β is the generic ppE modification at the corresponding PN order for a given theory, and α_{theory} is the theory-specific modifying parameter. The interested reader can find the mappings $\beta(\alpha_{\text{theory}})$ between each theory and the ppE formalism, and more in-depth explanations of their motivations, in Appendix B.

This mapping between ppE constraints and theory-specific constraints changes the scaling relations between the theory-specific bound and different source parameters, with many of the conclusions made by examining the generic constraints changing quite drastically. This is because the Jacobian typically depends on source parameters, like $\mathcal{M}$, η , χ_1 , and χ_2 , and this can strongly enhance the constraining power of one population of BBHs over another. No general trend can be ascertained across multiple modified theories since

each coupling is different, so we will examine each theory in turn. As we will see, constraints on different theory-specific parameters scale differently with SNR, chirp mass, etcetera, impacting how the cumulative bound improves with stacking and how dependent the bound is on small numbers of loud sources. To examine this in more detail, we will focus on a single detector network (HLVKIO8) with a single population model (SPOPS 0) to try and isolate the pertinent effects for each theory.

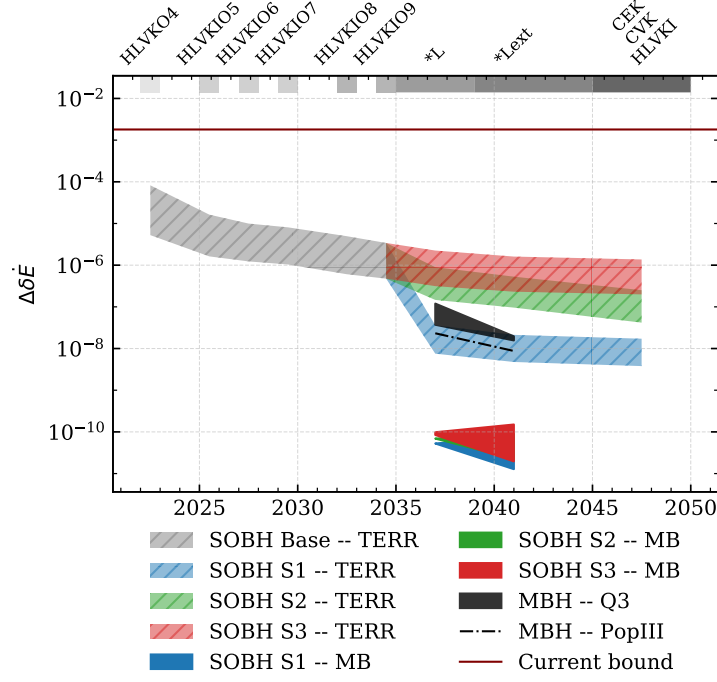


Figure 3.11: Projected cumulative constraints on generic dipolar radiation for the detector networks and population models examined in this paper. The multiband sources outperform all other source classes by at least ~ 2 orders of magnitude, with MBH sources and the most optimistic terrestrial scenario performing comparably.

Generic Dipole Radiation

Dipole radiation is absent in GR, since in Einstein’s theory GWs are sourced by the time variation of the quadrupole moment of the stress-energy tensor. Therefore, any observation of dipole radiation would indicate a departure from GR. Dipole radiation must be sourced by additional channels of energy loss, due to the presence of new (scalar, vector or tensor) propagating degrees of freedom. By the balance law, these new channels of energy loss affect the time variation of the binding energy E , and therefore dipole effects generically enter the GW Fourier phase at -1 PN (to leading order) [161]. While many theories predict specific forms of dipole radiation, we can constrain any process leading to dipole radiation by the time rate

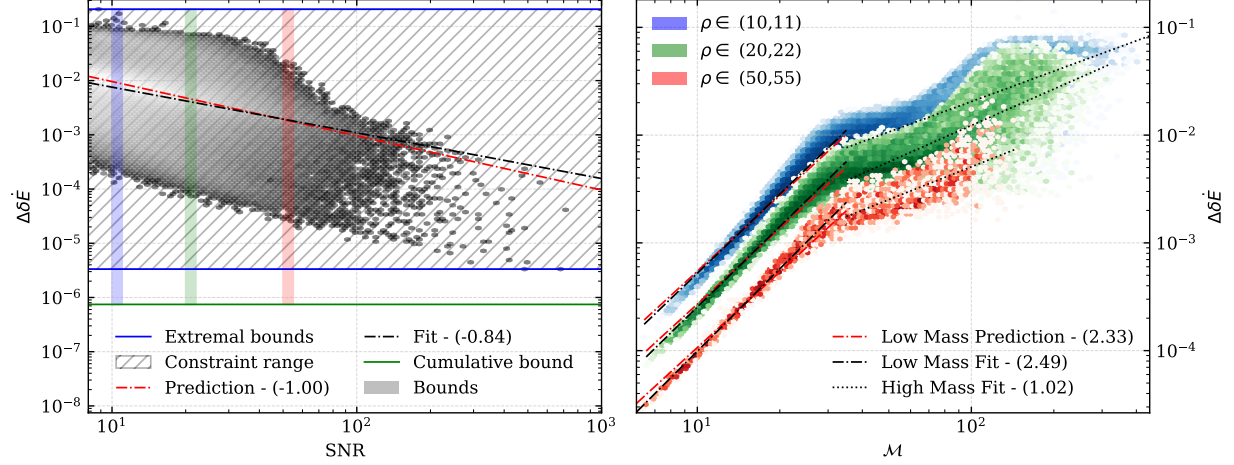


Figure 3.12: Result of the scaling analysis outlined in Sec. 3.6.2 performed on the data synthesized with the HLVKIO8 network and the SPOPS 0 population. The left panel shows a heat map of the constraint on $\delta\dot{E}$ versus the SNR of the source. The solid blue lines correspond to the strongest and weakest single-source constraint, and the area between these two bounds is shown in hatching. The cumulative bound from the entire catalog is shown as the solid green line. The power-law fit to the data in the left panel is shown as the solid black curve, and our prediction for the scaling is shown as the solid red curve. The right panel shows three distinct slices of the catalog, with ranges in SNR from 10 to 11 (blue), 20 to 22 (green), and 50 to 55 (red). These ranges are highlighted in the left panel. The right panel shows the density of the constraint versus the chirp mass, with empirical trends shown in black and predicted trends shown in red. There is a noticeable transition point in the distribution, so low-mass and high-mass systems were analyzed separately. The powers used in all trend lines are shown in the legend. For trend lines, the (logarithmic) offset for the predicted scaling relations has been adjusted to coincide with the empirically fit offset, to better compare the slopes of the trends. Of particular interest is the strong trend relating the SNR and the bound, as well as the tight correlation between chirp mass and constraint for low-mass systems, which seems to taper off for high-mass systems.

of change of the binding energy, $\dot{E}$.

We show in Appendix B, the Jacobian in this specific class of modifications scales as

$$\left(\frac{\partial\beta}{\partial\delta\dot{E}}\right)^2 \propto \eta^{4/5}, \quad (3.38)$$

where $\delta\dot{E} = \dot{E} - \dot{E}_{\text{GR}}$ is the variation in $\dot{E}$ due to dipole radiation: see Eq. (B.1). This implies that the scaling relations found earlier for generic ppE modifications should not change much when we translate them into constraints on dipole radiation.

These constraints are shown in Fig. 3.11. As dipole radiation is a negative PN effect, multiband sources will contribute significantly, improving bounds by at least two orders of magnitude over any other detector network or population class. LISA observations of MBH binaries are still highly competitive, outpacing the terrestrial-only network in all cases except the most optimistic detector schedule. Furthermore, the different terrestrial networks see a wide variation, as the difference between the typical SNRs between the networks

are quite large. After thirty years of GW measurements, our models suggest an improvement of 3–9 orders of magnitude over existing constraints, depending on source populations and detector characteristics, but a 9-orders-of-magnitude improvement is only possible with multiband events. All of these trends are consistent with the analysis presented in Sec. 3.6.1, with constraints on this negative PN order effect benefitting from the low initial frequency and low chirp masses of LISA multiband sources. This is because dipole radiation approximately scales like a generic ppE modification in terms of SNR and chirp mass, meaning that most of the analysis from above is still valid in this case.

To better understand the numerical results presented in Fig. 3.11, we can look at our analytical approximation of $\Delta\delta\dot{E}$ using the methods from the previous section. After mapping the bound on the generic β to $\delta\dot{E}$, expanding in $\epsilon = \mathcal{M}f_{\text{low}}$, and setting the upper frequency to the ISCO frequency, we have the approximation

$$\Delta\delta\dot{E} \approx \frac{112\sqrt{2}}{\eta^{2/5}} \frac{(\mathcal{M}\pi f_{\text{low}})^{7/3}}{\rho}. \quad (3.39)$$

Results related to this approximation are shown in Fig. 3.12. The left panel shows a density map of the bounds on $\delta\dot{E}$ versus the SNR of the source, with a numerical fit overlaid showing the SNR scaling trend in black. Our $1/\rho$ scaling prediction, shown in red, matches the numerics very well.

The right panel shows a density plot of the bound on $\delta\dot{E}$ versus chirp mass. To isolate the impact of the chirp mass on the attainable bound on $\delta\dot{E}$, we restrict ourselves to thin slices in different ranges of SNR (the ranges are highlighted in the top panel). This is to insulate our results from the fact that the SNR typically scales with the mass, causing a nonlinear relationship between the mass, SNR, and constraint. To ensure that the scaling does not change for different ranges of SNR, we have separately analyzed three different ranges. For lower mass systems, we see good agreement with the analytically predicted $\mathcal{M}^{7/3}$ scaling relationship, but around $\mathcal{M} \sim 30M_{\odot}$ we see a sharp transition, and our approximations fail.

The impact of these different scaling relations can be seen in the range of constraints and the cumulative constraint shown in Fig. 3.12. In the left panel, we have plotted the strongest and weakest constraint as solid blue lines, bounding the parameter space of single-source bounds. The cumulative bound for this one network-population combination is shown as a green line, near the bottom of the panel. As is evident in the figure, the improvement of the cumulative bound over the most stringent bound is marginal. This can be explained by the huge range of single-source bounds, covering five orders of magnitude, consistent with the analysis performed in Sec. 3.6.1.

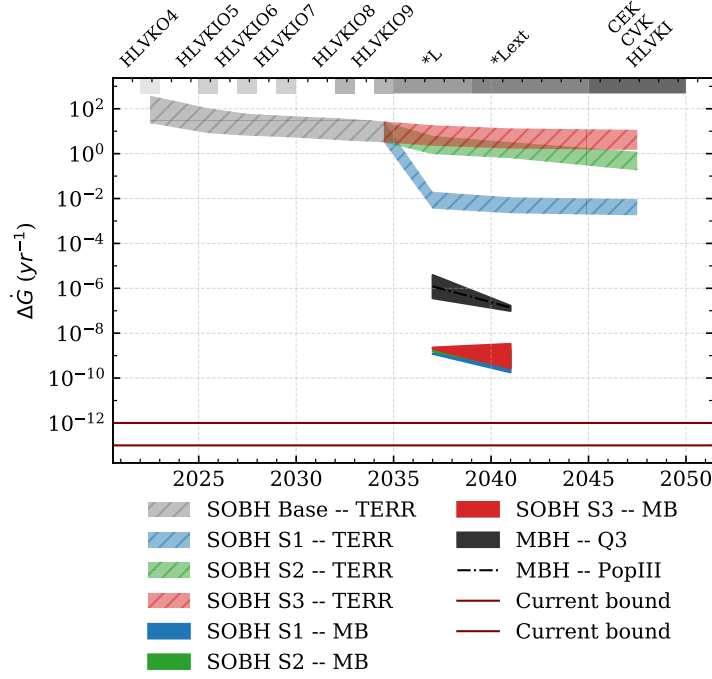


Figure 3.13: Projected cumulative constraints on the time derivative of the gravitational constant $\dot{G}$ for the detector networks and population models examined in this paper. Multiband sources outperform all other source classes by $\sim 1 - 2$ orders of magnitude, with MBH sources performing the next best. SOBHs observed by the terrestrial network alone perform the worst, but with Scenario 1 outperforming Scenarios 2 and 3 due to the high SNR of the observations in the former network.

Local Position Invariance – Variable G Theories

If the gravitational constant G were time-dependent, we would observe anomalous acceleration in the inspiral of BBHs [260]. At leading order, this affects the GW Fourier phase at -4 PN. From the transformation in Appendix B, the Jacobian to map from the generic ppE modification to the parameter $\dot{G}$ itself is

$$\left(\frac{\partial\beta}{\partial\dot{G}}\right)^2 \propto \left(\frac{\mathcal{M}}{1+z}\right)^2. \quad (3.40)$$

The mapping now includes a chirp mass-dependent factor, which can vary by orders of magnitude between source classes. From this scaling with chirp mass, and the fact that this modification enters at a highly negative PN order (-4 PN), we expect that the best sources will be those that are seen at the widest separations (like multi-band sources) and have the largest chirp mass.

Our predictions for the constraints on $\dot{G}$ can be seen in Fig. 3.13. Multiband constraints again outperform all other source classes and detector configurations, as expected. However, because the Jacobian is proportional to $\mathcal{M}^2$, MBH sources seen by LISA are not far behind. Comparatively, the terrestrial-only bounds trail

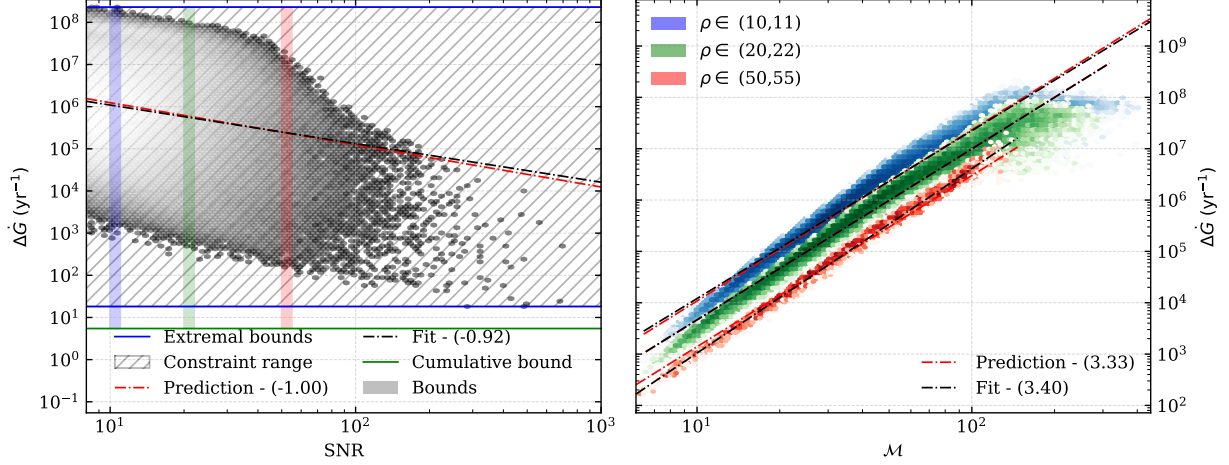


Figure 3.14: Result of the scaling analysis outlined in Sec. 3.6.2 performed on the data synthesized with the HLVKIO8 network and the SPOPS 0 population. The plotting style is the same as in Fig. 3.12. The left panel shows a heat map of the constraint on $\dot{G}$ versus the SNR of the source. The right panel shows the density of the constraint versus $\mathcal{M}$, with empirical trends shown in black and predicted trends shown in red. Again, the strong trend relating the SNR and the bound agrees well with the prediction, and there seems to be a tight correlation between $\mathcal{M}$ and constraint, well approximated by our analysis in Sec. 3.6.2.

significantly behind both of these source classes, by as much as three orders of magnitude. There is also a wide separation between the three different terrestrial-only observation scenarios. This suggests that the cumulative bound does not benefit too much from large catalogs, but instead is dominated by a small number of favorable observations.

A variable G modification presents the first departure from our analysis on the scaling of generic results. MBH sources receive a sizeable benefit over the SOBH sources due to the Jacobian factor between parameters. Consequently, constraints on this particular modification benefit greatly from the inclusion of LISA in the GW network, both in the form of multiband and MBH observations.

Even after thirty more years of GW detections with the most ideal networks, our models indicate that the bounds will still fall far short of the current constraints on $\dot{G}$ coming from cosmology. These constraints, however, are qualitatively different from those considered here. Cosmological constraints assume a Newton constant that is linearly dependent on time in the entire cosmological history of the Universe, i.e. that $G \rightarrow G(t) \sim G_{\text{BBN}} + \dot{G}_{\text{BBN}}t$, where t is time from the Big Bang until today, and where G_{BBN} and $\dot{G}_{\text{BBN}}$ are constants. Our $\dot{G}$ constraints only assume a linear time dependence *near* the BBH merger, i.e. that $G \rightarrow G(t) \sim G_{t_c} + \dot{G}_{t_c}(t - t_c)$ for $t < t_c$ where t_c is the time of coalescence, G_{t_c} and $\dot{G}_{t_c}$ are constants, and $G(t)$ relaxes back to G_{t_c} in a few horizon light-crossing times. In our stacking analysis, we are implicitly assuming that $\dot{G}_{t_c}$ is the same for all sources in all catalogs. Therefore, it is not strictly fair to compare cosmological and GW bounds.

We can again repeat the analysis from Sec. 3.6.1 to better understand the relationship between the bound on $\dot{G}$ and various source parameters. Making the approximations outlined in Sec. 3.6.1, we can approximately rewrite the constraint on $\dot{G}$ as

$$\Delta\dot{G} \approx \frac{32763}{5} \sqrt{\frac{6}{5}} \frac{(\pi \mathcal{M} f_{\text{low}})^{13/3} (1+z)}{\mathcal{M} \rho}, \quad (3.41)$$

where we obtain the expected extra dependence on the chirp mass from the Jacobian transformation. Results pertinent to this approximation are shown in Fig. 3.14. The left panel shows a heat map of the $\dot{G}$ constraints against the SNR for the sources in the HLVKIO8 network and the SPOPS 0 model. The right panel shows a heat map of the constraint on $\dot{G}$ against the chirp mass, for different slices in the SNR. Notably, the scaling of the constraint on $\dot{G}$ with respect to the chirp mass matches well with our prediction of $\mathcal{M}^{10/3}$, which differs from the generic constraint by a factor of $\mathcal{M}^{-1}$ due to the Jacobian factor. Again, we see a large spread in the magnitude of the constraint, ranging over ~ 6 orders of magnitude. This leads to a marginal improvement of the cumulative bound over the strongest bound from a single observation, further hampering the terrestrial-only networks, in agreement with our analysis in Sec. 3.6.1. After accounting for the modified scaling due to the Jacobian, the scaling relations and techniques from Sec. 3.6.1 generally hold for predicting constraints on variable G theories.

Lorentz Violation – Noncommutative Gravity

If a commutation relation is enforced between momentum and position, as in quantum mechanics, the leading order effect occurs at 2PN. Predictions for the constraints on the scale of the noncommutative relation are shown in Fig. 3.15. The Jacobian of the transformation found in Appendix B is given by

$$\left(\frac{\partial \beta}{\partial \Lambda^2} \right)^2 \propto \eta^{-4/5} (2\eta - 1). \quad (3.42)$$

The Jacobian only introduces source-dependent terms of $\mathcal{O}(1)$, and as such, bounds on Λ^2 should generally follow the scaling trends found in Sec. 3.6.1. Given that this modification comes at 2PN, we would expect the terrestrial-only source catalogs to constrain non-commutative gravity the strongest: the power of large catalogs is enhanced, and the effect of LISA observations of the early inspiral is less relevant for positive PN effects.

The bounds predicted by our models are shown in Fig. 3.15. As expected, the terrestrial networks contribute the most to any future bound on non-commutative gravity. Even when just considering the three terrestrial-only scenarios, the differences are minimal. Furthermore, the other source classes (MBH and

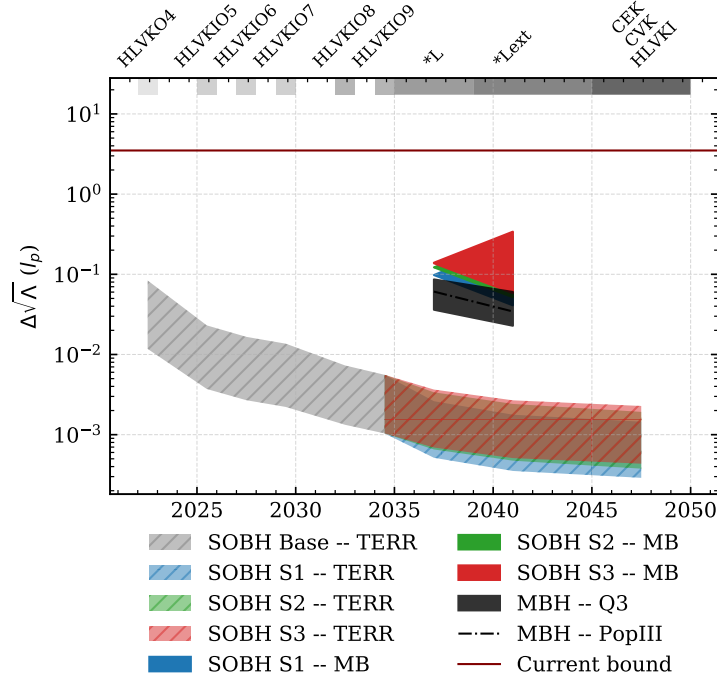


Figure 3.15: Projected cumulative constraints on $\sqrt{\Lambda}$ for the detector networks and population models examined in this paper. Terrestrial-only catalogs, with their populations of millions of sources, seem to dominate any future constraint on this particular deviation, with an improvement by 1–2 orders of magnitude over any other source classification. This conclusion seems independent of the particular terrestrial scenario we pick, with comparable performance from all three.

multiband) perform almost identically. All of these trends further solidify our conclusion that the key to future constraints on this particular modification is large catalogs of observations, as opposed to single, favorable sources. Future constraints from all source classes should improve by 1–3 orders of magnitude over present constraints.

Continuing our analysis to explore the more subtle trends we are seeing, we can repeat the analysis outlined in Sec. 3.6.1. This gives us the following approximation for the variance on $\sqrt{\Lambda}$:

$$\Delta\sqrt{\Lambda} \approx \left(\frac{32768}{1875}\right)^{1/8} \frac{\eta^{1/5} (\pi\mathcal{M}f_{\text{low}})^{1/12}}{(1-2\eta)^{1/4} \rho^{1/4}}. \quad (3.43)$$

Although the bound on Λ^2 scales as expected from Sec. 3.6.1, approximating our bound on $\sqrt{\Lambda}$ given our constraint on Λ^2 introduces modifications to the trends we would not have expected from a straightforward extrapolation from constraints on generic modifications. Namely, we see that the bound should generically scale with the SNR as $\rho^{-1/4}$, and the constraint should scale with the chirp mass as $\mathcal{M}^{1/12}$.

Pertinent trends related to this approximation are shown in Fig. 3.16, where the HLVKIO8 network and

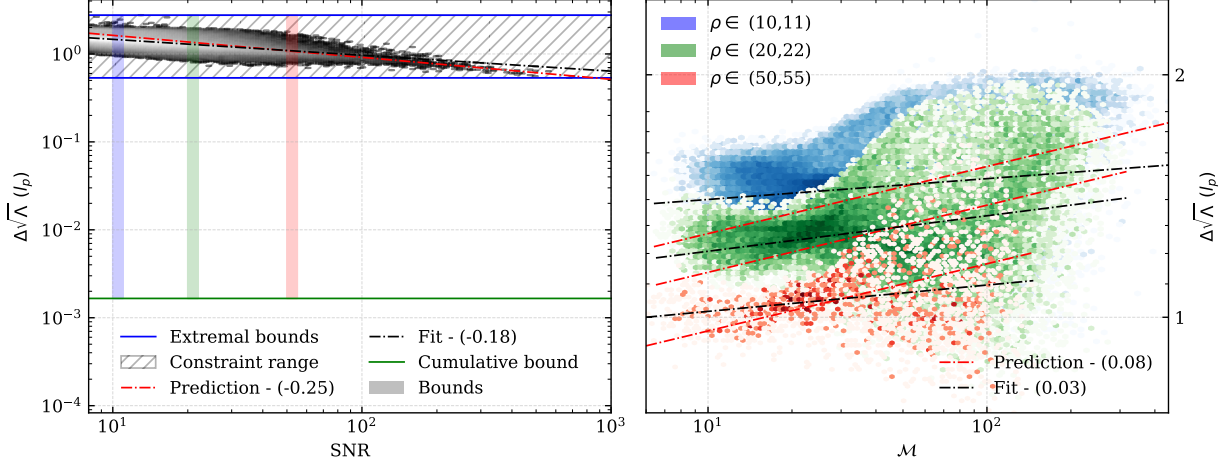


Figure 3.16: Result of the scaling analysis outlined in Sec. 3.6.2 performed on the data synthesized with the HLVKIO8 network and the SPOPS 0 population. The plotting style is the same as in Fig. 3.12. The left panel shows a heat map of the constraint on $\sqrt{\Lambda}$ versus the SNR of the source. The right panel shows the density of the constraint versus the chirp mass, with empirical trends shown in black and predicted trends shown in red. The small range of constraints from the catalog lead to considerable enhancements of the cumulative bound when stacking observations, and the weak scaling with chirp mass and moderate scaling with SNR further benefit SOBH sources over other source classes.

the SPOPS 0 model were used to do the analysis. The left panel shows a heat map in the constraint-SNR plane, with the extremal, single source bounds shown as solid blue lines. The cumulative bound for only this network-population combination is shown as the solid green line. Our predicted trend for the constraint with respect to the SNR is shown in red, while the empirically determined trend is shown in black. The right panel shows a heat map in the constraint-chirp mass plane, where we have separately analyzed three different slices of sources with specific SNRs, denoted by the colors red, blue, and green.

In the left panel of Fig. 3.16, we can see that our approximation for the relation between the constraint and the SNR does fairly well relative to the empirically determined trend. Furthermore, we see that the range of constraints is considerably tighter than even the generic constraints at 2PN. The largest and smallest bound for non-commutative gravity are separated by one order of magnitude, leading to a significant improvement of the cumulative bound over the tightest single-observation bound. This feature further explains to some degree the discrepancy between LISA sources and terrestrial-only sources in Fig. 3.15.

In the right panel of Fig. 3.16, we see much wider distributions in the constraint-chirp mass plane, as compared to the previously analyzed modifications. Our predicted trends are moderately accurate, although with noticeably lower accuracy. This is consistent with the fact the constraint scales very weakly with chirp mass ($\mathcal{M}^{1/12}$), and other correlations are widening the distribution and complicating the relation.

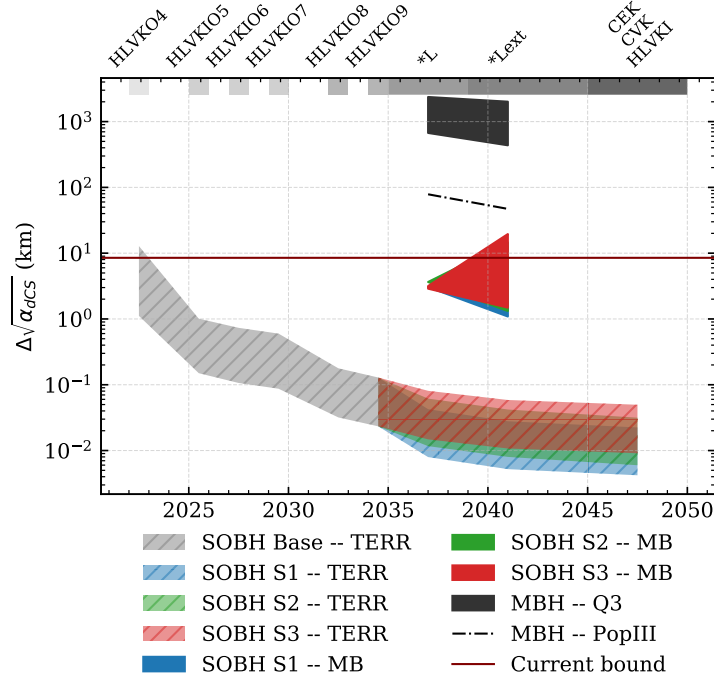


Figure 3.17: Projected cumulative constraints on $\sqrt{\alpha_{\text{dcs}}}$ for the detector networks and population models examined in this paper. Terrestrial-only catalogs, with their populations of millions of sources, dominate any future constraint on this particular deviation, with an improvement of 2-5 orders of magnitude over other source classification. This conclusion is independent of the terrestrial scenario we pick, with comparable performance from all three. Multiband sources, with their low chirp masses, seem to perform the next best.

Parity Violation – Dynamical Chern Simons

One of the fundamental tenets of GR is the parity invariance of the gravitational action. Dynamical Chern-Simons (dCS) gravity includes a parity-odd, second-order curvature term in the action, known as the Pontryagin density, coupled to a scalar field through a dimensionful parameter α_{dcs} . The fact that the Pontryagin density is parity-odd necessarily restricts the scalar field to also be odd in vacuum, making it an axial field. The leading-order effect in the GW phase sourced by these deviations enters at 2PN order. In Appendix B we recall that the following mapping holds:

$$\left(\frac{\partial \beta}{\partial \alpha_{\text{dcs}}^2} \right)^2 \propto \frac{[\hat{m}_1 s_2^{\text{dcs}} - \hat{m}_2 s_1^{\text{dcs}}]^4 \eta^{8/5}}{(1+z)^{-8} \mathcal{M}^8}, \quad (3.44)$$

where s_i^{dcs} is the BH sensitivity, defined in Eq. (B.9), and $\hat{m}_i = m_i/\mathcal{M} = \eta^{-3/5}(1 \pm \sqrt{1-4\eta})/2$ for the larger (+) and smaller (−) mass. Here, we have only shown the Jacobian to leading order in spin, and we have transformed the mass components to explicitly show the chirp mass dependence. As the mass ratio and spin factors are bounded to a magnitude of $\mathcal{O}(1)$, the dependence of the Jacobian on $\mathcal{M}^{-8}$ should have

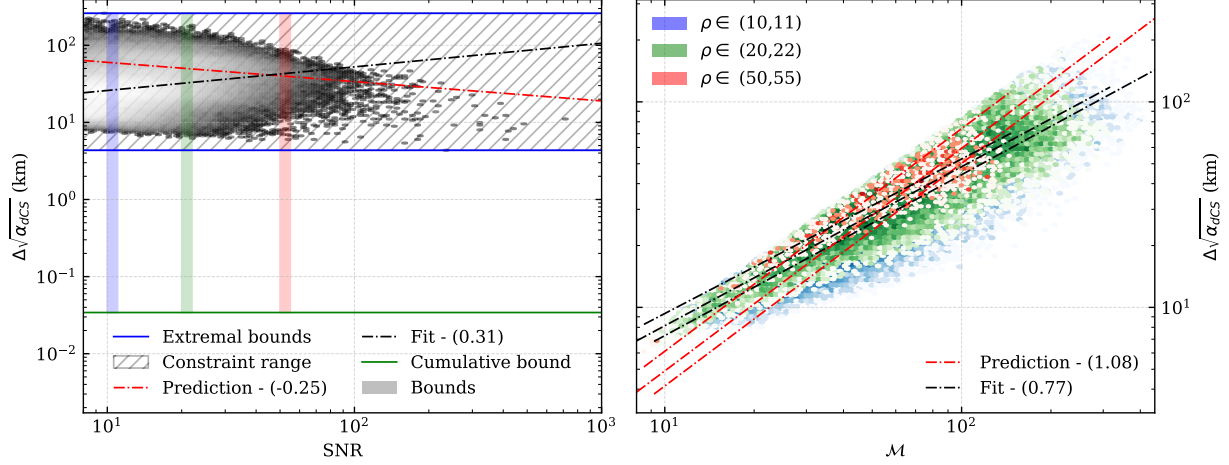


Figure 3.18: Result of the scaling analysis outlined in Sec. 3.6.2 performed on the data synthesized with the HLVKIO8 network and the SPOPS 0 population. The plotting style is the same as in Fig. 3.12. The left panel shows a heat map of the constraint on $\sqrt{\alpha_{dcs}}$ versus the SNR of the source. The right panel shows the density of the constraint versus the chirp mass, with empirical trends shown in black and predicted trends shown in red. Our prediction for the SNR scaling is considerably less accurate than for previous theories, presumably from covariances with other source parameters and competing scaling trends with the chirp mass. The tight range of constraints and large improvement of the cumulative bound over all other single source constraints, seen in the left panel, indicate strong dependence on the total number of sources in the catalog.

the most significant effect on $\Delta\alpha_{dcs}$ and *strongly* favor low-mass systems, suggesting that SOBHS would be considerably more effective than MBHs. Furthermore, as this is a positive PN modification, we would expect to see a sizeable benefit from large catalogs, given the analysis in Sec. 3.6.1, and the impact of LISA observations of the early inspiral should be considerably less important. All of these factors point to the terrestrial-observation only scenarios outperforming LISA detections of MBH sources and LISA-terrestrial joint detections of multiband sources.

Our predictions for the constraints on the strength of this coupling are shown in Fig. 3.17. Indeed, terrestrial-only detections perform the best at constraining dCS modifications to GW, with bounds up to ~ 2 orders of magnitude tighter than multiband sources and ~ 4 -5 orders of magnitude better than MBH sources. As expected, MBH sources detected by LISA are severely inhibited by the particular Jacobian for this specific modification. Furthermore, we also see little variation between the three terrestrial scenarios, indicating that a significant weight lies with the size of the catalogs, as opposed to the source properties of a select minority of favorable observations. As the power of constraining this particular modification to GR benefits strongly from large numbers of sources, we can expect to slowly push the current bound down by ~ 3 orders of magnitude, with minimal dependence on the actual detector schedule, over the course of the next thirty years.

Further analysis using the techniques in Sec. 3.6.1 leads to the following approximate form of the variance:

$$\begin{aligned} \Delta\sqrt{\alpha_{\text{dCS}}} &\approx \left(\frac{3584\sqrt{6}}{5\pi}\right)^{1/4} \frac{(\pi\mathcal{M}f_{\text{low}})^{1/12}\mathcal{M}}{(1+z)\eta^{1/5}\rho^{1/4}} \\ &\times \left(|3015\chi_2^2\hat{m}_1^2 - 5250\chi_1\chi_2\hat{m}_1\hat{m}_2 + 3015\chi_1^2\hat{m}_2^2\right. \\ &\quad \left.- 14(\hat{m}_2s_1^{\text{dCS}} - \hat{m}_1s_2^{\text{dCS}})^2\right)^{-1/4}. \end{aligned} \quad (3.45)$$

Beyond the additional terms coming from the Jacobian of the parameter transformation, we now see additional deviations from our analysis on generic modifications in Sec. 3.6.1. Raising the bound on α_{dCS}^2 to the one-fourth power to obtain our further approximated bound on $\sqrt{\alpha_{\text{dCS}}}$ has introduced new dependence of the constraint on all the source parameters of interest. Namely, the dependence on ρ has been amended to scale as $\rho^{-1/4}$, and the dependence on the chirp mass is now $\mathcal{M}^{13/12}$.

Results related to this analysis are shown in Fig. 3.18, derived from data produced with the HLVKIO8 network and the SPOPS 0 model. The left panel shows a heat map of the sources in the catalog in the $\Delta\alpha_{\text{dCS}}$ -SNR plane, with the extremal bounds shown in blue, and the cumulative bound (for this single catalog) shown in green. The right panel shows a heat map of the sources in the $\Delta\alpha_{\text{dCS}}$ - $\mathcal{M}$ plane for three slices in SNR-range (in red, blue, and green). The trends we have predicted are shown in red, while the empirically determined trends are shown in black, for both panels.

Starting in the left panel, the range in single-observation constraints on $\sqrt{\alpha_{\text{dCS}}}$ is quite small. The tight range of the constraints (just 1–2 orders of magnitude between the strongest and weakest constraints) helps to explain the enhanced effectiveness of the terrestrial networks at constraining this modification, as the constraint scales favorably with large numbers of observations. This is explicitly seen by the sizable improvement of the cumulative constraint over the constraint coming from the strongest single observation.

Furthermore, in the left panel, we see that our prediction for the SNR trend does not accurately reflect what we observe in the synthetic data. This is in stark contrast with non-commutative gravity, where the modification enters at the same PN order and predicts identical scaling with respect to the SNR. Notably, this deviation also occurs in EdGB gravity, detailed below, which has a similarly complicated Jacobian. The primary differences between the modification introduced by dCS and non-commutative gravity are (i) the scaling of the constraint with respect to the chirp mass, and (ii) covariances between the modified gravity coupling constant and all other sources parameters (such as the spins and mass ratio).

For difference (i), we can examine the right panel of Fig. 3.18, where we see moderate agreement with our predicted scaling trend for the chirp mass and much tighter correlations for dCS than for non-commutative gravity. Not only is the trend more accurately predicted, but the scaling with chirp mass in dCS, as compared

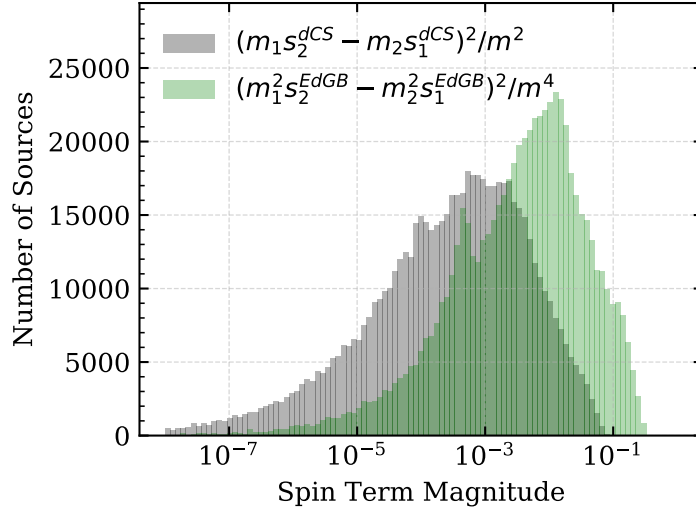


Figure 3.19: Histogram of spin-related terms contributing to the relevant Fisher element for dCS and EdGB. The sources were taken from the catalog derived from the HLVKIO8 network and SPOPS 0 population model. For dCS, this only includes the term to first order in spin. The wide range of magnitudes that this term can take (5–6 orders of magnitude) helps to explain the breakdown of our ability to predict trends concerning the constraints on these theories. From Fig. 3.18 we see that the SNR and chirp mass only span a range of 1 or 2 orders of magnitude, and as such, the trends we would expect to see for these parameters could be completely washed out by this additional spin-dependent term, which we have neglected in our simple analysis.

with non-commutative gravity, is considerably stronger ($\mathcal{M}^{13/12}$ as opposed to $\mathcal{M}^{1/12}$). Considering there is a negative correlation between the constraint and the SNR, a positive correlation between the constraint and the chirp mass, and a positive correlation between the SNR and chirp mass, a shift in the different trends as significant as that found in dCS may lead to the observed deterioration in our predictions.

For difference (ii), the mild agreement of the chirp mass scaling in the right panel suggests that covariances between parameters are degrading the accuracy of all of our approximations, not just the SNR. To further explore this idea, we can look at the typical range of values that the other source-dependent terms from the Jacobian in Eq. (3.44) can take. For the final bound from a given source, the magnitude of these additional terms in an absolute sense is important, but in terms of the trends we expect to see, the range of values these terms can take is the quantity of interest. If certain sources with comparable SNR and chirp mass have Jacobian transformations that span several orders of magnitude because of these additional terms, our simple analytical approximations cannot be expected to accurately match the synthetic data. A histogram of the spin- and mass ratio-dependent terms for both dCS and EdGB are shown in Fig. 3.19, where we do indeed see a non-negligible range of values. Figure 3.18 shows that the SNR and chirp mass both span approximately 1–2 orders of magnitude for this particular catalog, while the complicated Jacobian factors that we have neglected in our analysis span approximately 4–5 orders of magnitude. A range this large can

easily erase any structure we would hope to see with our simple approximations, and helps to explain why our simple analytical approximation fails for dCS (and for EdGB, as we will discuss below).

Between these two factors, our ability to predict scaling trends of the constraint on $\sqrt{\alpha_{\text{dCS}}}$ as a function of source parameters has moderate success with regards to the chirp mass, but is definitely degraded in general when compared with the same analysis for general modifications. The dCS example provides direct evidence that conclusions derived from generic constraints may be highly misleading when focusing on a particular modified theory.

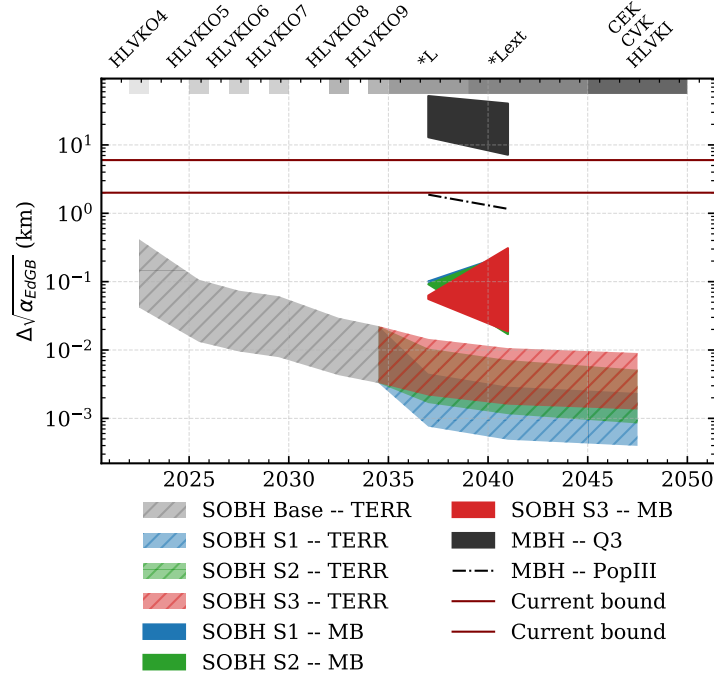


Figure 3.20: Projected cumulative constraints on $\sqrt{\alpha_{\text{EdGB}}}$ for the detector networks and population models examined in this paper. Terrestrial-only catalogs, with their populations of millions of sources, seem to most efficiently constrain EdGB, but multiband sources are not far behind. The modified scaling of the constraint with SNR and chirp mass work in favor of terrestrial networks, but the fact that EdGB produces a negative PN modification to leading order benefits multiband sources. MBHs are not effective at constraining EdGB, and will not contribute much to future bounds on this theory.

Quadratic Gravity – Einstein-dilaton-Gauss-Bonnet

Similar to dCS, Einstein-dilaton-Gauss-Bonnet (EdGB) gravity is also quadratic in curvature at the level of the action. In this case, a scalar field is coupled to the Gauss-Bonnet invariant through a dimensionful coupling constant α_{EdGB} . In contrast to dCS, the scalar field in EdGB is parity-even in vacuum (because the Gauss-Bonnet invariant is also parity-even), and the leading order correction to the GW phase comes at

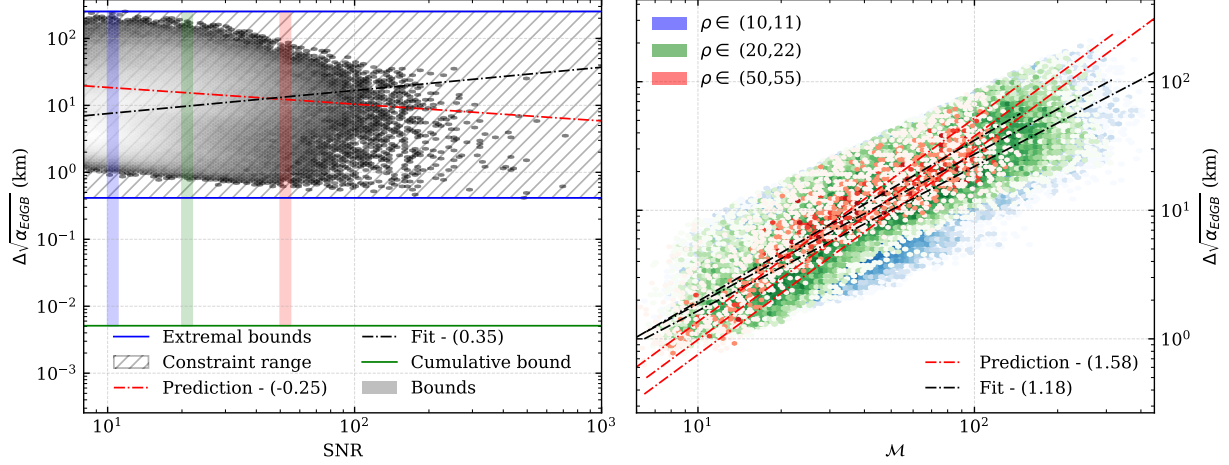


Figure 3.21: Result of the scaling analysis outlined in Sec. 3.6.2 performed on the data synthesized with the HLVKIO8 network and the SPOPS 0 population. The plotting style is the same as in Fig. 3.12. The left panel shows a heat map of the constraint on $\sqrt{\alpha_{\text{EdGB}}}$ versus the SNR of the source. The right panel shows the density of the constraint versus the chirp mass, with empirical trends shown in black and predicted trends shown in red. Because of the small range in single-observation constraints (about 1–2 orders of magnitude), the cumulative bound greatly benefits from large numbers of observations, despite this being a negative PN effect that would typically be dominated by a small cadre of favorable sources. The predicted trend for the constraint-SNR relationship fails, presumably due to covariances introduced through the Jacobian. The predicted trend for the constraint- $\mathcal{M}$ relationships performs fairly well, as the correlation is enhanced through the Jacobian.

–1PN order, because the dominant modification to the generation of GWs is the introduction of dipolar radiation. The Jacobian for this particular theory is

$$\left(\frac{\partial \beta}{\partial \alpha_{\text{EdGB}}^2} \right)^2 \propto \frac{[\hat{m}_2^2 s_1^{\text{EdGB}} - \hat{m}_1^2 s_2^{\text{EdGB}}]^4 \eta^{12/5}}{(1+z)^{-8} \mathcal{M}^8}, \quad (3.46)$$

where s_i^{EdGB} is the BH sensitivity defined in Eq. (B.4), and we again use the mass parameters $\hat{m}_i = m_i/\mathcal{M} = \eta^{-3/5}(1 \pm \sqrt{1-4\eta})/2$ for the larger (+) and smaller (–) mass. Given the new dependencies on source parameters introduced by the Jacobian, we would expect to see SOBH sources receive a sizeable boost due to the chirp mass scaling. Furthermore, this is a negative PN effect, which already tends to favor small chirp masses (cf. Sec. 3.6.1). Both of these considerations imply that multiband and terrestrial networks should outperform LISA MBH sources.

Constraints on $\sqrt{\alpha_{\text{EdGB}}}$ are shown in Fig. 3.20. Indeed, we see SOBH sources of all kinds outperforming MBH sources. Within the SOBH source classes, terrestrial networks outperform multiband sources by 1–2 orders of magnitude. While multiband sources benefit from long early inspiral observations from LISA, which encodes much information for a negative PN effect, the large catalogs of sources in the terrestrial-only catalogs are enhanced by the modified dependence on the SNR, discussed below. As a further consequence

of the adjusted SNR dependence, we also see fairly minor variations between the three terrestrial network scenarios. After approximately thirty years of observations, our models indicate that we could see ~ 2 –4 orders of magnitude improvement on previous constraints on $\sqrt{\alpha_{\text{EdGB}}}$. This conclusion is fairly robust under variations of the terrestrial network.

Analyzing the constraints on $\sqrt{\alpha_{\text{EdGB}}}$ with the machinery of Sec. 3.6.1, we obtain the following approximation on the variance of the coupling parameter:

$$\Delta\sqrt{\alpha_{\text{EdGB}}} \approx \left(\frac{903168}{25\pi^6}\right)^{1/8} \frac{(\pi\mathcal{M}f_{\text{low}})^{7/12} \mathcal{M}}{(1+z)\eta^{3/10}\rho^{1/4}} \times (\hat{m}_2^2 s_1^{\text{EdGB}} - \hat{m}_1^2 s_2^{\text{EdGB}})^{-1/2}. \quad (3.47)$$

We now see additional modifications to the dependencies on source parameters, beyond the Jacobian shown above. Just as in the cases of dCS and non-commutative gravity, we must transform from α_{EdGB}^2 to $\sqrt{\alpha_{\text{EdGB}}}$, which forces the constraint to scale with $\rho^{-1/4}$ and $\mathcal{M}^{19/12}$.

Trends related to this approximation are shown in Fig. 3.21, produced from our simulations based on HLVKIO8 and SPOPS 0. The left panel shows a heat map of all the sources in the $\Delta\sqrt{\alpha_{\text{EdGB}}}$ -SNR plane, with extremal single-source constraints shown in blue, and the cumulative constraint for this catalog shown in green. The right panel shows a heat map in the $\Delta\sqrt{\alpha_{\text{EdGB}}}$ - $\mathcal{M}$ plane, for three different slices of SNR, shown as blue, green, and red.

In the left panel, we again see that our prediction for the SNR scaling is not accurate. Just as in dCS gravity, this discrepancy lies in covariances complicating the relationships beyond the point where our simple approximations are valid. For comparison, we can examine what we found for generic dipole radiation constraints in Sec. 3.6.2, where we saw a much better agreement with our predictions for the constraint-SNR relationship. Referring again to the histogram in Fig. 3.19, we see that the terms related to the BH sensitivity in EdGB span several decades, washing out the trends we would expect to see from the analysis of Sec. 3.6.1. As a by product, these complications lead to a tight range in single-observation constraints, spanning 1–2 orders of magnitude. This in turn leads to a large enhancement for terrestrial networks: cumulative bounds from tightly grouped populations of constraints benefit from large numbers of sources, which is not typically expected from a modification at -1PN .

In the right panel, we see moderate agreement between our prediction for the $\Delta\sqrt{\alpha_{\text{EdGB}}}$ - $\mathcal{M}$ relationship, but again, covariances seem to degrade the quality of simple analytical scaling relationships between the constraint and the source parameters. In contrast, for generic dipole radiation we see a much tighter correlation between the constraint and the chirp mass. The difference between the two trends further confirms

our explanation: more complex Jacobians tend to complicate the source parameter-constraint relation we identified in Sec. 3.6.1.

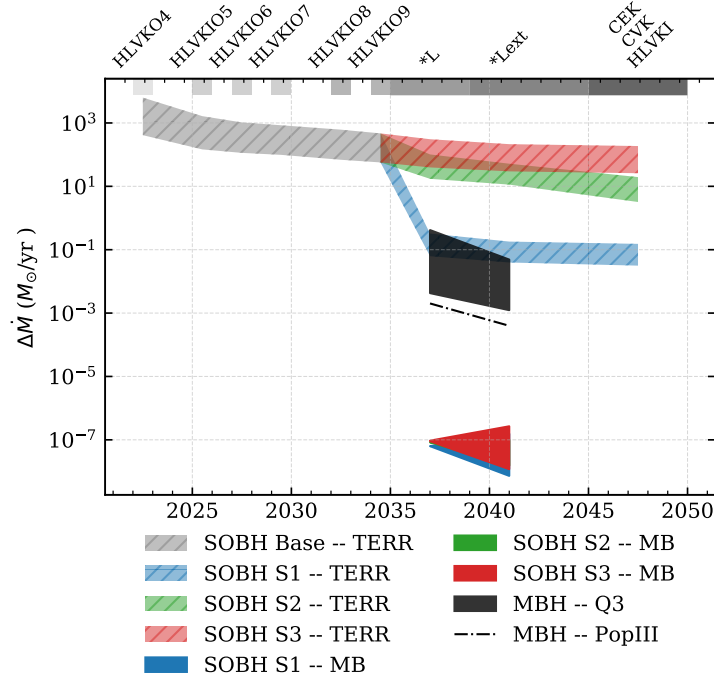


Figure 3.22: Projected cumulative constraints on the rate of black hole evaporation $\dot{M}$, for the detector networks and population models examined in this paper. Our models predict multiband sources to perform the best from the three classes of sources examined in this paper, followed next by MBH observations by LISA. Terrestrial-only observations from the most optimistic scenario are competitive with LISA’s MBH sources, but the other two scenarios considered in this work trail behind by 2-3 orders of magnitude.

Black Hole Evaporation

In the case of BH evaporation, the modification first enters the GW phase at -4PN order. The Jacobian from the ppE parameter to this particular process, as shown in Appendix B, is given by

$$\left(\frac{\partial\beta}{\partial\dot{M}}\right)^2 \propto \left[\frac{3 - 26\eta + 34\eta^2}{\eta^{2/5}(1 - 2\eta)}\right]^2. \quad (3.48)$$

As the Jacobian only depends on the system parameters through the symmetric mass ratio (bounded to $(0, 0.25]$), no parameters specific to a given system will induce large changes in the attainable bound. This fact leads us to the conclusion that the driving factors in the constraint magnitude will be the chirp mass (benefitting SOBH sources) and the SNR (benefitting LISA MBH sources and the most sensitive ground-based detector networks). Furthermore, as this modification also enters at a highly negative PN order, multiband

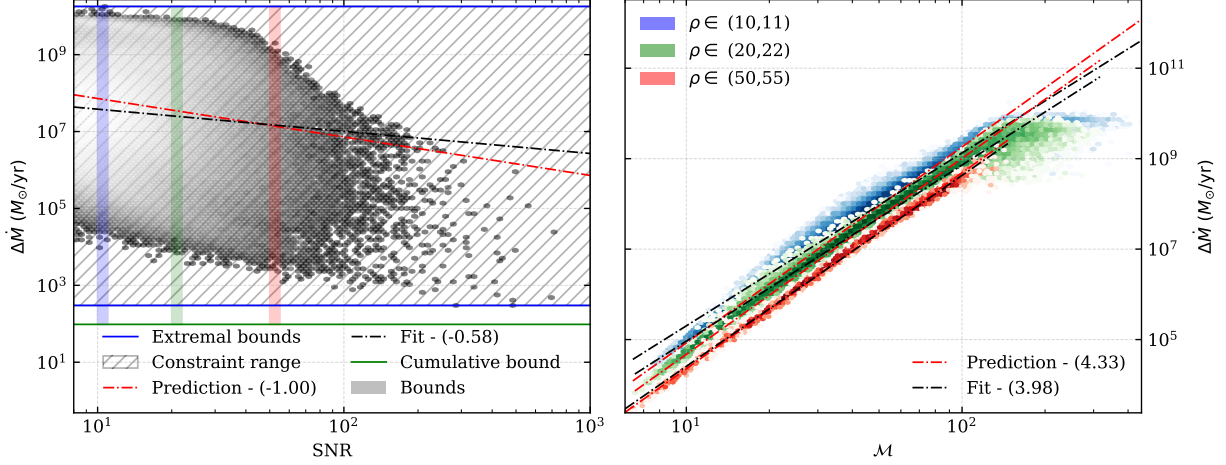


Figure 3.23: Result of the scaling analysis outlined in Sec. 3.6.2 performed on the data synthesized with the HLVKIO8 network and the SPOPS 0 population. The plotting style is the same as in Fig. 3.12. The left panel shows a heat map of the constraint on $\dot{M}$ versus the SNR of the source. The right panel shows the density of the constraint versus the chirp mass, with empirical trends shown in black and predicted trends shown in red. The wide distribution of constraints in this catalog indicate that the benefit of large catalogs is minimal, and the total bound is dominated by a select few, highly favorable observations. The distribution of the sources in the $\Delta\dot{M}$ - M plane is to a very good approximation linear, showing a tight correlation between the two quantities. The $\Delta\dot{M}$ -SNR relationship also agrees fairly well with our predictions.

sources can also be expected to perform competitively.

Constraints on the rate of BH evaporation are shown in Fig. 3.22. As expected, multiband sources constrain BH evaporation the tightest, with MBH sources from LISA’s catalog trailing by 4-6 orders of magnitude. The most sensitive terrestrial network scenario examined in this paper is also competitive with the LISA MBH sources, but the other two scenarios we have considered fall behind by 2-3 orders of magnitude.

By using the machinery of Sec. 3.6.1, we obtain the following approximate form of the bound on $\dot{M}$:

$$\Delta\dot{M} \approx \frac{425984}{5} \sqrt{\frac{6}{5}} \frac{(f_{\text{low}} \pi M)^{13/3} \eta^{2/5}}{\rho} \left| \frac{1 - 2\eta}{3 - 26\eta + 34\eta^2} \right|. \quad (3.49)$$

The Jacobian does not depend on the total mass and the phase modification scales linearly with the modifying parameter, so we see a scaling relation as expected from Sec. 3.6.1.

Results related to this approximation are shown in Fig. 3.23. The left panel depicts a heat map of the sources in the HLVKIO8 network and the SPOPS 0 population model in the $\Delta\dot{M}$ -SNR plane. The solid blue lines correspond to the strongest and weakest constraints coming from single observations, while the green line represents the cumulative bound for the entire catalog. The right panel shows a heat map in the $\Delta\dot{M}$ - M plane for different slices of SNR (in red, blue, and green). The empirically determined scaling trends are shown in black, while our predictions for the trends are shown in red.

The left panel of Fig. 3.23 shows good agreement between the trends predicted by our simple, analytic calculations and the data from our fully numerical treatment. The wide distribution in constraints coming from single sources in the catalog indicates weak scaling with the size of the catalog, giving a relative boost in power to the smaller source populations in the MBH LISA and MB catalogs. This conclusion is supported by the very modest improvement of the cumulative bound for the catalog over the strongest single-source constraint. In the right panel, we see good agreement with our predicted chirp mass scaling relation. The correlation between the chirp mass and the constraint is quite tight for this particular modification, due to the strong scaling and the highly negative PN order (reducing correlations that widen the distribution).

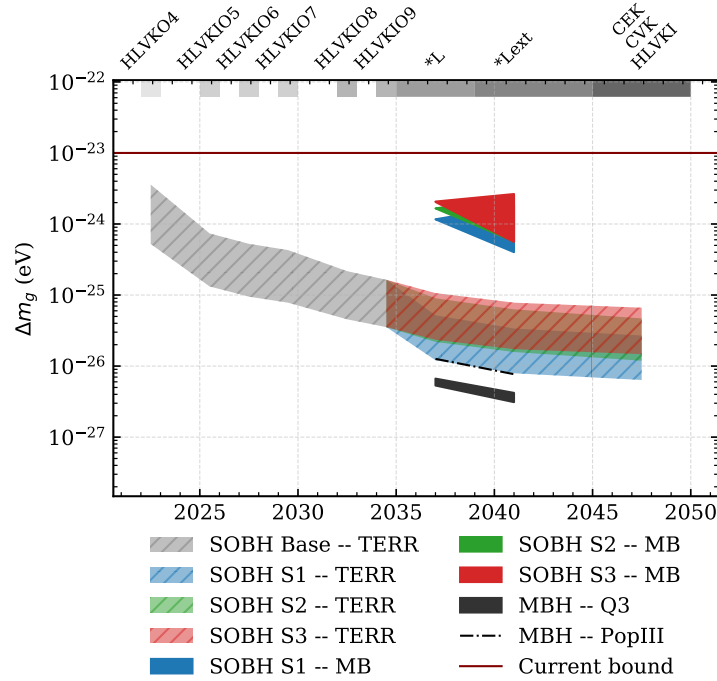


Figure 3.24: Projected cumulative constraints on the mass of the graviton, m_g , for the detector networks and population models examined in this paper. Our models show that MBH sources observed by LISA will perform the best at constraining this modification, but only slightly better than the terrestrially-observed only sources. Multiband sources perform the worst, as they received no benefits from the Jacobian and already perform only moderately well for positive PN order effects.

Modified Dispersion – Massive Graviton

If the graviton were massive, contrary to what is predicted when considering GR as the classical limit of a quantum theory of gravity, the leading order effect would enter the GW phase at 1PN. The Jacobian of the

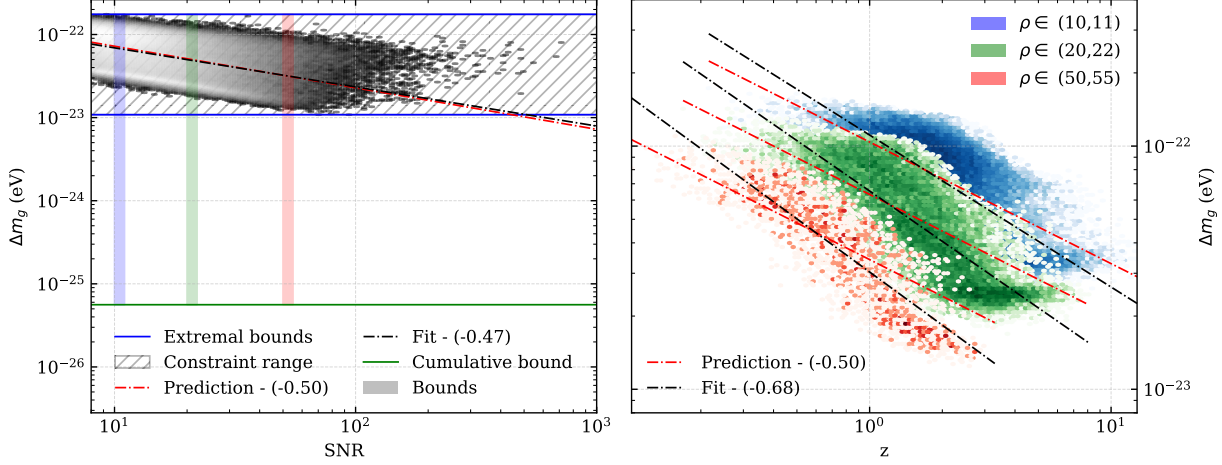


Figure 3.25: Result of the scaling analysis outlined in Sec. 3.6.2 performed on the data synthesized with the HLVKIO8 network and the SPOPS 0 population. The plotting style is the same as in Fig. 3.12. The left panel shows a heat map of the constraint on m_g versus the SNR of the source. The right panel shows the density of the constraint versus the redshift z , with empirical trends shown in black and predicted trends shown in red. Because of the narrow range of constraints in the catalog and the large enhancement of the cumulative bound over the strongest single observation, stacking observations is quite efficient for this modification. The right panel shows that there is indeed a trend in the Δm_g - z relation (although the distributions are moderately wide) which would favor sources far from Earth, and would primarily benefit MBH sources.

transformation from the ppE framework to this particular modification is

$$\left(\frac{\partial \beta}{\partial m_g^2}\right)^2 \propto \left(\frac{\mathcal{M} D_0}{1+z}\right)^2, \quad (3.50)$$

where the quantity D_0 is a new cosmological distance defined in Appendix B. We get modified scaling with the chirp mass, and similarly to the variable- G mapping, this Jacobian causes the constraint to inversely scale with the mass. As a result, this new mass factor will benefit MBHs over SOBHs. Furthermore, we now have strong dependence on the distance to the source, D_0 , where constraints from farther sources will be enhanced as compared to those sources closer to Earth (see e.g. [236]). These facts benefit LISA MBH sources, which therefore should provide the best constraints.

This is confirmed in Fig. 3.24. The MBH sources observed by LISA do indeed perform the best, but only marginally. The effectiveness of stacking is seen to still be quite high for this particular modification, as the three terrestrial scenarios all perform comparably. Furthermore, as this is a positive PN effect, terrestrial networks receive a boost from the generic scaling effects discussed in Sec. 3.6.1. Multiband sources perform the worst, as they receive little benefit from early inspiral observation, they typically have low mass, and are located at low redshifts. Ultimately, we can expect to improve on the current bound on m_g by 2–3 orders of magnitude over the next thirty years, and this conclusion is robust under variations of the terrestrial detector

schedule. This improvement will be insufficient to rule out a massive graviton as a possible explanation of the late-time acceleration of the Universe: in a cosmological context, the graviton would need a mass of the order of the inverse of the Hubble constant, H_0^{-1} , which is of the order of 10^{-30} eV, much smaller than our predicted final constraints.

To explore these relations deeper, we can apply our approximation from Sec. 3.6.1, giving us the following approximation for the constraint on m_g :

$$\Delta m_g \approx \frac{h}{\pi} \left(\frac{5}{2}\right)^{1/4} \sqrt{\frac{(1+z) \pi f_{\text{low}}}{D_0 \rho}}. \quad (3.51)$$

This approximation has produced a notably different scaling relation than what has been seen previously. Namely, the constraint no longer scales with the chirp mass, as the Jacobian factor has cancelled the chirp mass dependence from the generic ppE scaling. While this final form of the constraint does not explicitly benefit MBH systems, generic constraints scale with the chirp mass as $\mathcal{M}$. The removal of this chirp mass dependence benefits MBH sources much more than SOBH sources. Also different from previous constraints, we have strong scaling with the distance to the source. For low redshifts, the distance parameter $D_0 \approx z H_0$ to lowest order in redshift. Extending this expansion to the constraint, the leading-order term should scale as $z^{-1/2}$ for low-redshift sources.

The results related to this approximation are shown in Fig. 3.25. The left panel shows a heat map of the sources in the catalog created from the HLVKIO8 network and SPOPS 0 population model in the Δm_g -SNR plane, with the solid blue lines denoting the extremal, single observation constraints. The solid green line represents the cumulative bound from this particular catalog. We see good agreement between our predicted scaling for the SNR, after accounting for the Jacobian above. There is a narrow range for the constraints, only spanning one order of magnitude between all sources. This leads to sizeable benefits for large catalogs, also evident from the overlap between the different terrestrial network scenarios.

The right panel shows a heat map of the sources in the Δm_g -redshift plane. We do indeed see a trend in this particular relationship, although the distributions are moderately wide. Our predictions for the scaling relation agrees fairly well with the synthetic data.

3.6.3 Effect of Precession on the Constraints

The differences between the two SOBH population models go beyond the size of the catalogs, which has been our focus so far. An aspect differentiating the SPOPS 0 and SPOPS 265 catalogs, that could have a large impact on our analysis, is the typical magnitude of the in-plane component of the binary's spins, which is the

cause of relativistic precession. The question we now address is whether the stronger constraints coming from the SPOPS 0 catalog over the SPOPS 265 catalog are entirely due to the larger catalog sizes, or if the difference in source parameter distributions also impacts the cumulative bounds attainable through GWs.

Previous work has shown that the inclusion of precessional effects can break degeneracies in various source parameters when considering a full MCMC analysis, allowing for significantly tighter constraints on various source properties [261]. To determine if this effect can be seen in our data, in Fig. 3.26 we show histograms of the individual source constraints on dCS and EdGB, using the two different catalogs (SPOPS 0 and SPOPS 265) and the CEK network. These two theories in particular were chosen because conventional thinking would suggest that they would be the most sensitive to precessional effects, due to the dependence of the ppE parameter on spins.

The figure shows little deviation between the two population models for these theories. The distribution changes slightly on the larger-constraint side of the histogram, but the difference is negligible when considering cumulative constraints. Furthermore, these small deviations in the distributions of constraints cannot be solely attributable to precessional effects, as the parameter distributions shown in Fig. 3.4 are all modified as well.

To explore the impact of precession on generic modifications in a more controlled environment, we did a direct comparison between systems with zero precession and “maximal” precession (in a sense to be defined shortly), but which are otherwise identical. The results of this analysis are shown in Fig. 3.27. The methodology we implemented to produce Fig. 3.27 began with a set grid in the total mass, ranging from $5 M_\odot$ to $20 M_\odot$, mass ratio in the range $[0.05, 1]$, and aligned-spin components for each BH ranging from -0.8 to 0.8 . With this grid of intrinsic source parameters, we populated the other extrinsic parameters using randomly generated numbers in the conventional ranges. The range on the luminosity distance was chosen such that the SNRs would range from ~ 20 to 150 . Once a set of full parameter vectors had been created, we calculated one set of Fishers for a fixed detector network with the in-plane component of the spin, χ_p , set to 0. Then, without changing any other parameters, the in-plane spin component was increased to $\chi_p = \sqrt{1 - \chi_1^2}$, which is approximately the maximal spin one can achieve while still maintaining a total spin magnitude less than 1. The top panel shows the mean constraint for both configurations as a solid line, with the 1σ interval of the distribution of constraints shown as the shaded region. In the bottom panel we compare the constraints from each configuration (precessing and non-precessing) for each individual source. The mean of this ratio is then plotted as a solid line, and the 1σ region is shown as the shaded region.

The conclusion from Fig. 3.27 is that precession seems to have a moderate influence, but one that could be easily washed out by other physical effects. In the most favorable scenario where the binary is maximally

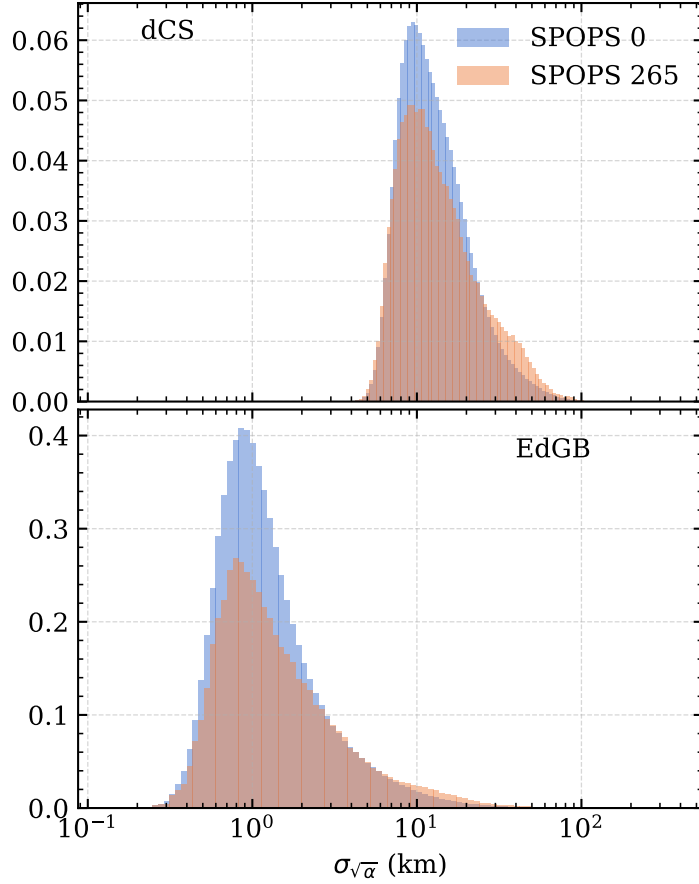


Figure 3.26: Distributions of single-source constraints on the GR-modifying parameters $\sqrt{\alpha}_{\text{dCS}}$ (top) and $\sqrt{\alpha}_{\text{EdGB}}$ (bottom) from the two population models SPOPS 0 (blue) and SPOPS 265 (orange) as detected by the CEK network. The histograms are normalized to provide a comparison of the shapes of the distributions, as opposed to the raw numbers of sources. We see that the distributions only diverge slightly, towards the larger-constraint side of the spectrum. This suggests that the larger precessional effects seen in the SPOPS 265 catalog do not significantly modify the typical constraints attainable by individual sources, or that any effect we may have seen was washed out by the differences in the distributions of other source parameters, such as the total mass and mass ratio. This lack of difference could also be an artifact of our waveform model (**IMRPhenomPv2**), which is not the most up-to-date waveform available, or of the Fisher approximation, which could be improved upon by a full MCMC analysis.

precessing, our analysis suggests an improvement of at most a factor of ~ 2 . Given previous work (see e.g. [261]), one may expect more significant improvements when considering even mild precession. While we do predict improvements from the use of precessing templates, our more restrained conclusions could be the result of two facets of our analysis. Our use of a more rudimentary statistical model, the Fisher matrix, does not capture all the more nuanced artifacts in the posterior surface, like a full MCMC analysis would. Furthermore, we here use the **IMRPhenomPv2** waveform, which is in some ways more limited in modeling precession with respect to the waveforms used in Ref. [261]. Future studies of precession could focus on these

two areas in particular.

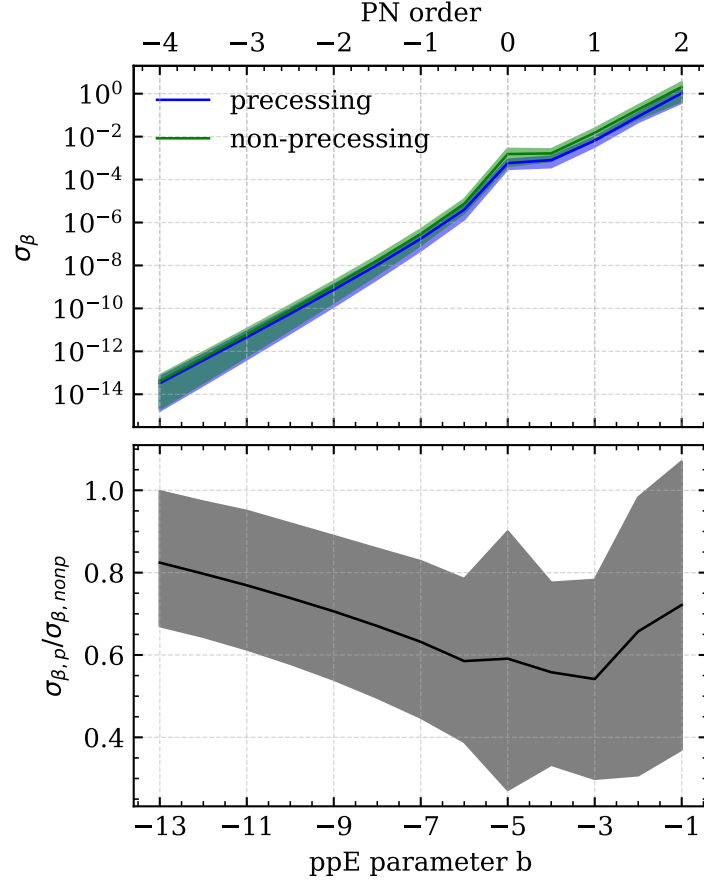


Figure 3.27: To create the data involved in this figure, we have created a set grid in parameter space with total mass ranging from $5 M_\odot$ to $20 M_\odot$, mass ratio in the range 0.05 to 1, and aligned-spin components for each binary ranging from -0.8 to 0.8 . The rest of the parameters were populated with random numbers in the usual ranges, and the luminosity distance was set such that the typical SNRs ranged from ~ 20 – 150 . We computed Fisher matrices for each set of parameters, with the in-plane component of the spin set to 0, and then we recomputed them setting the in-plane component of the spin $\chi_p = \sqrt{1 - \chi_1^2}$, so that the binary is approximately “maximally” precessing. The top panel shows the distribution of the bounds for the two binary subsets – precessing (blue) and non-precessing (green) – as a function of PN order. The solid line denotes the average of the synthetic catalog, while the shaded region denotes the 1σ interval. The lower panel shows the ratio $\sigma_{\beta,p}/\sigma_{\beta,nonp}$. Each ratio is calculated for a single parameter set, and the mean of these ratios is shown as a solid black line, with the 1σ spread shown by the shading. Even in this more extreme comparison, the improvement in constraint as the result of larger precession effects only amounts to a factor of ~ 2 . However, more drastic difference may be possible if we performed a full MCMC analysis, or if we used different waveform models.

3.7 Conclusions

In this work, we have constructed forecasts of what constraints can be placed on a variety of modifications to GR, both generic and theory-specific, using astrophysical population models and the most current projections for detector development over the next thirty years. Our analysis spans several topics of interest to the GW community concerned with tests of GR.

We investigate what fundamental physics can be done with a variety of source populations (heavy-seed MBHs, light-seed MBHs, terrestrially observed SOBHs, and multiband SOBHs) and plans for detector development. All of these aspects are connected to what fundamental science is achievable. Ours is the first robust study of this breadth and scope that is capable of quantifying the effects of detector development choices and astrophysical uncertainties.

We identify trends and scaling relationships of constraints for individual GW observations, studying how they evolve with PN order and how they depend on the target source class (MBHs, terrestrially observed SOBHs, and multiband SOBHs). We also quantify the effect of combining constraints from a full, synthetic catalog, appropriately informed by robust population models. We find that the effectiveness of stacking observations is a PN-dependent conclusion. The techniques developed here have important implications for the future of GW-based tests of GR, especially in the era of 3g detectors. The two components of our analysis (individual scaling and studies of the stacking of multiple observations) combine to create a full picture of some of the most important aspects involved in testing GR with GWs. We hope that this information will be valuable in driving design choices for future detector development.

We map our generic constraints to theory-specific constraints, where we analyze specific parameters in viable, interesting theories. Repeating some of the scaling analysis done in previous sections leads, in some cases, to a reversal of the conclusions drawn for generic modifications. This reinforces the need to incorporate theory-specific waveforms in future analyses, when available.

This work opens up several new avenues of research. We focused on BBH systems, neglecting future contributions from neutron star-neutron star and neutron star-BH binaries. These binaries have much longer inspiral signals relative to typical BH mergers observed by the LVC, and they could provide crucial information concerning early inspiral, negative PN effects. Beyond the signal length, neutron stars are sometimes treated on unequal footing in the context of specific theories, such as scalar-tensor gravity, EdGB and dCS. This could provide other insights into specific theories that do not affect BBH mergers.

Because of the scale of the catalogs involved we used simple Fisher matrix forecasts, running $\sim 10^8$ Fisher matrix calculations. A more thorough analysis using MCMC, or other more robust data analysis techniques, could provide more information about some of the trends we have identified. An MCMC population study on

the scale of this work is currently intractable, but even an analysis of a subset of sources could be enlightening.

Our work has focused on estimating only a single PN modification at a time, but any modified theory of gravity will correct the waveform at all orders in a PN expansion. Recent work studied how constraints are affected when one attempts to simultaneously constrain ppE deformations that enter at multiple orders [136]. Here we have chosen to limit ourselves to a single parameter at a time for the following reasons. While allowing for multiple parameters to vary in a completely independent way at several PN orders is a more robust and general framework, this treatment is probably overly pessimistic. Past work [116, 119] showed that, indeed, varying multiple generic parameters simultaneously drastically lowers our ability to constrain them. However, in the context of a given, physically motivated theory there should be some relation between the different ppE modifications. Any PN expansion should converge in the appropriate domains, ensuring a hierarchy on the size of the modifications. Moreover, the modification at each PN order should at least depend on the coupling parameters of the theory, ensuring that no two PN orders are totally independent from each other. These criteria suggest that the overall bound on a given modification, in the context of a physically motivated theory, should not be significantly weakened by the inclusion of higher-order corrections (except in the most unfortunate of fine-tuning scenarios). Therefore our conclusions should be robust under the inclusion of higher-order PN corrections to the waveform.

Our investigation of the effects of precession on modified GR constraints could be improved in at least three ways. While we did include a full inspiral/merger/ringdown model of precession by implementing IMRPhenomPv2 [106, 169, 202], more recent and complex waveform models (such as IMRPhenomPv3 [262], IMRPhenomXPHM [263] or SEOBNRv4PHM [264]) could encode more information in the signal, helping to break degeneracies. A more robust statistical analysis, such as a full MCMC, could explore the posterior space more thoroughly, shedding light on the effects of precession. Last but not least, the astrophysical SOBH models considered here only allow for isolated field formation under restrictive assumptions. Dynamical formation generally predicts a larger fraction of precessing systems [265], and it is important to consider other pathways for producing BBHs with large misaligned spins even within the isolated formation channel [237, 266].

Acknowledgments We thank Vishal Baibhav, Davide Gerosa, Gabriela González, Bangalore Sathyaprakash, Sashwat Tanay and Kaze Wong for many useful discussions on various aspects of this work. N.Y. acknowledges support from NSF Grants No. PHY-1759615, PHY-1949838 and NASA ATP Grant No. 17-ATP17-0225. E.B. is supported by NSF Grants No. PHY-1912550 and AST-2006538, NASA ATP Grants No. 17-ATP17-0225 and 19-ATP19-0051, and NSF-XSEDE Grant No. PHY-090003. This work has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 690904. The authors would like to acknowledge networking support by the GWverse COST

Action CA16104, “Black holes, gravitational waves and fundamental physics.” This work made use of the Illinois Campus Cluster, a computing resource that is operated by the Illinois Campus Cluster Program (ICCP) in conjunction with the National Center for Supercomputing Applications (NCSA) and which is supported by funds from the University of Illinois at Urbana-Champaign. It also used computational resources at the Maryland Advanced Research Computing Center (MARCC). The following software libraries were used at various stages in the analysis for this work, in addition to the packages explicitly mentioned above: `GSL` [1], `numpy`, `scipy`, `filltex` [4].

Chapter 4

Are Parametrized Tests of General Relativity with Gravitational Waves Robust to Unknown Higher Post-Newtonian Order Effects?

This chapter is based on: Perkins, S.¹ and Yunes, N.² (2022). *Are Parametrized Tests of General Relativity with Gravitational Waves Robust to Unknown Higher Post-Newtonian Order Effects?*. Submitted to *Phys. Rev. D.*, *arXiv:2201.02542*

Rev. D., arXiv:2201.02542

Abstract Gravitational wave observations have great potential to reveal new information about the fundamental nature of gravity, but extracting that information can be difficult. One popular technique is the parametrized inspiral test of general relativity (a realization of the parametrized post-Einsteinian framework), where the gravitational waveform, as calculated in Einstein’s theory as a series expansion in the orbital velocity, is parametrically deformed at a given set of orders in velocity. However, most current approaches usually only analyze the data while considering a single, specific modification at a time. Are then constraints placed with a single modification robust to our ignorance of higher post-Newtonian order corrections? We show here that for a wide class of theories, specifically those that admit a post-Newtonian expansion, single-parameter tests are indeed robust. In particular, through a series of full Bayesian parameter estimation studies on several different sets of synthetic data, we show that single-parameter constraints are not degraded but rather are improved by the inclusion of multiple parameters, provided one includes information about the mathematical structure of the series. We then exemplify this with a specific theory of gravity, shift-symmetric scalar Gauss-Bonnet theory, where the waveform has been calculated to higher post-Newtonian orders than leading. We show that the inclusion of these higher order terms strengthens single-parameter constraints, instead of weakening them, and that the strengthening is very mild. This analysis therefore provides strong evidence that single-parameter post-Einsteinian tests of general relativity are robust to ignorance of high post-Newtonian order terms in the general relativistic deformations.

¹Carried out the calculations in the paper; wrote the MCMC software; wrote the software related to waveform modelling; conducted the numerical experimentation; wrote the first draft of the manuscript

²Conceived of the project; supervised and advised on the calculations and experimentation; carried out parts of PN calculations in collaboration with first author; edited the manuscript

4.1 Introduction

As the field of gravitational wave (GW) astrophysics matures, the growing amount of data is being used in a variety of ways and in different contexts to solve problems in topics like fundamental physics, astrophysics, and cosmology. One critical subject that has drastically benefited by the observation of GWs is the study of the gravitational interaction. General relativity (GR), our best theory of gravity to date, has some shortcomings, if we may call them that [51, 52]. These include the incompatibility of GR with quantum mechanics (c.f. Ref. [84]) and the singularity problem [267, 268], as examples. Aside from these shortcomings, the theory also struggles to adequately explain certain phenomena observed in our universe without the inclusion of additional structure, like the late time acceleration of the universe (unless one includes an “unnaturally” small cosmological constant) [77, 78], the rotation curves of galaxies (unless one includes dark matter particles that have so far been undetected via direct experiment) [64, 65], and the matter-antimatter asymmetry of the universe (unless one includes new parity-violating interactions that satisfy the Sakharov conditions) [269]. Modifications to GR in the ultraviolet or in the infrared may potentially resolve some of these theoretical and observational issues, which is why they have attracted attention lately in light of new GW data.

If we are searching for or attempting to constrain “alternatives” to GR, GW observations are an excellent place to look [112, 270]. The dynamics of the system that generates them are highly relativistic (reaching characteristic speeds up to a substantial fraction of the speed of light), and involve large gravitational potentials and fields. However, deducing information about fundamental physics from GW data is a nuanced business, where the information encoded inside the waveform can, in some cases, be very faint and buried in noise. One possible philosophy is to take a maximally “ignorant” approach, where one seeks to involve as little outside information as possible. This approach has its appeal in terms of its perceived robustness (uninfluenced by any external, *a priori* information), but the exact method with which one encodes this “ignorance” is not a trivial or well defined matter. Furthermore, a maximally uninformed approach can minimize and sometimes even erase the strength of any inferences about fundamental physics one might hope to extract [61]. While it would be appealing to conduct these analyses independently of our priors, this seems to be a quixotic approach.

For example, one maximally uninformed approach would be to reconstruct the signal with an orthogonal basis of wavelets, as is normally done with `BayesWave` [90, 271]. The fact that this method is totally agnostic to what physics produced the signal allows one to reconstruct the signal to almost arbitrary precision, which is useful in many contexts. However, such an approach usually provides little information about the theory of gravity that led to the signal that is being reconstructed, or more specifically about the fundamental physics involved in the generation and propagation of GW in our universe (although there have been recent attempts

to tie in to certain topics in fundamental physics [272]). Given this, it’s useful to remember that general tests like “residual tests,” in which one studies the signal-to-noise ratio contained in the difference between the data and the best re-constructed signal [115, 273], are inherently a consistency check of the reconstruction procedure, and should not be interpreted as direct tests of the underlying physics. The value of such tests is only as great as the accuracy of the underlying reconstruction models, regardless of what physics those models are trying to represent.

Meanwhile, a less uninformative approach, though still agnostic, might attempt to model the signal as a GW produced in GR, but with some *small deformation* that is represented parametrically. This type of approach defines *the parametrized post-Einsteinian (ppE) formalism* [107, 115–117], which is the general framework that the “parametrized inspiral tests of GR,” used by the LIGO/Virgo collaboration (LVC), is built from [57–61, 118–122]. In this approach, deformations from GR are incorporated directly at the level of the model for the signal (i.e. the GW template or GW model) that one will use to compare against the data. Moreover, these deformations are represented parametrically through a particular basis (i.e. in the inspiral, a polynomial in orbital velocity), with the polynomial exponent of the leading post-Newtonian (PN) order term encoding the type of modification one is considering, and the polynomial coefficient the strength of the deformation.

The resulting functional form of the deformation of the signal is not arbitrary. In fact, one can easily show that such a representation derives from parametrically deforming the binding energy of the binary, the rate of change of this binding energy, or the dispersion relation of the propagating GW through a polynomial in velocity [107]. Such an analysis also reveals that there is a *mapping* between deformations in these physical quantities and those that are introduced in the waveform, thus allowing for constraints on the waveform deformations to then be *mapped back* to constraints on theoretical physics [107, 109, 111, 112]. Moreover, a polynomial in velocity is a natural basis for expansions in the inspiral phase, where the PN expansion is expected to hold to approximate the solution to the field equations. Obviously, such a representation need not apply in the merger, and would have to be changed in the ringdown, as discussed extensively in [107], but we are here focusing on inspiral tests only.

The initial and simplest ppE proposal was to include a single deformation at a time [107]. As explained above, the deformation is represented by adding a term of the form βv^b in the GW Fourier phase (and a similar term in the Fourier amplitude), where β is a real parameter one is attempting to estimate, v is the orbital velocity of the binary (which is connected to the GW frequency via Kepler’s third law), and b is a fixed constant. One attempts to measure or constrain β (sometimes called the ppE amplitude coefficient) because it determines the strength of the GR deformation, and it is connected to the coupling constant of

modified GR theories. The quantity b (sometimes called the ppE exponent coefficient) is not a parameter, but rather it is a real number that determines the type of modification one is considering; therefore, one chooses a value of b *a priori*, before carrying out parameter estimation or model selection. By picking a set of such b constants, one can then derive posterior probability distributions for each β (corresponding to each value of b that one chose), to then construct the so-called “violin plots” the LVC generates in their testing GR papers [57, 58, 60, 61].

However, as explained when the ppE framework was first developed [107], deformations to GR in the inspiral should not consist of a single PN modification. Rather, one expects modifications to enter at some leading PN order (a so-called “controlling factor” of an asymptotic expansion), multiplied by an entire PN series in velocity, i.e. $\beta_0 v^b(1 + \sum_{i=n} \beta_n v^n)$. One way to incorporate this idea into an inspiral test in GR would be to truncate the series at some PN order, allow every β_n coefficient to be independent, and attempt to estimate all of them simultaneously [119, 135]. As argued in [116], and then verified in [61], this approach is doomed to fail because of large covariances between the β_n parameters that prevents one from estimating any one of them with any precision (i.e. the marginalized posteriors of β_n that one recovers are very similar to the priors one chooses for them). If allowing for multiple, independent deformation terms at once erases all useful information one may glean from the GW signal, are constraints one obtains with a single deformation at a time robust?

At first sight, one may think the answer to this question is no. After all, we know that a modified gravity signal will not contain just a single modification in the phase. Therefore, forcing a template to be of this form ought to introduce uncontrolled systematic error. Moreover, the velocities at merger are close to 0.3–0.5 the speed of light, so PN terms at higher-than-leading order need not be negligible. Therefore, by neglecting these higher order terms one could be ignoring “degeneracies” that could deteriorate constraints if they were properly modeled. If this were the case, single-parameter ppE tests or single-parameter parametrized inspiral tests could over-confidently predict and place constraints that are more stringent than what we have any right in claiming.

We are of course not the first to study the concerns mentioned above. To our knowledge, one of the first studies to do this for aLIGO observations appeared in an Appendix [112], where the authors investigated how constraints on Brans-Dicke theory are improved as higher PN order terms are added to the inspiral phase in the extreme mass-ratio limit. The authors found that constraints on the Brans-Dicke coupling parameter are very stable to the inclusion of these higher PN order terms, with relative fractional changes of at most a few tens of percent. A more recent study of this same topic was by [123], where constraints on a specific theory of gravity, shift-symmetric scalar Gauss-Bonnet (ssGB), were studied with the O2 and O3 GW catalog. In

this work, the authors modeled the higher PN order terms as unknown, and then marginalized over them; they found that the inclusion of these higher order terms only changed the constraints again by only tens of percent.

In spite of these studies, there has not yet been a dedicated study of the concerns mentioned above, which is what motivated this paper. As we will show in detail here, and in agreement with the previous work described above, the pessimistic viewpoint expressed above regarding single- versus multiple-parameter tests of GR is completely unwarranted. For a multitude of reasons, which we will elaborate on in this work, constraints obtained by using single-parameter models are both robust and reliable. In particular, we will show that although systematic error is incurred when one uses a single-parameter ppE model, the error is small and tends to predict a *less* stringent constraint than what one would obtain if one included higher PN order corrections subject to a reasonable PN prior; in this sense, single-parameter ppE and parametrized inspiral tests are *conservative*. Moreover, we will show that the improvement one gains by including higher PN order terms in the GR deformation makes constraints only slightly more stringent (of at most one order of magnitude in the most extreme cases).

We arrive at these conclusions through a detailed and full Bayesian analysis of a multitude of synthetic signals. We first carry out a fully Bayesian Markov-chain Monte-Carlo (MCMC) parameter estimation study, in which we inject a synthetic GR signal in stationary and Gaussian noise, and then we extract it with single-parameter ppE models, as well as multi-parameter models. For the latter, we develop a new prior based on the theoretical foundations of PN theory [95, 274], which ensures that the modifications one introduces lead to a non-GR PN series that has similar or better convergence properties than that of GR. If the modified theory of gravity one is considering accepts a PN expansion, then the prior we develop is guaranteed to be valid. It is the development and use of this prior that restricts the impact of the covariances found in previous studies [61, 119, 135].

We then examine our conclusions under the light of a specific example: constraints on ssGB theory. Higher-than-leading PN order waveforms have been recently calculated in this theory [275], thus allowing us to verify our prior. With this at hand, we carry out again a fully Bayesian test of GR, injecting synthetic GR signals in stationary and Gaussian noise, and extracting them with the new, high-PN-order ssGB model, the leading-PN-order ssGB model, a single-parameter ppE model, and a multiple-parameter ppE model with the PN-inspired prior developed above. In all cases, we map constraints on the parameters we search over to constraints on the coupling constant of ssGB gravity. We find that the the leading PN order ssGB model, the single-parameter ppE model, and the multiple-parameter ppE model with the PN-inspired prior all lead to roughly the same constraints, in agreement with previous work [123]. The new, high-PN-order ssGB model

leads to constraints that are stronger (not weaker) than the other constraint, but only by a factor of roughly 3 (at 90% confidence), because this model adds new physical information (at higher PN orders). Therefore, our analysis proves that, at least in this theory (and very likely in all theories of this type), higher-PN order corrections to the ssGB waveform improve constraints instead of deteriorating them, and the improvement is only very mild.

The conclusions we arrive at in this paper are admittedly strong, so they come with a couple of caveats. One of them is that the signal one analyzes is dominated by the inspiral (and not the merger) phase of the coalescence. If one can only hear the last few cycles of coalescence and the ringdown, then one is outside the regime of validity of the PN approximation, rendering some of the arguments presented above invalid. In practice, this implies that our conclusions apply only to binaries of sufficiently small total mass, with the maximum mass allowed dependent on the characteristics of the instrument (and in particular, the low frequency seismic wall of the noise spectral density of the detector); for the Advanced LIGO detectors at design sensitivity, the validity of our conclusions require that at least the total mass of the binary be less than roughly $40M_{\odot}$, so that at least the frequency of the innermost stable circular orbit (of the effective one-body problem) be above in the sensitivity bucket of aLIGO at 100 Hz. Another caveat is that we consider modifications to GR that lead to “persistent” effects in the inspiral, as opposed to modifications that turn on suddenly during the inspiral. The latter can occur in theories with additional length scales that lead to screening (like massive gravity theories [165]), or theories with additional fields that activate during the inspiral due to non-linear effects (like dynamical scalarization [276–279] or vectorization [280, 281]). As shown in [282], however, even for such theories a single-parameter ppE test is sufficient to detect such sudden deviations (at the cost of deteriorating the effectiveness of the test).

The rest of this paper will present the details of the calculations that have led us to the conclusions we described above, and it is structured as follows. We begin by discussing some of basics of Bayesian inference with GWs in Sec. 4.2.1. This is followed by a discussion of the current methodology for testing GR in the inspiral phase of GW binaries in Sec. 4.2.2. We continue with a discussion of a proposed improvement on those methods in Sec. 4.3, in which we outline a framework informed by the PN formalism. After this, we expand on the experimental design we use to test this new framework in Sec. 4.4. With the experiment summarized, we move on to discuss the results and implications of those experiments in Sec. 4.5. In Sec. 4.6, we investigate alternative parametrizations to determine the robustness of our conclusions. From here, we focus on a specific theory of gravity, ssGB, to determine how realistic our conclusions are in the context of actual modifications to GR in Sec. 4.7. Finally, we summarize our findings and propose future research in Sec. 4.8. Throughout this work, we will work in geometric units, where $G = 1 = c$.

4.2 Testing GR with GWs

Inferences about fundamental physics from GW data begins with matched filtering and Bayesian signal analysis, which we review in Sec. 4.2.1. We then move on to one of the currently implemented methodologies used to test GR with GW data (the LVC implementation of the ppE framework, which they dub a parametrized inspiral test) in Sec. 4.2.2.

4.2.1 Bayesian Analysis of GW Data

The foundation of any Bayesian analysis is, of course, Bayes' theorem. In the context of GW analysis, this can be written as

$$p(\boldsymbol{\theta}|D) = \frac{p(\boldsymbol{\theta})p(D|\boldsymbol{\theta})}{p(D)}, \quad (4.1)$$

where $p(\boldsymbol{\theta}|D)$ is the posterior probability of the source parameters $\boldsymbol{\theta} = \{\theta_1, \theta_2, \dots, \theta_N\}$ given some dataset D , $p(\boldsymbol{\theta})$ is the prior probability of $\boldsymbol{\theta}$, $p(D|\boldsymbol{\theta})$ is the likelihood that the data D is described by the model and parameter vector $\boldsymbol{\theta}$, and $p(D)$ is the evidence for the model. The posterior distribution is the quantity of interest in this study, as we are primarily interested in deducing the values additional, beyond-GR parameters can take while still being consistent with data observed in our Universe. The evidence can also be of interest when calculating Bayes' factors between models, but we will not be focused on model selection in this study. The information gained from the current experiment at hand is encoded in the likelihood distribution, while the beliefs about the different parameters held *before the experiment began* is embodied by the prior distribution.

In the case of inference with GW data, the likelihood function is calculated by matched-filtering. With this method, a template response function defined by a model and a set of parameters is compared against the data. This function can be succinctly written in the Fourier domain (assuming Gaussian, stationary noise) as

$$\log p(D|\boldsymbol{\theta}) = -\frac{1}{2} \sum_i^N (D_i - h_i | D_i - h_i), \quad (4.2)$$

for N detectors, where h_i is the response template of the i -detector given a GW metric perturbation (the perturbation template contracted on to the i -th detector $h_i = h_+ F_{+,i} + h_\times F_{\times,i}$) and the inner product is defined as

$$(A|B) = 4 \operatorname{Re} \int_0^\infty \frac{AB^*}{S(f)} df, \quad (4.3)$$

where $(*)$ denotes complex conjugation and $S(f)$ is the one-sided power spectral density of the noise in the detector.

As a general rule in Bayesian inference, one should be careful in how the prior information is specified. Either in the case of encoding information from past experiments or through mathematical intuition, or in how one attempts to encode one’s “ignorance” about the parameter in question, an overly restrictive prior can lead to biased posterior distributions. Standard choices used in GW science include a uniform prior in volume, a uniform prior on the component masses and the spin magnitudes, with a prior for the spin directions, sky location, and binary orientation uniform on the unit two-sphere. These choices of prior encode a certain amount of basic information through certain assumptions we have made about these parameters (e.g. that the universe is uniform and isotropic on large scales, or that the masses of compact objects cannot be negative), but are considered to be “uninformative.”

Ultimately, we seek to determine the form of the posterior defined in Eq. (4.1), because the degree to which the final marginalized posterior differs from the prior determines how much new information we have gained by performing the experiment. Once the posterior has been calculated, we can deduce many useful quantities, such as the confidence regions and maximum likelihood values for various parameters. However, without making any simplifying assumptions, this function is not analytic. Therefore, we are typically forced to explore these distributions through sampling techniques like Markov-Chain Monte-Carlo (MCMC) algorithms. Analyses such as these produce samples from the distribution of interest, which can then be binned into a histogram representation of the distribution.

Throughout this work, we will sample the likelihood, and therefore construct the posterior, through an MCMC exploration of the parameter space using software written for and implemented in previous works [123]. The Bayesian exploration of parameter space is based on the Metropolis-Hasting algorithm, and implements a variety of techniques to optimize the sampling efficiency, which we will briefly outline here. A critical component of the sampler is the use of parallel-tempering [283, 284], in which multiple MCMC chains are run in parallel. The first chain in the series samples from an unmodified posterior distribution, but all of the other chains in the series sample from a “tempered” distribution, in which the likelihood is raised to some power, $1/T_i$, where $T_i \geq 1$. The modified posterior is given by

$$p_i(\boldsymbol{\theta}|D) \propto p(\boldsymbol{\theta})p(D|\boldsymbol{\theta})^{1/T_i}, \quad (4.4)$$

for some “temperature” T_i . Periodically, the chains are allowed to exchange information by proposing a swap of the parameter coordinates of two of the chains. Each temperature ladder has a certain set of $\{T_i\}$ comprised of 15 (30) separate chains for the 2g (3g) injections (where the 2g and 3g injection parameters are discussed in Sec. 4.4). The temperatures run from $T_0 = 1$ (untempered) to $T_{14} \text{ (29)} \approx \infty$ (effectively sampling

the prior). To increase efficiency, we run 14 (5) copies of the above set of chains in parallel, so that we have multiple chains sampling at each temperature, T_i . All chains in the ensemble are allowed to swap with all other chains to improve convergence. Samples are then harvested from the $T = 1$ chains, which are sampling from the unmodified, target posterior. While this does introduce some correlations between the samples from the cold chains, we harvest more than enough samples to mitigate the impact of this, as discussed below. Furthermore, correlations of this type are analogous to those that would be present with other ensemble samplers, which base proposal distributions on the position of other chains in the ensemble.

In terms of the proposals we have implemented, we used two types: steps drawn from a random normal distribution centered at the current location and steps along the eigenvectors of the Fisher information matrix. We “burn-in” approximately 80,000 (28,000) samples per chain, during which we allow the sampler to optimize itself to the target distribution by tuning the step widths, continuously updating the Fisher matrix, and optimizing the temperature spacing of the various chains using the algorithm by Vouden et.al. [285]. After this initial burn-in phase, we then fix all the parameters of the sampler and continue to run for an additional 100 independent samples per untempered chain, as determined by the auto-correlation length, to make sure that indeed the chains have exited the burn-in phase. Then, we proceed to harvest samples that we will use to build the distribution, while thinning out by the autocorrelation length. We sample until we have harvested approximately 4,000 independent samples per untempered chain. This produces on the order of 56,000 (20,000) total, independent samples per 2g (3g) injections. With the samples drawn, we calculate the $1\text{-}\sigma$ confidence intervals by calculating the 16% and 84% quantiles from the data using `Numpy` routines [2], and taking the difference then dividing by 2.

4.2.2 Current Parametric Tests of GR

With the general framework outlined above, we can now discuss some of the current methodologies used in searches for physics beyond GR in GW data. In the course of this paper, we will always work with small deformations away from GR, consistent with our experience that GR describes gravity extremely well up to (and possibly beyond) the most extreme energy scales we have observed so far. In the spirit of this philosophy, we use a base GR waveform model (`IMRPhenomD` [106, 169]) and append modifications on top through a phenomenological framework. We could have used a more advanced base GR model, like `IMRPhenomPv3` [286], but doing so will probably not impact our conclusions qualitatively.

The LVC parametrized inspiral tests are a realization of the ppE framework [112], restricted to single-parameter deformations and deformations that enter at half-integer PN orders, and thus, there are two “standard” ways to implement them. In the first implementation, the modifications are added as fractional

deviations at each PN order directly to the GW phase. Explicitly, let us write the GW Fourier phase of the dominant mode as as [95, 106, 169]

$$\begin{aligned}\Psi_{\text{GW}}(f) = & \frac{3}{128\eta} v^{-5} \left(1 + \sum_{i=2}^7 p_i(\Xi) v^i \right) \\ & + \frac{3}{128\eta} v^{-5} \left(\sum_{i=5}^6 p_{i,l}(\Xi) \log(v^3) v^i \right),\end{aligned}\quad (4.5)$$

where $v \equiv (\pi m f)^{1/3}$, $m \equiv m_1 + m_2$ is the total mass for a binary with component masses m_1 and m_2 , and f is the GW frequency of the dominant mode, while $p_i \equiv p_{i,\text{GR}}(1 + \delta_i^{\bar{i}} \delta p_i)$ and $p_{i,l} \equiv p_{i,l,\text{GR}}(1 + \delta_i^{\bar{j}} \delta p_{i,l})$. The parameters $p_{i,\text{GR}}$ are the coefficients of the PN expansion in GR at the $(i/2)$ -th PN order, and $p_{i,l,\text{GR}}$ are the coefficients proportional to logarithmic terms; these coefficients are purely functions of the source parameters, Ξ , and not of frequency. The parameters δp_i and $\delta p_{i,l}$ are the non-GR deformation parameters that one attempts to estimate or constrain [118–122]. The constants $\bar{i}$ and $\bar{j}$ determine the PN order of the single deformation that is turned on, with the Kronecker delta guaranteeing that all other deformations are turned off. This is the parametrization that has been largely adopted by the LVC in their parametrized inspiral tests of GR (c.f. [57–61]).

A second implementation is to write the GW Fourier phase of the dominant mode as

$$\begin{aligned}\Psi_{\text{GW}}(f) = & \frac{3}{128\eta} v^{-5} \left(1 + \sum_{i=2}^7 p_{i,\text{GR}}(\Xi) v^i \right) \\ & + \frac{3}{128\eta} v^{-5} \left(\sum_{i=5}^6 p_{i,l,\text{GR}}(\Xi) \log(v^3) v^i \right) \\ & + \beta \mathcal{U}^{(b-5)},\end{aligned}\quad (4.6)$$

where $\mathcal{U} = (\pi \mathcal{M} f)^{1/3}$, $\mathcal{M} = \eta^{3/5} m$ is the chirp mass of the binary, with $\eta = m_1 m_2 / m^2$ the symmetric mass ratio, β is a deformation parameter and b is a real number one fixes before carrying out parameter estimation. This is closer to the original ppE framework [107] and it is employed by researchers because it allows for a slightly more direct mapping to constraints on specific modified theories [109, 115–117]. As it should be painstakingly clear by comparing Eqs. (4.5) and (4.6), these two implementations are equivalent to each other with $\bar{i} = b - 5$, $\bar{j} = 0$, a reparametrization between δp_i and β , and the transformation of their priors, as shown in detail in [112].

Regardless of the preferred implementation, both of the ones presented above focus on a single modification at a time (e.g. a single β or δp_i or $\delta p_{i,l}$ at a time). This is more a choice made out of practicality than one made because of a physically-motivated reason. In reality, solving for the phase of the gravitational

waveform, even in GR, leads to a series with an infinite number of terms when using the PN formalism, as shown in Eqs. (4.5) and (4.6). As these generic parametrizations represent unknown corrections to these terms in the PN series (whether in the GW phase directly or in the energy flux, binding energy or dispersion relations), there should also be an infinite number of deformations when the calculation is done for any specific modified theory of gravity. One would therefore expect a Fourier waveform phase of the form of Eq. (4.5) but without the Kronecker deltas in the definitions of p_i and $\delta p_{i,l}$ for the first implementation. For the second implementation shown in Eq. (4.6), the natural generalization would be

$$\begin{aligned}\Psi_{\text{GW}}(f) = & \frac{3}{128\eta} v^{-5} \left(1 + \sum_{i=2}^7 p_i(\Xi) v^i \right) \\ & + \frac{3}{128\eta} v^{-5} \left(\sum_{i=5}^6 p_{i,l}(\Xi) \log(\pi m f) v^i \right) \\ & + \beta \mathcal{U}^{(b-5)} \left(1 + \sum_{i=1}^7 \beta_i \mathcal{U}^i \right),\end{aligned}\tag{4.7}$$

where we now include β along with higher terms in the PN expansion, β_i

To be maximally robust, one would then think that one should simultaneously fit for *all* the δp_i and $\delta p_{i,l}$ or all the β_i and β parameters at all known PN orders (for waveforms used in current parameter estimation analysis, this would go up to 3.5PN order). However, when this was attempted in the past, the constraints on any one δp_i or β_i parameter quickly degraded to the point that the posterior devolved into the prior [61], and no new information could be gleaned about the underlying physics. This was predicted back in [116] and then verified recently in [61]. As is commonly understood, this is due to overfitting [119, 135, 136]. By increasing the dimensions of the parameter space and allowing a large enough prior range that significant degeneracies surface, any insight into the individual modifications is washed out by covariances. But if information is lost when including multiple deformations, are the constraints found while using a single deformation robust, as these covariances are completely neglected in this restricted case? If all one wishes to establish is whether a deviation *exists*, then previous work has shown that indeed a single-term deformation is sufficient [116]. Here, however, we are not concerned with detecting a deviation, but rather determining the accuracy to which deformation coefficients can be estimated, and thus, constraints can be placed.

4.3 Improved Parametric Tests of GR

From the standpoint of testing for new physics, is the situation truly so bleak? There are some critical details that are completely ignored by the overly agnostic approach of varying over all deformation parameters

simultaneously, and which one could may be use to perform robust parametric tests. We discuss the critical, additional information we will incorporate into our analysis in Sec. 4.3.1 and the details of the exact implementation in Sec. 4.3.2.

4.3.1 Restrictions Based on Mathematical Structure of Modified Theories

There exists at least two pieces of basic information that one can infer and use about the exact form of the deformations introduced into the GW Fourier phase by real modified theories of gravity.

The first piece of information concerns the PN structure of the deformations. Let us consider theories in which the early inspiral waveform can be calculated with the PN approximation, as we do in GR, where the phase is written as a series in the orbital velocity. If this is the case, there is a hierarchy in the magnitude of the modifications at each PN order, so that the PN series is convergent in certain regimes of parameter space. For example, using the parametrization of Eq. (4.5)

$$\delta p_i p_{i,\text{GR}} v^i \gg \delta p_{i+1} p_{i+1,\text{GR}} v^{i+1}, \quad (4.8)$$

while using the parametrization of Eq. (4.7)

$$\beta_i v^i \gg \beta_{i+1} v^{i+1}. \quad (4.9)$$

These conditions then lead to the asymptotic inequalities $\delta p_{i+1} v \ll \delta p$ and $\beta_{i+1} v \ll \beta_i$. The implication here is that allowing the δp_i or the β_i parameters at each PN order to vary over completely independent ranges is not representative of any PN-compatible theory of gravity beyond GR. For example, if the 2 PN deformation coefficient δp_4 or β_4 had a prior range $(-10^7, +10^7)$, then there would be many choices for which the 2PN terms would completely dominate over the 1PN and the Newtonian term, rendering the PN approximation inaccurate in the regime of velocities of interest to ground-based detectors.

Whether the PN series is convergent, asymptotic, or divergent has not been formally established due to the lack of an exact solution one can compare against. For comparable masses, Ref. [95, 287] computed the radius of convergence of the PN series using the PN binding energy, and up to 2PN order the author found it to be $v/c \approx 1/2$. In the extreme mass-ratio limit, Ref. [288] have shown the PN series is asymptotic, and determined the optimal radius of convergence as a function of PN order to be $v/c \approx (1/5, 1/3)$ depending on the PN order studied. Either way, the growth of the coefficients of the PN approximation cannot grow with PN order too rapidly if the solution is to remain accurate in the late inspiral (at velocities of roughly $1/3c$). This is true not just in GR, but also in modified theories of gravity, since it is a mathematical statement

about the validity of a series solution.

The second piece of information one can use relates to the fact that the mathematical structure of all physical theories guarantee that each modified GR term will be a function of the coupling constants of the theory. Let us restrict attention to modified theories that have a continuous GR limit, i.e. they reduce to GR continuously as their coupling constants tend to a specific value. Then, for small deformation from GR, each modified GR term can be Taylor expanded in the coupling constants of the theory about their GR values. To make this concrete, let us focus on theories with a single coupling constant, and let this constant be called ℓ . Then, Taylor expanding the deformation coefficients at any given PN order, we have

$$\delta p_i(\ell, \Xi) = \delta \bar{p}_i(\Xi) \ell^p, \quad (4.10)$$

or

$$\beta_i(\ell, \Xi) = \bar{\beta}_i(\Xi) \ell^p, \quad (4.11)$$

where $\delta \bar{p}_i(\Xi)$ and $\bar{\beta}_i(\Xi)$ are functions of the source parameter vector Ξ only (like the masses, spins, etc), and p is a positive real number. Note that although the functional dependence on the source parameters may change with PN order, the functional dependence on the coupling parameter will not in the small deformation limit (as otherwise the Taylor expansion of the full series would not be well-defined). One can now introduce a new parameter $\gamma = \ell^p$ to write all deformation coefficients as linear functions in γ to leading-order in small deformations. This conclusion is not only valid for theories with a single coupling constant, but rather, it can be generalized to theories with multiple couplings if there is one that dominates due to others being already constrained by other experiments or observations.

In summary, the PN expansion of most theories of gravity should give a GW Fourier phase Ψ_{GW} that takes the form

$$\begin{aligned} \Psi_{\text{GW}}(f, \gamma, \Xi) &= \Psi_{\text{GR}}(f, \Xi) + \gamma \delta \psi_0(\Xi) v^b \\ &\times [1 + \delta \psi_2(\Xi) v^2 + \delta \psi_3(\Xi) v^3 + \dots], \end{aligned} \quad (4.12)$$

where we use the new parameter $\delta \psi_i$ to denote the deformation amplitudes because we now choose to use the PN expansion parameter $v = (\pi m f)^{1/3}$; this is a slight deviation from the notation of the standard ppE PN expansion, which uses the parameter $\mathcal{U} = (\pi \mathcal{M} f)^{1/3}$. Here, Ψ_{GR} is the GW phase in GR, and γ generically represents the coupling constant in a given theory of gravity. The subscripts i in $\delta \psi_i$ represent the order in v or the $i/2$ PN order at which the deformation enters the GW Fourier phase relative to the leading PN order

term in the non-GR sector. The 0PN modification could enter the phase at any PN order relative to GR, as controlled by the b parameter. As required by the first piece of information presented above, the range of the $\delta\psi_i$ functions above must satisfy $\delta\psi_{i+1} v \ll \delta\psi_i$. As required by the second piece of information, the coupling constant of the modified theory shows up as a pre-factor that multiplies all deformations, so that in the GR limit, the non-GR terms vanish. When explicitly attempting to search for physics beyond GR, the real information comes from the first term in the series, where the coupling constant γ appears. Therefore, estimates on our ability to discern new information about physics should be quantified purely in terms of constraints on this constant, regardless of the number of higher order terms we include.

One of the first things one notices from Eq. (4.12) is that the $\delta\psi_i$ are not constants but rather functions of the source parameters Ξ . Given a specific modified theory, one knows what these functions are, and therefore, the $\delta\psi_i$ are not new parameters. In fact, in this case, the only new parameter to the model is simply the coupling constant γ . If one is attempting to carry out an agnostic test of GR, however, one is not privy to the functional form of $\delta\psi_i$. In this case, one can replace Eq. (4.12) via

$$\begin{aligned} \Psi_{\text{GW}}(f, \gamma, \Xi) &= \Psi_{\text{GR}}(f, \Xi) + \gamma \delta\bar{\psi}_0 v^b \\ &\times [1 + \delta\bar{\psi}_2 v^2 + \delta\bar{\psi}_3 v^3 + \dots] , \end{aligned} \quad (4.13)$$

$$\delta\bar{\psi}_i = \frac{\int \delta\psi_i d^N \Xi}{\int d^N \Xi} , \quad (4.14)$$

where N is the dimensionality of the source parameter space and $d^N \Xi$ is the parameter space volume factor. Clearly then, the coefficients $\delta\bar{\psi}_i$ are now constants that become new independent parameters to search over. Moreover, since γ enters multiplied by $\delta\bar{\psi}_0$, these two parameters are 100% degenerate, so one must employ the reparametrization $\bar{\gamma} = \gamma \delta\bar{\psi}_0$. The waveform parameter space is then enlarged to $\Xi \cup \Theta$, where $\Theta = \{\bar{\gamma}, \delta\bar{\psi}_2, \delta\bar{\psi}_3, \dots\}$. As before, the GR deviations are all still controlled by a single parameter $\bar{\gamma}$, while the range of the deformation coefficients is restricted by $\delta\bar{\psi}_{i+1} v \ll \delta\bar{\psi}_i$.

4.3.2 Implementation of Improvements

in Parametric Inspiral Tests

Using the representation of Eq. (4.13) and enforcing convergence of the series, we can now attempt to include multiple deformations at once. However, care must be taken in the way one enforces convergence, as the criteria $\delta\bar{\psi}_{i+1} v \ll \delta\bar{\psi}_i$ depends on the orbital velocity v .

To do this, we will implement a specialized prior that ensures the series of $\delta\bar{\psi}_i$ are convergent at a

“reasonable” value of v . That is, we choose the prior to be

$$|\delta\bar{\psi}_i| > |\delta\bar{\psi}_{(i+1)}(v_{\text{eval}})|, \quad (4.15)$$

for $i \in [2, 3, 4, 6, 7]$, where $v_{\text{eval}} = (\pi m f_{\text{eval}})^{1/3}$; the exact value of f_{eval} and thus v_{eval} that we choose are discussed below. For the 1PN term, the prior must be handled separately, and we choose it to be

$$|\delta\bar{\psi}_2(v_{\text{eval}})|^2 < 1, \quad (4.16)$$

as this is the next-to leading order (NLO) term due to our omission of the 0.5PN term.

We must now choose the value of the velocity, or equivalently the frequency, at which to evaluate this prior. We do so by choosing the GW frequency and orbital velocity that would correspond to the waves emitted by a binary at an orbital separation of $r_{12,\text{eval}} = 100m$ at leading-order in PN theory. Note that $r_{12,\text{eval}}$ is *not* the formal radius of convergence of the series, assuming the series were convergent. Rather, this quantity represents the orbital separation at which we are confident that adding higher order PN terms improves the PN approximation to the (unknown) exact answer. More aggressive (smaller) choices of the orbital separation could be viable options as well, but we will remain conservative in this work, as we will explain below. Using the Newtonian version of Kepler’s third law for quasi-circular orbits, we then have

$$v_{\text{eval}} = \left(\frac{m}{r_{12,\text{eval}}} \right)^{1/2} = 10^{-1}, \quad (4.17)$$

$$f_{\text{eval}} = \frac{(r_{12,\text{eval}}/m)^{-3/2}}{(\pi m)} \approx \frac{3 \times 10^{-4}}{m}. \quad (4.18)$$

This choice in frequency is inspired by the convergence properties of the GW phase in GR, as studied in [288] and revisited below. GR is the obvious model for us to base our new prior on, as it is much better understood than any modified theory and it is our null hypothesis.

As we now show, with this choice of separation (or equivalently frequency or velocity) the coefficients of the PN expansion in GR present a convergent structure for most mass ratios. The phase in GR (via the Taylor F2 waveform [289–291], accurate to 3.5PN) was already presented in Eq. (4.5), and the exact forms of the p_i and $p_{i,l}$ parameters can be found in various places in the literature (c.f. [95, 106, 169]). For a non-spinning binary in GR, the ratios $p_i/(p_{i+1}v_{\text{eval}})$ only depend on the symmetric mass ratio η alone. If these ratios are larger than unity, then the PN coefficient exhibit a convergent structure³. Figure 4.1 presents

³This is in the sense that v_{eval} would then be within the radius of convergence of the series if the series were convergent. If the series is asymptotic, then this would be an indication that one is evaluating the series in the regime where the asymptotic series is a good approximation to the exact answer.

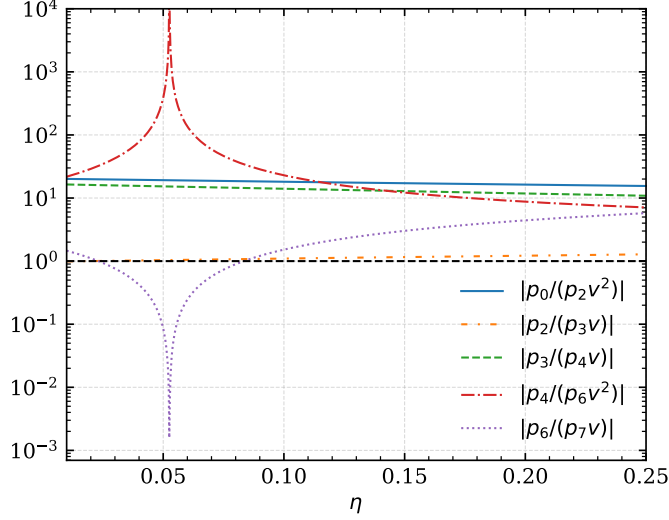


Figure 4.1: Ratio of the PN coefficients of the Fourier GW phase (Eq. (4.5)) for a non-spinning binary as functions of the symmetric mass ratio η . The ratios are evaluated at an orbital separation of $r_{\text{eval}} = 100m$, corresponding to $v_{\text{eval}} = 0.1$. Each line corresponds to a different pair of coefficients, and the dotted black line simply identifies the threshold we use for convergence ($|p_i/(p_{i+1}u_{\text{eval}})| > 1$). The spikes shown here correspond to when the coefficient $p_6 = 0$ at $\eta \sim 0.053$. Observe that all ratios are above unity for symmetric mass ratios $\eta > 0.09$, indicating that our choice of v_{eval} has led to a convergent structure of PN coefficients.

this ratio as a function of η , where we observe that our choice of v_{eval} leads to a convergent structure of the series for all $\eta > 0.09$. The spikes at around $\eta \sim 0.053$ are because p_6 vanishes, but this occurs outside the comparable-mass regime, which is what we focus on here.

As the parametrization in Eq. (4.13) has not been used extensively in previous literature, we will also study a related parametrization, closely associated with the standard ppE formalism. Let us then denote absolute deviations as $\bar{\Delta}_i$, and relative deviations as $\delta\bar{\psi}_i$ (as introduced in Eq. (4.13)) for the $i/2$ -th PN order. These new $\bar{\Delta}_i$ parameters then enter the GW phase at various PN orders via

$$\begin{aligned} \Psi_{\text{GW}}(f, \gamma, \mathbf{X}\mathbf{i}) &= \Psi_{\text{GW},GR}(f, \Xi) + \bar{\gamma} v^b + \\ &\quad \bar{\Delta}_2 v^{(2+b)} + \bar{\Delta}_3 v^{(3+b)} + \dots, \end{aligned} \quad (4.19)$$

where each modification $(\bar{\gamma}, \bar{\Delta}_i)$ now explicitly depends on the coupling constant of the theory. The mapping between Eq. (4.19) and Eq. (4.13) is simply given by

$$\bar{\Delta}_i = \delta\bar{\psi}_i \times \bar{\gamma}, \quad (4.20)$$

and $\bar{\gamma}$ is unchanged.

The priors we have defined above, although simple to state mathematically, are non-trivial. To illustrate this, we sampled from the prior directly using a variation of the sampling methods outlined in Sec. 4.2.1. In effect, we drew each of the deformation parameters from a uniform distribution $U(-B, +B)$ with some boundaries $-B$ and $+B$. For the leading order and NLO deformation, $|B| = 100$, while for the other, higher order parameters, $|B|$ is set to a value large enough to ensure the final distributions are not affected by $|B|$. With these random draws, we excluded any sample that violated Eq. (4.15). Although each parameter began as a uniform distribution, our prior in Eq. (4.15) imposed very non-trivial structure to the histogrammed distributions, as the boundary of the prior for each deformation depends on the value of other deformations in the series.

The priors on $\delta\bar{\psi}_i$ with 6 deformations and beginning at 0PN order relative to GR (proportional to v^{-5}) are shown in the left panels of Fig. 4.2, while those transformed to the absolute deformations $\bar{\Delta}_i$ are in the right panel. Observe that the prior on the leading PN order term in the series (which encodes most of the information we are interested in) is indeed flat, because Eq. (4.15) does not affect it. Observe also that the larger PN order the relative deviation, the larger the range of the prior. This range, however, is indeed *finite*, since if it were any larger, then the relative deviations would begin to affect the convergent structure of the PN series. Finally, observe that although the priors on $\delta\bar{\psi}_i$ seem to be pushing modifications away from GR ($\delta\bar{\psi}_i = 0$), this is just an artifact of looking at relative deviations; the priors on the absolute modifications $\bar{\Delta}_i$ are fully consistent with GR ($\bar{\Delta}_i = 0$).

4.4 Experimental Design

To determine the impact of these higher order deformations on our inferences about fundamental physics, we now conduct a series of parameter estimation analyses on synthetic signals (injections) with a variety of different recovery models. Each of the recovery models has the same base GR waveform (IMRPhenomD), but they differ in the different number of phase deformations overlaid on top of that base model. For each model, the modifications are added in ascending order beyond leading, beginning with the 1PN term (relative to the leading order deformation), and increasing by 0.5PN order up to 3.5PN order relative to GR, or relative to the leading order term, whichever criteria is met first. We skip terms of 2.5PN order relative to GR in all models, as they are 100% degenerate with the coalescence phase (which is an arbitrary constant). When skipping terms in the series due to these total-degeneracies, we also skip them in the enforcement of the prior. For example, if we were to skip the $\delta\bar{\psi}_i$ term in the series because it was degenerate with the coalescence

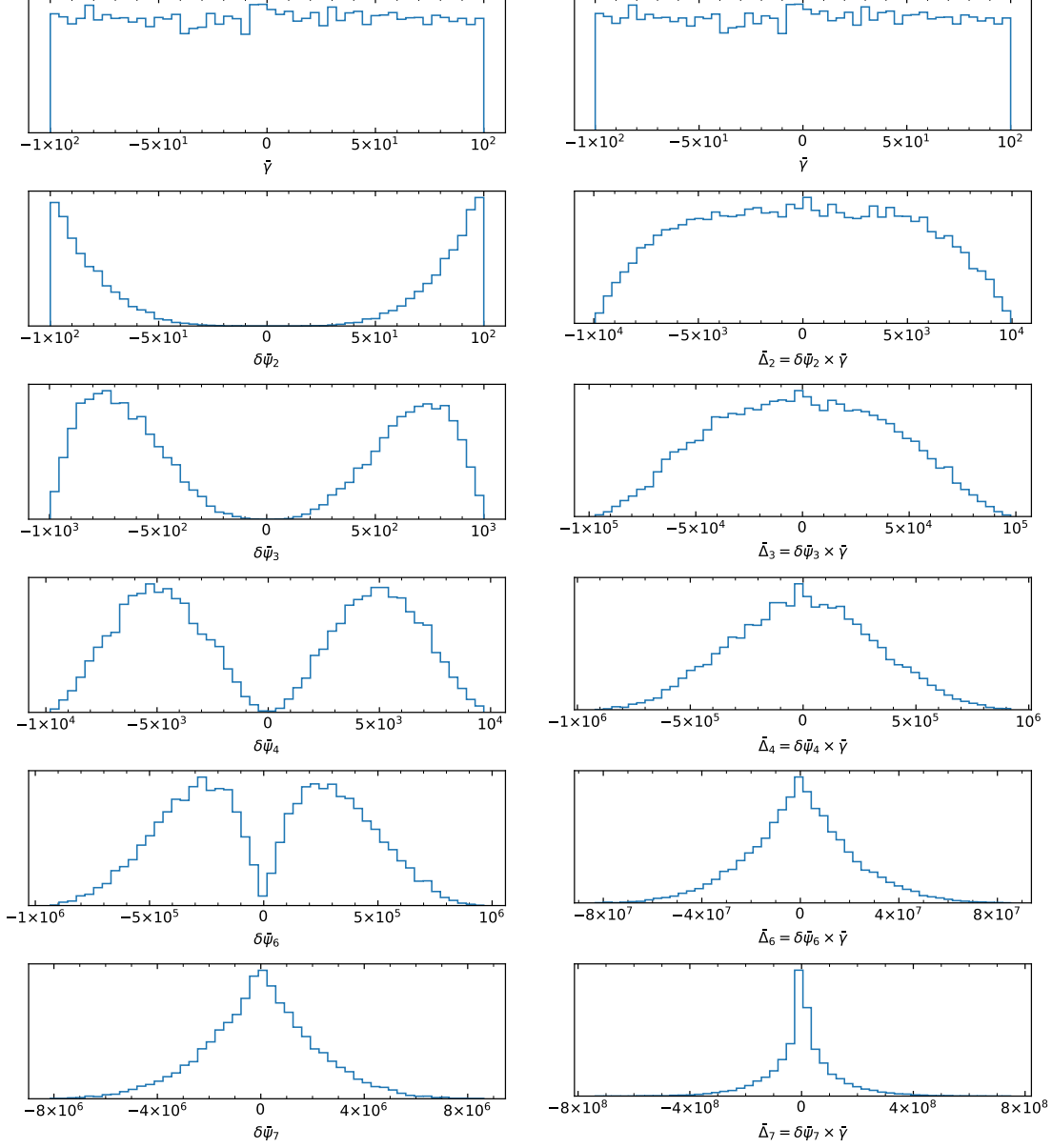


Figure 4.2: The one dimensional, marginalized prior on a set of deformations. The leading order deformation enters at Newtonian order relative to GR, and each of the subsequent deformations are labeled as the $i/2$ PN order relative to this leading order deformation. The priors for this parametrization are non-trivial, and can be misleading. They appear to disallow the GR limit $\delta\bar{\psi}_i = 0$ for several of the different PN orders. This is resolved when looking at the transformed priors of the right panel, rewritten in terms of the absolute modifications, $\bar{\Delta}_i$.

phase, we would update our prior for $\delta\bar{\psi}_{i-1}$ and $\delta\bar{\psi}_{i+1}$ to satisfy $\delta\bar{\psi}_{i-1} \geq \delta\bar{\psi}_{i+1}v^2$. Therefore, every model is uniquely defined by the PN order of the leading deformation (relative to GR) and the number of subsequent terms.

We will specifically target leading order GR deformations that enter at $[-1, 0, 1, 1.5, 2, 3, 3.5]$ PN orders,

relative to Newtonian order in GR, and we include up to 5 additional deformations (again, never exceeding 3.5PN order relative to Newtonian order in GR). For example, one model might have a leading order deformation at 1PN order relative to Newtonian order in GR, with 3 higher PN order deformations at [2, 3, 3.5]PN orders (again relative to Newtonian order in GR). As another example, another model might have a leading order deformation that enters at 1.5PN order relative to the Newtonian term in GR, with only a single higher PN order deformation. This would leave us with deformations at [1.5, 3]PN order relative to Newtonian order in GR.

As expected, all inferences will depend on the astrophysical source that produced the GW we are assuming has been detected, as well as the detector network that observes the signals. To attempt to explore these aspects as much as possible (while still keeping the scope of this work computational tractable), we focus on two sources: a “heavy” source and a “light” source, whose properties are listed in Table 4.1. The spin configurations also differ slightly, with one binary having both spins aligned with the orbital angular momentum, while the other binary has one spin anti-aligned. In both cases, we focus on binary black hole inspirals, and do not consider neutron star inspirals or mixed binaries. We expect the qualitative conclusions we will find to also hold for those systems.

For the detector networks, we also focus on two configurations. For the first configuration, and as a proxy for a 2g network, we use a network comprised of LIGO Hanford [198], LIGO Livingston [198], and Virgo [292]. For their sensitivities, we use analytic approximations to the aLIGO design sensitivity [227] and the first phase of Advanced Virgo’s sensitivity estimate [227]. For the second configuration, and as a proxy for a 3g network, we use a network comprised of Cosmic Explorer (CE) [200] and the Einstein Telescope (ET) [201]. For their sensitivities, we use the first phase of the CE noise curve [230] and the ET-D configuration of the ET noise curve [231]. To reduce the impact of the uncertainties concerning the noise curves we use, the luminosity distances of the injected sources are all scaled such that the SNR is exactly 20 (when observed by the entire 2g detector network). The distances are then left fixed at these values when we transitioned to the 3g network, so as to isolate the impact of a pure boost in SNR.

With all these considerations in mind, we then have 21 separate analyses (one for each combination of multi-deformation-parameter ppE model) for each detector network and source combination, for which there are four combinations. This comes out to a total of 84 separate Bayesian parameter estimation studies, whose results we summarize next.

Source Identifier	$m_1 (M_\odot)$	$m_2 (M_\odot)$	χ_1	χ_2	D_L (Mpc)	SNR_{2g}	SNR_{3g}
Heavy	25	25	0.3	-0.1	1613	20	517.798
Light	5	5	0.3	0.1	466	20	496.355

Table 4.1: Choices of GW sources for injection campaigns. The source parameters are the following: m_1 and m_2 are the masses of the larger and smaller black holes, respectively, D_L is the luminosity distance from Earth to the source, χ_i is the aligned, dimensionless spin of the i -th black hole, SNR_{2g} and SNR_{3g} are the SNRs of the source as measured by a 2g and 3g detector network respectively. The “heavy” source and the light “source” both have an SNR of 20 (for the 2g detector network), but have different total masses and different spin configurations. All source parameters were kept the same between the analyses involving the 2g and the 3g detector networks.

4.5 Bayesian Results

Let us now present the results of our experiments. Because the leading PN order deformation contains all the relevant constants controlling the magnitude of GR deviations, we focus on constraints on this leading-order term, presenting results both visually and in tabular form.

The results for the 2g network injections are shown in Fig. 4.3 and Table 4.2 for the “heavy” and “light” sources. The top panels show the marginalized 1σ constraint on the leading PN order deformation as a function of the number of terms included in the series, beginning with 1 (only the leading PN order deformation) and up to 6 total terms. The lower panel shows the *strengthening factor*, which we define here as

$$(\text{strengthening factor}) = \frac{\sigma_{\text{Single}}}{\sigma_{\text{Multiple}}} . \quad (4.21)$$

If this number is larger (smaller) than unity, then adding higher PN order deformations strengthens (weakens) the constraint one obtains by using a single parameter deformation. The bottom panels of Fig. 4.3 show the strengthening factor as a function of PN order.

Two main conclusions can be drawn from these figures and tables. First, observe how all curves in the top panels trend downwards (have a negative slope). This implies that adding higher PN order corrections to the modified sector *strengthens* the constraints one would have gotten if one included only a single parametric deformation. This is confirmed in the bottom panels, which shows the strengthening factor is always greater than one. In this sense, single-parameter constraints are therefore *conservative*. Second, observe how the slopes of the curves are small. This implies that the strengthening one obtains is somewhat mild, with improvements of only up to a factor of 40 in the most extreme case, as shown in the bottom panels of the figure.

These results and conclusions may seem counter-intuitive and, in fact, opposite to what one would expect. Adding additional, independent parameters to a model has been shown to increase degeneracies, and therefore

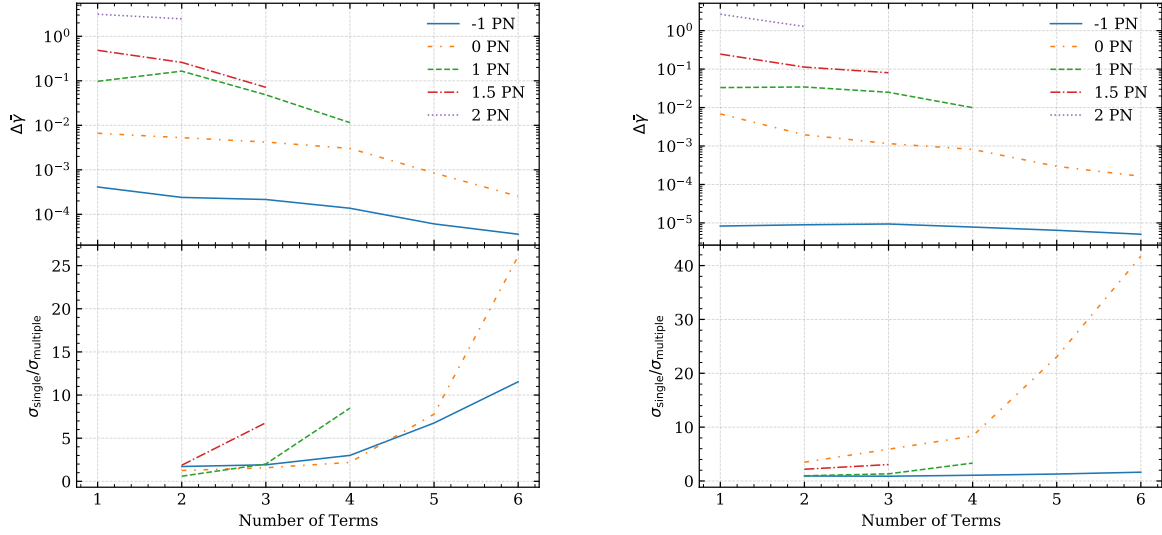


Figure 4.3: Marginalized 1σ constraints on the leading PN order deformations for the “heavy” source (left) and the “light” source (right) using the 2g network (see Sec. 4.4 for details) as a function of the number of sub-leading PN order terms added in the modified sector (in ascending order). Each line corresponds to modifications that start at different leading PN order. The functional form of GW phase is given in Eq. (4.13), and the priors used are those presented in Sec. 4.3. The top panels show the 1σ constraint on the leading PN order deformation, while the lower panel shows the strengthening factor, as defined in Eq. (4.21). Observe how the lines in the upper panels trend downward, which means that constraints with just a single parameter deviation are *conservative* (i.e. adding higher PN order corrections strengthens the constraint obtained with a single parametric deviation). Moreover, observe how the slope of the lines are small, which means that the strengthening of the constraint is mild, with improvements of at most roughly one order of magnitude.

Terms/LO	-1	0	1	1.5	2
1	4.1×10^{-4}	6.6×10^{-3}	9.7×10^{-2}	4.8×10^{-1}	3.1
2	2.4×10^{-4}	5.3×10^{-3}	1.6×10^{-1}	2.6×10^{-1}	2.5
3	2.2×10^{-4}	4.2×10^{-3}	4.8×10^{-2}	7.1×10^{-2}	
4	1.4×10^{-4}	3.0×10^{-3}	1.1×10^{-2}		
5	6.1×10^{-5}	8.5×10^{-4}			
6	3.6×10^{-5}	2.5×10^{-4}			

Terms/LO	-1	0	1	1.5	2
1	8.3×10^{-6}	6.8×10^{-3}	3.3×10^{-2}	2.5×10^{-1}	2.7
2	8.9×10^{-6}	2.0×10^{-3}	3.4×10^{-2}	1.1×10^{-1}	1.3
3	9.4×10^{-6}	1.2×10^{-3}	2.5×10^{-2}	8.0×10^{-2}	
4	7.8×10^{-6}	8.2×10^{-4}	9.9×10^{-3}		
5	6.4×10^{-6}	3.0×10^{-4}			
6	5.1×10^{-6}	1.6×10^{-4}			

Table 4.2: Marginalized 1σ constraints on the leading PN order deformations for the “heavy” source (left) and the “light” source (right) using the 2g network (see Sec. 4.4 for details). The columns represent the PN order (relative to the Newtonian term in GR) at which the GR deformation is first introduced. The rows corresponds to the number of PN corrections that are added on top of the leading PN order one in ascending PN order. The functional form of the phase is given in Eq. (4.13), where the priors are presented in Sec. 4.3. Observe the (order of magnitude) consistency of the constraints as one includes more and more sub-leading PN order deviations (i.e. as one moves down the column for a fixed leading-order term). This indicates that higher PN order corrections to the current modified gravity ppE waveforms will not invalidate current bounds placed on modified theories with only leading PN order deformations.

deteriorate our ability to estimate any given parameter [61, 116, 119, 135]. The reason this does not happen here is because of the series structure of the deformations in the model. Equation (4.13) is different from how other work has modeled deviations, because the strength of all deviations is here controlled by the leading PN order term (as expected from most modified theories of gravity). If an MCMC chain keeps the leading PN order deformation small, then the overall modification at each PN order ($\bar{\Delta} = \delta\bar{\psi}_i \times \bar{\gamma}$) can remain small even if the chain visits large values of $\delta\bar{\psi}_i$. This opens up more of the prior volume for chain exploration, which can thus be preferred more by the posterior distribution. In other words, there are more “states” for the model to take with small leading PN order deformation than with large leading PN order deformations, putting “pressure” on the leading coefficient to remain small.

But if the higher PN order terms were allowed to explore infinitely large values, then this would put infinite “pressure” on the leading PN order coefficient, forcing its posterior to be artificially tight around zero. Whether this happens or not depends on the prior one chooses on the higher PN order parameters. Naively, one may think that the most conservative prior is one that is infinitely wide and flat (this would correspond to our PN-based priors but evaluated at an infinite orbital separation.) Choosing such a flat, uniform prior, however, would allow the magnitude of the higher PN order deformation parameters to increase indefinitely. In turn, this would put infinite “pressure” on the leading PN order coefficient and push its

posterior to zero. The result would be an overly confident or overly aggressive constraint that is generated not by the information contained in the data, but by the prior choice. Our PN-based prior prevents this from happening. By ensuring the higher order deformations are not infinitely large (as otherwise the PN series would break down), the prior volume remains a reasonable size, and the “pressure” on the leading PN order deformation, sourced by this prior volume, is therefore kept to a reasonable amount. This then means that the constraint on the leading PN order parameter is governed by the information contained in the data (and the mathematical requirement that the PN approximation be valid in the inspiral for the modified theory), and not by the imposition of an overly restrictive prior.

Given how important the PN prior is to prevent overly aggressively constraints on the leading PN order term, one may wonder whether these are robust to variations of the details associated with this prior. The main quantity we can vary to change the PN prior is the choice of velocity v_{eval} (or orbital separation $r_{12,\text{eval}}$) at which the prior is evaluated. Figure 4.4 shows the constraints on $\bar{\gamma}$ and the strengthening factor as a function of the number of higher PN order terms kept in the modified sector, but this time obtained with PN-based priors evaluated at 3 different orbital radii. Observe that in all cases the constraints improve as you add higher PN order terms in the modified sector. Therefore, the first main conclusion of our paper (i.e. that single-parameter constraints are conservative) is robust to modifications in the PN-based prior.

Figure 4.4 also shows that the degree of importance of the higher PN order terms does depend on the PN-based prior. Indeed, the strengthening factor is larger, the larger the value of $r_{12,\text{eval}}$ that one chooses for the PN-based prior. This is consistent with our explanation above that as $r_{12,\text{eval}} \rightarrow \infty$, then the prior on the higher PN order terms becomes flat and infinite, and, thus, $\Delta\bar{\gamma} \rightarrow 0$ because of the infinite “pressure” created by the higher PN order terms. The choice of $r_{12,\text{eval}}$ should then be the smallest value of the orbital separation (or the orbital velocity) at which one expects the PN approximation to still be valid in the inspiral of compact binaries in the modified theory. Without specifying a particular modified theory, one can therefore not choose $r_{12,\text{eval}}$ precisely. This is why we used the properties of the PN series in GR to set $r_{12,\text{eval}} = 100m$ in this paper, since we are sure that for such large values of $r_{12,\text{eval}}$, the PN series in GR is still a good approximation. Had we chosen a smaller value, our conclusions about the importance of the higher PN order terms would have been even stronger (i.e. we would have concluded the higher PN order terms are even less important than stated so far).

Our conclusions thus far are the following. First, single-parameter tests of GR are conservative, and would become stronger if higher PN order terms are included in the modified sector. Second, the improvement of these higher PN order terms is relatively mild, with enhancements of at most 1–2 orders of magnitude, depending on the inspiral signal observed. Given these conclusions, we then infer that current constraints on

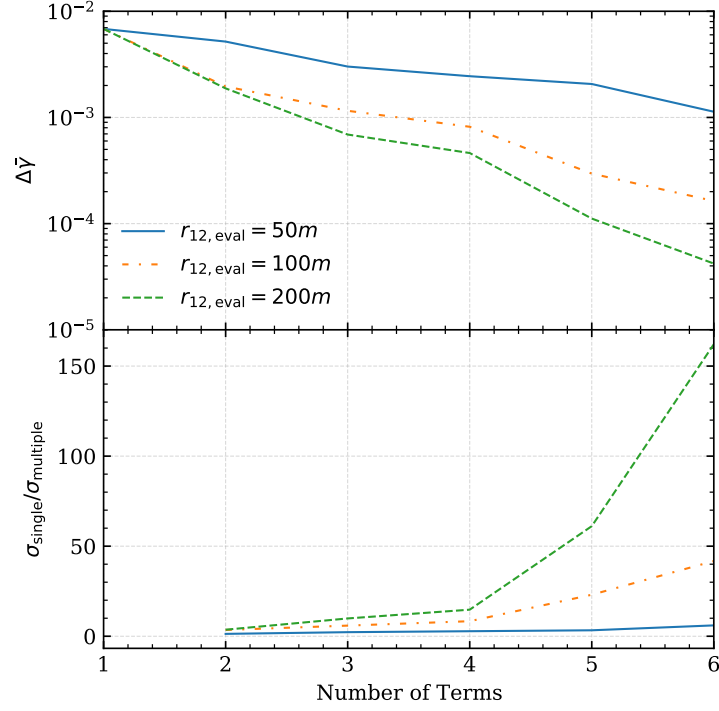


Figure 4.4: Marginalized 1σ constraint on $\bar{\gamma}$ for a modification that first enters at 0PN order as a function of number of PN terms kept in the modified sector, using 3 different choices of $r_{12, \text{eval}}$ for the PN prior. In all cases, we here focus on the light source and the 2g detector network. Observe that in all cases the constraint on $\bar{\gamma}$ becomes stronger the more PN order terms are added. Observe also that the larger we choose $r_{12, \text{eval}}$ to be, the stronger the “pressure” on $\bar{\gamma}$ and thus the stronger the constraint.

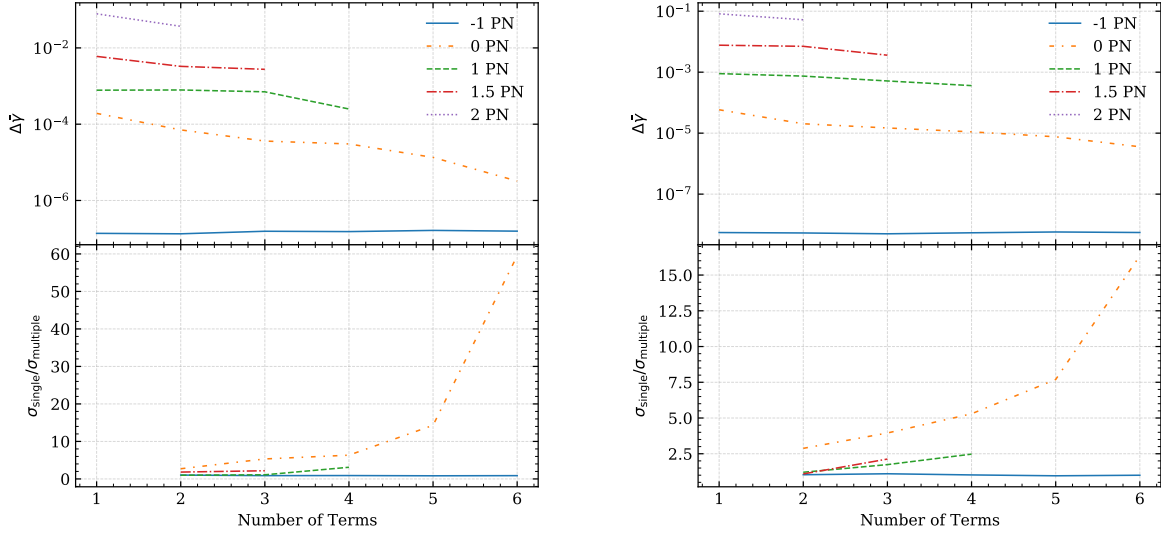


Figure 4.5: Same as Fig. 4.3, but for a 3g network, using a “heavy” source (left) and a “light” source (right). Observe that the trends found with the 2g detector networks continue when considering 3g detector networks.

modified theories of gravity are robust to uncertainties in waveform modeling related to unknown, higher PN order corrections. But are these conclusions robust also to an improvement in the detector’s sensitivity, or alternatively, to an increase in the signal-to-noise ratio of the signals detected? Motivated by this, we repeated the analysis described above for the proxy of a 3g network described in Sec. 4.4, and we arrived at very similar conclusions, as shown in Fig. 4.5 and Table 4.3. Observe that the increased SNR of the injections does not change the trends we have described above for the less sensitive 2g networks. While the bounds become stronger because of the more sensitive detectors, the inferences made with single-parameter models are still robust to uncertainties in the higher PN order terms.

4.6 Alternative Parametrizations

One may wonder how robust the conclusions presented in Sec. 4.3 are to the exact form of the parametrization we implemented in that section. Therefore, before continuing on to a specific theory of gravity, let us briefly examine two other reasonable parameterizations to determine how our conclusions are affected. In what follows, we will re-analyze the light system with the 2g network, exactly as defined earlier, using 6 deformations and starting with the Newtonian-order term relative to GR, i.e. terms at $[0,1,1.5,2,3,3.5]$ PN orders relative to GR. The difference, however, will be in the exact form of the parametrization of the deformations at each of those orders and how the prior for those deformations is imposed.

The first model uses the same functional form as the majority of our work here, defining the deformations

Terms/LO	-1	0	1	1.5	2
1	1.4×10^{-7}	1.9×10^{-4}	7.8×10^{-4}	6.0×10^{-3}	7.8×10^{-2}
2	1.3×10^{-7}	7.1×10^{-5}	7.9×10^{-4}	3.3×10^{-3}	3.7×10^{-2}
3	1.6×10^{-7}	3.6×10^{-5}	7.1×10^{-4}	2.7×10^{-3}	
4	1.5×10^{-7}	3.0×10^{-5}	2.5×10^{-4}		
5	1.6×10^{-7}	1.3×10^{-5}			
6	1.6×10^{-7}	3.2×10^{-6}			

Terms/LO	-1	0	1	1.5	2
1	5.5×10^{-9}	5.8×10^{-5}	9.0×10^{-4}	7.7×10^{-3}	8.2×10^{-2}
2	5.3×10^{-9}	2.0×10^{-5}	7.4×10^{-4}	7.0×10^{-3}	5.2×10^{-2}
3	5.0×10^{-9}	1.5×10^{-5}	5.2×10^{-4}	3.6×10^{-3}	
4	5.4×10^{-9}	1.1×10^{-5}	3.6×10^{-4}		
5	5.7×10^{-9}	7.6×10^{-6}			
6	5.5×10^{-9}	3.6×10^{-6}			

Table 4.3: Same as Table 4.2, but for a 3g network, using the “heavy” source (top) and the “light” source (bottom). Once more, the trends found with the 2g detector networks continue when considering a 3g detector network.

in the phase as a series of coefficients exactly as shown in Eq. (4.13), but with a *different* prior. We modify the prior to have simple, fixed boundaries, as opposed to the more complicated prior used in the main body of this work (previously, we enforced the convergence criteria at every point in parameter space through Eq. (4.15) and Eq. (4.16)). In particular, the prior for the new model will have fixed boundaries on $\bar{\gamma}$ and the $\delta\bar{\psi}_i$ ’s with progressively larger ranges, namely

$$\begin{aligned}
|\bar{\gamma}| &< 10^{-1}, \\
|\delta\bar{\psi}_2| &< 10^2, \\
|\delta\bar{\psi}_3| &< 10^3, \\
|\delta\bar{\psi}_4| &< 10^4, \\
|\delta\bar{\psi}_6| &< 10^6, \\
|\delta\bar{\psi}_7| &< 10^7.
\end{aligned} \tag{4.22}$$

This prior ensures that the convergence criteria is *generally* satisfied but not guaranteed, while removing some of the complicated structure from our original prior (shown in Fig 4.2). By reanalyzing our synthetic data with this new model, we will quantify how strong of an impact our enforcement of a strict convergence criteria had on our results above.

We note in passing that the prior range on $\bar{\gamma}$ never played much of a role in our analysis, as the prior bounds on the higher-order parameters, $\delta\bar{\psi}_i$ ’s, were independent of $\bar{\gamma}$. As we will see below, that will no

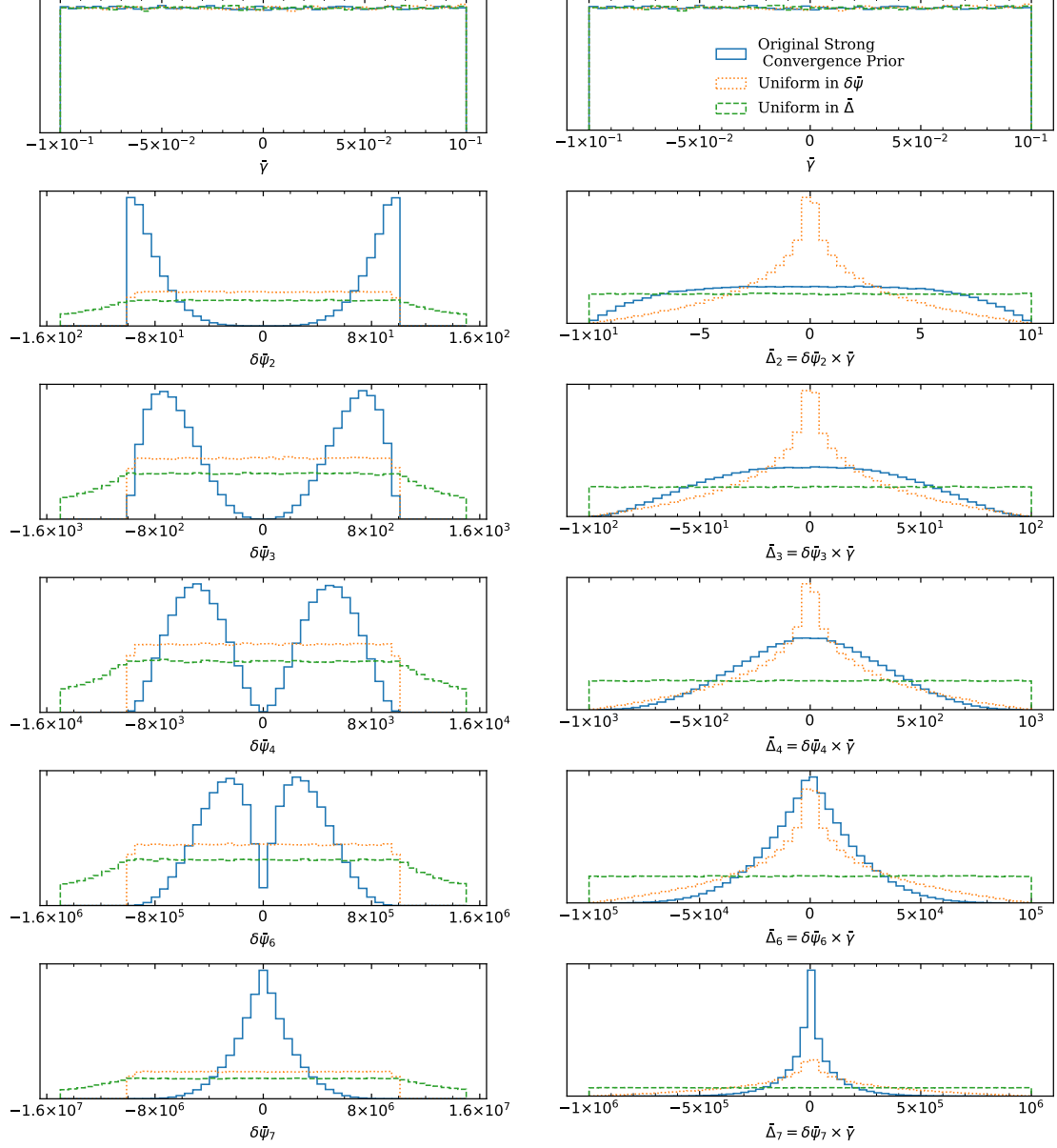


Figure 4.6: The one dimensional, marginalized prior on a set of deformations in the three models considered in Sec. 4.6. The leading-order deformation enters at Newtonian order relative to GR, and each of the subsequent deformations are labeled as the $i/2$ PN order relative to this leading order deformation. Included in this figure is the original model described in Sec. 4.3, shown as a solid blue line. The first alternative model uses the $\delta\bar{\psi}_i$ parametrization described in Eq. (4.13) but with simple, uniform ranges with fixed (increasingly larger with higher PN order) boundaries, shown as a dotted orange line. The second alternative model is described by a series of $\tilde{\Delta}_i$ parameters with uniform distributions with fixed (increasingly larger with higher PN order) boundaries, defined in Eq. (4.19), shown as a dashed green line. Note that the range of the prior on the $\delta\bar{\psi}_i$ parameters in the second alternative parametrization extends beyond the frame of the figure, but the range was restricted for visual purposes.

longer be the case, and we therefore updated our prior bound on $\bar{\gamma}$ to a much more reasonable range of $|\bar{\gamma}| < 10^{-1}$ (informed by our original experiments). We have verified that we obtain the same results with this new prior range and with the original parametrization.

The second parametrization we investigate is written in the form of Eq. (4.19), where our model is parametrized by $\bar{\gamma}$ and the series of $\bar{\Delta}_i$'s. In other words, we are moving from parametrizing our deformations as a relative series of terms to working with a model described by the *absolute* deformations. With this parametrization, we enforce similar priors as our first alternative parametrizations, with fixed boundaries of successively larger sizes, namely

$$\begin{aligned}
|\bar{\gamma}| &< 10^{-1}, \\
|\bar{\Delta}_2| &< 10^1, \\
|\bar{\Delta}_3| &< 10^2, \\
|\bar{\Delta}_4| &< 10^3, \\
|\bar{\Delta}_6| &< 10^5, \\
|\bar{\Delta}_7| &< 10^6.
\end{aligned} \tag{4.23}$$

With this model, we will investigate the impact that our choice of parameterization had on our conclusions of the previous section.

As the prior distribution is the major reason for testing these alternative parametrizations, we show all three prior distributions on $\bar{\gamma}$, $\delta\bar{\psi}_i$, and $\bar{\Delta}_i$ in Fig. 4.6. From this figure, we can see that the first alternative parametrization produces the strongest prior on the absolute deformations, $\bar{\Delta}_i$, for low orders, but the original prior still places the most prior weight at small values of $\bar{\Delta}_i$ for large PN orders. Our first alternative parametrization does alleviate the “pressure” on the $\delta\bar{\psi}_i$ parameters seen in the original parametrization, which was pushing low PN deformations away from zero (and away from GR) and pushing high PN deformations towards zero. The second alternative parametrization has a uniform distribution in the absolute deformations, or the $\bar{\Delta}_i$ parameters, giving a variation of an uninformative prior. From this figure, we would expect comparable results from the original parametrization and the first alternative model, as the prior distributions are relatively similar. The second alternative parametrization, uniform in $\bar{\Delta}_i$, will most likely result in the most degradation between the six-deformation and single-deformation model, as the prior is the least informative and most similar to the current, state-of-the art methodology.

We now use these two additional parametrizations to analyze the “light” source, as seen by the 2g detector network, and compare the constraints on $\bar{\gamma}$ to the posterior coming from a model that has a single deformation

at Newtonian order and to the constraint coming from the original parametrization with six deformations. As all three models (the original model and the two alternative parametrizations) reduce to the same form as the number of deformations is taken to one, this will provide a consistent metric for comparison. The results for those analyses are shown in Fig. 4.7, and as you can observe, they are consistent with the expectations presented above. The two models described by a relative series of deformations at higher orders (the original model and the first alternative model) produce comparable results. The second alternative model (uniform in $\bar{\Delta}_i$) produces a constraint that is mildly weaker than the original, single deformation model. We therefore conclude that the conclusions presented in Sec. 4.3 are robust to our parameterization and choice of priors.

4.7 A Concrete Example: ssGB theory

The methodology we have proposed here seems to accomplish the purpose it was designed for: allowing for realistic uncertainty in our limited modeling while illustrating the current robustness of the bounds placed on modified theories of gravity. A critical question, however, is the accuracy to which it actually relates to known, interesting theories of gravity actively being studied. One particular theory currently of interest is scalar Gauss Bonnet, or sGB, which is inspired by low-energy limits of higher energy theories [138, 293, 294]. sGB contains a scalar field, ϕ , that couples to a curvature invariant called the Gauss Bonnet invariant, $\mathcal{G}$. The Gauss Bonnet invariant is defined by $\mathcal{G} = R^2 - 4R_{\mu\nu}R^{\mu\nu} + R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma}$, where $R_{\mu\nu\rho\sigma}$, $R_{\mu\nu}$, and R are the Riemann tensor, Ricci tensor, and Ricci scalar, respectively. The full action can be written as [275]

$$S = \frac{1}{16\pi} \int d^4x \sqrt{-g} [R - 2\partial^\mu \phi \partial_\mu \phi + \alpha f(\phi) \mathcal{G}] , \quad (4.24)$$

where g is the determinant of the metric, α is the (dimensionful) coupling constant of the theory, and $f(\phi)$ is a coupling function. The theory reduces to GR minimally coupled to the scalar field in the $\alpha \rightarrow 0$. Therefore, constraints on this theory typically relate to the magnitude of α . Certain choices of the coupling function $f(\phi)$ result in different flavors of sGB. In particular, an interesting choice motivated by string theory is Einstein-dilaton-Gauss-Bonnet, where $f(\phi) = e^{2\phi/4}$ [138]. Another option, which will be the focus of this paper, is shift-symmetric sGB or ssGB, obtained by making the choice of $f(\phi) = 2\phi$ [295–297], which results from a small-field expansion of Einstein-dilaton-Gauss-Bonnet gravity [296].

The GW phase in ssGB was recently calculated beyond the leading PN order modification (at -1PN order, relative to Newtonian order in GR) in [275], giving us an example theory to explore. In their work, they classified the inspiral into two distinct regimes: that which is dominated by the emission of dipolar radiation by the scalar field (which is absent in GR), and that which is dominated by quadrupolar radiation. The

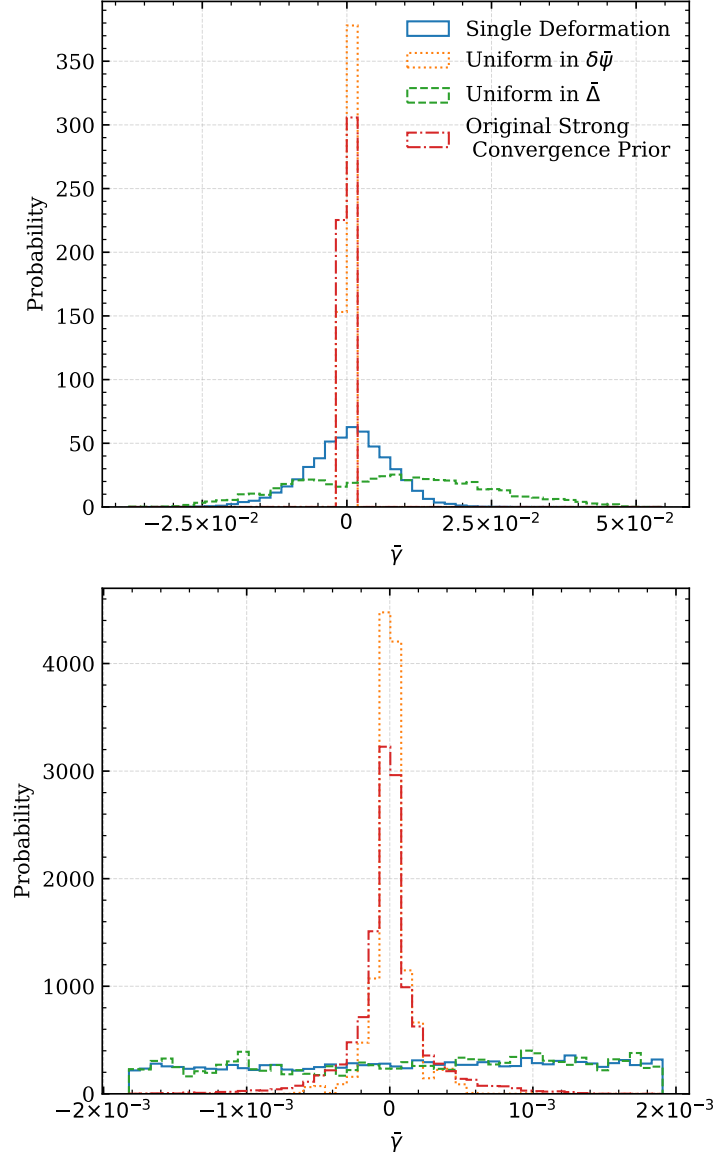


Figure 4.7: The final, marginalized posterior distributions on $\bar{\gamma}$ coming from the three models discussed in Sec. 4.6 using the synthetic data from the “light” source as observed by the 2g detector network. The top panel shows a larger range of $\bar{\gamma}$ and the lower panel zooms-in the range of $\bar{\gamma}$ to present details in the two narrower distributions. The probabilities on the y-axis are normalized to the shown range of $\bar{\gamma}$ in both panels. The distribution coming from the model using a single deformation parameter at Newtonian order is shown as a solid blue line. The distribution coming from the model described by the first alternative parametrization with six deformations (with a prior uniform in $\delta\bar{\psi}_i$) is shown as the dotted orange line. The distribution coming from the model described by the second alternative parametrization with six deformations (with a prior uniform in $\bar{\Delta}_i$) is shown as the dashed green line. The distribution coming from the model described by the original parametrization with six deformations (with a prior that strictly enforces our notion of convergence) is shown as the dotted-dashed red line. Note that the first alternative parametrization and the original parametrization are almost totally overlapping in the top panel.

boundary between dipolar and quadrupolar radiation is typically well before the binary reaches a separation of $100m$, so the regime of interest to ground-based GW detectors is the quadrupolar driven one. To evaluate the convergence properties of the series, we take the expression for the GW phase derived in Eq. (93)–(96) of [275] and isolate terms involving deviations from GR, i.e. terms that involve α , the coupling constant in ssGB. This gives the following form for the phase modification

$$\Psi_{\text{GW}} - \Psi_{\text{GW,GR}} = \delta\psi_{\text{LO,ssGB}} v^{-7} + \delta\psi_{\text{NLO,ssGB}} v^{-5} + \delta\psi_{\text{NNLO,ssGB}} v^{-3}, \quad (4.25)$$

where the coefficients take the form

$$\delta\psi_{\text{LO,ssGB}} = -\frac{5}{7168} \zeta \frac{(4\eta - 1)}{\eta^5}, \quad (4.26)$$

$$\delta\psi_{\text{NLO,ssGB}} = -\frac{5}{688128} \zeta \frac{(685 - 3916\eta + 2016\eta^2)}{\eta^5}, \quad (4.27)$$

$$\delta\psi_{\text{NNLO,ssGB}} = \frac{5}{387072} \zeta \frac{(1 - 2\eta)^2 (995 + 952\eta)}{\eta^5}, \quad (4.28)$$

where $\zeta \equiv \alpha^2/m^4$ is the dimensionless coupling constant of the theory⁴. We have verified that this expression reduce exactly to the leading PN order results first obtained in [110, 112] in the non-spinning limit⁵. Observe that the phase deformation has the exact structure that we anticipated in Eq. (4.13). Namely, the above phase deformation can be rewritten as

$$\begin{aligned} \Psi_{\text{GW}} - \Psi_{\text{GW,GR}} &= \zeta \delta\psi_{\text{LO,ssGB}}(\eta) v^{-7} \\ &\times [1 + \delta\psi'_{\text{NLO,ssGB}}(\eta) v^2 + \delta\psi'_{\text{NNLO,ssGB}}(\eta) v^4], \end{aligned} \quad (4.29)$$

with the following re-definitions of the coefficients

$$\delta\psi'_{\text{NLO,ssGB}} = \frac{\delta\psi_{\text{NLO,ssGB}}}{\delta\psi_{\text{LO,ssGB}}} = \frac{685 - 3916\eta + 2016\eta^2}{96(4\eta - 1)} \quad (4.30)$$

$$\delta\psi'_{\text{NNLO,ssGB}} = \frac{\delta\psi_{\text{NNLO,ssGB}}}{\delta\psi_{\text{LO,ssGB}}} = -\frac{(1 - 2\eta)^2 (995 + 952\eta)}{54(4\eta - 1)}. \quad (4.31)$$

The phase modification has therefore been written as a series in v , all proportional to the coupling constant α appearing in the coefficient of the overall controlling factor of the series $\delta\psi_{\text{LO,ssGB}}$.

With this in hand, we can now compare the above results to the general framework we developed in the

⁴The expressions presented here are not valid in the $\eta \ll 1$ regime, because then the curvature of the small black hole becomes very large, and the effective field theory treatment used to derive these expressions breaks down.

⁵Note that the results of [275] are formally complete only to 1PN order, implying that the v^{-3} coefficient has a 2PN correction coming from the dipolar term that has not yet been computed.

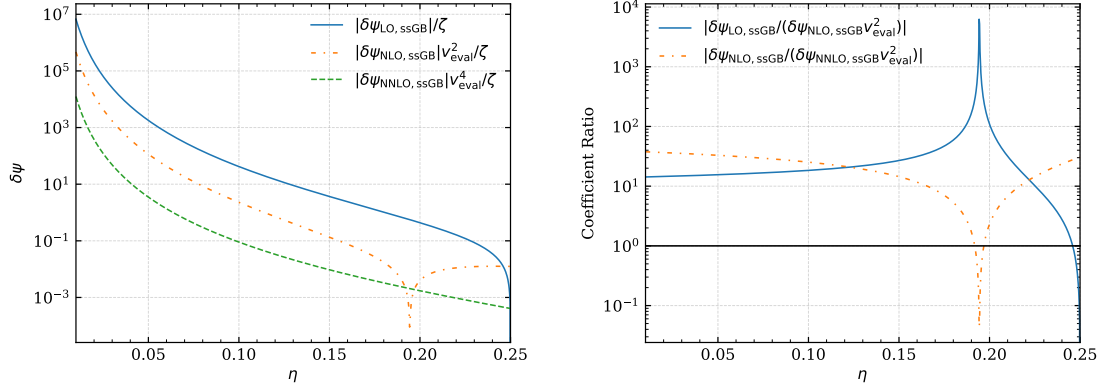


Figure 4.8: Absolute value of the terms in the phase deformation at different PN order (left) and ratio of terms (right) in ssGB as a function of symmetric mass ratio, evaluated at $v_{\text{eval}} = 0.1$. Observe that, for this choice of v_{eval} , the leading PN order term is larger than the next-to-leading order one, which in turn is larger than the next-to-next-to-leading order, except at $\eta \sim 0.19$, where the 1PN term vanishes identically. Observe also that the ratios are all larger than unity, except again for a specific value of η at which the 1PN term vanishes.

previous section when studying the convergence properties of the PN series in modified theories. To do so, we need to look at the ratios of these coefficients, which we show below. For the LO and NLO terms, we arrive at the following expression for their ratio

$$\frac{\delta\psi_{\text{LO,ssGB}}}{\delta\psi_{\text{NLO,ssGB}}v_{\text{eval}}^2} = \frac{96(4\eta - 1)}{[685 + 4\eta(-979 + 504\eta)]v_{\text{eval}}^2}. \quad (4.32)$$

For the NLO and NNLO terms, we evaluate their ratio to

$$\frac{\delta\psi_{\text{NLO,ssGB}}}{\delta\psi_{\text{NNLO,ssGB}}v_{\text{eval}}^2} = \frac{9[685 + 4\eta(-979 + 504\eta)]}{16(1 - 2\eta)^2(995 + 952\eta)v_{\text{eval}}^2}. \quad (4.33)$$

As we are only considering non-spinning binaries, these ratios are only a function η , so we can easily plot them to determine their convergence properties. The absolute deviations (divided by ζ), as shown in Eqs. (4.26), (4.27), and (4.28) and including appropriate values of v_{eval} , are shown in the left panel of Fig. 4.8, while the ratios of these coefficients, as defined in Eqs. (4.32) and (4.33), are shown in the right panel.

Figure 4.8 illustrates that the PN expansion of ssGB theory conforms, on average, to the criteria we have outlined in Sec. 4.3. This is most easily evidenced by the right panel of this figure, which shows that the ratio of coefficients is always above unity, except for a specific value of η at which the 1PN term $\delta\psi_{\text{NLO,ssGB}}$ vanishes. Since this happens at a single point, averaging over all values of η , it is clear that the convergence criteria is satisfied.

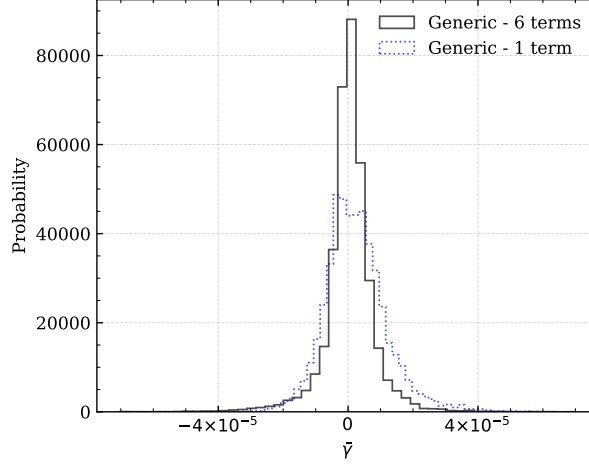


Figure 4.9: For reference, we show the two distributions on $\bar{\gamma}$ coming from the two generic deformation models (models 3 and 4). This figure shows that the distributions on the generic parameter $\bar{\gamma}$ for these two models are comparable, with a slight enhancement when using the 6 deformation variation. This same information is shown graphically in Fig. 4.3 and tabulated in Table 4.2. When compared with Fig. 4.10, we see that this relation persists. Even once the posteriors are transformed to theory-specific constants, constraints are not meaningfully changed when using these two models. Furthermore, note that the posterior on $\bar{\gamma}$ is fully consistent with GR (the model that truly describes the injected data), which is further evidence that the deviation away from GR in Fig 4.10 for the generic models is an artifact of the transformation.

But are our conclusions from Sec. 4.5 still valid in this specific theory? To answer this question, we completed another series of Bayesian studies, but now within the context of ssGB theory. We used the “light” injection from Table 4.1, and recovered with a IMRPhenomD model with one of the following phase deformations appended to its inspiral phase:

- Model 1: only the leading PN order deformation in ssGB theory (defined in Eq. (4.26)),
- Model 2: all the phase deformation terms in ssGB theory (defined in Eq. (4.27) and Eq. (4.28))
- Model 3: a single ppE deformation term (as in Eq. (4.13) but with $\delta\bar{\psi}_{i>0} = 0$),
- Model 4: 6-parameter ppE deformations (as in Eq. (4.13) but with $\delta\bar{\psi}_{i>6} = 0$).

For models 3 and 4, after carrying out a Bayesian parameter estimation study, we mapped the constraints on $\bar{\gamma}$ to constraints on α through

$$\bar{\gamma} = \delta\psi_{\text{LO,ssGB}}, \quad b = -7, \quad (4.34)$$

to enable comparisons with the results using models 1 and 2. For reference, the posteriors for $\bar{\gamma}$ for models 3 and 4 are shown in Fig. 4.9. This figure simply reiterates in more detail the conclusions from Sec. 4.5 (Fig. 4.3 and Table 4.2). Namely, the difference between using a single deformation and six deformations within the model established in Sec. 4.3 is a mild enhancement of constraints.

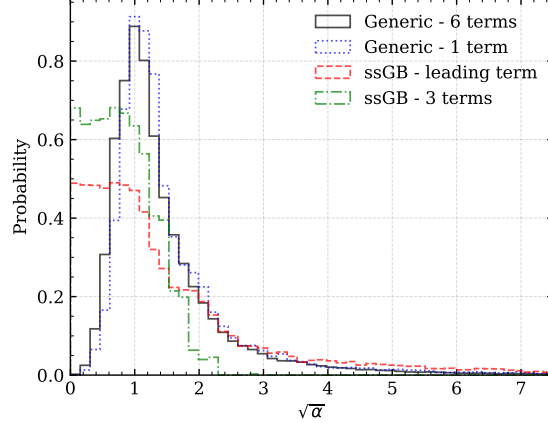


Figure 4.10: Marginalized posterior distributions on $\sqrt{\alpha}$ for a GR injection extracted with the four models described in the text. We here considered the “light” source, defined by Table 4.1, and a 2g detector network. When carrying out the Bayesian studies with the multi-parameter ppE model, we employed the PN prior with $r_{12,\text{eval}} = 100$, as done in the rest of this paper. The constraints on $\bar{\gamma}$ obtained with models 3 and 4 were mapped to constraints on $\sqrt{\alpha}$ to enable comparisons with the results obtained with models 1 and 2. Observe that clearly the posterior distributions are all consistent with each other (modulo singularities in the transformation at $\alpha = 0$, discussed in detail in [123, 124]). This shows clearly that in ssGB theory, leading PN order constraints are conservative and sufficient to place bounds on the theory.

Marginalized posterior distributions of $\sqrt{\alpha}$ obtained using these four models are shown in Fig. 4.10. While the transformation leaves certain artifacts in the posterior distributions because of singularities in the transformation (as discussed in great detail in [123] and [124]), the upper limits on $\sqrt{\alpha}$ obtained with the four models are consistent with each other. The 90% confidence upper limit on $\sqrt{\alpha}$ with a single deformation to the phase (the red curve in Fig. 4.10) is 5.0 km, while that obtained using the full, three term modification to the GW phase (the green curve in Fig. 4.10) is 1.5 km, leading to a strengthening factor of about 3. The additional information incorporated into the waveform through the higher order deformation only serve to improve the constraint on $\sqrt{\alpha}$. This comparison provides further evidence that constraints derived from leading PN order deformations are robust to future work on deriving higher PN order corrections.

The bias in the distribution on $\sqrt{\alpha}$ coming from the generic parametrizations is a known issue [123, 124], and is not of much concern in the present context. The issue fundamentally lies with the Jacobian of the transformation between the two parametrizations, and causes the derived prior distribution on $\sqrt{\alpha}$ to go to zero in the GR limit when mapping the generic parametrization to ssGB. When transforming distributions between two different basis, one must account for the Jacobian of the transformation as

$$p(\theta_1) = p(\theta_2) \frac{d\theta_2}{d\theta_1}, \quad (4.35)$$

In the present case, we are using one parametrization $\theta_1 \equiv [\Xi \cup \sqrt{\alpha}]$ in ssGB and one parametrization $\theta_2 \equiv [\Xi \cup \bar{\gamma} \cup \sum_i \delta\bar{\psi}_i]$ in the generic framework. From Eq. (4.35), we know we need the Jacobian, $d\theta_1/d\theta_2$, to transform from parametrization 2 to parametrization 1. Here lies the issue, as the component of the Jacobian related to $\sqrt{\alpha}$ and $\bar{\gamma}$ is given as

$$\frac{d\bar{\gamma}}{d\sqrt{\alpha}} \propto (\alpha)^{3/2}, \quad (4.36)$$

using the relation in Eq. (4.26). As this expression goes to 0 in the $\sqrt{\alpha} \rightarrow 0$ limit, the prior will have zero weight for the GR limit. Now the discrepancy is clear: the lack of agreement with the posterior on $\sqrt{\alpha}$ coming from the generic parametrization is not an indication that GR is lacking, but instead an inherent flaw of the parametrization. While this is an indication of a failing of the parametrization, the fact that the prior seems to be widening the constraint on $\sqrt{\alpha}$ instead of artificially shrinking it indicates that any constraints derived from this method related to *upper limits on $\sqrt{\alpha}$* are actually *conservative*.

Graphically, this is illustrated in Fig. 4.11 where the posteriors on $\sqrt{\alpha}$ coming from mapping constraints from the generic models are plotted along side samples from the full, derived prior on $\sqrt{\alpha}$. The samples from the prior were drawn from a uniform distribution for $\bar{\gamma}$ and the usual priors for the source parameters, then mapped to $\sqrt{\alpha}$ in the same way as the analysis using models 3 and 4. The original posterior and prior are completely consistent with GR, as shown in this figure, so the lack of support in the GR limit for $\sqrt{\alpha}$ must come from the transformation itself.

4.8 Conclusions

We have here studied whether the inclusion of higher PN order terms in the modified gravity deformations to the GW phase of inspiraling binaries affects the constraints one can place on these theories. First, we focus on a particular implementation of multi-parameter phase deformation, in which an overall controlling factor contains all of the coupling-constant information of the modified theory. Then, we develop a novel, PN-based prior to impose on parametric deviations, which ensures that the terms added obey certain convergence criteria, as they must if they derive from a PN expansion in the inspiral phase (even in a modified theory).

Our analysis shows that the constraints placed on modified theories with single-parameter ppE waveforms are robust and reliable. More specifically, the inclusion of higher PN order terms in the inspiral phase do not weaken the constraints we can place with single-parameter ppE models. In fact, the inclusion of these terms actually improves the constraints on coupling constants of modified theories, and typically the strengthening of the bound is mild. We verified that these conclusions with an array of Bayesian parameter estimation studies, in which we injected synthetic GR signals and extracted with a variety of single- and multi-parameter

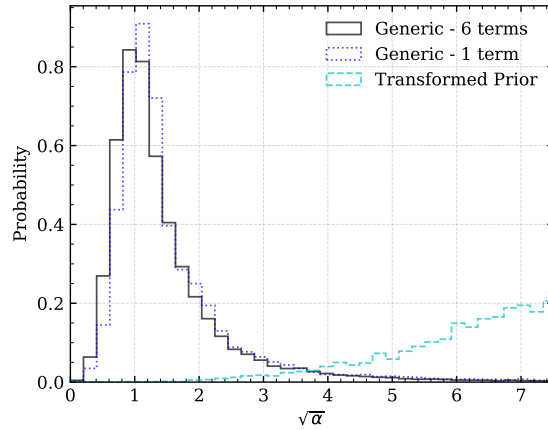


Figure 4.11: Marginalized posterior distributions on $\sqrt{\alpha}$ for a GR injection extracted with the two generic models described in the text. One of the models uses a single deformation while the second model uses a series of six deformations. We here considered the “light” source, defined by Table 4.1, and a 2g detector network. These two data sets are models 3 and 4, described in the text and shown in Fig. 4.10. For comparison, we also show the derived prior on $\sqrt{\alpha}$, calculated by taking a uniform prior on $\tilde{\gamma}$ and the usual priors on source parameters, then mapping it to $\sqrt{\alpha}$ with the same prescription as was used for transforming the two generic model constraints. This figure illustrates the interesting behavior around $\sqrt{\alpha} \rightarrow 0$ (the GR limit). The fact that the derived prior on $\sqrt{\alpha}$, as transformed from a uniform prior on $\tilde{\gamma}$, disallows $\sqrt{\alpha} = 0$ explains the bias in the posteriors for models 3 and 4. The issue is related to the Jacobian of the transformation, discussed in the text, and is not of serious concern in analysis such as these.

ppE models. We further considered a specific theory, ssGB gravity, to exemplify our findings with a concrete set of deformations to the phase. As expected from our generic analysis, the inclusion of higher PN order terms in the ssGB inspiral phase does not weaken bounds obtained with leading order waveforms. Instead, the higher PN order terms improve the bounds on the ssGB coupling parameter, but only by about a factor of 3. These results are consistent with a very recent analysis of ssGB theory with the same higher PN order model we use here on real aLIGO/Virgo data [298].

One can compare our methodology to other techniques to measure multiple phase deformations simultaneously. One such methodology is through the use of a principal component analysis (PCA) on tests of GR [299–301]. In this method, the phase deformations are combined through certain linear transformations, so that covariances are minimized. To do this, samples are first drawn from the posterior distribution with a waveform model that includes multiple ppE phase deformations, and then this posterior distribution is decomposed into its eigenvectors. Constraints are placed on linear combinations of deformations that produce the tightest constraints (the eigenvectors with the largest eigenvalues). Effectively, this just equates to finding a parametrization that optimizes the constraint you can place on these deformations, but it comes with serious drawbacks related to tests of fundamental physics. All information is lost about the physical meaning

of these deformations, as this basis has no connection to calculations performed in modified gravity. This can still be an effective consistency test of GR (in the same category of “residual tests” discussed in Sec. 4.1), but it does not provide information about fundamental physics beyond this specific type of test. Our approach, instead, attempts to stay close to theoretical physics, using physical insight to concentrate our attention on the “relevant” deformation, the controlling factor $\bar{\gamma}$ that encodes the coupling constants of the theory, at the expense of losing less relevant information about the higher order corrections (which are just functions of the system parameters). With this, we still retain the mapping between theories of modified gravity and $\bar{\gamma}$, allowing constraints in the latter to lead to constraints on the coupling constants of modified theories.

These results encourage the use of leading PN order deformations in the inspiral phase to constrain theories of gravity beyond GR. The results will certainly improve with the inclusion of higher PN order deformations, but the bounds will not degrade through their inclusion. This does, however, come with the caveat that we have only considered theories of gravity that allow for well-behaved, series solutions to the field equations, both in the PN expansion and the expansion in the coupling. There are theories that do not conform to these criteria, such as theories of gravity exhibiting spontaneous scalarization [278, 279, 282, 302, 303] or other abrupt, discontinuous transformations. Future work could focus on those theories to attempt to develop a generic framework that can also be applied to them.

Acknowledgements

The authors would like to thank Walter del Pozzo and Emanuele Berti for useful comments on the first draft of this manuscript. We would also like to especially thank Jonathan Gair for detailed comments on the first draft, which led us to add Sec. VI to the final version. This work was partially supported by the Center for AstroPhysical Surveys (CAPS) at the National Center for Supercomputing Applications (NCSA), University of Illinois at Urbana-Champaign. S. P. acknowledges support from the Illinois Center for Advanced Studies of the Universe (ICASU)/CAPS Graduate Fellowship. N. Y. acknowledges support from the Simons Foundation through Grant No. 896696. This work made use of the Illinois Campus Cluster, a computing resource that is operated by the Illinois Campus Cluster Program (ICCP) in conjunction with the National Center for Supercomputing Applications (NCSA) and which is supported by funds from the University of Illinois at Urbana-Champaign.

Chapter 5

Fundamental Physics Implications on Higher-Curvature Theories from the Binary Black Hole Signals in the LIGO-Virgo Catalog GWTC-1

This chapter is based on: *Nair, R.¹, Perkins, S.², Silva, H. O.³ and Yunes, N.⁴ (2019). *Fundamental Physics Implications on Higher-Curvature Theories from the Binary Black Hole Signals in the LIGO-Virgo Catalog GWTC-1. Phys. Rev. Lett. 123, 191101**

Abstract Gravitational-wave astronomy offers not only new vistas into the realm of astrophysics, but also opens an avenue for probing, for the first time, general relativity in its strong-field, nonlinear and dynamical regime, where the theory's predictions manifest themselves in their full glory. We present a study of whether the gravitational-wave events detected so far by the LIGO-Virgo scientific collaborations can be used to probe higher-curvature corrections to general relativity. In particular, we focus on two examples, Einstein-dilaton-Gauss-Bonnet and dynamical Chern-Simons gravity. We find that the two events with a low-mass $m \approx 7M_{\odot}$ BH (GW151226 and GW170608) place stringent constraints on Einstein-dilaton-Gauss-Bonnet gravity, $\alpha_{\text{EdGB}}^{1/2} \lesssim 5.6$ km, whereas dynamical Chern-Simons gravity remains unconstrained by the gravitational-wave observations analyzed.

5.1 Introduction

General relativity (GR) remains our most accurate theory for the gravitational interaction [51]. The centennial theory has passed a plethora of tests ranging from those carried out in the weak-gravitational field and low-velocity regime of our Solar System, to those performed in the extreme, nonlinear and highly-dynamical regime of plunging and merging compact objects, such as neutron stars (NSs) and black holes (BHs) [57, 107]. The agreement between the observations and predictions is dazzling. In turn, any new observation

¹Carried out and supervised many of the calculations in the paper (particularly those related to statistics); wrote the software related to transformations of the constraints; wrote part of the first draft of the manuscript

²Carried out many of the calculations in the paper (particularly those related to statistics); wrote the software related to waveform modelling; wrote the software related to the Fisher analysis; wrote sections of the first draft of the manuscript

³Carried out and supervised many of the calculations in the paper (particularly those related to dynamics of BBH and PN calculations); wrote most of the first draft of the manuscript

⁴Conceived of the project; supervised and advised on the calculations and experimentation; edited the manuscript

that may hint toward a failure of GR will require us to revisit its foundations. Experimental tests of GR not only allow us to place its foundational principles on solid ground, but they also allow us to constrain (or even rule out) contending theories that violate one or more of its pillars. Such contending theories have been developed to address certain outstanding mysteries in recent observations [52, 304], such as the enigmatic late-time acceleration of the Universe [196, 197], the matter-antimatter asymmetry in our Universe [269, 305] and the rotation curve of galaxies [64, 65].

One broad class of modifications to GR that arise naturally in attempts to unify gravity with quantum mechanics are quadratic gravity theories [296]. This class of theories is characterized by the presence of an additional scalar degree of freedom (violating the GR pillar that gravity is mediated by a single metric tensor) coupled to a higher-order curvature scalar. Two preeminent examples of such theories are Einstein-dilaton-Gauss-Bonnet (EdGB) and dynamical Chern-Simons (dCS) gravity [137]. Both of these emerge naturally in the context of grand unified theories (string theory in particular) in the low-energy limit upon dimensional reduction. Phenomenologically, they predict BHs that carry a nontrivial scalar field, resulting in a violation of the strong equivalence principle.

Aside from these theoretical motivations, are EdGB and dCS gravity consistent with experimental tests? Within the confines of our Solar System, the parameterized-post-Newtonian parameters of EdGB gravity are identical to those of GR [306], and therefore the theory survives all experimental tests in this regime. In contrast, dCS gravity contains a nonzero (different from GR) parameter that leads to modifications in the Lense-Thirring precession of spinning bodies [113, 307]. Solar System experiments such as LAGEOS [308] and Gravity Probe B [309] can place constraints on the dCS coupling parameter, but due to the weak curvatures in the Solar System, these constraints are extremely weak [310]. Exquisitely accurate binary-pulsar observations suffer the same fate. The post-Keplerian motion of NS binaries in EdGB and dCS gravity is very similar to that in GR, because the scalar field sourced by such stars is suppressed relative to that created by BHs, which means that constraints with present day binary pulsar observations are not possible [114, 296].

This leaves us with gravitational wave (GW) observations as a last resort. In recent years, considerable effort has been made in modeling the inspiral [110, 311, 312], merger [313–316] and ringdown [317, 318] phases of compact binaries in these two theories. One could then imagine comparing such waveform models against the GW data to determine how small the EdGB and dCS coupling parameters must be in order to be consistent with statistical noise. We build on these efforts and use the constraints on GR deviations obtained by the LIGO-Virgo collaboration (LVC) [18] to analyze whether these two theories can be constrained with the binary BH events detected during the first two observation runs of the LVC. More specifically, we will consider the binary BH events in the LIGO-Virgo Catalog GWTC-1 GW150914 [319, 320], GW151226 [321],

GW170104 [139], GW170608 [322] and GW170814 [323] for which the posteriors on theory-independent GR modifications, obtained through a Markov-Chain-Monte-Carlo (MCMC) exploration of the parameter space, have been made public [57, 324].

5.2 Quadratic gravity

dCS and (decoupled) EdGB theories are defined in vacuum by the Lagrangian density [296]

$$\mathcal{L}_{\text{dCS}} = \kappa R - \frac{1}{2} \nabla_\mu \vartheta_{\text{dCS}} \nabla^\mu \vartheta_{\text{dCS}} + \frac{\alpha_{\text{dCS}}}{4} \vartheta_{\text{dCS}} {}^*RR, \quad (5.1)$$

$$\mathcal{L}_{\text{EdGB}} = \kappa R - \frac{1}{2} \nabla_\mu \vartheta_{\text{EdGB}} \nabla^\mu \vartheta_{\text{EdGB}} + \alpha_{\text{EdGB}} \vartheta_{\text{EdGB}} \mathcal{G}, \quad (5.2)$$

where $\kappa \equiv (16\pi)^{-1}$, g is the determinant of the metric $g_{\mu\nu}$, ${}^*RR = R_{\nu\mu\rho\sigma} {}^*R^{\mu\nu\rho\sigma}$ is the Pontryagin density (constructed in terms of the Riemann tensor and its dual), $\mathcal{G} = R^2 - 4R_{\mu\nu}R^{\mu\nu} + R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma}$ is the Gauss-Bonnet density (where R and $R_{\mu\nu}$ are the Ricci scalar and tensor), and we have used geometric units, in which $c = 1 = G$. These quadratic-in-curvature scalars are coupled⁵ to a massless scalar (pseudo-scalar) field ϑ_{EdGB} (ϑ_{dCS}) through the coupling constants α_{EdGB} (α_{dCS}), with units of $(\text{length})^2$.

To ensure the perturbative well-posedness of these theories, we work in the small-coupling approximation, in which modifications to GR are *small deformations*. This is a justified assumption given the agreement of GR with various observations, GW events included. It is convenient to define the dimensionless parameter $\zeta_{\text{dCS,EdGB}} \equiv \alpha_{\text{dCS,EdGB}}^2 / (\kappa \ell^4)$, where ℓ is the typical mass scale of a system. For the small-coupling approximation to be valid we must have $\zeta_{\text{dCS,EdGB}} < 1$ or $\alpha_{\text{dCS,EdGB}}^{1/2} / m_s \lesssim 0.5$ where m_s is the smallest mass scale involved in the problem. Note that 0.5 is a rough threshold which we use as a proxy for the validity of the approximation.

Consistency with Solar System experiments (in dCS) and with low-mass x-ray binary observations (in EdGB) impose the upper bounds $\alpha_{\text{dCS}}^{1/2} \leq \mathcal{O}(10^8 \text{ km})$ [137, 310] and $\alpha_{\text{EdGB}}^{1/2} \leq \mathcal{O}(2 \text{ km})$ [214].

How can the GWs emitted by BH binaries in these theories be different from GR's predictions? In both theories, BHs support a nontrivial scalar field – dipolar in dCS [325] and monopolar in EdGB [138] – which results in the emission of scalar quadrupole (in dCS) and scalar dipole (in EdGB) radiation during the inspiral. This additional channel for binding energy loss results in modification to the GW phase, which appear at 2PN⁶ (for dCS) and -1PN (for EdGB) order. In dCS gravity, the scalar field also introduces a

⁵In EdGB, the coupling to the Gauss-Bonnet density is usually of exponential form. We here work in the decoupling (effective field theory) limit, in which the exponential is expanded to linear order [296].

⁶In the PN formalism, quantities of interest such as the conserved energy, flux etc. can be written as expansions in (v/c) , where v is the characteristic speed of the binary system and c is the speed of light. $\mathcal{O}((v/c)^n)$ corrections counting from the Newtonian (leading order GR) term are referred to as $(n/2)$ PN-order terms [95, 96].

quadrupolar correction to the binary BH spacetime, introducing 2PN corrections to the binding energy, which in turn affect the GW phase evolution at the same PN order. Hereafter, we use these facts, together with the estimates of the GW model parameters and the posterior distributions released in [18, 57], to investigate how well (if at all) the observed GW events in the LVC catalog can be used to constrain these theories.

5.3 Order of magnitude constraints

It is illuminating to start with a simple order-of-magnitude calculation to assess if the binary BH events detected by LIGO-Virgo can place any constraints on dCS and EdGB gravity. Consider the Fourier domain gravitational waveform $\tilde{h} = A(f) \exp[i\Psi(f)]$, and for simplicity we assume that the spins of the compact objects are (anti)aligned to the orbital angular momentum. Under these assumptions, the leading-order modification to the Fourier phase $\Psi(f)$ takes on the parameterized post-Einsteinian (ppE) form [107] $\Psi = \Psi_{\text{GR}} + \beta (\pi \mathcal{M} f)^b$, where $b_{\text{dCS}} = -1/3$ in dCS gravity (a 2PN correction) and $b_{\text{EdGB}} = -7/3$ in EdGB gravity (a -1PN correction). The amplitude coefficient β is

$$\begin{aligned} \beta_{\text{dCS}} = & -\frac{5}{8192} \frac{\zeta_{\text{dCS}}}{\eta^{14/5}} \frac{(m_1 s_2^{\text{dCS}} - m_2 s_1^{\text{dCS}})^2}{m^2} \\ & + \frac{15075}{114688} \frac{\zeta_{\text{dCS}}}{\eta^{14/5}} \frac{1}{m^2} \left(m_2^2 \chi_1^2 - \frac{305}{201} m_1 m_2 \chi_1 \chi_2 + m_1^2 \chi_2^2 \right) \end{aligned} \quad (5.3)$$

in dCS gravity⁷ [326] and

$$\beta_{\text{EdGB}} = -\frac{5}{7168} \frac{\zeta_{\text{EdGB}}}{\eta^{18/5}} \frac{(m_1^2 s_2^{\text{EdGB}} - m_2^2 s_1^{\text{EdGB}})^2}{m^4}, \quad (5.4)$$

in EdGB gravity [110], where $\mathcal{M} = (m_1 m_2)^{3/5} / (m_1 + m_2)^{1/5}$ is the chirp mass, $\eta = m_1 m_2 / m^2$ (with $m = m_1 + m_2$) is the symmetric mass ratio, $\chi_{s,a} = (\chi_1 \pm \chi_2)/2$ are the symmetric and antisymmetric dimensionless spin combinations with $\chi_i = \vec{S}_i \cdot \hat{L} / m_i^2$ the projections of dimensional spin angular momenta

⁷Our expression for β_{dCS} is different from that presented, e.g. in [109, 112]. First, we corrected an error in the rate of scalar radiation emission $d\delta E^{(\partial)}/dt$, which propagates to the final expression for β_{dCS} [114]. Second, we do not expand the charge s_i^{dCS} to leading order in χ_i as has been done in the past. The reason is the following: the binding energy contribution to β_{dCS} in Eq. (5.3) only contains the quadrupole moment to $O(\chi_i^2)$. In principle, there will be a $O(\chi_i^4)$ correction to it, which will also enter at 2PN order and has not been calculated yet. Thus, unlike in the EdGB case, we cannot calculate the dCS correction at 2PN order to also all orders in the spins. To estimate how robust our bounds are to the absence of this quadrupolar contribution, we include the full expression for s_i^{dCS} , in the calculation of β , as a proxy for the missing $O(\chi_i^4)$ term. We checked that all our results are unaffected by using Eq. (5.6) or its leading order in spin expansion.

$\vec{S}_i$ in the direction of the orbital angular momentum $\hat{L}$ and

$$s_i^{\text{dCS}} = \frac{2 + 2\chi_i^4 - 2(1 - \chi_i^2)^{1/2} - \chi_i^2[3 - 2(1 - \chi_i^2)^{1/2}]}{2\chi_i^3}, \quad (5.5)$$

$$s_i^{\text{EdGB}} = \frac{2[(1 - \chi_i^2)^{1/2} - 1 + \chi_i^2]}{\chi_i^2}, \quad (5.6)$$

are the dimensionless spin and mass-dependent BH scalar charges, to all orders in spin, in both theories [112, 326, 327]. Although β_{dCS} has uncontrolled remainders of $\mathcal{O}(\chi^4)$, β_{EdGB} is valid to all orders in the spin. We can obtain an order-of-magnitude bound on $\zeta_{\text{dCS,EdGB}}$ using the best-fit parameters from GW170608 and doing a crude Fisher matrix analysis⁸. Given that the event is consistent with GR, we can ask how large $\zeta_{\text{dCS,EdGB}}$ can be and yet remain consistent with the event. For sufficiently high signal-to-noise ratio (SNR) ρ , the accuracy at which a parameter θ^a of the GW model can be estimated from the Cramer-Rao bound [125] $\Delta\theta^a = \sqrt{(\Gamma^{-1})^{aa}}$ where the Fisher matrix is

$$\Gamma_{ab} \equiv 4 \text{Re} \int_{f_{\min}}^{f_{\max}} \frac{\partial_a \tilde{h}(f) \partial_b \tilde{h}^*(f)}{S_n(f)} df, \quad (5.7)$$

and the asterisk stands for complex conjugation. The partial derivatives are taken with respect to the model parameters θ^i and $S_n(f)$ is the spectral noise density of the detector. The integration limits denote the lower and upper cut-off frequencies at which the detector operates. For a rough estimate, it suffices to neglect correlations between model parameters, and thus, Γ_{ab} is approximately diagonal. With this, one then finds that the variance satisfies $(\Delta\zeta)^2 = 1/\Gamma_{\zeta\zeta}$, which can be evaluated analytically assuming white noise. This matrix element is dominated by the lower limit of integration $f_{\min}$, and thus, one finds that

$$(\Delta\alpha_{\text{dCS,EdGB}})^{1/2} \gtrsim \left(1 - \frac{3b_{\text{dCS,EdGB}}}{2}\right)^{1/8} \frac{(\pi\hat{\mathcal{M}}f_{\min})^{-b_{\text{dCS,EdGB}}/4}}{(16\pi\hat{\rho})^{1/4}} \frac{\hat{m}}{\hat{\beta}_{\text{dCS,EdGB}}^{1/4}}. \quad (5.8)$$

where the overhead hat stands for the best-fit values, with $\zeta_{\text{dCS,EdGB}}$ set to unity in $\hat{\beta}_{\text{dCS,EdGB}}$. As the individual spins χ_i could not be resolved for the events we are considering, we assign $\chi_1 = \chi_{\text{eff}}(m/m_1)$ and $\chi_2 = 0$ to proceed. Using $f_{\min} = 10$ Hz and the SNR $\hat{\rho}$ and median values for m_1 , m_2 and χ_{eff} , we obtain $(\Delta\alpha_{\text{dCS}})^{1/2} \approx 1.1$ km and $(\Delta\alpha_{\text{EdGB}})^{1/2} \approx 1.0$ km at 90% credibility. These bounds agrees well with the forecast made in [326] for dCS and in [115] for EdGB.

⁸We use this particular event as an example because it will allows us to compare our analytical estimate with more robust calculations later.

5.4 Fisher-estimated constraints on LIGO-Virgo data

We also perform a fully numerical calculation of the Fisher matrix, by modeling the binaries with the phenomenological waveform template IMRPhenomD [106, 169]. We make similar assumptions for the fiducial parameters as we made to obtain the order of magnitude constraints and consider 5 GW events, GW150914, GW151226, GW170104, GW170608 and GW170814 (cf. Table III in [18]). The bounds obtained for the two most constraining events, GW151226 and GW170608, are shown in Table 5.1 and they are in good agreement with our order-of-magnitude calculation for both theories.. The Fisher-estimated constraints for dCS gravity are not shown because they violate the small coupling approximation, as we will discuss in more detail below.

System	Method	$\alpha_{\text{EdGB}}^{1/2}$ [km]	$\alpha_{\text{dCS}}^{1/2}$ [km]
Current	Frequentist	2	10^8
GW151226	estimate	0.9	0.5
	Fisher	6.0	—
	Bayesian	5.7	—
GW170608	estimate	1.0	1.1
	Fisher	3.9	—
	Bayesian	5.6	—

Table 5.1: Current constraints on EdGB and dCS gravity from low-mass x-ray binary and Solar System observations respectively, with the Fisher-estimated constraints, and Bayesian constraints using LVC (testing GR) posteriors for GW151226 and GW170608 [57, 324]

5.5 Bayesian-estimated constraints on LIGO-Virgo data

The LVC recently released constraints on model-independent deviations from GR to check consistency of the GW events with GR predictions [57, 324]. The model used to capture these deviations is a variant of IMRPhenomPv2 [169, 328–330], where parameterized relative shifts in the PN coefficients of the Fourier phase of IMRPhenomPv2 are introduced, namely

$$\phi_i \rightarrow \phi_i (1 + \delta\phi_i) , \quad (5.9)$$

with $\delta\phi_i$ then treated as additional *free* parameters in the model. This modification is nothing but an implementation of the ppE framework [107, 117], as shown explicitly in [112], with the mapping

$$\beta_{\text{dCS}} = \frac{3}{128} \phi_4 \delta\phi_4 \eta^{-4/5} , \quad (5.10a)$$

$$\beta_{\text{EdGB}} = \frac{3}{128} \delta\phi_{-2} \eta^{2/5} , \quad (5.10b)$$

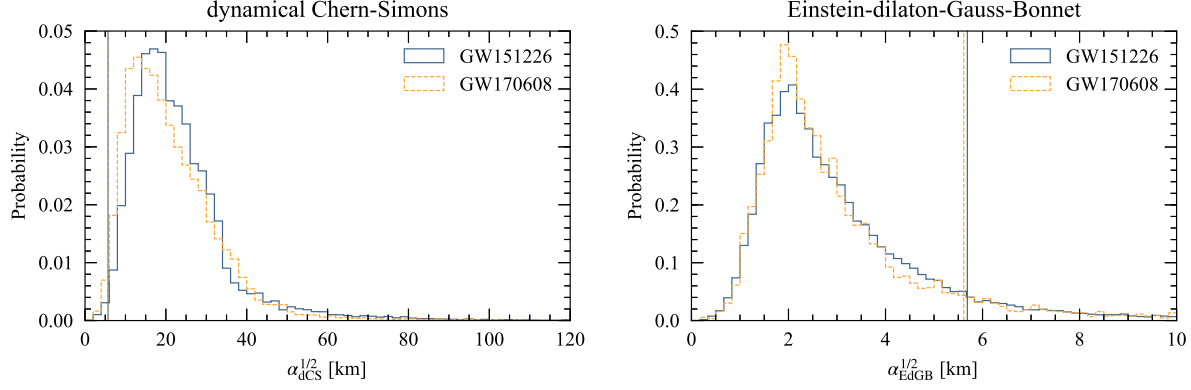


Figure 5.1: Posterior distributions of $\alpha_{\text{dCS}}^{1/2}$ (left panel) and $\alpha_{\text{EdGB}}^{1/2}$ (right panel) obtained using GW151226 and GW170608. For the GW events shown in both panels, $m_2/M_\odot = 7.7^{+2.2}_{-2.6}$ (GW151226) and $m_2/M_\odot = 7.6^{+1.3}_{-2.1}$ (GW170104) at 90% credibility. This implies that the small-coupling approximation is valid only when $\alpha_{\text{dCS,EdGB}}^{1/2} \lesssim 5.6$, shown as vertical lines in the plots. For dCS gravity (left-panel) we see that most of the support of the posterior distributions of these two events lays *passed* the bounds set by the small-coupling approximation. Consequently, one cannot place constraints on $\alpha_{\text{dCS}}^{1/2}$ with these two events. For EdGB gravity (right-panel) most ($> 90\%$) of the posteriors' support lays *within* the bound, therefore allowing us to constrain the theory with these two events. For the other three events, which contain a large m_2 ($\gtrsim 13 M_\odot$) BH [18], the vertical lines are pushed towards the left, leaving most of the posterior's support outside the small-coupling approximation bound. We stress that the location of the peaks in the posteriors *are not an indication of a deviation from GR*. Instead, as detailed in the main text, the lack of support at zero is an artifact of the choice of the sampling variable $\delta\phi_i$.

where ϕ_4 is the GR coefficient of the Fourier phase at 2PN order (cf. Appendix B in [106]). Since the predictions from both dCS and EdGB theories can be mapped to the ppE framework, one can propagate the LIGO-Virgo bounds on $\delta\phi_{-2}$ and $\delta\phi_4$ to constraints on the dCS and EdGB coupling constants. More specifically, we use the posteriors provided by the LVC on $\delta\phi_{-2}$ and $\delta\phi_4$ to first obtain constraints on β_{dCS} and β_{EdGB} , which we then translate into constraints on $\alpha_{\text{dCS}}^{1/2}$ and $\alpha_{\text{EdGB}}^{1/2}$ using Eqs. (5.3)-(5.4).

The 90% constraints on $\alpha_{\text{dCS}}^{1/2}$ and $\alpha_{\text{EdGB}}^{1/2}$ are shown in Table 5.1 for the two most constraining events (GW151226 and GW170608) and the corresponding posterior distributions are shown in Fig. 5.1. The Fisher estimates, although quite close to the constraints using posteriors derived from GW data, are over-optimistic since they assume a Gaussian posterior around the peak, which we see in Fig. 5.1 is not correct. Moreover, since the Fisher analysis is a point estimate, it is difficult to gauge its robustness. On the other hand, a MCMC exploration of the posterior surface helps us evaluate *explicitly* how much support the posterior distributions have in the regions of validity set by the small-coupling approximation.

Constraints on quadratic gravity theories that employ the small-coupling approximation are robust only provided the former satisfy the requirements of the latter. For the systems considered, this translates to $\alpha_{\text{dCS,EdGB}}^{1/2} \lesssim 5.6$ km, which is shown with vertical lines in Fig. 5.1. For dCS gravity (left panel of Fig. 5.1),

more than 99% of the posterior distribution of $\alpha_{\text{dCS}}^{1/2}$ lies *beyond* this region of validity for GW151226 and GW170608 and for all the other events we considered.

Consequently, *we cannot place constraints on dCS gravity with the events for which the posteriors samples obtained by LIGO-Virgo have been released.*

For EdGB, the situation is strikingly different. As one can observe in the right panel of Fig. 5.1, more than 90% of the posterior distribution falls *within* the requirements of the small-coupling approximation for the GW151226 and GW170608 events. *This implies that a 90% bound of $\alpha_{\text{EdGB}}^{1/2} \lesssim 5.6 \text{ km}$ is statistically meaningful and can be placed on EdGB gravity using these two events.* This is not the case for the other events (GW150914, GW170104 and GW170814), for which constraints would violate the small coupling approximation.

We emphasize that the location of the peaks in the posteriors of Fig. 5.1 *do not indicate a deviation from GR*. Rather, the lack of support at zero is an artifact of the choice of the sampling variable $\delta\phi_i$ and its functional dependence on $\alpha_{\text{dCS,EdGB}}^{1/2}$. A uniform prior in $\delta\phi_i$ translates to a non-uniform prior on $\alpha_{\text{dCS/EdGB}}^{1/2}$ with almost no support near $\alpha_{\text{dCS,EdGB}}^{1/2} = 0$. One can re-weight the $\alpha_{\text{dCS/EdGB}}^{1/2}$ posteriors with the priors to obtain better estimates, albeit at the cost of introducing binning errors close to $\alpha_{\text{dCS/EdGB}}^{1/2} = 0$.

Alternatively, this issue could be avoided by sampling directly in $\alpha_{\text{dCS,EdGB}}$ instead of in the generic parameter $\delta\phi_i$. We expect that this would shift our 90% bound to the left, thereby improving our bounds, and hence our constraints are *conservative and robust* to changes in the sampling variable.

The fact that GW151226 and GW170608 have more constraining power than their cousins is not surprising. These two events were produced by binaries in which the secondary BH had the lowest mass ($m_2 \approx 7M_\odot$) of all events in the catalog. Quadratic gravity theories introduce new length scales, and deviations from GR are thus proportional to the curvature scale, which for BH binaries scales inversely with the square of the lowest mass, m_2^{-2} . Hence one can expect the largest deviations for GW151226 and GW170608 and thus, the strongest constraints. In dCS gravity, the modifications enter at 2PN order, and thus, they are much more weakly constrained than the EdGB modifications, which enter at -1PN order. This deterioration in the constraint then implies that a large percentage of the posterior weight is outside the regime of validity of the small coupling approximation, rendering the constraint invalid.

5.6 Fundamental physics implications

Our results dramatically constrain EdGB gravity, essentially confining deviations from GR due to this theory down to the horizon scale of stellar mass BHs. These constraints are competitive with those obtained in [214]

($\alpha_{\text{EdGB}}^{1/2} \lesssim 2$ km at 95% confidence level) from the orbital decay on the BH low-mass x-ray binary A0620-00, which probes the theory in a different energy scale. Our constraints, however, have the advantage of being robust to astrophysical systematics, unlike those placed in [214] which require assumptions about the mass transfer efficiency and the specific angular momentum carried by stellar winds.

The constraint we have placed on (decoupled) EdGB gravity is stringent, limiting this type of quantum-inspired violation of the strong equivalence principle, the strength of the scalar monopole charge carried by black holes, and the possibility of using EdGB gravity to explain the late-time acceleration of the universe. However, our constraints do not directly apply to other functional couplings between the Gauss-Bonnet density and a scalar field. For example, in models where BHs acquire charges through spontaneous scalarization [278, 279, 331–333], BHs are identical to GR unless they fall within certain mass intervals (at fixed coupling parameter of the theory) and thereby can (in principle) mimic binary BH mergers in GR.

Our results also have important implications for restricting parity-violation in the gravitational interaction. Recently, a broad class of ghost-free, parity-violating theories, which in four-dimensions requires the presence of a massless scalar field, was presented [334]. In [335–337], these theories were tested against the exquisite constraint obtained on the speed of GW propagation from the binary NS event GW170817/GRB 170817A, which estimated that c_{GW} is the same as the speed of light in vacuum to one part in 10^{15} . dCS gravity is the only ghost-free, parity-violating theory in four-dimensions that is consistent with this constraint [88, 335]. Therefore, our results combined with those by [335], leave dCS as the single subclass of the broad set of parity-violating theories of gravity which remains consistent with observations.

Future work could focus on constraints on other modified theories within the broad class of quadratic gravity models [296]. Alternatively, one could include GW amplitude corrections due to EdGB and dCS gravity to determine whether GW constraints become stronger [109]. Finally, one could study how well future ground-based and space-based detectors could constraint quadratic gravity theories, or the type of system that would be ideal to place constraints the hitherto evasive dCS gravity.

Acknowledgements This work was supported by NASA grants No. NNX16AB98G and No. 80NSSC17M0041. We thank Kent Yagi for discussions and checking Eq. (5.3). We thank Alejandro Cárdenas-Avendaño and Katerina Chatziioannou for helpful discussions and we thank Sandipan Sengupta for useful comments on the draft.

Chapter 6

Improved gravitational-wave constraints on higher-order curvature theories of gravity

This chapter is based on: *Perkins, S.¹, Nair, R.², Silva, H. O.³ and Yunes, N.⁴ (2021). Improved gravitational-wave constraints on higher-order curvature theories of gravity. *Phys. Rev. D* 104, 024060*

Abstract Gravitational wave observations of compact binaries allow us to test general relativity (and modifications thereof) in the strong and highly-dynamical field regime of gravity. Here we confront two extensions to general relativity, dynamical Chern-Simons and Einstein-dilaton-Gauss-Bonnet theories, against the gravitational wave sources from the GWTC-1 and GWTC-2 catalogs by the LIGO-Virgo Collaboration. By stacking the posterior of individual events, we strengthen the constraint on the square root of the coupling parameter in Einstein-dilaton-Gauss-Bonnet gravity to $\sqrt{\alpha_{\text{EdGB}}} < 1.7$ km, but we are unable to place meaningful constraints on dynamical Chern-Simons gravity. Importantly, we also show that our bounds are robust to (i) the choice of general-relativity base waveform model, upon which we add modifications, (ii) unknown higher post-Newtonian order terms in the modifications to general relativity, (iii) the small-coupling approximation, and (iv) uncertainties on the nature of the constituent compact objects.

6.1 Introduction

General Relativity (GR) is our best description of gravity to date, fully consistent with all experiments performed to date. This is no small feat, as the precision and energy scales at which we now probe the gravitational interaction have been rapidly increasing. With the LIGO-Virgo collaboration (LVC) directly detecting gravitational waves (GWs) [18] through binary black hole (BBH), neutron star-black hole (NSBH), and neutron star-neutron star (NSNS) mergers, and the Event Horizon Telescope collaboration resolving the shadow of a supermassive black hole (BH) [194], we are testing our best theory of gravity like never before.

¹Carried out most of the calculations in the paper (particularly those related to statistics); wrote the software related to waveform modelling; wrote the software related to the Fisher analysis; wrote the MCMC software; wrote a majority of the first draft of the manuscript

²Supervised many of the calculations in the paper (particularly those related to statistics); supervised the application of the MCMC software and numerical experimentation; wrote part of the first draft of the manuscript

³Supervised many of the calculations in the paper (particularly those related to dynamics of BBH and PN calculations); wrote sections of the first draft of the manuscript

⁴Conceived of the project; supervised and advised on the calculations and experimentation; edited the manuscript

With such impressive experimental support achieved by GR, one might wonder why resources should be expended on exploring alternative gravity theories. From an observational standpoint, open questions like the existence or nature of dark matter [64, 65] and dark energy [196, 197], the quantum mechanical description of gravity, and the puzzle of the matter-antimatter asymmetry [269, 305] are being widely studied. For some of these questions, the problem might be resolved without additional forms of matter or energy, but rather by modifying the gravitational theory. This can come in the form of the addition of auxiliary fields and higher curvature terms to the gravitational action [52, 154, 304]. Beyond our desire to describe unexplained phenomena, there are purely null-test reasons to consider as well. The fundamental tenants of GR can be succinctly summarized by Lovelock’s theorem [338, 339]. While GR is the only theory which satisfies the stipulations of this theorem, there is no fundamental reason why the true theory of gravity has to respect these criteria.

A very interesting class of theories from these perspectives are known as quadratic theories of gravity [296]. This class of theories introduce terms to the gravitational action that are quadratic in curvature invariants, composed of quantities like the Ricci scalar, Ricci tensor, and Riemann tensor. Even though some of these theories lead to higher-than-second order field equations, these higher order terms can be tamed when treating them as arising from an effective field theory [340]. In this work, we focus on dynamical Chern-Simons (dCS) [137] and Einstein-dilaton-Gauss-Bonnet (EdGB) [138] theories, which have an extra scalar degree of freedom coupled to the Pontryagin density and Gauss-Bonnet density, respectively.

Modifications to the gravitational action, like those introduced in dCS and EdGB, will produce modified field equations and, in turn, result in dynamics that differ from those predicted in GR. These difference can become imprinted in observables, like the GWs produced by inspiraling and merging binaries, which opens an avenue for discerning the true description of gravity obeyed by Nature [112, 327, 341]. Since LVC’s first detection, the veritable treasure trove of information about strong field gravity contained in these signals has resulted in a gold rush of sorts, with a multitude of recent works looking for “smoking-gun” signatures of deviations from GR (see e.g. [58, 139, 195, 323, 342–347]), including our previous work specifically focused on the subject of quadratic theories of gravity [124].

Theory	Coupling	New constraint	Prior constraint
EdGB	Gauss-Bonnet [Eq. (6.2a)]	1.7 km	2–5.6 km [124, 214]
dCS	Pontryagin [Eq. (6.2b)]	–	8.5 km [223]

Table 6.1: Current state of quadratic theories of gravity and the bounds resulting from this work. The columns are, in order, the theory, the topological invariant that couples non-minimally to the scalar field, the constraints as the result of this work, and the previous strongest constraints on these theories.

In this work, we first extend and confirm assertions made in our previous work [124] by exploring the

posterior surface of the parameter space through extensive numerical analysis of current LVC data and through a parametrization that is naturally tailored to this problem. This removes any sampling artifacts that were introduced in our previous work as the result of recycling the theory agnostic analysis by the LVC. With this in hand, we then place constraints on beyond-GR modifications through the combination of multiple detections using a fully Bayesian “stacking” of posterior probability density functions. In all of these constraints, we ensure that our bounds are consistent with the small-coupling approximation, which is a necessary requirement for the waveforms used to be valid. The key results of this analysis are summarized in Table 6.1, which show that the stacked constraints are about 3 times stronger than previous GW constraints [124] and about 15% stronger than other model-dependent astrophysical bounds [214]. Therefore, the stacked constraints found in this paper are the strongest yet placed on this theory.

But are these constraints we find through stacking robust? We explore this question by studying the impact of unknown, higher-order post-Newtonian (PN) corrections introduced by beyond-GR modifications, as well as on systematic biases that could be introduced by using different GR base waveform models. The lack of information about higher PN order modifications has been of concern recently [108, 136, 348, 349]. We show, however, that given basic assumptions about the nature of any mathematically consistent PN modification to the non-GR waveform, our results are robust. This means that, although the calculation of higher PN order terms in the non-GR waveform may be useful to better understand the late-inspiral and merger, these corrections are not necessary to place constraints on the theory because the constraints are dominated by the leading-order PN terms. Concerning the impact of the specific choice of waveform on our analysis, we compared the constraints obtained using two different GR waveform base models, onto which leading-order modifications are appended. We find that the impact of these systematics on the constraints we place are negligible. These results therefore prove that the stacked constraints we obtained here for EdGB gravity are robust to unknown PN corrections in both the GR and the non-GR sectors.

Can one place stronger constraints than these joint constraints by looking at particular events in the GWTC-2 catalog? Indeed, as pointed out in [88], there are some binary systems that, if detected, would be ideal to place constraints on quadratic gravity theories. In particular, the detection of a mixed (spinning BH and neutron star (NS)) binary could lead to much more stringent constraints than those obtained from BBHs alone. There are three events in the GWTC-2 catalog that could potentially be used for such tests: GW190425 [350], GW190426 [351], and GW190814 [352]. Unfortunately, GW190425 is very likely to be a NS binary and not a mixed binary, in spite of the lack of an electromagnetic counterpart, because there are no known formation channels for BBHs in this mass range [350]. If GW190425 was produced by a NS binary, then monopole scalar charge would not be present [296, 353, 354], thus preventing scalar dipole emission, and

thereby suppressing non-GR modifications. Event GW190426 is probably a mixed binary, but unfortunately again, its false alarm rate (FAR) is extremely high (1.4 yr^{-1} - the highest rate in the entire catalog) and its signal-to-noise ratio (SNR) is not large enough (10.1) [351], leading to very wide and somewhat uninformative posteriors on the extracted parameters. In fact, the LVC analysis concerned with tests of GR *excludes* this event due to its low significance. Finally, GW190814 can potentially be a BBH or NSBH binary, but the conclusions one arrives at differ drastically between the two assumptions. Lacking a definite answer on the exact nature of this event, the results subsequently derived from this event are suspect. We will show in this paper how all of this plays out (when considering EdGB theory in particular) and why these three events, among others, can therefore not be robustly used to place constraints on these theories.

We emphasize that the prior assumptions on the composition of a binary can greatly affect the posterior conclusions obtained from parameter estimation, and this fact is not new. Indeed, imagine we would like to constrain the equation of state of supranuclear matter, instead of testing GR, and imagine we insisted on using GW190814 to do this. Through a Bayesian analysis, the LVC determined that this event was produced by an asymmetric binary with masses $\sim 23M_{\odot}$ and $\sim 2.6M_{\odot}$. The heavier object is obviously a BH. If we assume the lighter object is also a BH, then we cannot place any constraints on the NS equation of state. However, if we assume the lighter object is a NS, then we can place very stringent constraints on the equation of state, since many of them do not allow for NSs that massive. Clearly, given that the data cannot identify the nature of the object, we cannot use this event to constrain the equation of state. Constraining quadratic gravity with this event (and other similar ones) suffers from the same problems, and thus such constraints are not robust.

The remainder of the paper is organized as follows. In Sec. 6.2, we will outline the basics of the two theories we will be considering, including the GW phase modification introduced by them. We will continue in Sec. 6.3 with a discussion of the methodology (Sec. 6.3.1), preliminary estimates of our constraints (Sec. 6.3.3), our considerations when selecting sources (Sec. 6.3.2 and Appendix ??), and our final constraints (Sec. 6.3.4 and 6.3.5) on these theories. Moreover, we establish the robustness of these results by analyzing the impact of waveform systematics and truncated expansions are discussed in Sec. 6.4.1 and Sec. 6.4.2, respectively. Finally, we present our conclusions in Sec. 6.5. Throughout this work, we use geometrical units $G = c = 1$.

6.2 Basics of Quadratic Gravity

We consider two higher-curvature modified theories of gravity, namely EdGB and dCS, which are described by the action

$$S = \int d^4x \sqrt{-g} [\kappa R - (1/2)(\nabla\vartheta)^2 + \mathcal{L}_{\text{EdGB,dCS}}] , \quad (6.1)$$

where $\kappa = 1/(16\pi)$, g is the determinant of the metric $g_{\mu\nu}$, and ϑ is a (pseudo) scalar field that couples to the Gauss-Bonnet $\mathcal{G}$ (Pontryagin ${}^*RR$) curvature invariants as⁵

$$\mathcal{L}_{\text{EdGB}} = \alpha_{\text{EdGB}} \vartheta_{\text{EdGB}} \mathcal{G} , \quad (6.2a)$$

$$\mathcal{L}_{\text{dCS}} = (\alpha_{\text{dCS}}/4) \vartheta_{\text{dCS}} {}^*RR , \quad (6.2b)$$

where

$${}^*RR = R_{\nu\mu\rho\sigma} {}^*R^{\mu\nu\rho\sigma} , \quad (6.3)$$

constructed from the Riemann tensor $R_{\mu\nu\rho\sigma}$ and its dual ${}^*R^{\mu\nu\rho\sigma} = (1/2)\epsilon^{\mu\nu\alpha\beta}R_{\alpha\beta}{}^{\rho\sigma}$, where $\epsilon_{\mu\nu\rho\sigma}$ is the totally antisymmetric Levi-Civita symbol, and

$$\mathcal{G} = R^2 - 4R_{\mu\nu}R^{\mu\nu} + R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma} , \quad (6.4)$$

where R and $R_{\mu\nu}$ are the Ricci scalar and tensor. Finally, α_{EdGB} and α_{dCS} are coupling constants, both with units of $[\text{length}]^2$. GR is recovered in the limit $\alpha_{\text{EdGB,dCS}} \rightarrow 0$.

To ensure the perturbative well-posedness of both theories, we work in the small-coupling approximation, in which modifications to GR are considered small. To establish the small-coupling approximation, it is convenient to define the dimensionless perturbative parameter

$$\zeta_{\text{dCS,EdGB}} = \alpha_{\text{dCS,EdGB}}^2 / (\kappa \ell^4) , \quad (6.5)$$

where ℓ is the typical mass scale of a system. For the small-coupling approximation to be valid, we must have $\zeta_{\text{dCS,EdGB}} < 1$. Here we set a rough threshold for the validity of the approximation by requiring that

$$\alpha_{\text{dCS,EdGB}}^{1/2} / m_s \lesssim 0.5 , \quad (6.6)$$

⁵Strictly speaking, the name ‘‘Einstein-dilaton-Gauss-Bonnet’’ gravity refers to an interaction Lagrangian $\propto \exp(\vartheta)\mathcal{G}$ (see e.g., [138]), which for small scalar field amplitudes reduces to the linearized coupling used here and oftentimes called ‘‘shift-symmetric’’ scalar-Gauss-Bonnet gravity in the literature.

where m_s is the smallest mass scale involved in the problem. Hereafter, we use α to denote both α_{dCS} and α_{EdGB} , where the distinction between these two should be obvious by context.

In both the theories we consider here, BHs support a nontrivial scalar field which is dipolar in dCS (see e.g. [325, 355]) and monopolar in EdGB (see e.g. [138, 295, 356]). This results in the emission of scalar quadrupole (in dCS) and scalar dipole (in EdGB) radiation during BBH inspirals. This additional channel for binding energy loss modifies the GW phase, with leading order correction appearing at 2PN⁶ (for dCS) and -1PN (for EdGB) order. In dCS gravity, the scalar field also results in a quadrupolar correction to the binary BH spacetime, introducing 2PN corrections to the binding energy, which in turn affect the GW phase evolution at the same PN order.

To be more precise, if we write the GW signal in the Fourier domain as

$$\tilde{h}(f) = A(f) \exp[i\Psi(f)], \quad (6.7)$$

both the theories modify the phase schematically as

$$\Psi = \Psi_{\text{GR}} + \beta u^b, \quad (6.8)$$

i.e., by an additional contribution relative to the GR phase Ψ_{GR} , where $u = (\pi\mathcal{M}f)^{1/3}$. Here, $\mathcal{M} = m\eta^{3/5}$ is the chirp mass, $\eta = m_1 m_2 / m^2$ is the symmetric mass ratio, and $m = m_1 + m_2$ is the total mass. Equation (6.8) is cast in the form suggested in the parametrized post-Einsteinian (ppE) formalism [107, 109, 117], where b corresponds to the PN order at which the leading order correction modifies the GR phase Ψ_{GR} and β is a constant that controls the amplitude of this modification. In dCS and EdGB gravity, the exponent b takes the values

$$b_{\text{dCS}} = -1, \quad \text{and} \quad b_{\text{EdGB}} = -7, \quad (6.9)$$

respectively. The amplitude coefficients β for these theories are [110, 326]

$$\begin{aligned} \beta_{\text{dCS}} = & -\frac{5}{8192} \frac{\zeta_{\text{dCS}}}{\eta^{14/5}} \frac{(m_1 s_2^{\text{dCS}} - m_2 s_1^{\text{dCS}})^2}{m^2} + \frac{15075}{114688} \frac{\zeta_{\text{dCS}}}{\eta^{14/5}} \\ & \times \frac{1}{m^2} \left(m_2^2 \chi_1^2 - \frac{350}{201} m_1 m_2 \chi_1 \chi_2 + m_1^2 \chi_2^2 \right), \end{aligned} \quad (6.10)$$

$$\beta_{\text{EdGB}} = -\frac{5}{7168} \frac{\zeta_{\text{EdGB}}}{\eta^{18/5}} \frac{(m_1^2 s_2^{\text{EdGB}} - m_2^2 s_1^{\text{EdGB}})^2}{m^4}, \quad (6.11)$$

⁶In the PN formalism, quantities of interest such as the conserved energy, flux, etc., can be written as expansions in v/c , where v is the characteristic speed of the binary and c is the speed of light. $\mathcal{O}((v/c)^n)$ corrections counting from the Newtonian (leading order GR) term are referred to as $(n/2)$ -PN order terms [95, 96].

where

$$\chi_i = (\vec{S}_i \cdot \hat{L})/m_i^2, \quad (6.12)$$

is the dimensionless spin parameter, obtained from the projection of the dimensionless spin angular momentum vector $\vec{S}_i$ onto the direction of the orbital angular momentum $\hat{L}$. Finally, s_i are the spin and mass dependent BH scalar charges, valid to all orders in spin [112, 326, 327],

$$s_i^{\text{dCS}} = \frac{2 + 2\chi_i^4 - 2\varsigma_i - \chi_i^2(3 - 2\varsigma_i)}{2\chi_i^3}, \quad (6.13)$$

$$s_i^{\text{EdGB}} = \frac{2\varsigma_i(1 - \varsigma_i)}{\chi_i^2}, \quad (6.14)$$

where $\varsigma_i = (1 - \chi_i^2)^{1/2}$. The expression for β_i^{dCS} has uncontrolled remainders of $\mathcal{O}(\chi^4)$ which have not been calculate yet, while the expression for β_i^{EdGB} is valid for all orders in spin. We remark that the amplitude $A(f)$ is also modified in both theories, but these are subdominant relative to the changes to the phase [203] and we do not consider them here. Note that the EdGB and dCS modifications are completely controlled by the ζ parameter (6.5) (or equivalently by $\sqrt{\alpha}$), and thus, this is the only parameter that needs to be constrained by the data.

6.3 Constraints on Quadratic Gravity

6.3.1 Bayesian inference in GW science

Scientific conclusions drawn from GW observations are typically constructed through Bayesian inference. One can express Bayes' theorem in the context of GW analysis as:

$$P(\boldsymbol{\theta}|D) = \frac{P(D|\boldsymbol{\theta})P(\boldsymbol{\theta})}{P(D)}, \quad (6.15)$$

where D is the data coming from the detector network and $\boldsymbol{\theta}$ is the vector of parameters uniquely defining the waveform. For this work, we use the IMRPhenomPv2 [106, 169, 202] waveform, defined in GR by the parameter vector $\boldsymbol{\theta} = \{\alpha', \delta, \psi, \iota, \phi_{\text{ref}}, t_c, D_L, \mathcal{M}, \eta, a_1, a_2, \cos \theta_1, \cos \theta_2, \phi_1, \phi_2\}$, where α' and δ define the right ascension and declination of the binary in the sky (not to be confused with the $\sqrt{\alpha}$ coupling parameter of the modified theory), ψ is the polarization angle defined with respect to the Earth centered coordinates (consistent with the definition in [357]), ι is the inclination angle of the binary's orbital angular momentum at a reference frequency of 20 Hz, ϕ_{ref} is the phase at a reference frequency, t_c is the time of coalescence, D_L is the luminosity distance, $\mathcal{M}$ is the detector frame chirp mass, η is the symmetric mass ratio, a_1 (a_2) is the

dimensionless spin magnitude of the larger (smaller) BH, θ_1 (θ_2) is the angle between the BH spin vector and the orbital angular momentum of the binary for the larger (smaller) BH, and ϕ_1 (ϕ_2) is the in plane orientation of the spin vector for the larger (smaller) BH. To consider dCS (EdGB) waveforms, we append one additional parameter, $\sqrt{\alpha_{\text{dCS}}}$ ($\sqrt{\alpha_{\text{EdGB}}}$), which has units of length and defines the typical length scale at which these modifications take effect.

The terms in Eq. (6.15) are identified as follows - $P(D|\theta)$ is the “likelihood” and describes the probability of observing the data D given the parameters θ of an assumed model. In GW studies, this quantity is calculated for a given waveform model describing the GW, and a noise model that describes the detector characteristics. $P(\theta|D)$ is the “posterior” which is the quantity of interest in all parameter estimation studies. It is the probability density on the parameters θ given the data D . $P(\theta)$ is the “prior” which represents our knowledge of the parameters θ before we analyze the data. This knowledge can be a result of a previous independent measurement or could be because of some theoretical/physical restrictions. For example, even without looking at the data, we can assume confidently that the mass of a BH is non-negative. For this work, we will use priors that are uniform in α' , $\sin \delta$, ϕ_{ref} , t_c , ψ , $\cos \iota$, a_i , $\cos \theta_i$, ϕ_i , the component masses, m_1 and m_2 , and $\sqrt{\alpha}$. The prior on the luminosity distance is uniform in volume ($\propto D_L^2$). Finally, $P(D) = \int P(D|\theta)P(\theta)d\theta$ is the Bayesian evidence, which is the probability of obtaining the data D , marginalized over all possible parameter values for θ , and serves as the normalization constant.

GW signals from the inspiral of compact binaries are extremely weak and the technique of matched filtering is used to extract these signals from the noisy data. Assuming Gaussian and stationary detector noise, this technique equates to defining the (log) likelihood function in Eq. (6.15) as the following

$$\ln p(D|\theta) \propto -(1/2) \sum_i^n (D_i - h|D_i - h) , \quad (6.16)$$

where h is the template response function defined in Eq. (6.7) contracted onto the i -th detector response matrix, and n is the number of detectors in the network. The noise weighted inner product is defined as

$$(A|B) = 4 \text{Re} \left[\int_0^\infty \frac{AB^*}{S(f)} df \right] , \quad (6.17)$$

where $*$ denotes complex conjugation and $S(f)$ represents the power spectral density (PSD) function of the detector.

So far, we have focused on inferring the properties of a single binary system, but how can we combine the information from multiple observations to enhance our constraints? A growing number of detections are joining the existing catalogs of GW sources available from the LVC. The statistical power associated with

a larger sample size can have a significant impact on the constraints on modified theories of gravity [111]. This power can be harnessed through various methods that combine the statistical analysis of single events to produce tighter cumulative bounds. These methods vary depending on the type of parameter being constrained. For example, recent work has been done on a hierarchical framework to update constraints on parameters that are *not* common between all the events being analyzed. The values these parameters take are then determined by their respective hyper distributions, where one would not expect each parameter to be the same between each source [358]. This approach is suitable for the combination of information about generic deviations, for example parameters introduced by the ppE framework or the generic deviations investigated by the LVC [58]. We will focus on a simpler case, in which the parameter being constrained has a single, true value in Nature common to all events. Both EdGB and dCS fall into this category, where deviations in the waveform are parametrized by a fixed value of the coupling constant $\sqrt{\alpha}$.

Consider a catalog of N detections comprised of data $\{D_i\}$ described by parameters $\{\theta_i\}$, where i runs from 1 to N . The joint posterior for the parameters describing all the events can be written as

$$p(\sqrt{\alpha}, \{\theta_i\} | \{D_i\}) = \frac{p(\sqrt{\alpha}, \{\theta_i\}) p(\{D_i\} | \sqrt{\alpha}, \{\theta_i\})}{p(\{D_i\})}, \quad (6.18)$$

where we have explicitly separated the parameter of interest $\sqrt{\alpha}$, the only common parameter in the set, from the N sets of source parameters $\{\theta_i\}$. To produce a cumulative bound on the parameter introduced by a modified theory of gravity, we need to marginalize over all parameters $\{\theta_i\}$, which includes quantities like the binary masses, spins, sky location, and orientation for each binary in the catalog. This leads to an intermediate expression for the marginal posterior distribution on $\sqrt{\alpha}$ given by

$$p(\sqrt{\alpha} | \{D_i\}) = \frac{p(\sqrt{\alpha})}{p(\{D_i\})} \int p(\{\theta_i\}) \times p(\{D_i\} | \sqrt{\alpha}, \{\theta_i\}) d\{\theta_i\}, \quad (6.19)$$

where we have assumed that the prior on the modified theory parameter is independent of the source parameters for all the observations and we note that the evidence, defined by the integral $p(\{D_i\}) = \int p(\{\theta_i\}, \sqrt{\alpha}) p(\{D_i\} | \{\theta_i\}) d\{\theta_i\} d\sqrt{\alpha}$, is just a normalizing constant that can be factored out of the integral.

Because each of the events are statistically independent, we can separate the integral by source, such that

$$\begin{aligned} p(\sqrt{\alpha}|\{D_i\}) &= \frac{p(\sqrt{\alpha})}{p(\{D_i\})} \prod_i \int p(\boldsymbol{\theta}_i) p(D_i|\sqrt{\alpha}, \boldsymbol{\theta}_i) d\boldsymbol{\theta}_i, \\ &= \frac{p(\sqrt{\alpha})}{p(\{D_i\})} \prod_i p(D_i|\sqrt{\alpha}), \end{aligned} \quad (6.20)$$

where $p(D_i|\sqrt{\alpha})$ is the partially-marginalized likelihood for the i -th event, marginalized over source parameters only.

6.3.2 Selection of Sources

Our choice of events used in the Bayesian analysis carried out in this paper was guided by several motivations. Namely, we are confident that low mass events will typically be the most effective at constraining quadratic theories of gravity like dCS and EdGB, as the curvature is higher for lighter systems. This idea is motivated by the form of the modifications shown in Eqs. (6.10) and (6.11). As the phase modification scales inversely with the total mass to the fourth power, smaller masses will drastically improve the constraint (all else being equal). This is still true when also considering the PN order of the modification, as the $\beta_{\text{dCS,EdGB}}$ phase factors in Eq. (6.10) and Eq. (6.11) are accompanied by powers of $u = (\pi\mathcal{M}f)^{1/3}$. The phase modification (relative to the Newtonian term in GR proportional to u^{-5}) becomes $\beta_{\text{dCS}}u^4$ and $\beta_{\text{EdGB}}u^{-2}$ in dCS (2PN) and EdGB (-1PN), respectively. In the case of EdGB, both components [the factor of u and the specific form of Eq. (6.11)] improve the constraining power of lighter systems over heavier ones. This is less obvious for dCS, where the modification enters at 2PN and introduces an additional factor of $u^4 = (\pi\mathcal{M}f)^{4/3}$ relative to the Newtonian order term in GR. This factor weakens constraints coming from light systems as compared with heavier systems. However, combining the dependence on u with β_{dCS} yields an overall dependence on mass of $m^{-8/3}$. This leads to lighter system still outperforming heavier ones, in spite of the high PN order in which the modification is introduced. This analytic reasoning is also supported by the numerical explorations performed in Refs. [88, 111]. Furthermore, highly spinning systems would produce tighter constraints than non-spinning binaries as the phase modification grows with higher spin systems in general. Hence systems with large effective aligned spins χ_{eff} were also preferred, where χ_{eff} is defined as

$$\chi_{\text{eff}} = (m_1\chi_1 + m_2\chi_2)/m. \quad (6.21)$$

Given all of this, there are several interesting sources in GWTC-1 and GWTC-2, but not all of these could be used to place constraints on dCS and EdGB gravity. To begin with, the highly asymmetric sources

GW190814 [352] and GW190412 [26] could be of interest, but systematics due to waveform mismodeling and the uncertain nature of the lighter compact object in GW190814 are of concern. These concerns (specifically in the case of GW190814) are explored more thoroughly in Appendix ??.

Another interesting source to analyze would be one of the first NSBH binaries observed to date, namely GW190426. Its extremely high false alarm rate (FAR) and its small SNR, however, makes any conclusions we might obtain from this data questionable. On top of the lack of confidence we have that the signal is of astrophysical origin at all, the broadening of the mass distributions due to new degeneracies with the coupling parameters makes the analysis more complicated. The widening of the component mass distributions can be severe enough to bring into question the nature of the lighter object, again introducing complications to our analysis similar to those seen with GW190814 and discussed in Appendix ??.

Finally, the source GW190425 [350] was also excluded. While the prospect that this binary is actually of mixed nature (formed by a BH and a NS) would be exciting, that possibility seems remote. According to the analysis performed by the LVC [350], no known formation channels would be able to produce a BH of such a small mass as that found for the larger component in this event. As it seems much more likely that this source is actually a pair of NSs, and NSNS binaries do not source dipolar radiation [296, 353, 354], we omitted this source as well.

In summary, we have excluded the following events from our analysis: GW190814 and GW190412 (because their mass ratios are large, leading to issues with waveform systematics and complications to the analysis), GW190426 (because of a very high FAR, suggesting it may be a noise artifact) and GW190425 (because it is likely a NS binary and not a mixed binary). This then leaves the following set of sources to be analyzed in this work (we drop additional identifier information for GWTC-2 sources for brevity in the rest of this work): GW151226, GW170608, GW190924 (GW190924_021846), GW190720 (GW190720_000836), GW190707 (GW190707_093326), and GW190728 (GW190728_064510). Some of the critical source parameters discussed above are shown for each of the binaries analyzed in this study in Fig. 6.1.

6.3.3 Fisher analysis predictions

It can be informative to determine an order of magnitude estimate of the constraints on $\sqrt{\alpha}$ before discussing (and performing) a full Bayesian analysis. To this end, we use the more computationally tractable method of predicting the error covariances through the Fisher Information Matrix, or the Fisher for short [257, 359, 360]. This approximation begins by expanding the likelihood function about the maximum likelihood values

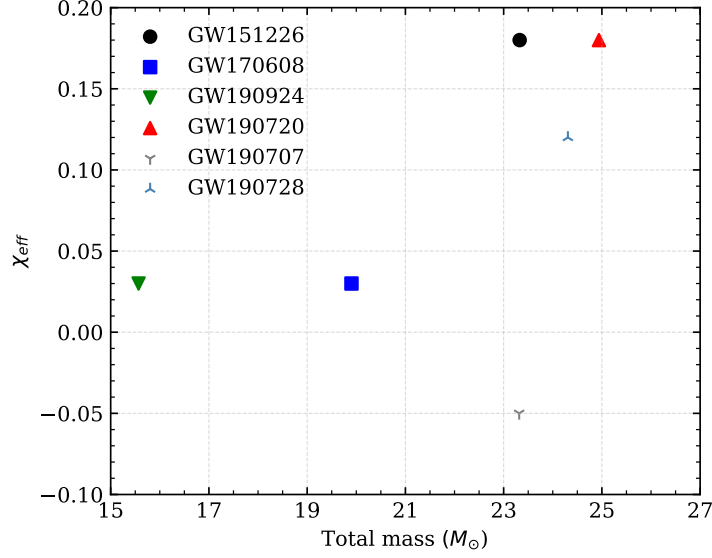


Figure 6.1: The detector frame total mass and the effective inspiral spin, χ_{eff} , of the events analyzed in this work. These values were inferred by LIGO assuming GR correctly describes the gravitational sector [18, 351].

θ^{ML} , resulting in the following expression

$$\ln p(D|\theta) \approx -(1/2) \Gamma_{ij} \Delta\theta^i \Delta\theta^j, \quad (6.22)$$

where $\Delta\theta^i = \theta^{i,\text{ML}} - \theta^i$ is the deviation away from the maximum likelihood value, and Γ_{ij} is the Fisher,

$$\Gamma_{ij} = (\partial_i h | \partial_j h) |_{\text{ML}}, \quad (6.23)$$

where $\partial_i = \partial/\partial\theta^i$. In the limit of large SNR and stationary Gaussian noise, the inverse of the Fisher gives the error covariance matrix Σ of the parameters and the diagonal elements of this covariance matrix give the root mean square error in the estimate of the parameters:

$$\sqrt{\langle (\Delta\theta^i)^2 \rangle} = \sqrt{\Sigma^{ii}}. \quad (6.24)$$

For the calculation of the Fisher, we use a restricted parameter set, mapping $\{a_1, a_2, \cos\theta_1, \cos\theta_2, \phi_1, \phi_2\}$ to $\{\chi_1, \chi_2, \chi_p, \phi_p\}$, where χ_i is the dimensionless spin defined in Eq. (6.12), χ_p is a certain projection of the BH spin onto the plane orthogonal to the binary's angular momentum, and ϕ_p is the angle of the perpendicular spin component χ_p in the plane. The in-plane component of the spin is defined as in [18], and

can be written as

$$\chi_p = \frac{1}{B_1 m_1^2} \max(B_1 S_{1\perp}, B_2 S_{2\perp}) , \quad (6.25)$$

where $B_1 = 2 + 3q/2$, $B_2 = 2 + 3/(2q)$, $q = m_2/m_1 < 1$ is the mass ratio and $S_{i\perp}$ is the projection of the spin of BH i on the plane orthogonal to the orbital angular momentum $\mathbf{L}$. We use a reduced parameter space because the Fisher can become ill-conditioned if the probability distributions have a high degree of covariance.

The Fisher matrix must be evaluated at certain parameter values, called the injection parameters, which correspond to a rough guess of the maximum likelihood values of the posterior (around which the likelihood is expanded in the Fisher matrix approximation). Initially, the injection values for the Fisher matrices were taken from the LVC's published inferences on the values of the source parameters in order to quickly pick the most competitive sources. However, in the course of validating our full analysis (detailed below), we derived posterior distributions for all the sources within GR. As the LVC does not publish many of the less informative parameters, like the polarization angle, reference phase, etc., we recycled the posteriors from our custom GR analysis for the calculation of the Fishers presented in Table 6.2 in order to have consistent injections. The GR parameters were the median values recovered by our Markov-Chain Monte Carlo (MCMC) analysis, and the coupling parameter was set to zero, corresponding to our null hypothesis that GR is correct.

The noise curves used in the construction of the Fisher matrices were an O2 approximate PSD for Hanford and Livingston, and an O3-O4 proxy PSD for Virgo [227]. The luminosity distances used in each Fisher were scaled such that the network SNR of the injection matched the SNR quoted by the LVC to account for minor differences in the detector noise. To accurately model the network with which each source was observed, we calculate a separate Fisher matrix for each detector and combine them for each source before inverting the final matrix. This gives a final Fisher for each source as

$$\Gamma_{\text{source}} = \sum_i^n \Gamma_{\text{source},i} , \quad (6.26)$$

where n is the number of detectors for each source, as reported by the LVC [18, 351]. Finally, the marginalized variance on the coupling constant from each of these covariance matrices, obtained from each source individually, are combined to construct a cumulative bound assuming a normal distribution. The process of inversion for the Fisher matrix plays the role of marginalization in this approximation, and the resulting components of the covariance matrix properly reflect uncertainties introduced through correlations. This

gives the following relation:

$$(\Delta\sqrt{\alpha_{\text{total}}})^{-2} = \sum_i^N (\Delta\sqrt{\alpha_i})^{-2}, \quad (6.27)$$

where $\Delta\sqrt{\alpha_{\text{total}}}$ is the final bound on the modifying parameter, $\Delta\sqrt{\alpha_i}$ is the standard deviation [defined by Eq. (6.24)] for the i -th source, and N represents the number of events. The details of this calculation are outlined in Sec. V and Appendix A of [111], and we refer the reader to that work for a more in-depth explanation of this analysis.

Source	$(\Delta\sqrt{\alpha_{\text{EdGB}}})$ km	$(\Delta\sqrt{\alpha_{\text{dCS}}})$ km
GW151226	2.51 (6.02)	77.18 (6.02)
GW170608	2.98 (6.14)	41.38 (6.14)
GW190924	1.24 (4.02)	35.49 (4.02)
GW190720	3.30 (7.06)	35.53 (7.06)
GW190707	3.19 (6.83)	73.78 (6.83)
GW190728	3.17 (7.14)	69.0 (7.14)
Combined	0.91	19.13

Table 6.2: Results from a Fisher projection, calculated using the median values for the source parameters from an MCMC analysis within GR. The results show the 90% confidence interval for the root of the coupling parameter in km for each source. The values in parenthesis are the upper limits on the validity of the small coupling approximation (6.6). The last row of the table shows the combined constraint assuming each posterior is normally distributed, consistent with the assumptions of a Fisher analysis.

The Fisher results presented in Table 6.2 suggest that stringent constraints on EdGB gravity may be possible for a variety of events. The same, however, is not true for dCS gravity, since the projected Fisher constraints all violate the small coupling approximation. Given this, we will next carry out a Bayesian analysis and compare our Bayesian constraints to the Fisher estimates of Table 6.2.

6.3.4 Bayesian analysis: single events

To sample from the posterior distribution for each source individually, we employed a MCMC analysis, which produces independent samples from the distribution of interest. These samples can then be binned to show a discrete approximation to the posterior probability density.

The final one-dimensional marginalized posteriors on $\sqrt{\alpha}$ for EdGB (dCS) are shown in Fig. 6.2 (Fig. 6.3). Each panel of the figures also shows the 90% confidence upper limit on the magnitude of $\sqrt{\alpha}$, as well as the approximate upper limit on the validity of the small-coupling approximation in Eq. (6.6), required for the waveforms to be valid, both overlaid as vertical lines. The approximate upper limit of the validity of the waveform was derived with the median value of the smaller mass from the same samples used to calculate the posterior distributions shown in Figs. 6.2 and 6.3. We used as priors uniform on $\sqrt{\alpha}$ between $[0, 20]$ km and $[0, 100]$ km for EdGB and dCS, respectively. The fact that the posteriors in $\sqrt{\alpha_{\text{EdGB}}}$ (i) differ considerably

from the priors and (ii) the 90% confidence upper limit is within the small-coupling approximation, give us confidence that these events can place a constraint on EdGB gravity. The robustness of these constraints is analyzed in Sec. 6.4.

As can be seen from Fig. 6.2, all the sources studied in this work independently place robust constraints on $\sqrt{\alpha_{\text{EdGB}}}$, ranging between 3 – 7 km. The situation is opposite for dCS and none of the sources satisfy the small coupling approximation. As no individual analysis is trustworthy, any results derived from stacking these dCS posteriors would also not be correct. As such, the strongest constraint to date on this theory, remains the one established in [223], $\sqrt{\alpha_{\text{dCS}}} < 8.5$ km, obtained by combining x-ray pulse profile observations of the millisecond pulsar PSR J0030+0451 by NICER [361, 362] with the NS tidal deformability inferred from the binary NS event GW170817 by the LVC [336, 363]. Any further discussion on the posteriors of $\sqrt{\alpha_{\text{dCS}}}$ (combined or not) are pure speculation, for the benefit of future studies that might be more fortunate in the sources available for analysis.

Importantly, the sampling artifacts of the first iteration of this work [124] have been removed through the MCMC analysis carried out here. As speculated in [124], the lack of support at zero for GW151226 and GW170608 for both EdGB and dCS *has disappeared with the use of the more natural sampling parameter $\sqrt{\alpha}$* . Critically, the 90% confidence regions of the different sources are consistent, indicating that, while the artifacts could be misleading, the conclusions reached in that work are substantiated.

6.3.5 Bayesian analysis: stacked events

In theory, the process of combining posteriors from independent experiments to create a total probability distribution on a single parameter is as simple as multiplying the marginalized likelihoods together, considering Eq. (6.20). To achieve this in practice two methods of combining posteriors are commonly employed, which we will briefly outline for clarity.

The first method begins by creating a histogram of the samples output by the MCMC analysis, which produces a discrete representation of the marginalized posterior on the parameter of interest. One can then fit an ansatz function or use a more general kernel density estimation (KDE) to approximate the histogram, producing an analytic representation to the discrete posterior distribution. This introduces two immediate sources of error, namely the possibility of a poor choice of fitting function and numerical errors in the fitting process directly. For distributions that have obvious candidate ansatz (for example Gaussian distributions), this can be an attractive method as one can enforce certain properties, such as no-where vanishing probability distributions and smooth distributions for the final posterior. When using KDE's, there is less flexibility to enforce these nice properties, like non-vanishing distributions or strict boundary conditions, but this method

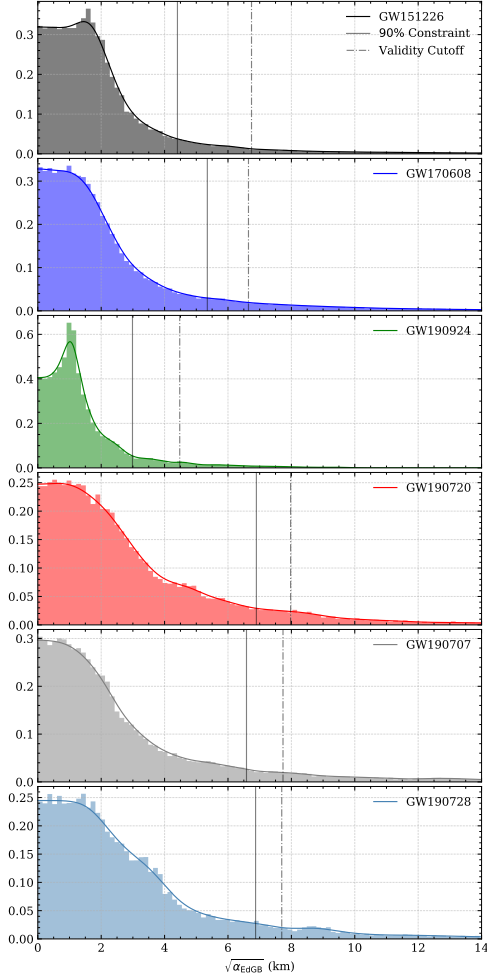


Figure 6.2: The individual constraints on the coupling constant of EdGB, $\sqrt{\alpha_{\text{EdGB}}}$. Each panel shows the marginalized posterior distribution of the square root of the coupling constant in EdGB. Overlaid is the 90% confidence value of the coupling constant, shown as the vertical solid line, with the upper limit of validity for the small coupling approximation shown as a vertical dashed line. On top of the discrete histogram, the KDE approximation used to determine the joint distribution from all the sources is also shown as a solid curve. All six sources shown satisfy the small coupling approximation at 90% confidence, resulting in a robust bound on $\sqrt{\alpha_{\text{EdGB}}}$.

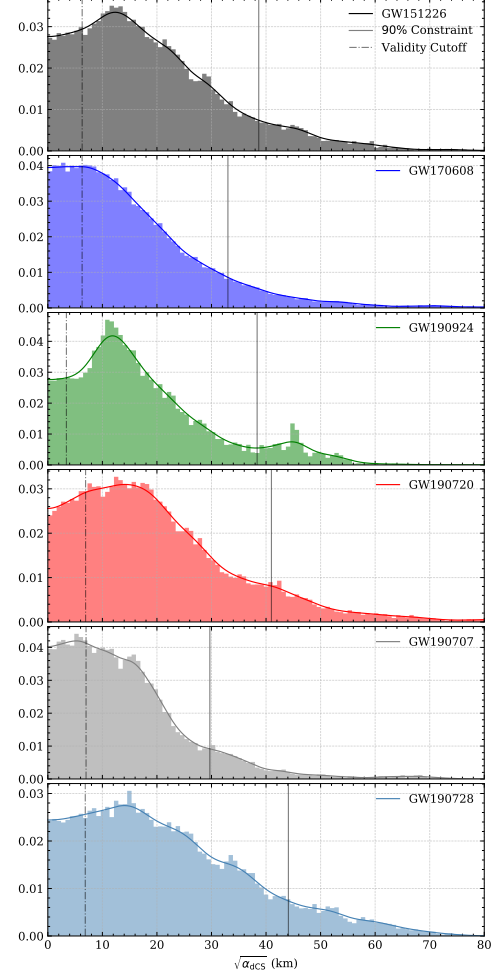


Figure 6.3: The individual constraints on the coupling constant of dCS, $\sqrt{\alpha_{\text{dCS}}}$. Each panel shows the marginalized posterior distribution of the square root of the coupling constant in dCS. Overlaid is the 90% confidence value of the coupling constant, shown as the vertical solid line, with the upper limit of validity for the small coupling approximation shown as a vertical dashed line. On top of the discrete histogram, the KDE approximation used to determine the joint distribution from all the sources is also shown as a solid curve. None of the sources shown satisfy the small coupling approximation, and as such, we can still not place a meaningful constraint on the dCS coupling parameter purely through GW observations.

provides more flexibility in terms of minimizing manual “tuning” during the analysis. This kind of technique is outlined and appropriately implemented in works like [50, 364], for example.

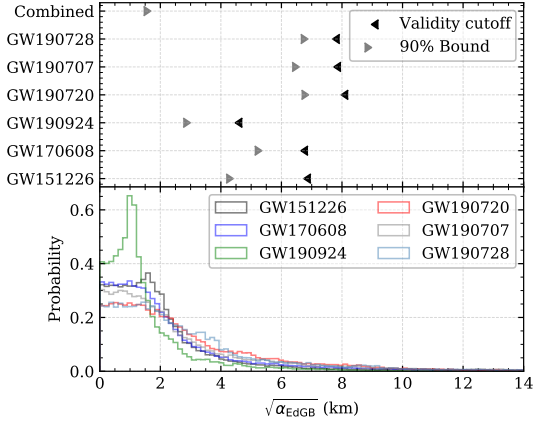


Figure 6.4: In the bottom panel, we show the histogram representation of the probability density of the value of the EdGB coupling constant, $\sqrt{\alpha_{\text{EdGB}}}$, for the six events we have chosen for this analysis. The top panel shows the 90% confidence constraint on the magnitude of the coupling parameter and the maximum value for validity of the small-coupling approximation for each of the individual events, with the cumulative constraint shown at the top. After combining the information of all the sources, we can achieve a constraint on $\sqrt{\alpha_{\text{EdGB}}}$ of less than 1.7 km at 90% confidence.

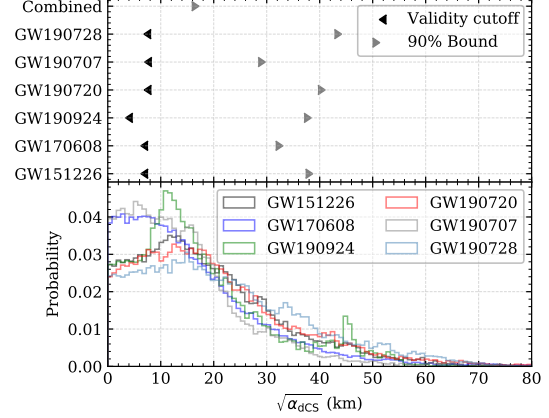


Figure 6.5: In the bottom panel, we show the histogram representation of the probability density of the value of the dCS coupling constant, $\sqrt{\alpha_{\text{dCS}}}$, for the six events we have chosen for this analysis. The top panel shows the 90% confidence constraint on the magnitude of the coupling parameter and the maximum value for validity of the small-coupling approximation for each of the individual events, with the cumulative constraint shown at the top. As no single constraint satisfies the small-coupling approximation, our cumulative bound is also untrustworthy. We only show the combined posterior distribution for illustrative purpose.

A second method is even more straightforward, where the individual posteriors are also initially approximated by a histogram. Instead of fitting a function to this discrete distribution, the histograms of each source are directly multiplied together (where, of course, the histograms must have the same bin sizes or be appropriately transformed). This method is convenient from the standpoint of how simple and easily implementable it can be. However, numerical noise and finite numbers of samples can cause issues, especially when computing quantities at the tails of the distribution, like 95% confidence intervals. For example, if one source has zero counts in a certain histogram bin, the final distribution will *always* have zero counts in the said bin, regardless of how much weight is assigned to that region by other sources. This makes this particular method sensitive to finite bin-size and finite sample-size effects. When calculating confidence intervals, smoothing functions can also be applied to the final distribution to minimize issues with convergence for numerical integration, which was a step implemented for this work.

To ensure maximum reliability, we have repeated the calculation twice, once with each method, and always quote the more pessimistic constraint on $\sqrt{\alpha}$ throughout this work. We discuss the difference of these techniques in the context of this work in Appendix ??.

The results of combining posteriors on $\sqrt{\alpha}$ in EdGB and dCS are shown in Fig. 6.4 and Fig. 6.5, respectively. We see a moderate improvement, as expected with a set of posteriors that are already comparable in constraining ability [111]. Our new bound on the coupling parameter in EdGB at 90% confidence is $\sqrt{\alpha_{\text{EdGB}}} < 1.7$ km. For dCS, as all the constituent posteriors in the set violate the small coupling approximation, *the cumulative bound is still not valid and is only shown for illustrative purposes and guidance for future efforts to constrain this elusive theory with GW observations.*

6.4 Robustness of Constraints

The constraints we have placed here are contingent on the reliability of each individual component in our analysis. Undoubtedly, the GW waveforms we have used are one of the most uncertain elements. There are two factors that could play a role in causing systematic biases in our conclusions. The first issue we address concerns the base model (that is, the waveform model in GR onto which our modifications are appended) implemented for this work, as there are many variations available today [365]. This is discussed below in Sec. 6.4.1. We then examine the issue of missing information at higher PN orders sourced by the contending theories, discussed below in Sec. 6.4.2. In both sections, we will focus on the bounds we have placed upon the coupling constant in EdGB, as we are still unable to place constraints on dCS gravity with GW observations alone.

6.4.1 Changing the GR base model

As mentioned earlier, we have used a widely adopted waveform as our base GR template, `IMRPhenomPv2` [202]. By base model, we mean that this is the model that we added the EdGB and dCS corrections to. However, there exist other templates that we could have used, such as some that include higher-order spin effects, higher harmonics, and various effective-one-body models. The custom software developed for this analysis currently lacks support for most of these other options but does include the `IMRPhenomD` [106, 169] template. To quickly infer an estimate of the impact of the choice of base waveform template on our results with the tools available, we have rerun our analysis with `IMRPhenomD` as our base GR model. The comparison between using an `IMRPhenomPv2` and `IMRPhenomD` base GR model will provide some insight into how sensitive our constraints are to the underlying GR base waveform.

The two base GR waveforms we will use, `IMRPhenomPv2` and `IMRPhenomD`, differ by the inclusion of precession effects. The `IMRPhenomD` waveform can only model spin-aligned and spin-anti-aligned binaries, whereas the `IMRPhenomPv2` waveform can deal with precession effects by “twisting up” a co-precessing

waveform (in this case **IMRPhenomD**) through frequency-dependent Euler angles [202]. This could impact our analysis in one of the two ways: the extra information imparted by precessional effects could break certain degeneracies between source parameters, resulting in tighter bounds. Alternatively, it could be that the increased dimension of the model (moving from two spin parameters to six) could cause a degradation in accuracy of recovered parameters due to new degeneracies.

Our results are shown in Fig. 6.6, which presents the marginalized posteriors on $\sqrt{\alpha_{\text{EdGB}}}$ using both an **IMRPhenomPv2** base model and a **IMRPhenomD** base model, and the GW151226 event as an example. As the figure shows, the posteriors lead to very similar constraints, with the **IMRPhenomPv2** base model leading to $\sqrt{\alpha_{\text{EdGB}}} \lesssim 4.5$ km and the **IMRPhenomD** base model leading to $\sqrt{\alpha_{\text{EdGB}}} \lesssim 3.5$ km. For this example, we see that the GW151226 event did not contain enough precessional information to lead to a better constraint; in fact, the increase in dimensionality of the parameter space worsens the constraint by about 20%. Such a deterioration of the constraint is very small, and using the more complex model leads to a more conservative (i.e. less stringent) constraint. This, therefore, increases our confidence that the constraints we have placed here are robust against the choice of base GR waveform model.

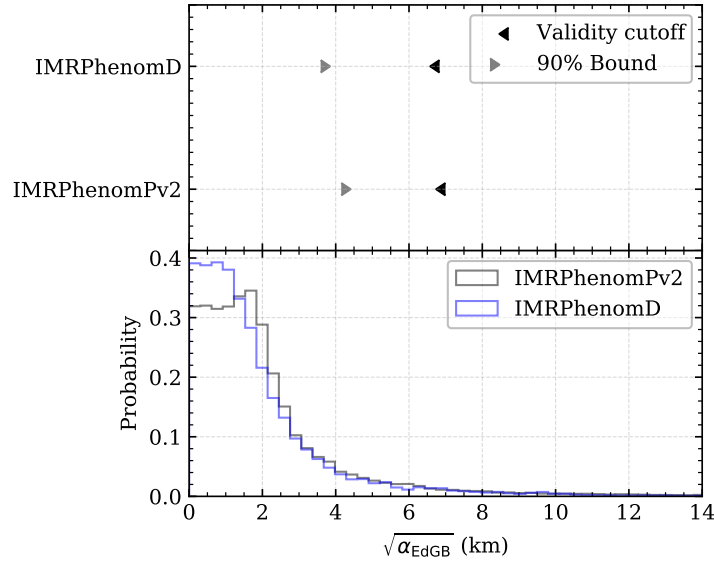


Figure 6.6: Comparison of the constraints obtained through two separate analyses of the event GW151226. In the first iteration, we used a more robust template, **IMRPhenomPv2**, which encodes information about precession. We then repeated the analysis using the base template **IMRPhenomD**, which only models spin-aligned and spin-anti-aligned binaries. Our results indicate that the constraints we have placed on the coupling constant of EdGB are robust to changes in the GR-template used to recover the parameters of the signal.

6.4.2 Adding higher-order PN order GR modifications

The waveforms we have used here only include the leading-order PN correction due to EdGB and dCS in the inspiral of the binary. As there are currently no *complete*, analytic expressions for higher-order corrections, despite ongoing efforts both with PN methods [108, 110, 311, 312, 349] and numerical relativity approaches [303, 316, 366–369], we are forced to work with a template which we know is lacking far behind the accuracy of the base (GR) template.

Given this uncertainty about the true waveform, there could be some concern that the inclusion of higher PN order corrections in the modifications to GR could degrade our bound through increased degeneracy between source parameters and the coupling parameter. However, it is reasonable to expect that this will not be the case when considering realistic higher PN order corrections [112], as there are some fundamental restrictions that these additional modifications must satisfy to remain mathematically consistent in PN theory.

There are two criteria that any modification to GR that admits a PN expansion must meet: (i) all corrections to the waveform must be linear in the expansion parameter ζ , and (ii) the expansion must be valid in the typical domain in which PN approximations are applied. The first condition must be true if one is working in the small coupling approximation. The second condition requires that the coefficient in the PN expansion of the modified gravity terms do not increase with PN order much more rapidly than how the coefficients grow in GR. Since these coefficients can grow in GR quite rapidly (especially in the extreme mass ratio limit), this second condition is not as stringent as it may seem at first.

With this in mind, the additional terms entering the waveform would need to have a very particular and unfortunate functional form to *increase* degeneracies, as opposed to decreasing them. This is because the only way for additional information in the waveform (introduced through the higher PN order corrections) to increase the ambiguity about source parameters is to have very precise cancellation between terms. As this seems unlikely given the additional information is proportional to a different power of velocity (or frequency), when higher PN order corrections are finally derived, they will only serve to improve our current constraints. Therefore, not including such higher PN order terms will lead to weaker or less stringent constraints, which we call here “conservative.”

Lacking a complete higher PN order template to compare against, we will here approximate the degree of systematic bias we may be incurring by truncating the non-GR modification at leading PN order. We will do so by appending reasonable, generic modifications at higher PN orders as place holders for the true modifications. These generic place holders can then be marginalized over, leaving a final cumulative bound on the coupling parameter that better reflects uncertainties in the higher PN order corrections omitted by

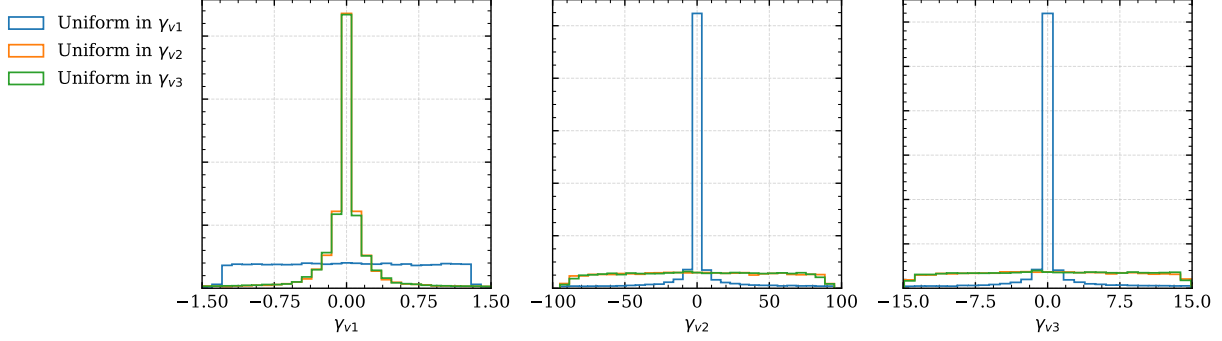


Figure 6.7: Comparison of priors for the different parametrization schemes outlined in Sec. 6.4.2. The left panel shows the priors of each parametrization transformed to the parametrization $\delta\phi_{v1}$, shown in Eq. (6.29). The middle panel shows each parametrization transformed to that shown in Eq. (6.30). Finally, the right panel shows the priors transformed to the parametrization shown in Eq. (6.32). From these figures, we can see that the first parametrization differs substantially from parametrizations 2 and 3, while parametrizations 2 and 3 are very similar.

our waveforms.

Marginalizing over the generic higher PN order term leads to a final constraint on the coupling parameter that reflects uncertainties in this new parameter, but even generic parametrizations that are marginalized over are not created equal. We have noted these differences in previous works [124], where it was shown that different generic parametrizations can result in different final constraints through the implicit difference in priors. While this is not always a significant issue, as re-weighting the samples with new priors can sometimes mitigate this effect, our previous work showed a particularly unfortunate case where the Jacobian of the transformation from one parametrization to another was singular.

This is not something that can be unambiguously resolved, and so in this work we will pick three well-motivated parametrizations and compare the final results. As outlined above, an important aspect that is shared by all three parametrizations is the fact that modifications at all orders should be proportional to the non-GR coupling parameter, in this case ζ_{EdGB} . This must be the case for any mathematically consistent, higher PN order expansion of the EdGB field equations.

We will use the variable γ to represent any higher PN order coupling of source parameters to the coupling parameter, which will be the variable that we will marginalize over by the end of this analysis. This is a conservative implementation, as this parameter would be replaced by a function of the conventional source parameters in the true expansion, just as is the case at -1PN in the standard EdGB waveform. As pointed out earlier, information introduced by this new coupling in the true expansion will almost certainly reduce our estimate of $\sqrt{\alpha_{\text{EdGB}}}$, not widen it. This ultimately can be seen from the fact that if higher PN order terms, proportional to the expansion parameter, are appended to the waveform, the waveform will deviate

further from the GR template for equal values of the expansion parameter. This should only serve to reduce uncertainty in the modifying variable.

Beyond these details, it is also important to note that the true expansion (as of yet unknown) would *not* increase the dimensionality of the parameter space. Any higher order corrections would only depend on the existing source parameters and the coupling constant, $\sqrt{\alpha_{\text{EdGB}}}$. Because of these considerations and the fact that our analysis here *does* add an additional parameter, our results here can be viewed as even more conservative.

All three parametrizations versions can be written as in the form

$$\delta\phi_{vi} = \phi_{\text{GW}} - \phi_{\text{GR}}, \quad (i = 1, 2, 3). \quad (6.28)$$

The first parametrization we will use modifies the GW phase as follows

$$\delta\phi_{v1} = \zeta_{\text{EdGB}} g_{\text{EdGB}}(\boldsymbol{\theta}_{\text{GR}}) u^{-7} + \zeta_{\text{EdGB}} \gamma_{v1} u^{-5}, \quad (6.29)$$

where $u = (\pi\mathcal{M}f)^{1/3}$ as before and $g_{\text{EdGB}}(\boldsymbol{\theta}_{\text{GR}}) = \beta_{\text{EdGB}}/\zeta_{\text{EdGB}}$ is the source dependent term in the leading order deviation due to EdGB defined in Eq. (6.11). This term depends on the parameters of the binary described within GR, $\boldsymbol{\theta}_{\text{GR}}$. The second term in the above equation therefore introduces a 1PN correction to the -1PN dipole EdGB correction to the GR waveform phase, with γ_{v1} assumed to be a constant number we will later marginalize over. This choice is motivated by the ppE framework, where the additional modification can be seen as an absolute deviation.

The second parametrization we will investigate modifies the GW phase in the following way

$$\begin{aligned} \delta\phi_{v2} &= \zeta_{\text{EdGB}} g_{\text{EdGB}}(\boldsymbol{\theta}_{\text{GR}}) u^{-7} \\ &+ \zeta_{\text{EdGB}} \gamma_{v2} g_{\text{EdGB}}(\boldsymbol{\theta}_{\text{GR}}) u^{-5}. \end{aligned} \quad (6.30)$$

Making the 0PN modification proportional to the same function as the leading PN order modification results in a relative deviation from the leading PN order term. For clarity, we can rewrite the phase slightly to give the overall phase modification as

$$\delta\phi_{v2} = \zeta_{\text{EdGB}} g_{\text{EdGB}}(\boldsymbol{\theta}_{\text{GR}}) u^{-7} [1 + \gamma_{v2} u^2], \quad (6.31)$$

which shows more clearly the motivation of this parametrization. As before, here γ_{v2} is also a constant we

will later marginalize over.

Finally, the third parametrization we will investigate modifies the GW phase as follows

$$\begin{aligned}\delta\phi_{v3} &= \zeta_{\text{EdGB}} g_{\text{EdGB}}(\boldsymbol{\theta}_{\text{GR}}) u^{-7} \\ &+ \zeta_{\text{EdGB}} \gamma_{v3} g_{\text{EdGB}}(\boldsymbol{\theta}_{\text{GR}}) \phi_{\text{GR,1PN}}(\boldsymbol{\theta}_{\text{GR}}) u^{-5},\end{aligned}\tag{6.32}$$

where the term $\phi_{\text{GR,1PN}}$ is the 1PN term appearing in the PN expansion in GR, $\phi_{\text{GR,1PN}} = 3715/756 + 55\eta/9$. This choice in parametrization is motivated by the idea that the proper expansion of the EdGB field equations might mirror the form taken by the same expansion in GR, meaning the generic scheme imposed here would match the true expansion in EdGB when $\gamma_{v3} \rightarrow 1$. This type of modification can be rewritten in the following form,

$$\delta\phi_{v3} = \zeta_{\text{EdGB}} g_{\text{EdGB}}(\boldsymbol{\theta}_{\text{GR}}) u^{-7} [1 + \gamma_{v3} \phi_{\text{GR,1PN}} u^2],\tag{6.33}$$

showing this relation more clearly, where once more γ_{v3} is a constant we will later marginalize over.

As we stated above, we will marginalize over the γ parameters, which control the size of the 1PN modifications considered here, relative to the leading PN order modification. But how large should we allow these γ parameters to be? The answer to this question is fixed by our choice of prior. We will here choose the priors on all of the γ parameters to be flat everywhere inside a fixed region that will be chosen by requiring that the PN approximation remain valid in the modified theory, i.e. requiring that the 1PN term is not much much larger than the leading PN order term. To ensure this, we set the prior to zero when the following condition is violated

$$|\zeta_{\text{EdGB}} g_{\text{EdGB}}(\boldsymbol{\theta}_{\text{GR}}) (\pi \mathcal{M} f_{\text{lim}})^{-7/3}| > |\delta\phi_{0\text{PN}} (\pi \mathcal{M} f_{\text{lim}})^{-5/3}|\tag{6.34}$$

where f_{lim} is an approximate upper limit on the regime of validity for the expansion, and $\delta\phi_{0\text{PN}}$ is the term proportional to u^{-5} shown in Eqs. (6.29), (6.30), and (6.32) for each parametrization, respectively. This translates directly to the following criteria,

$$|\gamma_{v1}| < |g_{\text{EdGB}}(\boldsymbol{\theta}_{\text{GR}}) (\pi \mathcal{M} f_{\text{limit}})^{-2/3}|,\tag{6.35a}$$

$$|\gamma_{v2}| < |(\pi \mathcal{M} f_{\text{limit}})^{-2/3}|,\tag{6.35b}$$

$$|\gamma_{v3}| < |(\phi_{\text{1PN}})^{-1} (\pi \mathcal{M} f_{\text{limit}})^{-2/3}|.\tag{6.35c}$$

To keep the analysis simpler, we evaluate the conditions in Eq. (6.34) at the best fit values for the source parameters coming from a separate parameter estimation analysis assuming GR. In this work, the frequency f_{lim} will be set to the value of the GW frequency emitted by the binary at a separation of $50M_{\odot}$. For each version of our parametrization, this corresponds to the allowed ranges γ_{vi} of $[-1.31, 1.31]$, $[-90.67, 90.67]$, and $[-14.40, 14.40]$, respectively. Because each implementation uses a different definition of γ_{vi} , a flat prior on γ_{vi} will not have the same impact on $\sqrt{\alpha_{\text{EdGB}}}$ and will result in slightly different constraints once the posterior has been marginalized over γ . This can be seen in Fig. 6.7, which shows a direct comparison of the mapping of a uniform prior from each parametrization to each of the other parametrizations.

The inferred value of $\sqrt{\alpha_{\text{EdGB}}}$ for these three different parametrization schemes applied to the GW151226 event are shown in Fig. 6.8. As is clearly shown in the figure, the results are indeed robust to higher PN order corrections to the waveform, at least to next-to-leading order. The physical reason for this is that the EdGB correction enters at -1PN order, so the low frequency part of the signal is what is dominating the constraint. In this low frequency regime, the higher PN order terms are indeed subdominant, and do not affect the constraint on the leading PN order term. Indeed, this has also been observed in GR parameter estimation, where the chirp mass (which enters at leading 0PN order) is fixed by the low-frequency part of the signal, with the symmetric mass ratio (which enters at 1PN order) mattering only at higher frequencies.

One other concern that could be raised is the lack of knowledge about the highly nonlinear regime of the merger of BBH in either of these theories of gravity. This part of the coalescence of BBH cannot be described perturbatively, but instead requires numerical simulations. In this work, we targeted the lowest mass sources from the events catalog for the reasons outlined in Sec. 6.3.4. Because of this, the merger typically falls outside the range of sensitivity for the current 2G network of detectors, and would not impact this study. It is true that future work, performed with higher SNR sources, would need to be more cognizant of these issues, but the current restrictions due to detector limitations has ensured this work is robust to these uncertainties. To illustrate our reasoning, we have shown in Fig. 6.9 the characteristic strain, given by $h_c = \sqrt{f} \sqrt{|h_+|^2 + |h_{\times}|^2}$, against a proxy noise curve for the LIGO detectors in the second observation run. This can be informative as the contribution to the SNR as a function of frequency for a source can be approximately seen as the ratio of the two curves. The majority of the SNR for the sources used in this work comes from the early inspiral, with many of the sources' actual mergers contributing very little.

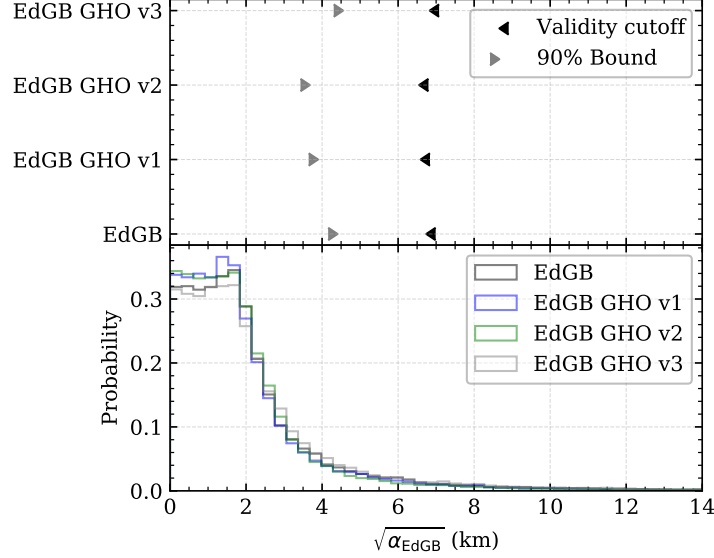


Figure 6.8: The results of analyzing the GW151226 event with EdGB and three different versions of EdGB with some generic, higher PN order modification. The histograms in the lower panel show the posterior probability on the value of the EdGB coupling constant, $\sqrt{\alpha_{\text{EdGB}}}$ in km. All three iterations include the -1PN effect, and the three versions labeled GH0 (generic higher order) incorporate some generic modification at 0PN (relative to the Newtonian term in GR), outlined in the text. The upper panel shows the 90% confidence constraint on $\sqrt{\alpha_{\text{EdGB}}}$, and the maximum value the coupling constant can take and still satisfy the small coupling approximation. We see good agreement between these four methods, giving us confidence that our results in this work will be robust to any additional modifications that are derived in the future.

6.5 Conclusions and Future Directions

In this work, we have conducted a full, numerical exploration of the posterior surface of the coupling constants in two viable extensions to GR (dCS and EdGB) for six GW sources from GWTC-1 [18] and GWTC-2 [351], utilizing the natural parameter-basis adapted to the problem. This custom analysis reduces the errors introduced in the recycling of results from generic analysis and provides the most robust and reliable results to date. To further verify the conclusions of this work, we investigated the impact of changing the base (GR) waveform (onto which our modifications were appended), as well as the impact of our lack of knowledge of higher PN order corrections that would be introduced to the waveform as the result of modifications to the field equations.

We inferred to 90% confidence that the square root of the coupling constant for EdGB is less than 1.7 km, i.e. $\sqrt{\alpha_{\text{EdGB}}} < 1.7$ km. Our work on EdGB gravity has produced a constraint on the coupling constant that is now the most stringent to date, as well as less prone to systematic error than previous analysis [214]. Unfortunately, dCS gravity has continued to evade our efforts to constrain its coupling constant through the exclusive use of GW observations. The constraints we place here, along with those of past works, have

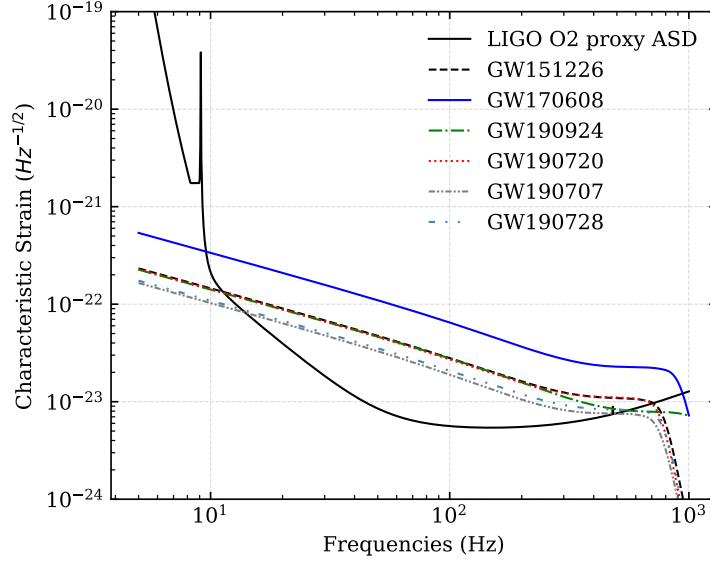


Figure 6.9: The characteristic strain compared against an example sensitivity curve representative of the LIGO detectors in O2. The contribution to the SNR from the merger is negligible for almost all the events, ensuring that our analysis is robust to our lack of knowledge of the highly nonlinear dynamics of the merger itself.

reliably restricted the length scale of the theory to roughly the size of the horizon of astrophysical BHs. Future work might improve on this bound through the use of merger-ringdown measurements [370] (for which, more work will be needed in perturbation theory [318, 371–374] and numerical simulations [303, 316, 366–369], but continued use of inspiral measurements will only serve to slowly push the limit further and further down in small increments. From the point of view of inspiral only measurements, the theory is now at the mercy of the statistics of large catalogs.

Our results open the door for many future directions of research. Of course, we are still hoping to constrain dCS gravity purely through GWs, but we must continue to wait for more favorable sources, such as those pointed out in [88]. Of particular interest would be a NSBH source with a reasonable SNR, as the unequal footing of NS and BH in dCS gravity works to the benefit of our efforts to constrain the theory. With sufficient SNR, a NSBH event would provide the best avenue for constraining dCS gravity with a single source.

Note added after completion: Right before we submitted this paper, Ref. [375] appeared in the arXiv. That work is different from the one presented here in various respects. First, Ref. [375] considered single events, instead of stacking as we do in this paper. Second, Ref. [375] did not consider the possible effect of waveform systematics on their constraints; neither due to higher PN order corrections in the non-GR part of the waveform, nor due to different models for the GR part of the waveform. We studied this in

great detail. Finally, and perhaps most importantly, Ref. [375] used two particular events in the GWTC-2 catalog, GW190814 and GW190425, both of which we purposely excluded from our analysis (see Sec. 6.3.2 and Appendix D).

We excluded GW190814 for two reasons: the uncertainties in the GR part of the waveform greatly affected non-GR constraints in our analysis and the uncertain nature of the lighter object. In fact, the former reason affected the tests of GR that the LVC carried out with GWTC-2, which is why the LVC made explicit mention of this event in their analysis (see Appendix C and, in particular, Fig. 19 of [58]). We excluded GW190425 because this event was very low-mass and, in fact, it was identified as a binary NS inspiral by the LVC (at $> 99\%$ confidence at the GraceDB website [376]). As is well-known from [296], NSs do not possess a monopole scalar charge in the version of EdGB gravity we considered in this work. Therefore, NS binaries in this theory do not introduce scalar dipole modifications to the GW phase. As a consequence, the only way the analysis of Ref. [375] is correct is if one assumes *a priori* that the binary was composed of BHs, which even with the analysis of Ref. [377] seems like a very strong assumption.

Acknowledgements This work made use of the Illinois Campus Cluster, a computing resource that is operated by the Illinois Campus Cluster Program (ICCP) in conjunction with the National Center for Supercomputing Applications (NCSA) and which is supported by funds from the University of Illinois at Urbana-Champaign. We acknowledge financial support from NSF Grant PHY-1759615 and PHY-1949838, NASA Grants 80NSSC18K1352, NNX16AB98G and 80NSSC17M0041 and NASA ATP Grant No. 17-ATP17-0225.

Chapter 7

Conclusions

The detection of GW radiation has opened up an entirely new avenue to study the gravitational interaction and to test our best gravitational theory to date, GR. In this dissertation, we have studied extensively a particular class of tests of GR, inspiral tests. We began by simulating future GW science with inspiral tests, preparing for future GW detectors and evaluating new methodologies. We then move on to use these tests to place constraints on viable, alternative theories of gravity. We have arrived at many conclusions, which are helping to shape current theoretical and observational work on tests of GR.

In Chapter 2, we studied the effect of giving the graviton a mass in the presence of a less studied, nuanced effect: screening mechanisms. In this work, we found that this mechanism could generically obscure the effects of a graviton mass on the propagation of GWs. We predicted the impact this effect may have on future constraints on the graviton mass using the Fisher matrix, with a variety of different classes of sources and different detectors. From there, we expanded our numerical investigation in Chapter 3 to consider many more theories of gravity with a much wider range of detectors. In this work, we considered seven viable, alternative theories of gravity or generic physical modifications, as well as theory-agnostic waveform models. Foundational to this work was the use of realistic, simulation-informed synthetic catalogs of GW sources, both for stellar-origin BHs and supermassive BHs. By modelling the catalogs of detected events from different global detector networks, we were able to conduct certain analyses that would have been inaccessible otherwise. From these investigations, we were able to make several predictions and studied many different trends in the way certain modifications can be constrained. In particular, we identified the impact of the number of sources and the type of sources on placing constraints on various modifications to GR. Within specific theories, as well as within a theory-agnostic framework, we were able to make predictions about the next three decades of fundamental science research with GW observations, as related to inspiral tests of GR.

We then moved on to more targeted, highly accurate methods of simulating GW observation in Chapter 4. In this chapter, we focus on the methodology used to conduct inspiral tests of GR. As the standard practice for testing GR with theory-agnostic waveform models is to only include a single deformation at a time, there is concern in the GW community about results from these analyses being overly optimistic. In this

work, we propose a new parametrization and prior, informed by the structure of a PN series solution. By incorporating more physics-based information into the analysis, we are able to show that the bounds being placed on theory-agnostic parameters in the waveform model used in these tests are robust, and that in fact, the bounds are expected to enhance with the derivation of more accurate models, not degrade.

From here, we transition to placing constraints on alternative theories of gravity with current data in Chapters 5 and 6. In Chapter 5, we transform results related to theory-agnostic parameters to bounds on constants in specific theories of gravity. We focus on two particular theories of gravity which fall into the class of theories known as quadratic theories of gravity, EdGB and dCS gravity. Because of limitations in the waveform modelling, we are unable to constrain dCS. However, we place robust constraints on EdGB, restricting the lengthscale of the system to 5.6km with GWs for the first time.

We improve on this constraint in Chapter 6, expanding the number of sources we analyze up to six and implementing theory-specific waveform models, as opposed to generic ones. This allows us to restrict the lengthscale of EdGB down to 1.7km, although dCS still eludes us in our attempts to constrain it. We also examine several possible sources of systematic error that could bias our analysis. We first determine if changes to the waveform modelling within GR (onto which our modifications from EdGB and dCS are appended) could bias our results. We find that whether we use a spin-precessing template or spin-aligned template does not significantly change our results. From there, we investigate how our restricted modelling within dCS and EdGB could introduce error into our analysis. At the time of publication, there was no work done on higher order corrections in EdGB, so we constructed a new waveform model that introduced a generic parameter 1PN order higher than the leading order correction. We repeated our analysis with this new parameter, and marginalized over it. The results from this auxiliary analysis were identical to the analysis done in the main body of this work, giving high confidence that our analysis is robust to a lack of knowledge about higher order corrections, consistent with Chapter 4.

In the era of GW detection, the ways in which one can test GR seem endless. We have thoroughly investigated one such class of methods, the inspiral tests of GR. We have examined how these tests might be effectively leveraged in the future, with new methodologies and new detector networks. We have also applied these tests to current data in order to constrain viable alternatives to gravity. There are many future avenues down which one could take the work laid out in this thesis. Of particular interest would be studies examining the effective use of the statistical power of the sizeable catalogs expected to be coming in the next decades. Furthermore, work on theories that exhibit behavior that does *not* conform to PN series solutions (such as spontaneous scalarization [278, 279, 282, 302, 303]) would be an interesting future investigation as well. Finally, utilizing other methodologies, such as Bayesian model selection, could potentially lead to

informative conclusions. New ways of analyzing data, through new statistical paradigms, might shed more light on the situation. The work to find GR's successor has been one hundred years in the making and still continues in earnest. However, the breakthrough needed to expand our understanding about gravity could always be buried in the signal of the next GW observation detection.

Appendix A

Bayesian Theory and Fisher Analysis Details

Signal analysis in GW science is usually based on Bayes' theorem:

$$p(\vec{\theta}, d) = \frac{p(d, \vec{\theta})p(\vec{\theta})}{p(d)}, \quad (\text{A.1})$$

where $p(\vec{\theta}, d)$ is the posterior probability of the vector of parameters $\vec{\theta}$ given some data set d . The quantity $p(\vec{\theta})$ is the prior information about the source parameters, reflecting any initial beliefs held before the data was taken. The evidence, $p(d)$, is the normalization of the posterior, which also generally holds valuable information about the signal, but will not be the focus of this work. The quantity $p(d, \vec{\theta})$ is the likelihood of the data, and describes the probability that one would see a data set d given some set of parameters $\vec{\theta}$. For GW data analysis, this is given by

$$p(d, \vec{\theta}) \propto \exp \left[-\frac{1}{2} \sum_i^{N_{\text{detector}}} (d_i - h_i | d_i - h_i) \right], \quad (\text{A.2})$$

for each data series d_i and detector response template h_i from the i -th detector, where the noise-weighted inner product is given by

$$(d - h | d - h) = 4 \text{Re} \left[\int \frac{(d - h)(d - h)^*}{S_n(f)} df \right]. \quad (\text{A.3})$$

To estimate the posterior using real data from LIGO, one would use a Markov Chain Monte Carlo [319, 378] to explore the parameter space of the signal. This would yield a set of independent samples from the posterior that quantifies not only the most likely values for the vector $\vec{\theta}$, but also includes information about our confidence in those estimates. This approach is the most reliable and accurate, but it is too computationally expensive for our purposes. Even the most optimized algorithms would take considerable computational resources to analyze the number of sources examined in this paper. We therefore turn to a commonly used approximation of the posterior to estimate the confidence intervals on $\vec{\theta}$ that is much more computationally tractable: the Fisher information matrix.

We calculate the Fisher matrices for each detector and combine them to construct a total Fisher matrix

for each source according to Eq. (3.22). To properly reflect the ability of a terrestrial network to localize sources in the sky, we incorporate a time delay between detectors that is α - and δ -dependent. That is, for each detector besides the reference detector, we append the following factor to the phase:

$$t_{c,i} \rightarrow t_{c,\text{ref}} + \delta t_{c,i}(\alpha, \delta), \quad (\text{A.4})$$

where $\delta t_{c,i}$ is defined as

$$\delta t_{c,i}(\alpha, \delta) = \frac{\mathbf{x}_{\text{ref}} \cdot \hat{\mathbf{x}}_{\text{source}}(\alpha, \delta) - \mathbf{x}_i \cdot \hat{\mathbf{x}}_{\text{source}}(\alpha, \delta)}{c}. \quad (\text{A.5})$$

The detector positions $\mathbf{x}_{\text{ref}}$ and $\mathbf{x}_i$ are in Earth-centered coordinates, the unit vector $\hat{\mathbf{x}}_{\text{source}}$ points to the source in the sky in the same coordinates, and we have reintroduced the speed of light c for clarity. The positions of the detectors in these Earth-centered coordinates were taken from **LALSuite** [234]. This procedure is neglected when considering LISA, as sky localization comes from the orbital motion of the satellites and long signal durations for space-based detectors.

An additional concern in the context of utilizing Fisher metrics with consistent parameters is the description of the binary's orientation. There are three coordinate systems that naturally arise in the description of terrestrial and space detectors. The natural coordinate system to use for LISA is the ecliptic coordinate system, specifically the parameters θ_j and ϕ_j , as these are the quantities that show up in LISA's response function. For terrestrial detectors, the polarization angle $\bar{\psi}$ and the inclination angle ι naturally arise in the response function, where the polarization angle is naturally defined in the equatorial coordinate system. Finally, the source properties themselves are stipulated in the source frame, aligned with the orbital angular momentum $\mathbf{L}$, and subsequently used to calculate the waveform. Any choice is valid as long as it is consistently enforced, so we chose to use the equatorial coordinates, and we accounted for the coordinate transformation in the calculation of the derivative of the response function. An equally simple solution would be to compute the Fisher matrices in their respective, natural coordinates, then use the Jacobian matrix to transform them as follows:

$$\Gamma_{i'j'} = \frac{\partial x^i}{\partial x^{i'}} \Gamma_{ij} \frac{\partial x^j}{\partial x^{j'}}, \quad (\text{A.6})$$

which is exactly how we transform our bounds on generic modifications to theory-specific modifications.

The actual transformation relies on the construction of an explicit rotation matrix between the different frames of reference. Transforming between ecliptic and equatorial coordinates is a trivial rotation by a constant angle, so we will instead just describe the transformation between the source frame and the equatorial system.

The first frame in question is the equatorial frame, which is the frame that defines the parameters θ_L , ϕ_L , α , and δ . From these quantities, one can construct two vectors: the direction of propagation $\hat{\mathbf{N}}$ (which points from the solar system to the source), and the direction of the orbital angular momentum $\hat{\mathbf{L}}$ at some reference frequency. These two vectors also define the inclination angle of the orbital angular momentum

$$\cos \iota = -\hat{\mathbf{L}} \cdot \hat{\mathbf{N}}, \quad (\text{A.7})$$

which will be needed in the next frame.

The second frame is the source frame, in which the waveform is naturally constructed. This frame is defined by a coordinate system with $\hat{\mathbf{L}} = \hat{\mathbf{z}}$, while the other two Cartesian axes are chosen such that the direction of propagation $-\hat{\mathbf{N}}$ (where $\hat{\mathbf{N}}$ points from the solar system to the source) lies in the x - z plane when the reference phase $\phi_{\text{ref}} = 0$. The vector $\hat{\mathbf{N}}$ is then rotated azimuthally by an angle ϕ_{ref} for nonzero reference phases. The angle between $\hat{\mathbf{L}}$ and $\hat{\mathbf{N}}$ in the source frame is just the inclination defined in Eq. (A.7), which fully specifies this vector in the second frame. Using these two vectors, we can construct a third, orthogonal vector as the cross product of these two, which we will call $\hat{\mathbf{K}} = \hat{\mathbf{L}} \times \hat{\mathbf{N}}$.

With three vectors in each frame, we can construct an explicit rotation matrix to transform any quantities from one frame to the other by the set of equations

$$\begin{aligned} \hat{\mathbf{L}}_{eq} &= \mathbf{R} \cdot \hat{\mathbf{L}}_{SF}, \\ \hat{\mathbf{N}}_{eq} &= \mathbf{R} \cdot \hat{\mathbf{N}}_{SF}, \\ \hat{\mathbf{K}}_{eq} &= \mathbf{R} \cdot \hat{\mathbf{K}}_{SF}, \end{aligned} \quad (\text{A.8})$$

where $\mathbf{R}$ is the unspecified rotation matrix and the subscripts “eq” and “SF” correspond to equatorial coordinates and source-frame coordinates, respectively. The system (A.8) can be inverted analytically, resulting in analytical expressions for the rotation matrix $\mathbf{R}$.

This rotation matrix allows us to transform any quantity between the two frames. This can be used to calculate the ecliptic angles of the total angular momentum $\hat{\mathbf{J}}$, which is needed for the LISA response function. The vector $\hat{\mathbf{J}}$ is easily constructed in the source frame, as the spins are defined in this frame and the orbital angular momentum already defines the coordinate system. The vector is simply rotated into the equatorial frame, and subsequently into the ecliptic frame, to compute the LISA response function.

We also need to specify the polarization angle for the terrestrial network. We simply use the relation [286]

$$\tan \bar{\psi} = \frac{\hat{\mathbf{J}} \cdot \hat{\mathbf{z}} - (\hat{\mathbf{J}} \cdot \hat{\mathbf{N}})(\hat{\mathbf{z}} \cdot \hat{\mathbf{N}})}{\hat{\mathbf{N}} \cdot (\hat{\mathbf{J}} \times \hat{\mathbf{z}})}, \quad (\text{A.9})$$

where $\hat{\mathbf{z}}$ is the unit vector of the equatorial coordinate system aligned with the axis of rotation of the Earth, defining a globally consistent polarization angle. These transformations allow us to use the vector of parameters outlined above, where all the quantities are consistently defined.

Once a combined Fisher for each source is calculated, the inversion of each Fisher results in the individual covariance matrices, which effectively acts as marginalization. We extract the variance of the ppE parameter β by taking the diagonal element $\sigma_{\beta\beta}$, which gives us a marginalized posterior on β for a single source. Finally, to combine the sources, we multiply the marginalized posteriors together (because each source is completely independent), which for a series of Gaussians becomes

$$\begin{aligned} p(\beta|\vec{\theta}) &\propto \prod_i^N \exp\left(-\frac{1}{2} \frac{\beta^2}{\sigma_{\beta,i}^2}\right) \\ &\propto \exp\left(-\frac{1}{2} \beta^2 \sum_i^N \frac{1}{\sigma_{\beta,i}^2}\right). \end{aligned} \quad (\text{A.10})$$

Therefore, our resulting bound on β is simply given by

$$\sigma_{\beta}^2 = \left(\sum_i^N \frac{1}{\sigma_{\beta,i}^2}\right)^{-1}. \quad (\text{A.11})$$

Appendix B

Mapping to Specific Theories

The main goal of this appendix is to map parameterized deviations, that do not necessarily have a physical interpretation, to specific parameters appearing in beyond-GR theories.

B.1 Dipole Radiation

In GR, the generation of GWs is sourced from the second time derivative of the mass quadrupole moment, resulting in quadrupolar radiation. This connection to the quadrupole moment is tied to the conservation of the stress energy tensor, rooted in the Bianchi identities (a purely geometrical constraint). If additional fields were added to the gravitational sector that were not subject to such energy conditions, one would generically expect dipolar radiation, providing an additional avenue of energy loss for the system. An additional channel for outgoing power would drive the binary to inspiral faster than what would be predicted by GR, and this faster inspiral would produce a measurable effect on the waveform.

To determine this effect on the waveform, we can write the time derivative of the gravitational binding energy of the system as [161]

$$\dot{E} = \dot{E}_{\text{GR}} + \delta\dot{E}, \quad (\text{B.1})$$

where $\dot{E}_{\text{GR}}$ is the GW power output in GR, and $\delta\dot{E}$ is our generic deviation. In terms of these parameters, our modification to the waveform becomes (in the language of ppE parameters) [161]

$$\beta_{\text{dipole}} = \frac{-3}{224} \eta^{2/5} \delta\dot{E}, \quad (\text{B.2})$$

where $\eta = m_1 m_2 / (m_1 + m_2)^2$ is the symmetric mass ratio of the binary system.

Of course, $\delta\dot{E}$ is written generically in Eq. (B.1). Once a specific theory has been selected, this term will be a function of the source parameters and of any fundamental constants of the theory in question. For example, in Einstein-dilaton Gauss-Bonnet gravity (EdGB) [137] the waveform modification can be calculated

to be [124]

$$\beta_{\text{EdGB}} = -\frac{5}{7168} \frac{\zeta_{\text{EdGB}}}{\eta^{18/5}} \frac{(m_1^2 s_2^{\text{EdGB}} - m_2^2 s_1^{\text{EdGB}})^2}{m^4}, \quad (\text{B.3})$$

$$s_i^{\text{EdGB}} = \frac{2 \left[(1 - \chi_i^2)^{1/2} - 1 + \chi_i^2 \right]}{\chi_i^2}, \quad (\text{B.4})$$

where ζ_{EdGB} is related to the coupling parameter of the theory α_{EdGB} by $\zeta_{\text{EdGB}} = 16\pi\alpha_{\text{EdGB}}^2(1+z)^4/m^4$, and $m = m_1 + m_2$ is the total redshifted mass of the system. The quantities s_i^{EdGB} given in Eq. (B.4) are the sensitivities of the BHs, and χ_i are the dimensionless, (anti-)aligned spin components of the i th BH.

Because of the approximations used to derive Eq. (B.3), this particular formula is only valid when $\sqrt{\alpha_{\text{EdGB}}} \leq m_s/2$, where m_s is the smallest length scale of the system (see e.g. [349]). For this work, the smallest length scale will be the mass of the smaller BH, m_2 .

B.2 Black Hole Evaporation

High-energy theories that might be candidates for quantum theories of gravity often involve the embedding of our four-dimensional spacetime in a higher-dimensional space, where the extra dimensions are often compactified. For example, Arkani-Hamed, Dimopoulos, and Dvali proposed a model which had implications for the hierarchy problem between the electroweak and Planck scale [379, 380]. Another set of models proposed by Randall and Sundrum (RS-I/II) [381, 382] postulate a braneworld model where the four-dimensional brane we occupy resides in a five-dimensional anti-de Sitter bulk spacetime. In RS-II, BHs were initially predicted to evaporate much faster as compared with analogous situations in four dimensions, with an evaporation rate given by [327, 383]

$$\frac{dm}{dt} = -2.8 \times 10^{-7} \left(\frac{1M_\odot(1+z)}{m} \right)^2 \left(\frac{l}{10\mu m} \right)^2 \frac{M_\odot}{\text{yr}}, \quad (\text{B.5})$$

where l is the lengthscale of the extra dimension and m is the detected mass. However, more recent work has shown that black holes in RS-II are actually stable and evaporation does not occur [384, 385].

Regardless of the physical origin of the evaporation, it is still interesting to consider its effect on the gravitational waveform. Let us imagine that either the volume or the area of a BH changes with time due to some quantum or classical extension of GR. The volume and the area are common geometric quantities associated with a BH, so it is plausible that if BH solutions become time-dependent, then it is these quantities that acquire the time dependence. Assume then that $dV/dt = c_V \ell^2$ or $dA/dt = c_A \ell$, where $c_{V,A}$ are dimensionless constants and ℓ is a new length that controls the scale at which time dependence

kicks in. If so, using that $V = (32\pi/3)m^3$ and $A = 16\pi m^2$ for a Schwarzschild BH, we then have that $dm/dt = [c_V/(32\pi)](\ell/m)^2$ or $dm/dt = [c_A/(32\pi)](\ell/m)$. On general grounds, then, one would expect $dm/dt \sim (\ell/m)^q$ with $q = 1$ or $q = 2$, depending on whether the time dependence acts on the area or the volume of the BH, and a constraint on dm/dt would then imply a constraint on the evaporation scale ℓ .

Regardless of the process that leads to evaporation, the waveform modification has the form [386]

$$\beta_{\text{BHE}} = \frac{25}{851968} \dot{M} \left(\frac{3 - 26\eta + 34\eta^2}{\eta^{2/5}(1 - 2\eta)} \right), \quad (\text{B.6})$$

where $\dot{M} = dM/dt = dm_1/dt + dm_2/dt$ is the anomalous evaporation rate.

B.3 Local Position Invariance Violation

In the case where Newton's gravitational constant is promoted to a time-dependent quantity, conspicuous additional accelerations could be experienced by binaries inspiralling together. This phenomenon could come about, for example, because the gravitational constant is tied to a background scalar field which evolves on cosmological timescales. This effect can be observed as alterations to the binding energy of the binary, and it has a mapping to the ppE framework [260]:

$$\beta_{\dot{G}} = \frac{-25}{65526} \frac{\dot{G}\mathcal{M}}{(1+z)}, \quad (\text{B.7})$$

where $\dot{G} = dG/dt$ is the time derivative of the gravitational constant and $\mathcal{M}$ is the redshifted chirp mass.

B.4 Parity Violation

Many attempts to unify quantum mechanics and gravity involve terms quadratic in curvature at the level of the action in the low-energy limit, as well as additional fields coupled to these higher-order terms. The strength of this coupling is determined by the coupling parameter of the theory, and therefore determines the magnitude of the effect on the waveform. EdGB (discussed above) is an example of this type of modification where the modifying parameter comes at a negative PN order because of dipolar radiation. EdGB, however, preserves parity because the term added to the action is parity-even, introducing a scalar field that is also parity-even. A quadratic theory that does not preserve parity is dynamical Chern-Simons (dCS) gravity [137], which incorporates an additional quadratic curvature term into the action that is parity-odd. In order to keep the action invariant under parity transformations, this odd-parity term must be coupled to an odd-parity

scalar field, leading to a variety of implications in different gravitational interactions [137].

This modification affects the waveform as follows [124]:

$$\beta_{\text{dCS}} = -\frac{5}{8192} \frac{\zeta_{\text{dCS}}}{\eta^{14/5}} \frac{(m_1 s_2^{\text{dCS}} - m_2 s_1^{\text{dCS}})^2}{m^2} + \frac{15075}{114688} \frac{\zeta_{\text{dCS}}}{\eta^{14/5}} \frac{(m_2^2 \chi_1^2 - \frac{350}{201} m_1 m_2 \chi_1 \chi_2 + m_1^2 \chi_2^2)}{m^2}, \quad (\text{B.8})$$

$$s_i^{\text{dCS}} \equiv \frac{2 + 2\chi_i^4 - 2(1 - \chi_i^2)^{1/2} - \chi_i^2 [3 - 2(1 - \chi_i^2)^{1/2}]}{2\chi_i^3}, \quad (\text{B.9})$$

where ζ_{dCS} is related to the coupling parameter by $\zeta_{\text{dCS}} = 16\pi\alpha_{\text{dCS}}^2(1+z)^4/m^4$. The quantity s_i^{dCS} given in Eq. (B.9) is the sensitivity of the i th BH in dCS.

As the result of the approximations involved in the derivation of the flux, Eq. (B.8) is only valid if $\sqrt{\alpha_{\text{dCS}}} \leq m_s/2$, where m_s is the smallest length scale of the system, just as in EdGB. Here we are interested in BBHs, and m_s is the mass of the smaller BH.

B.5 Lorentz Violation

Noncommutative gravity promotes the coordinates in GR to operators with a nontrivial commutation relation defined by $[\hat{x}^\mu, \hat{x}^\nu] = i\theta^{\mu\nu}$, where $\theta^{\mu\nu}$ is a real, constant antisymmetric tensor [109, 224]. This tensor plays a role analogous to the role of Planck's constant in quantum mechanics, and it defines a length scale at which there is a fundamental uncertainty between physical parameters.

Defining the quantity $\Lambda^2 = \theta^{0i}\theta_{0i}/(l_p t_p)^2$, where l_p and t_p are the Planck length and time, respectively, one can derive the modification to the waveform as [109, 224]

$$\beta_{\text{NC}} = -\frac{75}{256} \eta^{-4/5} (2\eta - 1) \Lambda^2. \quad (\text{B.10})$$

In this parameterization, $\sqrt{\Lambda}$ defines the energy scale of noncommutativity, relative to the Planck scale.

B.6 Modified Dispersion

Another assumption made by GR is that gravitons are massless. If this is not assumed, the leading-order correction to the measured GW signal would come about through the propagation of GW [112, 131]. The graviton would be ascribed a massive-particle dispersion relation $E^2 = p^2 + m_g^2$, where E is the graviton energy, p is the graviton momentum, and m_g is the graviton mass. With a nonlinear relation between

energy and momentum, one would expect that the group velocity would become frequency-dependent. This introduces an additional term in the GW phase [131]:

$$\beta_{\text{MG}} = \pi^2 \frac{D_0}{1+z} \frac{\mathcal{M}_z}{\lambda_{\text{MG}}^2}, \quad (\text{B.11})$$

$$D_0 \equiv (1+z) \int_0^z \frac{1}{H(z')} \frac{dz'}{(1+z')^2}, \quad (\text{B.12})$$

where D_0 is a new cosmological distance similar to the luminosity distance, and λ_g is the Compton wavelength of the graviton, related to the mass by $\lambda_g = h/m_g$. To evaluate the Hubble parameter $H(z)$ we use the cosmological parameters inferred from the Planck Collaboration [180] and software from the **Astropy** python package [7, 8].

Appendix C

Inspiral/merger/ringdown vs. inspiral waveforms

Concerning the deviations away from GR that we have injected into the waveforms, we examine two families of modifications: those that affect GW propagation and those that modify GW generation. The difference between these two mechanisms arises from our lack of knowledge about the dynamics of BBHs close to merger in modified theories of gravity. To reflect this ignorance, we include the modification due to generation effects in the inspiral portion of the waveform only. Propagation effects are under no such shroud as the mechanism responsible acts in the low-curvature regions between galaxies and should equally affect the waveform across the entire frequency range. We therefore include modifications due to propagation effects in the entire waveform. As we are only ever looking at one effect at a time, these two families of effects are never examined concurrently. To incorporate these modifications, we utilize the ppE methodology [107, 115–117]. In the case of precessing systems, the modifications are treated slightly differently. For generational effects, we append a phase modification to the waveform in the coprecessing frame, where the physics of GW generation are approximately the same as those for a nonprecessing binary. The waveform is then “twisted-up” in the usual fashion for IMRPhenomPv2 waveforms, but with the modified coprecessing waveform. For propagation effects, we append the modification to the waveform at all frequencies, after the waveform has been transformed to the inertial frame. In equations:

$$\tilde{h}_{\text{coprec,gen}} = \begin{cases} \tilde{h}_{\text{coprec,GR}} e^{i\beta(\mathcal{M}\pi f)^{-b/3}} & f < 0.018m \\ \tilde{h}_{\text{coprec,GR}} & 0.018m < f \end{cases} \quad (\text{C.1})$$

$$\tilde{h}_{\text{inertial,prop}} = \tilde{h}_{\text{inertial,GR}} e^{i\beta(\mathcal{M}\pi f)^{-b/3}}. \quad (\text{C.2})$$

A comparison between the two methods is shown in Fig. C.1, which illustrates that the difference is small. Because of this, we used the full inspiral-merger-ringdown modification in all of this paper.

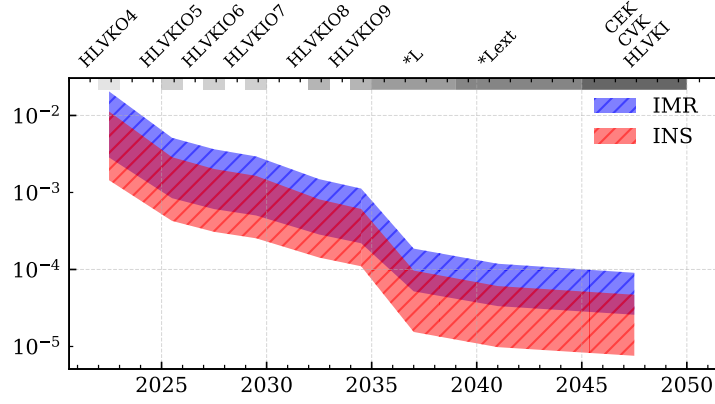


Figure C.1: Comparison between the constraints on β at 1.5PN predicted by using the generation modification (INS), as opposed to the propagation modification (IMR). We used the same catalogs and networks (SOBH Base and SOBH S1) in both cases. The difference is negligible when considering the order-of-magnitude constraints of interest in this work.

Appendix D

Concerns over source classification

Unique GW events can offer new and exciting avenues to test fundamental physics, but they do not come without their own complications. For example, binaries near the mass gap, where one component might be a heavy NS or a light BH, can potentially provide a strong test of GR. However, such tests would undoubtedly be hindered if one were unsure of the classification of each component object (BH or NS). In many cases, one can be confident in the exact nature of the binary, for example in the case of an electromagnetic counterpart or a robust detection of the tidal deformability of one of the objects. Unfortunately, two intriguing events we considered do not fall in this category, and we opted to not include them in the main body of this work, as explained in Sec. 6.3.2. In this Appendix, we will discuss some implications of these complications.

Let us first discuss GW190814, which was found to be a highly asymmetric binary. The uncertain nature of the lighter compact object in the binary poses some complications to our analysis. This is because a NS and a BH have different properties in dCS and EdGB theories. In particular, BHs in EdGB gravity support monopole scalar charges (see e.g. [138, 295, 356]), while NSs do not [296, 354]. This means that if one knew that the binary was mixed, one could use NSBH waveforms with EdGB modifications, setting the sensitivity s_2^{EdGB} in Eq. (6.11) to zero. On the other hand, if one assumes that the binary was composed of two BHs, then one must use a BBH waveform with EdGB modifications that allow both sensitivities s_1^{EdGB} and s_2^{EdGB} in Eq. (6.11) to be non-zero. Therefore, our lack of knowledge about the exact composition of the lighter object can significantly affect any parameter estimation in EdGB, hence casting doubt over any constraint placed on the theory from such systems.

As we mentioned above, for the GW190814 event there are two reasonable possibilities: that the source is a BBH or a NSBH binary. There are various reasons to believe either one over the other (see e.g. [352, 387–389]), and as such, any analysis making conclusions based on the *a priori* assumption of one classification over the other should carefully consider both possibilities. To show how the assumptions on the binary constituents can dramatically affect the constraints on EdGB, we analyzed GW190814 for *both BBH and mixed binary scenarios*. The posteriors on $\sqrt{\alpha_{\text{EdGB}}}$ are shown in Fig. D.1. We can immediately see that the prior assumption of the source composition seriously impacts our constraints, with over an order of magnitude

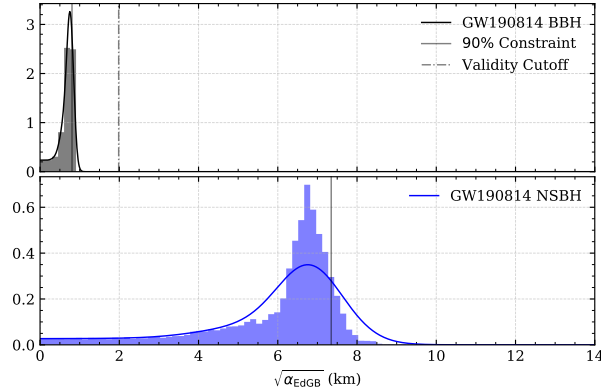


Figure D.1: The posterior distributions on $\sqrt{\alpha_{\text{EdGB}}}$ for the GW190814 event. The top panel was derived from an analysis that assumed GW190814 was a BBH system, while the lower panel shows results that were derived assuming the source is a NSBH system. The small coupling criteria for this source, when considering it as a NSBH source, is approximately 18 km, and lies beyond the edge of the plot. As is immediately apparent, the upper limit on $\sqrt{\alpha_{\text{EdGB}}}$ based on each of these assumptions differs by an order of magnitude.

difference between the two bounds. This fact casts doubt over any constraint on this particular theory using this event.

Moreover, possible waveform systematics can clearly be seen in our posteriors. The LVC analysis of GW190814 led to posteriors on a -1PN order modification to GR that excluded GR to 90% confidence [58]. We performed a similar Bayesian analysis on GW190814 both with a generic -1PN order modification and with an EdGB modification. 5 In both cases we find that the posteriors on the non-GR parameter excludes GR, in agreement with the LVC analysis. This result is not believed to be a real deviation from GR, but rather a consequence of waveform systematics and covariances between parameters. While one could ignore these deviations and calculate upper limits on $\sqrt{\alpha_{\text{EdGB}}}$ regardless, the obvious impact of waveform systematics is of note [390]. This stark difference in the two posteriors along with the issues of waveform systematics, in a context where neither interpretation of the event GW190814 is unquestionable, led us to omit this source from the main body of this work.

In the case of GW190425 [350], the situation becomes even more stark. While this event might lead to stringent constraints on one or both of these theories if it were considered a NSBH binary, there is the strong possibility that this source is actually a NSNS binary. In the latter case, one cannot meaningfully constrain either EdGB or dCS. This is an even more extreme case where the difference between the constraints when using one assumption over the other is infinite. Considering the significant impact of the prior assumption on the composition of the binary on final constraint conclusions, and the current evidence suggesting this is actually a NSNS binary [350], we neglected this source as well.

Appendix E

Methods for combining posteriors

To combine posteriors on a shared parameter from multiple observations, there are two commonly employed methods of calculating the joint posterior, as outlined in Sec. 6.3.5. One involves the fitting of some ansatz function or KDE approximation to the discrete, marginalized likelihoods, then multiplying the analytic fits together for the cumulative constraint. Meanwhile, the other method involves the direct multiplication of the discrete, marginalized likelihoods together, then working with the final histogram directly.

Both have their benefits and drawbacks, and we present both in Fig. E.1 and Fig. E.2 for dCS and EdGB, respectively. For the former method, we have used a KDE to estimate the marginalized likelihoods for each individual event, and for the latter method, we smoothed the *final* histogram with a Savitzky-Golay filter with a polynomial order of 5 and a window of 7 before calculating confidence intervals for numerical stability.

As noted in the main body of this paper, one drawback to the KDE method is the lack of ability to account for hard cutoffs, for example the boundary at $\sqrt{\alpha} = 0$. The KDE approach led to a joint posterior that rapidly dropped to zero for small $\sqrt{\alpha}$, and to remedy this, we fit the KDE to a distribution described by the set of samples twice as large as the original set coming from the MCMC. We used the distribution $\{s_i\} \cup -\{s_i\}$, which amounts to reflecting the distribution across 0. We then renormalized the distribution numerically in the range $\sqrt{\alpha} \in [0, \infty]$. This effectively solved the boundary issue in this particular use case, as the KDE approximation smoothly approached a final value at $\sqrt{\alpha} = 0$.

The final result from the two methods agreed well for the six sources considered in this work. To be maximally cautious, we only use the worse of the two methods when quoting 90% confidence constraints on EdGB.

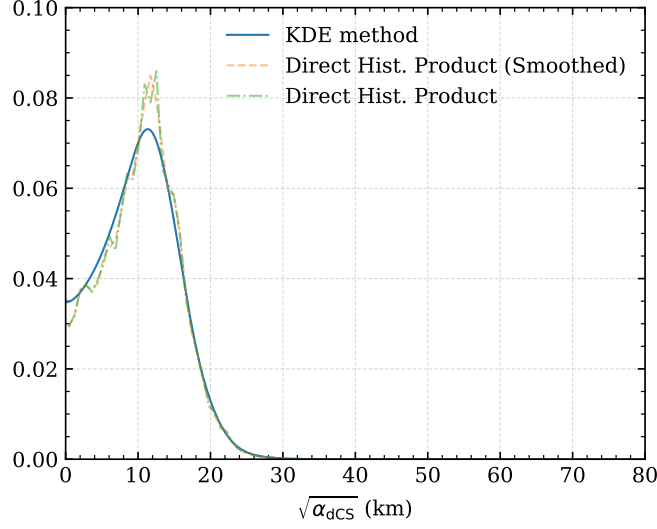


Figure E.1: Comparison of the two different methods for combining posterior distributions from independent experiments into a single, joint distribution on $\sqrt{\alpha_{\text{dCS}}}$. The two methods in question involve the fitting of functions to the individual distributions (KDE method) and the direct product of the histograms (Direct Hist. Product), as well as a smoothed version of the latter used for the calculation of the confidence intervals. While there seems to be a bias towards higher values of α_{dCS} , any conclusion from this distribution should be tempered, as this analysis still doesn't provide reliable results, due to violations to the small coupling approximation.

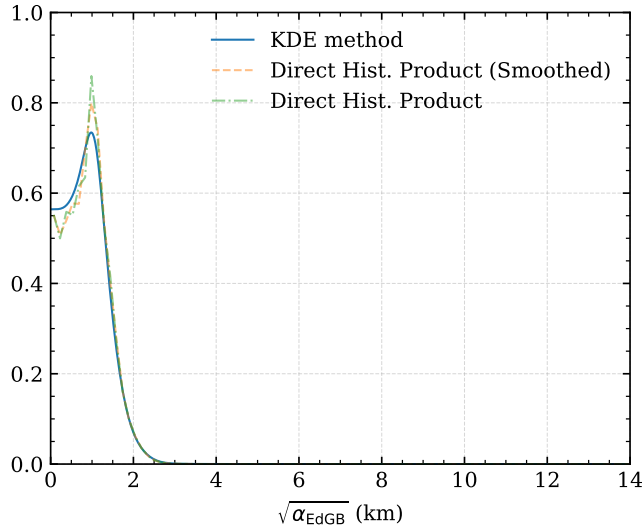


Figure E.2: Comparison of the two different methods for combining posterior distributions from independent experiments into a single, joint distribution on $\sqrt{\alpha_{\text{EdGB}}}$. The two methods in question involve the fitting of functions to the individual distributions (KDE method) and the direct product of the histograms (Direct Hist. Product), as well as a smoothed version of the latter used for the calculation of the confidence intervals.

Bibliography

- [1] Brian Gough. *GNU scientific library reference manual*. Network Theory Ltd., 2009.
- [2] Charles R. Harris et al. “Array programming with NumPy”. In: *Nature* 585.7825 (Sept. 2020), pp. 357–362. DOI: [10.1038/s41586-020-2649-2](https://doi.org/10.1038/s41586-020-2649-2). URL: <https://doi.org/10.1038/s41586-020-2649-2>.
- [3] Pauli Virtanen et al. “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python”. In: *Nature Methods* 17 (2020), pp. 261–272. DOI: [10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2).
- [4] Davide Gerosa and Michele Vallisneri. “filltex: Automatic queries to ADS and INSPIRE databases to fill LaTeX bibliography”. In: *The Journal of Open Source Software* 2 (May 2017), p. 222. DOI: [10.21105/joss.00222](https://doi.org/10.21105/joss.00222).
- [5] J. D. Hunter. “Matplotlib: A 2D graphics environment”. In: *Computing in Science & Engineering* 9.3 (2007), pp. 90–95. DOI: [10.1109/MCSE.2007.55](https://doi.org/10.1109/MCSE.2007.55).
- [6] Daniel Foreman-Mackey. “corner.py: Scatterplot matrices in Python”. In: *The Journal of Open Source Software* 1.2 (June 2016), p. 24. DOI: [10.21105/joss.00024](https://doi.org/10.21105/joss.00024). URL: <https://doi.org/10.21105/joss.00024>.
- [7] Astropy Collaboration et al. “Astropy: A community Python package for astronomy”. In: *A&A* 558, A33 (Oct. 2013), A33. DOI: [10.1051/0004-6361/201322068](https://doi.org/10.1051/0004-6361/201322068). arXiv: [1307.6212](https://arxiv.org/abs/1307.6212) [[astro-ph.IM](https://arxiv.org/abs/1307.6212)].
- [8] Astropy Collaboration et al. “The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package”. In: *AJ* 156.3, 123 (Sept. 2018), p. 123. DOI: [10.3847/1538-3881/aabc4f](https://doi.org/10.3847/1538-3881/aabc4f). arXiv: [1801.02634](https://arxiv.org/abs/1801.02634) [[astro-ph.IM](https://arxiv.org/abs/1801.02634)].
- [9] Gaël Guennebaud, Benoît Jacob, et al. *Eigen v3*. <http://eigen.tuxfamily.org>. 2010.
- [10] Andreas Griewank, David Juedes, and Jean Utke. “Algorithm 755: ADOL-C: A Package for the Automatic Differentiation of Algorithms Written in C/C++”. In: *ACM Trans. Math. Softw.* 22.2 (June 1996), pp. 131–167. ISSN: 0098-3500. DOI: [10.1145/229473.229474](https://doi.org/10.1145/229473.229474). URL: <https://doi.org/10.1145/229473.229474>.

- [11] Leonardo Dagum and Ramesh Menon. “OpenMP: An Industry-Standard API for Shared-Memory Programming”. In: *IEEE Comput. Sci. Eng.* 5.1 (Jan. 1998), pp. 46–55. ISSN: 1070-9924. DOI: [10.1109/99.660313](https://doi.org/10.1109/99.660313). URL: <https://doi.org/10.1109/99.660313>.
- [12] M. Frigo and S.G. Johnson. “The Design and Implementation of FFTW3”. In: *Proceedings of the IEEE* 93.2 (2005), pp. 216–231. DOI: [10.1109/JPROC.2004.840301](https://doi.org/10.1109/JPROC.2004.840301).
- [13] Wolfram Research Inc. *Mathematica, Version 12.1*. URL: <https://www.wolfram.com/mathematica>.
- [14] B. P. Abbott et al. “Observation of Gravitational Waves from a Binary Black Hole Merger”. In: *Phys. Rev. Lett.* 116.6 (2016), p. 061102. DOI: [10.1103/PhysRevLett.116.061102](https://doi.org/10.1103/PhysRevLett.116.061102). arXiv: [1602.03837](https://arxiv.org/abs/1602.03837) [gr-qc].
- [15] B. P. Abbott et al. “LIGO: The Laser interferometer gravitational-wave observatory”. In: *Rept. Prog. Phys.* 72 (2009), p. 076901. DOI: [10.1088/0034-4885/72/7/076901](https://doi.org/10.1088/0034-4885/72/7/076901). arXiv: [0711.3041](https://arxiv.org/abs/0711.3041) [gr-qc].
- [16] F. Acernese et al. “Advanced Virgo: a second-generation interferometric gravitational wave detector”. In: *Class. Quant. Grav.* 32.2 (2015), p. 024001. DOI: [10.1088/0264-9381/32/2/024001](https://doi.org/10.1088/0264-9381/32/2/024001). arXiv: [1408.3978](https://arxiv.org/abs/1408.3978) [gr-qc].
- [17] T. Akutsu et al. “Overview of KAGRA: Detector design and construction history”. 2020. arXiv: [2005.05574](https://arxiv.org/abs/2005.05574) [physics.ins-det].
- [18] B. P. Abbott et al. “GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs”. In: *Phys. Rev. X* 9.3 (2019), p. 031040. DOI: [10.1103/PhysRevX.9.031040](https://doi.org/10.1103/PhysRevX.9.031040). arXiv: [1811.12907](https://arxiv.org/abs/1811.12907) [astro-ph.HE].
- [19] R. Abbott et al. “GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run”. In: (Nov. 2021). arXiv: [2111.03606](https://arxiv.org/abs/2111.03606) [gr-qc].
- [20] R. Abbott et al. “The population of merging compact binaries inferred using gravitational waves through GWTC-3”. In: (Nov. 2021). arXiv: [2111.03634](https://arxiv.org/abs/2111.03634) [astro-ph.HE].
- [21] R. Abbott et al. “Population Properties of Compact Objects from the Second LIGO-Virgo Gravitational-Wave Transient Catalog”. In: *Astrophys. J. Lett.* 913.1 (2021), p. L7. DOI: [10.3847/2041-8213/abe949](https://doi.org/10.3847/2041-8213/abe949). arXiv: [2010.14533](https://arxiv.org/abs/2010.14533) [astro-ph.HE].
- [22] B. P. Abbott et al. “Binary Black Hole Population Properties Inferred from the First and Second Observing Runs of Advanced LIGO and Advanced Virgo”. In: *Astrophys. J. Lett.* 882.2 (2019), p. L24. DOI: [10.3847/2041-8213/ab3800](https://doi.org/10.3847/2041-8213/ab3800). arXiv: [1811.12940](https://arxiv.org/abs/1811.12940) [astro-ph.HE].

- [23] R. Abbott et al. “Observation of Gravitational Waves from Two Neutron Star–Black Hole Coalescences”. In: *Astrophys. J. Lett.* 915.1 (2021), p. L5. DOI: [10.3847/2041-8213/ac082e](#). arXiv: [2106.15163 \[astro-ph.HE\]](#).
- [24] R. Abbott et al. “GW190521: A Binary Black Hole Merger with a Total Mass of $150M_{\odot}$ ”. In: *Phys. Rev. Lett.* 125.10 (2020), p. 101102. DOI: [10.1103/PhysRevLett.125.101102](#). arXiv: [2009.01075 \[gr-qc\]](#).
- [25] R. Abbott et al. “GW190814: Gravitational Waves from the Coalescence of a 23 Solar Mass Black Hole with a 2.6 Solar Mass Compact Object”. In: *Astrophys. J. Lett.* 896.2 (2020), p. L44. DOI: [10.3847/2041-8213/ab960f](#). arXiv: [2006.12611 \[astro-ph.HE\]](#).
- [26] R. Abbott et al. “GW190412: Observation of a Binary-Black-Hole Coalescence with Asymmetric Masses”. In: *Phys. Rev. D* 102.4 (2020), p. 043015. DOI: [10.1103/PhysRevD.102.043015](#). arXiv: [2004.08342 \[astro-ph.HE\]](#).
- [27] B. P. Abbott et al. “GW190425: Observation of a Compact Binary Coalescence with Total Mass $\sim 3.4M_{\odot}$ ”. In: *Astrophys. J. Lett.* 892.1 (2020), p. L3. DOI: [10.3847/2041-8213/ab75f5](#). arXiv: [2001.01761 \[astro-ph.HE\]](#).
- [28] B. P. Abbott et al. “GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral”. In: *Phys. Rev. Lett.* 119.16 (2017), p. 161101. DOI: [10.1103/PhysRevLett.119.161101](#). arXiv: [1710.05832 \[gr-qc\]](#).
- [29] Benjamin P. Abbott et al. “GW170104: Observation of a 50-Solar-Mass Binary Black Hole Coalescence at Redshift 0.2”. In: *Phys. Rev. Lett.* 118.22 (2017). [Erratum: *Phys.Rev.Lett.* 121, 129901 (2018)], p. 221101. DOI: [10.1103/PhysRevLett.118.221101](#). arXiv: [1706.01812 \[gr-qc\]](#).
- [30] F. J. Fattoyev et al. “GW190814: Impact of a 2.6 solar mass neutron star on the nucleonic equations of state”. In: *Phys. Rev. C* 102.6 (2020), p. 065805. DOI: [10.1103/PhysRevC.102.065805](#). arXiv: [2007.03799 \[nucl-th\]](#).
- [31] Hung Tan, Jacquelyn Noronha-Hostler, and Nico Yunes. “Neutron Star Equation of State in light of GW190814”. In: *Phys. Rev. Lett.* 125.26 (2020), p. 261104. DOI: [10.1103/PhysRevLett.125.261104](#). arXiv: [2006.16296 \[astro-ph.HE\]](#).
- [32] V. De Luca et al. “GW190521 Mass Gap Event and the Primordial Black Hole Scenario”. In: *Phys. Rev. Lett.* 126.5 (2021), p. 051101. DOI: [10.1103/PhysRevLett.126.051101](#). arXiv: [2009.01728 \[astro-ph.CO\]](#).

- [33] G. Raaijmakers et al. “Constraints on the Dense Matter Equation of State and Neutron Star Properties from NICER’s Mass–Radius Estimate of PSR J0740+6620 and Multimessenger Observations”. In: *Astrophys. J. Lett.* 918.2 (2021), p. L29. DOI: [10.3847/2041-8213/ac089a](https://doi.org/10.3847/2041-8213/ac089a). arXiv: [2105.06981](https://arxiv.org/abs/2105.06981) [[astro-ph.HE](#)].
- [34] Michael Zevin et al. “Exploring the Lower Mass Gap and Unequal Mass Regime in Compact Binary Evolution”. In: *Astrophys. J. Lett.* 899 (2020), p. L1. DOI: [10.3847/2041-8213/aba74e](https://doi.org/10.3847/2041-8213/aba74e). arXiv: [2006.14573](https://arxiv.org/abs/2006.14573) [[astro-ph.HE](#)].
- [35] Antonios Tsokaros, Milton Ruiz, and Stuart L. Shapiro. “GW190814: Spin and equation of state of a neutron star companion”. In: *Astrophys. J.* 905.1 (2020), p. 48. DOI: [10.3847/1538-4357/abc421](https://doi.org/10.3847/1538-4357/abc421). arXiv: [2007.05526](https://arxiv.org/abs/2007.05526) [[astro-ph.HE](#)].
- [36] Amanda M. Farah et al. “Bridging the Gap: Categorizing Gravitational-Wave Events at the Transition Between Neutron Stars and Black Holes”. In: (Nov. 2021). arXiv: [2111.03498](https://arxiv.org/abs/2111.03498) [[astro-ph.HE](#)].
- [37] B. P. Abbott et al. “GW170817: Measurements of neutron star radii and equation of state”. In: *Phys. Rev. Lett.* 121.16 (2018), p. 161101. DOI: [10.1103/PhysRevLett.121.161101](https://doi.org/10.1103/PhysRevLett.121.161101). arXiv: [1805.11581](https://arxiv.org/abs/1805.11581) [[gr-qc](#)].
- [38] Soumi De et al. “Tidal Deformabilities and Radii of Neutron Stars from the Observation of GW170817”. In: *Phys. Rev. Lett.* 121.9 (2018). [Erratum: *Phys.Rev.Lett.* 121, 259902 (2018)], p. 091102. DOI: [10.1103/PhysRevLett.121.091102](https://doi.org/10.1103/PhysRevLett.121.091102). arXiv: [1804.08583](https://arxiv.org/abs/1804.08583) [[astro-ph.HE](#)].
- [39] Andreas Bauswein et al. “Identifying a first-order phase transition in neutron star mergers through gravitational waves”. In: *Phys. Rev. Lett.* 122.6 (2019), p. 061102. DOI: [10.1103/PhysRevLett.122.061102](https://doi.org/10.1103/PhysRevLett.122.061102). arXiv: [1809.01116](https://arxiv.org/abs/1809.01116) [[astro-ph.HE](#)].
- [40] Deep Chatterjee et al. “Cosmology with Love: Measuring the Hubble constant using neutron star universal relations”. In: *Phys. Rev. D* 104.8 (2021), p. 083528. DOI: [10.1103/PhysRevD.104.083528](https://doi.org/10.1103/PhysRevD.104.083528). arXiv: [2106.06589](https://arxiv.org/abs/2106.06589) [[gr-qc](#)].
- [41] B. P. Abbott et al. “A Gravitational-wave Measurement of the Hubble Constant Following the Second Observing Run of Advanced LIGO and Virgo”. In: *Astrophys. J.* 909.2 (2021). [Erratum: *Astrophys.J.* 923, 279 (2021)], p. 218. DOI: [10.3847/1538-4357/abdc7](https://doi.org/10.3847/1538-4357/abdc7). arXiv: [1908.06060](https://arxiv.org/abs/1908.06060) [[astro-ph.CO](#)].
- [42] Rachel Gray et al. “Cosmological inference using gravitational wave standard sirens: A mock data analysis”. In: *Phys. Rev. D* 101.12 (2020), p. 122001. DOI: [10.1103/PhysRevD.101.122001](https://doi.org/10.1103/PhysRevD.101.122001). arXiv: [1908.06050](https://arxiv.org/abs/1908.06050) [[gr-qc](#)].

- [43] Alejandro Cardenas-Avendano et al. “Experimental Relativity with Accretion Disk Observations”. In: *Phys. Rev. D* 100.2 (2019), p. 024039. DOI: [10.1103/PhysRevD.100.024039](https://doi.org/10.1103/PhysRevD.100.024039). arXiv: [1903.04356](https://arxiv.org/abs/1903.04356) [gr-qc].
- [44] Samuel E. Gralla. “Can the EHT M87 results be used to test general relativity?” In: *Phys. Rev. D* 103.2 (2021), p. 024023. DOI: [10.1103/PhysRevD.103.024023](https://doi.org/10.1103/PhysRevD.103.024023). arXiv: [2010.08557](https://arxiv.org/abs/2010.08557) [astro-ph.HE].
- [45] T. Damour and N. Deruelle. “General relativistic celestial mechanics of binary systems. I. The post-Newtonian motion.” In: *Ann. Inst. Henri Poincaré Phys. Théor* 43.1 (Jan. 1985), pp. 107–132.
- [46] Thibault Damour and J. H. Taylor. “Strong-field tests of relativistic gravity and binary pulsars”. In: *Phys. Rev. D* 45 (6 Mar. 1992), pp. 1840–1868. DOI: [10.1103/PhysRevD.45.1840](https://doi.org/10.1103/PhysRevD.45.1840). URL: <https://link.aps.org/doi/10.1103/PhysRevD.45.1840>.
- [47] Toral Gupta et al. “New binary pulsar constraints on Einstein-æther theory after GW170817”. In: *Class. Quant. Grav.* 38.19 (2021), p. 195003. DOI: [10.1088/1361-6382/ac1a69](https://doi.org/10.1088/1361-6382/ac1a69). arXiv: [2104.04596](https://arxiv.org/abs/2104.04596) [gr-qc].
- [48] Nicolas Yunes and Scott A. Hughes. “Binary Pulsar Constraints on the Parameterized post-Einsteinian Framework”. In: *Phys. Rev. D* 82 (2010), p. 082002. DOI: [10.1103/PhysRevD.82.082002](https://doi.org/10.1103/PhysRevD.82.082002). arXiv: [1007.1995](https://arxiv.org/abs/1007.1995) [gr-qc].
- [49] Remya Nair and Nicolás Yunes. “Improved binary pulsar constraints on the parametrized post-Einsteinian framework”. In: *Phys. Rev. D* 101.10 (2020), p. 104011. DOI: [10.1103/PhysRevD.101.104011](https://doi.org/10.1103/PhysRevD.101.104011). arXiv: [2002.02030](https://arxiv.org/abs/2002.02030) [gr-qc].
- [50] Alejandro Cardenas-Avendano, Sourabh Nampalliwar, and Nicolas Yunes. “Gravitational-wave versus X-ray tests of strong-field gravity”. In: *Class. Quant. Grav.* 37.13 (2020), p. 135008. DOI: [10.1088/1361-6382/ab8f64](https://doi.org/10.1088/1361-6382/ab8f64). arXiv: [1912.08062](https://arxiv.org/abs/1912.08062) [gr-qc].
- [51] Clifford M. Will. “The Confrontation between General Relativity and Experiment”. In: *Living Rev. Rel.* 17 (2014), p. 4. DOI: [10.12942/lrr-2014-4](https://doi.org/10.12942/lrr-2014-4). arXiv: [1403.7377](https://arxiv.org/abs/1403.7377) [gr-qc].
- [52] Emanuele Berti et al. “Testing General Relativity with Present and Future Astrophysical Observations”. In: *Class. Quant. Grav.* 32 (2015), p. 243001. DOI: [10.1088/0264-9381/32/24/243001](https://doi.org/10.1088/0264-9381/32/24/243001). arXiv: [1501.07274](https://arxiv.org/abs/1501.07274) [gr-qc].
- [53] R. V. Pound and G. A. Rebka. “Gravitational Red-Shift in Nuclear Resonance”. In: *Phys. Rev. Lett.* 3 (9 Nov. 1959), pp. 439–441. DOI: [10.1103/PhysRevLett.3.439](https://doi.org/10.1103/PhysRevLett.3.439). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.3.439>.

- [54] R. V. Pound and G. A. Rebka. “Apparent Weight of Photons”. In: *Phys. Rev. Lett.* 4 (7 Apr. 1960), pp. 337–341. DOI: [10.1103/PhysRevLett.4.337](https://doi.org/10.1103/PhysRevLett.4.337). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.4.337>.
- [55] R. V. Pound and J. L. Snider. “Effect of Gravity on Nuclear Resonance”. In: *Phys. Rev. Lett.* 13 (18 Nov. 1964), pp. 539–540. DOI: [10.1103/PhysRevLett.13.539](https://doi.org/10.1103/PhysRevLett.13.539). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.539>.
- [56] J. H. Taylor, L. A. Fowler, and P. M. McCulloch. “Measurements of general relativistic effects in the binary pulsar PSR1913 + 16”. In: *Nature* 277.5696 (Feb. 1979), pp. 437–440. DOI: [10.1038/277437a0](https://doi.org/10.1038/277437a0).
- [57] B. P. Abbott et al. “Tests of General Relativity with the Binary Black Hole Signals from the LIGO-Virgo Catalog GWTC-1”. In: *Phys. Rev. D* 100.10 (2019), p. 104036. DOI: [10.1103/PhysRevD.100.104036](https://doi.org/10.1103/PhysRevD.100.104036). arXiv: [1903.04467 \[gr-qc\]](https://arxiv.org/abs/1903.04467).
- [58] R. Abbott et al. “Tests of general relativity with binary black holes from the second LIGO-Virgo gravitational-wave transient catalog”. In: *Phys. Rev. D* 103.12 (2021), p. 122002. DOI: [10.1103/PhysRevD.103.122002](https://doi.org/10.1103/PhysRevD.103.122002). arXiv: [2010.14529 \[gr-qc\]](https://arxiv.org/abs/2010.14529).
- [59] R. Abbott et al. “Tests of General Relativity with GWTC-3”. Dec. 2021. arXiv: [2112.06861 \[gr-qc\]](https://arxiv.org/abs/2112.06861).
- [60] B. P. Abbott et al. “Tests of General Relativity with GW170817”. In: *Phys. Rev. Lett.* 123.1 (2019), p. 011102. DOI: [10.1103/PhysRevLett.123.011102](https://doi.org/10.1103/PhysRevLett.123.011102). arXiv: [1811.00364 \[gr-qc\]](https://arxiv.org/abs/1811.00364).
- [61] B. P. Abbott et al. “Tests of general relativity with GW150914”. In: *Phys. Rev. Lett.* 116.22 (2016). [Erratum: *Phys.Rev.Lett.* 121, 129902 (2018)], p. 221101. DOI: [10.1103/PhysRevLett.116.221101](https://doi.org/10.1103/PhysRevLett.116.221101). arXiv: [1602.03841 \[gr-qc\]](https://arxiv.org/abs/1602.03841).
- [62] Dimitrios Psaltis et al. “Gravitational Test Beyond the First Post-Newtonian Order with the Shadow of the M87 Black Hole”. In: *Phys. Rev. Lett.* 125.14 (2020), p. 141104. DOI: [10.1103/PhysRevLett.125.141104](https://doi.org/10.1103/PhysRevLett.125.141104). arXiv: [2010.01055 \[gr-qc\]](https://arxiv.org/abs/2010.01055).
- [63] Sebastian H. Völkel et al. “EHT tests of the strong-field regime of General Relativity”. 2020. arXiv: [2011.06812 \[gr-qc\]](https://arxiv.org/abs/2011.06812).
- [64] Yoshiaki Sofue and Vera Rubin. “Rotation curves of spiral galaxies”. In: *Ann. Rev. Astron. Astrophys.* 39 (2001), pp. 137–174. DOI: [10.1146/annurev.astro.39.1.137](https://doi.org/10.1146/annurev.astro.39.1.137). arXiv: [astro-ph/0010594](https://arxiv.org/abs/astro-ph/0010594).
- [65] Gianfranco Bertone and Dan Hooper. “History of dark matter”. In: *Rev. Mod. Phys.* 90.4 (2018), p. 045002. DOI: [10.1103/RevModPhys.90.045002](https://doi.org/10.1103/RevModPhys.90.045002). arXiv: [1605.04909 \[astro-ph.CO\]](https://arxiv.org/abs/1605.04909).

- [66] Constantinos Skordis and Tom Zlosnik. “New Relativistic Theory for Modified Newtonian Dynamics”. In: *Phys. Rev. Lett.* 127.16 (2021), p. 161302. DOI: [10.1103/PhysRevLett.127.161302](#). arXiv: [2007.00082 \[astro-ph.CO\]](#).
- [67] Jose A. R. Cembranos. “Dark Matter from R2-gravity”. In: *Phys. Rev. Lett.* 102 (2009), p. 141301. DOI: [10.1103/PhysRevLett.102.141301](#). arXiv: [0809.1653 \[hep-ph\]](#).
- [68] Ali H. Chamseddine and Viatcheslav Mukhanov. “Mimetic Dark Matter”. In: *JHEP* 11 (2013), p. 135. DOI: [10.1007/JHEP11\(2013\)135](#). arXiv: [1308.5410 \[astro-ph.CO\]](#).
- [69] Frederico Arroja et al. “Cosmological perturbations in mimetic Horndeski gravity”. In: *JCAP* 04 (2016), p. 042. DOI: [10.1088/1475-7516/2016/04/042](#). arXiv: [1512.09374 \[gr-qc\]](#).
- [70] L. Sebastiani, S. Vagnozzi, and R. Myrzakulov. “Mimetic gravity: a review of recent developments and applications to cosmology and astrophysics”. In: *Adv. High Energy Phys.* 2017 (2017), p. 3156915. DOI: [10.1155/2017/3156915](#). arXiv: [1612.08661 \[gr-qc\]](#).
- [71] Alessandro Casalino et al. “Mimicking dark matter and dark energy in a mimetic model compatible with GW170817”. In: *Phys. Dark Univ.* 22 (2018), p. 108. DOI: [10.1016/j.dark.2018.10.001](#). arXiv: [1803.02620 \[gr-qc\]](#).
- [72] Dario Bettoni, Mattia Colombo, and Stefano Liberati. “Dark matter as a Bose-Einstein Condensate: the relativistic non-minimally coupled case”. In: *JCAP* 02 (2014), p. 004. DOI: [10.1088/1475-7516/2014/02/004](#). arXiv: [1310.3753 \[astro-ph.CO\]](#).
- [73] S. Mendoza et al. “MOND as the weak-field limit of an extended metric theory of gravity”. In: *AIP Conf. Proc.* 1458.1 (2012). Ed. by Jose Beltran Jimenez et al., pp. 483–486. DOI: [10.1063/1.4734465](#). arXiv: [1202.3629 \[gr-qc\]](#).
- [74] Massimiliano Rinaldi. “Mimicking dark matter in Horndeski gravity”. In: *Phys. Dark Univ.* 16 (2017), pp. 14–21. DOI: [10.1016/j.dark.2017.02.003](#). arXiv: [1608.03839 \[gr-qc\]](#).
- [75] George Koutsoumbas et al. “Unification of Dark Matter - Dark Energy in Generalized Galileon Theories”. In: *JCAP* 02 (2018), p. 003. DOI: [10.1088/1475-7516/2018/02/003](#). arXiv: [1704.08640 \[gr-qc\]](#).
- [76] Alberto Diez-Tejedor, Francisco Flores, and Gustavo Niz. “Horndeski dark matter and beyond”. In: *Phys. Rev. D* 97.12 (2018), p. 123524. DOI: [10.1103/PhysRevD.97.123524](#). arXiv: [1803.00014 \[gr-qc\]](#).

- [77] Adam G. Riess et al. “Observational evidence from supernovae for an accelerating universe and a cosmological constant”. In: *Astron. J.* 116 (1998), pp. 1009–1038. DOI: [10.1086/300499](#). arXiv: [astro-ph/9805201](#).
- [78] S. Perlmutter et al. “Measurements of Ω and Λ from 42 high redshift supernovae”. In: *Astrophys. J.* 517 (1999), pp. 565–586. DOI: [10.1086/307221](#). arXiv: [astro-ph/9812133](#).
- [79] Christos Charmousis et al. “General second order scalar-tensor theory, self tuning, and the Fab Four”. In: *Phys. Rev. Lett.* 108 (2012), p. 051101. DOI: [10.1103/PhysRevLett.108.051101](#). arXiv: [1106.2000 \[hep-th\]](#).
- [80] Christos Charmousis et al. “Self-tuning and the derivation of a class of scalar-tensor theories”. In: *Phys. Rev. D* 85 (2012), p. 104040. DOI: [10.1103/PhysRevD.85.104040](#). arXiv: [1112.4866 \[hep-th\]](#).
- [81] Claudia de Rham et al. “Cosmic Acceleration and the Helicity-0 Graviton”. In: *Phys. Rev. D* 83 (2011), p. 103516. DOI: [10.1103/PhysRevD.83.103516](#). arXiv: [1010.1780 \[hep-th\]](#).
- [82] G. D’Amico et al. “Massive Cosmologies”. In: *Phys. Rev. D* 84 (2011), p. 124046. DOI: [10.1103/PhysRevD.84.124046](#). arXiv: [1108.5231 \[hep-th\]](#).
- [83] Claudia de Rham, Gregory Gabadadze, and Andrew J. Tolley. “Resummation of Massive Gravity”. In: *Phys. Rev. Lett.* 106 (2011), p. 231101. DOI: [10.1103/PhysRevLett.106.231101](#). arXiv: [1011.1232 \[hep-th\]](#).
- [84] Assaf Shomer. “A Pedagogical explanation for the non-renormalizability of gravity”. Sept. 2007. arXiv: [0709.3555 \[hep-th\]](#).
- [85] J. Polchinski. *String theory. Vol. 1: An introduction to the bosonic string*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, 2007. ISBN: 9780511252273, 9780521672276, 9780521633031. DOI: [10.1017/CB09780511816079](#).
- [86] J. Polchinski. *String theory. Vol. 2: Superstring theory and beyond*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, 2007. ISBN: 9780511252280, 9780521633048, 9780521672283. DOI: [10.1017/CB09780511618123](#).
- [87] Y. Fujii and K. Maeda. *The scalar-tensor theory of gravitation*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, 2007. ISBN: 9780521037525, 9780521811590, 9780511029882. DOI: [10.1017/CB09780511535093](#). URL: <http://www.cambridge.org/uk/catalogue/catalogue.asp?isbn=0521811597>.

- [88] Stephon H. Alexander and Nicolas Yunes. “Gravitational wave probes of parity violation in compact binary coalescences”. In: *Phys. Rev. D* 97.6 (2018), p. 064033. DOI: [10.1103/PhysRevD.97.064033](https://doi.org/10.1103/PhysRevD.97.064033). arXiv: [1712.01853](https://arxiv.org/abs/1712.01853) [gr-qc].
- [89] Nathan K. Johnson-McDaniel et al. “Investigating the relation between gravitational wave tests of general relativity”. In: *Phys. Rev. D* 105.4 (2022), p. 044020. DOI: [10.1103/PhysRevD.105.044020](https://doi.org/10.1103/PhysRevD.105.044020). arXiv: [2109.06988](https://arxiv.org/abs/2109.06988) [gr-qc].
- [90] Neil J. Cornish and Tyson B. Littenberg. “BayesWave: Bayesian Inference for Gravitational Wave Bursts and Instrument Glitches”. In: *Class. Quant. Grav.* 32.13 (2015), p. 135012. DOI: [10.1088/0264-9381/32/13/135012](https://doi.org/10.1088/0264-9381/32/13/135012). arXiv: [1410.3835](https://arxiv.org/abs/1410.3835) [gr-qc].
- [91] Tyson B. Littenberg and Neil J. Cornish. “Bayesian inference for spectral estimation of gravitational wave detector noise”. In: *Phys. Rev. D* 91.8 (2015), p. 084034. DOI: [10.1103/PhysRevD.91.084034](https://doi.org/10.1103/PhysRevD.91.084034). arXiv: [1410.3852](https://arxiv.org/abs/1410.3852) [gr-qc].
- [92] Abhirup Ghosh et al. “Testing general relativity using golden black-hole binaries”. In: *Phys. Rev. D* 94.2 (2016), p. 021101. DOI: [10.1103/PhysRevD.94.021101](https://doi.org/10.1103/PhysRevD.94.021101). arXiv: [1602.02453](https://arxiv.org/abs/1602.02453) [gr-qc].
- [93] Abhirup Ghosh et al. “Testing general relativity using gravitational wave signals from the inspiral, merger and ringdown of binary black holes”. In: *Class. Quant. Grav.* 35.1 (2018), p. 014002. DOI: [10.1088/1361-6382/aa972e](https://doi.org/10.1088/1361-6382/aa972e). arXiv: [1704.06784](https://arxiv.org/abs/1704.06784) [gr-qc].
- [94] James Healy, Carlos O. Lousto, and Yosef Zlochower. “Remnant mass, spin, and recoil from spin aligned black-hole binaries”. In: *Phys. Rev. D* 90.10 (2014), p. 104004. DOI: [10.1103/PhysRevD.90.104004](https://doi.org/10.1103/PhysRevD.90.104004). arXiv: [1406.7295](https://arxiv.org/abs/1406.7295) [gr-qc].
- [95] Luc Blanchet. “Gravitational Radiation from Post-Newtonian Sources and Inspiralling Compact Binaries”. In: *Living Rev. Rel.* 17 (2014), p. 2. DOI: [10.12942/lrr-2014-2](https://doi.org/10.12942/lrr-2014-2). arXiv: [1310.1528](https://arxiv.org/abs/1310.1528) [gr-qc].
- [96] Thibault Damour and Alessandro Nagar. “The Effective-One-Body Approach to the General Relativistic Two Body Problem”. In: *Lect. Notes Phys.* 905 (2016), pp. 273–312. DOI: [10.1007/978-3-319-19416-5_7](https://doi.org/10.1007/978-3-319-19416-5_7).
- [97] Andrea Taracchini et al. “Modeling the horizon-absorbed gravitational flux for equatorial-circular orbits in Kerr spacetime”. In: *Phys. Rev. D* 88 (2013). [Erratum: *Phys.Rev.D* 88, 109903 (2013)], p. 044001. DOI: [10.1103/PhysRevD.88.044001](https://doi.org/10.1103/PhysRevD.88.044001). arXiv: [1305.2184](https://arxiv.org/abs/1305.2184) [gr-qc].

- [98] Yi Pan et al. “Inspirational-merger-ringdown waveforms of spinning, precessing black-hole binaries in the effective-one-body formalism”. In: *Phys. Rev. D* 89.8 (2014), p. 084006. DOI: [10.1103/PhysRevD.89.084006](#). arXiv: [1307.6232 \[gr-qc\]](#).
- [99] Stanislav Babak, Andrea Taracchini, and Alessandra Buonanno. “Validating the effective-one-body model of spinning, precessing binary black holes against numerical relativity”. In: *Phys. Rev. D* 95.2 (2017), p. 024010. DOI: [10.1103/PhysRevD.95.024010](#). arXiv: [1607.05661 \[gr-qc\]](#).
- [100] Emanuele Berti, Vitor Cardoso, and Clifford M. Will. “On gravitational-wave spectroscopy of massive black holes with the space interferometer LISA”. In: *Phys. Rev. D* 73 (2006), p. 064030. DOI: [10.1103/PhysRevD.73.064030](#). arXiv: [gr-qc/0512160](#).
- [101] Olaf Dreyer et al. “Black hole spectroscopy: Testing general relativity through gravitational wave observations”. In: *Class. Quant. Grav.* 21 (2004), pp. 787–804. DOI: [10.1088/0264-9381/21/4/003](#). arXiv: [gr-qc/0309007](#).
- [102] Charles W. Misner, K. S. Thorne, and J. A. Wheeler. *Gravitation*. San Francisco: W. H. Freeman, 1973. ISBN: 9780716703440, 9780691177793.
- [103] C. V. Vishveshwara. “Scattering of Gravitational Radiation by a Schwarzschild Black-hole”. In: *Nature* 227 (1970), pp. 936–938.
- [104] S. Chandrasekhar and Steven L. Detweiler. “The quasi-normal modes of the Schwarzschild black hole”. In: *Proc. Roy. Soc. Lond. A* 344 (1975), pp. 441–452. DOI: [10.1098/rspa.1975.0112](#).
- [105] Clifford M. Will. “On the unreasonable effectiveness of the post-Newtonian approximation in gravitational physics”. In: *Proc. Nat. Acad. Sci.* 108 (2011), p. 5938. DOI: [10.1073/pnas.1103127108](#). arXiv: [1102.5192 \[gr-qc\]](#).
- [106] Sebastian Khan et al. “Frequency-domain gravitational waves from nonprecessing black-hole binaries. II. A phenomenological model for the advanced detector era”. In: *Phys. Rev. D* 93.4 (2016), p. 044007. DOI: [10.1103/PhysRevD.93.044007](#). arXiv: [1508.07253 \[gr-qc\]](#).
- [107] Nicolas Yunes and Frans Pretorius. “Fundamental Theoretical Bias in Gravitational Wave Astrophysics and the Parameterized Post-Einsteinian Framework”. In: *Phys. Rev. D* 80 (2009), p. 122003. DOI: [10.1103/PhysRevD.80.122003](#). arXiv: [0909.3328 \[gr-qc\]](#).
- [108] Banafsheh Shiralilou et al. “Nonlinear curvature effects in gravitational waves from inspiralling black hole binaries”. Dec. 2020. arXiv: [2012.09162 \[gr-qc\]](#).

- [109] Sharaban Tahura and Kent Yagi. “Parameterized Post-Einsteinian Gravitational Waveforms in Various Modified Theories of Gravity”. In: *Phys. Rev. D* 98.8 (2018). [Erratum: Phys.Rev.D 101, 109902 (2020)], p. 084042. DOI: [10.1103/PhysRevD.98.084042](https://doi.org/10.1103/PhysRevD.98.084042). arXiv: [1809.00259](https://arxiv.org/abs/1809.00259) [gr-qc].
- [110] Kent Yagi et al. “Post-Newtonian, Quasi-Circular Binary Inspirals in Quadratic Modified Gravity”. In: *Phys. Rev. D* 85 (2012). [Erratum: Phys.Rev.D 93, 029902 (2016)], p. 064022. DOI: [10.1103/PhysRevD.85.064022](https://doi.org/10.1103/PhysRevD.85.064022). arXiv: [1110.5950](https://arxiv.org/abs/1110.5950) [gr-qc].
- [111] Scott E. Perkins, Nicolás Yunes, and Emanuele Berti. “Probing Fundamental Physics with Gravitational Waves: The Next Generation”. In: *Phys. Rev. D* 103.4 (2021), p. 044024. DOI: [10.1103/PhysRevD.103.044024](https://doi.org/10.1103/PhysRevD.103.044024). arXiv: [2010.09010](https://arxiv.org/abs/2010.09010) [gr-qc].
- [112] Nicolas Yunes, Kent Yagi, and Frans Pretorius. “Theoretical Physics Implications of the Binary Black-Hole Mergers GW150914 and GW151226”. In: *Phys. Rev. D* 94.8 (2016), p. 084002. DOI: [10.1103/PhysRevD.94.084002](https://doi.org/10.1103/PhysRevD.94.084002). arXiv: [1603.08955](https://arxiv.org/abs/1603.08955) [gr-qc].
- [113] Stephon Alexander and Nicolas Yunes. “Parametrized post-Newtonian expansion of Chern-Simons gravity”. In: *Phys. Rev. D* 75 (2007), p. 124022. DOI: [10.1103/PhysRevD.75.124022](https://doi.org/10.1103/PhysRevD.75.124022). arXiv: [0704.0299](https://arxiv.org/abs/0704.0299) [hep-th].
- [114] Kent Yagi et al. “Isolated and Binary Neutron Stars in Dynamical Chern-Simons Gravity”. In: *Phys. Rev. D* 87 (2013). [Erratum: Phys. Rev.D93,no.8,089909(2016)], p. 084058. DOI: [10.1103/PhysRevD.87.084058](https://doi.org/10.1103/PhysRevD.87.084058), [10.1103/PhysRevD.93.089909](https://doi.org/10.1103/PhysRevD.93.089909). arXiv: [1302.1918](https://arxiv.org/abs/1302.1918) [gr-qc].
- [115] Neil Cornish et al. “Gravitational Wave Tests of General Relativity with the Parameterized Post-Einsteinian Framework”. In: *Phys. Rev. D* 84 (2011), p. 062003. DOI: [10.1103/PhysRevD.84.062003](https://doi.org/10.1103/PhysRevD.84.062003). arXiv: [1105.2088](https://arxiv.org/abs/1105.2088) [gr-qc].
- [116] Laura Sampson, Neil Cornish, and Nicolas Yunes. “Gravitational Wave Tests of Strong Field General Relativity with Binary Inspirals: Realistic Injections and Optimal Model Selection”. In: *Phys. Rev. D* 87.10 (2013), p. 102001. DOI: [10.1103/PhysRevD.87.102001](https://doi.org/10.1103/PhysRevD.87.102001). arXiv: [1303.1185](https://arxiv.org/abs/1303.1185) [gr-qc].
- [117] Katerina Chatziioannou, Nicolas Yunes, and Neil Cornish. “Model-Independent Test of General Relativity: An Extended post-Einsteinian Framework with Complete Polarization Content”. In: *Phys. Rev. D* 86 (2012). [Erratum: Phys.Rev.D 95, 129901 (2017)], p. 022004. DOI: [10.1103/PhysRevD.86.022004](https://doi.org/10.1103/PhysRevD.86.022004). arXiv: [1204.2585](https://arxiv.org/abs/1204.2585) [gr-qc].
- [118] Chandra Kant Mishra et al. “Parametrized tests of post-Newtonian theory using Advanced LIGO and Einstein Telescope”. In: *Phys. Rev. D* 82 (2010), p. 064010. DOI: [10.1103/PhysRevD.82.064010](https://doi.org/10.1103/PhysRevD.82.064010). arXiv: [1005.0304](https://arxiv.org/abs/1005.0304) [gr-qc].

- [119] K.G. Arun et al. “Testing post-Newtonian theory with gravitational wave observations”. In: *Class. Quant. Grav.* 23 (2006), pp. L37–L43. DOI: [10.1088/0264-9381/23/9/L01](#). arXiv: [gr-qc/0604018](#).
- [120] T. G. F. Li et al. “Towards a generic test of the strong field dynamics of general relativity using compact binary coalescence”. In: *Phys. Rev. D* 85 (2012), p. 082003. DOI: [10.1103/PhysRevD.85.082003](#). arXiv: [1110.0530 \[gr-qc\]](#).
- [121] Michalis Agathos et al. “TIGER: A data analysis pipeline for testing the strong-field dynamics of general relativity with gravitational wave signals from coalescing compact binaries”. In: *Phys. Rev. D* 89.8 (2014), p. 082001. DOI: [10.1103/PhysRevD.89.082001](#). arXiv: [1311.0420 \[gr-qc\]](#).
- [122] Jeroen Meidam et al. “Parametrized tests of the strong-field dynamics of general relativity using gravitational wave signals from coalescing binary black holes: Fast likelihood calculations and sensitivity of the method”. In: *Phys. Rev. D* 97.4 (2018), p. 044033. DOI: [10.1103/PhysRevD.97.044033](#). arXiv: [1712.08772 \[gr-qc\]](#).
- [123] Scott E. Perkins et al. “Improved gravitational-wave constraints on higher-order curvature theories of gravity”. Apr. 2021. arXiv: [2104.11189 \[gr-qc\]](#).
- [124] Remya Nair et al. “Fundamental Physics Implications for Higher-Curvature Theories from Binary Black Hole Signals in the LIGO-Virgo Catalog GWTC-1”. In: *Phys. Rev. Lett.* 123.19 (2019), p. 191101. DOI: [10.1103/PhysRevLett.123.191101](#). arXiv: [1905.00870 \[gr-qc\]](#).
- [125] Lee Samuel Finn and David F. Chernoff. “Observing binary inspiral in gravitational radiation: One interferometer”. In: *Phys. Rev. D* 47 (1993), pp. 2198–2219. DOI: [10.1103/PhysRevD.47.2198](#). arXiv: [gr-qc/9301003 \[gr-qc\]](#).
- [126] Ali H. Chamseddine and Mikhail S. Volkov. “Cosmological solutions with massive gravitons”. In: *Phys. Lett. B* 704 (2011), pp. 652–654. DOI: [10.1016/j.physletb.2011.09.085](#). arXiv: [1107.5504 \[hep-th\]](#).
- [127] Kurt Hinterbichler. “Cosmology of Massive Gravity and its Extensions”. In: *Proceedings, 51st Rencontres de Moriond, Cosmology session: La Thuile, Italy, March 19-26, 2016*. [223(2017)]. 2016, pp. 223–232. arXiv: [1701.02873 \[astro-ph.CO\]](#).
- [128] Philippe Brax et al. “A Unified Description of Screened Modified Gravity”. In: *Phys. Rev. D* 86 (2012), p. 044015. DOI: [10.1103/PhysRevD.86.044015](#). arXiv: [1203.4812 \[astro-ph.CO\]](#).
- [129] Moritz Plattscher et al. “Long Range Effects in Gravity Theories with Vainshtein Screening”. 2018. arXiv: [1809.05318 \[astro-ph.CO\]](#).

- [130] Timothy Clifton et al. “Modified gravity and cosmology”. In: *Physics Reports* 513.1 (2012). Modified Gravity and Cosmology, pp. 1–189. ISSN: 0370-1573. DOI: <https://doi.org/10.1016/j.physrep.2012.01.001>. URL: <http://www.sciencedirect.com/science/article/pii/S0370157312000105>.
- [131] Clifford M. Will. “Bounding the mass of the graviton using gravitational wave observations of inspiralling compact binaries”. In: *Phys. Rev. D* 57 (1998), pp. 2061–2068. DOI: [10.1103/PhysRevD.57.2061](https://doi.org/10.1103/PhysRevD.57.2061). arXiv: [gr-qc/9709011](https://arxiv.org/abs/gr-qc/9709011) [gr-qc].
- [132] Saeed Mirshekari, Nicolas Yunes, and Clifford M. Will. “Constraining Generic Lorentz Violation and the Speed of the Graviton with Gravitational Waves”. In: *Phys. Rev. D* 85 (2012), p. 024041. DOI: [10.1103/PhysRevD.85.024041](https://doi.org/10.1103/PhysRevD.85.024041). arXiv: [1110.2720](https://arxiv.org/abs/1110.2720) [gr-qc].
- [133] A.I. Vainshtein. “To the problem of nonvanishing gravitation mass”. In: *Physics Letters B* 39.3 (1972), pp. 393–394. ISSN: 0370-2693. DOI: [https://doi.org/10.1016/0370-2693\(72\)90147-5](https://doi.org/10.1016/0370-2693(72)90147-5). URL: <http://www.sciencedirect.com/science/article/pii/0370269372901475>.
- [134] Furqan Dar et al. “Scalar Gravitational Radiation from Binaries: Vainshtein Mechanism in Time-dependent Systems”. 2018. arXiv: [1808.02165](https://arxiv.org/abs/1808.02165) [hep-th].
- [135] K. G. Arun et al. “Probing the non-linear structure of general relativity with black hole binaries”. In: *Phys. Rev. D* 74 (2006), p. 024006. DOI: [10.1103/PhysRevD.74.024006](https://doi.org/10.1103/PhysRevD.74.024006). arXiv: [gr-qc/0604067](https://arxiv.org/abs/gr-qc/0604067).
- [136] Anuradha Gupta et al. “Multiparameter tests of general relativity using multiband gravitational-wave observations”. In: *Phys. Rev. Lett.* 125.20 (2020), p. 201101. DOI: [10.1103/PhysRevLett.125.201101](https://doi.org/10.1103/PhysRevLett.125.201101). arXiv: [2005.09607](https://arxiv.org/abs/2005.09607) [gr-qc].
- [137] Stephon Alexander and Nicolas Yunes. “Chern-Simons Modified General Relativity”. In: *Phys. Rept.* 480 (2009), pp. 1–55. DOI: [10.1016/j.physrep.2009.07.002](https://doi.org/10.1016/j.physrep.2009.07.002). arXiv: [0907.2562](https://arxiv.org/abs/0907.2562) [hep-th].
- [138] P. Kanti et al. “Dilatonic black holes in higher curvature string gravity”. In: *Phys. Rev. D* 54 (1996), pp. 5049–5058. DOI: [10.1103/PhysRevD.54.5049](https://doi.org/10.1103/PhysRevD.54.5049). arXiv: [hep-th/9511071](https://arxiv.org/abs/hep-th/9511071) [hep-th].
- [139] Benjamin P. Abbott et al. “GW170104: Observation of a 50-Solar-Mass Binary Black Hole Coalescence at Redshift 0.2”. In: *Phys. Rev. Lett.* 118.22 (2017). [Erratum: *Phys. Rev. Lett.* 121, no. 12, 129901 (2018)], p. 221101. DOI: [10.1103/PhysRevLett.118.221101](https://doi.org/10.1103/PhysRevLett.118.221101), [10.1103/PhysRevLett.121.129901](https://doi.org/10.1103/PhysRevLett.121.129901). arXiv: [1706.01812](https://arxiv.org/abs/1706.01812) [gr-qc].
- [140] Claudia de Rham, Lavinia Heisenberg, and Raquel H. Ribeiro. “Quantum Corrections in Massive Gravity”. In: *Phys. Rev. D* 88 (2013), p. 084058. DOI: [10.1103/PhysRevD.88.084058](https://doi.org/10.1103/PhysRevD.88.084058). arXiv: [1307.7169](https://arxiv.org/abs/1307.7169) [hep-th].

- [141] Kurt Hinterbichler, Mark Trodden, and Daniel Wesley. “Multi-field galileons and higher co-dimension branes”. In: *Phys. Rev. D* 82 (2010), p. 124018. DOI: [10.1103/PhysRevD.82.124018](https://doi.org/10.1103/PhysRevD.82.124018). arXiv: [1008.1305 \[hep-th\]](https://arxiv.org/abs/1008.1305).
- [142] Minjoon Park. “Quantum Aspects of Massive Gravity”. In: *Class. Quant. Grav.* 28 (2011), p. 105012. DOI: [10.1088/0264-9381/28/10/105012](https://doi.org/10.1088/0264-9381/28/10/105012). arXiv: [1009.4369 \[hep-th\]](https://arxiv.org/abs/1009.4369).
- [143] Minjoon Park. “Quantum Aspects of Massive Gravity II: Non-Pauli-Fierz Theory”. In: *JHEP* 10 (2011), p. 130. DOI: [10.1007/JHEP10\(2011\)130](https://doi.org/10.1007/JHEP10(2011)130). arXiv: [1011.4266 \[hep-th\]](https://arxiv.org/abs/1011.4266).
- [144] Lee Samuel Finn and Patrick J. Sutton. “Bounding the mass of the graviton using binary pulsar observations”. In: *Phys. Rev. D* 65 (2002), p. 044022. DOI: [10.1103/PhysRevD.65.044022](https://doi.org/10.1103/PhysRevD.65.044022). arXiv: [gr-qc/0109049 \[gr-qc\]](https://arxiv.org/abs/gr-qc/0109049).
- [145] Richard Brito, Vitor Cardoso, and Paolo Pani. “Massive spin-2 fields on black hole spacetimes: Instability of the Schwarzschild and Kerr solutions and bounds on the graviton mass”. In: *Phys. Rev. D* 88.2 (2013), p. 023514. DOI: [10.1103/PhysRevD.88.023514](https://doi.org/10.1103/PhysRevD.88.023514). arXiv: [1304.6725 \[gr-qc\]](https://arxiv.org/abs/1304.6725).
- [146] Alfred S. Goldhaber and Michael Martin Nieto. “Mass of the graviton”. In: *Phys. Rev. D* 9 (4 Feb. 1974), pp. 1119–1121. DOI: [10.1103/PhysRevD.9.1119](https://doi.org/10.1103/PhysRevD.9.1119). URL: <https://link.aps.org/doi/10.1103/PhysRevD.9.1119>.
- [147] C. Talmadge et al. “Model-Independent Constraints on Possible Modifications of Newtonian Gravity”. In: *Phys. Rev. Lett.* 61 (10 Sept. 1988), pp. 1159–1162. DOI: [10.1103/PhysRevLett.61.1159](https://doi.org/10.1103/PhysRevLett.61.1159). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.61.1159>.
- [148] Jeremy Sakstein. “Tests of Gravity with Future Space-Based Experiments”. In: *Phys. Rev. D* 97.6 (2018), p. 064028. DOI: [10.1103/PhysRevD.97.064028](https://doi.org/10.1103/PhysRevD.97.064028). arXiv: [1710.03156 \[astro-ph.CO\]](https://arxiv.org/abs/1710.03156).
- [149] Shantanu Desai. “Limit on graviton mass from galaxy cluster Abell 1689”. In: *Phys. Lett. B* 778 (2018), pp. 325–331. DOI: [10.1016/j.physletb.2018.01.052](https://doi.org/10.1016/j.physletb.2018.01.052). arXiv: [1708.06502 \[astro-ph.CO\]](https://arxiv.org/abs/1708.06502).
- [150] Akshay Rana et al. “Bounds on graviton mass using weak lensing and SZ effect in galaxy clusters”. In: *Phys. Lett. B* 781 (2018), pp. 220–226. DOI: [10.1016/j.physletb.2018.03.076](https://doi.org/10.1016/j.physletb.2018.03.076). arXiv: [1801.03309 \[astro-ph.CO\]](https://arxiv.org/abs/1801.03309).
- [151] Sajal Gupta and Shantanu Desai. “Limit on graviton mass using stacked galaxy cluster catalogs from SPT-SZ, Planck-SZ and SDSS-redMaPPer”. In: *Annals Phys.* 399 (2018), pp. 85–92. DOI: [10.1016/j.aop.2018.09.017](https://doi.org/10.1016/j.aop.2018.09.017). arXiv: [1810.00198 \[astro-ph.CO\]](https://arxiv.org/abs/1810.00198).
- [152] Clifford M. Will. “Solar system versus gravitational-wave bounds on the graviton mass”. In: *Class. Quant. Grav.* 35.17 (2018), 17LT01. DOI: [10.1088/1361-6382/aad13c](https://doi.org/10.1088/1361-6382/aad13c). arXiv: [1805.10523 \[gr-qc\]](https://arxiv.org/abs/1805.10523).

- [153] Claudia de Rham et al. “Graviton Mass Bounds”. In: *Rev. Mod. Phys.* 89.2 (2017), p. 025004. DOI: [10.1103/RevModPhys.89.025004](#). arXiv: [1606.08462 \[astro-ph.CO\]](#).
- [154] Nicolás Yunes and Xavier Siemens. “Gravitational-Wave Tests of General Relativity with Ground-Based Detectors and Pulsar Timing-Arrays”. In: *Living Rev. Rel.* 16 (2013), p. 9. DOI: [10.12942/lrr-2013-9](#). arXiv: [1304.3473 \[gr-qc\]](#).
- [155] Eugeny Babichev and Cédric Deffayet. “An introduction to the Vainshtein mechanism”. In: *Class. Quant. Grav.* 30 (2013), p. 184001. DOI: [10.1088/0264-9381/30/18/184001](#). arXiv: [1304.7240 \[gr-qc\]](#).
- [156] Eugeny Babichev and Marco Crisostomi. “Restoring general relativity in massive bigravity theory”. In: *Phys. Rev. D* 88.8 (2013), p. 084002. DOI: [10.1103/PhysRevD.88.084002](#). arXiv: [1307.3640 \[gr-qc\]](#).
- [157] Giga Chkareuli and David Pirtskhalava. “Vainshtein Mechanism In Λ_3 - Theories”. In: *Phys. Lett. B* 713 (2012), pp. 99–103. DOI: [10.1016/j.physletb.2012.05.030](#). arXiv: [1105.1783 \[hep-th\]](#).
- [158] Anuradha Samajdar and K. G. Arun. “Projected constraints on the dispersion of gravitational waves using advanced ground- and space-based interferometers”. In: *Phys. Rev. D* 96.10 (2017), p. 104027. DOI: [10.1103/PhysRevD.96.104027](#). arXiv: [1708.00671 \[gr-qc\]](#).
- [159] Clifford M. Will and Nicolas Yunes. “Testing alternative theories of gravity using LISA”. In: *Class. Quant. Grav.* 21 (2004), p. 4367. DOI: [10.1088/0264-9381/21/18/006](#). arXiv: [gr-qc/0403100 \[gr-qc\]](#).
- [160] Emanuele Berti, Alessandra Buonanno, and Clifford M. Will. “Testing general relativity and probing the merger history of massive black holes with LISA”. In: *Class. Quant. Grav.* 22 (2005), S943–S954. DOI: [10.1088/0264-9381/22/18/S08](#). arXiv: [gr-qc/0504017 \[gr-qc\]](#).
- [161] Katie Chamberlain and Nicolas Yunes. “Theoretical Physics Implications of Gravitational Wave Observation with Future Detectors”. In: *Phys. Rev. D* 96.8 (2017), p. 084039. DOI: [10.1103/PhysRevD.96.084039](#). arXiv: [1704.08268 \[gr-qc\]](#).
- [162] Alberto Nicolis, Riccardo Rattazzi, and Enrico Trincherini. “The Galileon as a local modification of gravity”. In: *Phys. Rev. D* 79 (2009), p. 064036. DOI: [10.1103/PhysRevD.79.064036](#). arXiv: [0811.2197 \[hep-th\]](#).
- [163] Jose Beltran Jimenez, Federico Piazza, and Hermano Velten. “Evading the Vainshtein Mechanism with Anomalous Gravitational Wave Speed: Constraints on Modified Gravity from Binary Pulsars”. In: *Phys. Rev. Lett.* 116.6 (2016), p. 061101. DOI: [10.1103/PhysRevLett.116.061101](#). arXiv: [1507.05047 \[gr-qc\]](#).

- [164] S. F. Hassan and Rachel A. Rosen. “Bimetric Gravity from Ghost-free Massive Gravity”. In: *JHEP* 02 (2012), p. 126. DOI: [10.1007/JHEP02\(2012\)126](https://doi.org/10.1007/JHEP02(2012)126). arXiv: [1109.3515](https://arxiv.org/abs/1109.3515) [hep-th].
- [165] Claudia de Rham. “Massive Gravity”. In: *Living Rev. Rel.* 17 (2014), p. 7. DOI: [10.12942/lrr-2014-7](https://doi.org/10.12942/lrr-2014-7). arXiv: [1401.4173](https://arxiv.org/abs/1401.4173) [hep-th].
- [166] L. S. Collaborations. *Ligo-t1400316-v4: Instrument science white paper*.
- [167] M Punturo et al. “The Einstein Telescope: a third-generation gravitational wave observatory”. In: *Classical and Quantum Gravity* 27.19 (2010), p. 194002. URL: <http://stacks.iop.org/0264-9381/27/i=19/a=194002>.
- [168] Pau Amaro-Seoane et al. “Laser Interferometer Space Antenna”. In: *arXiv e-prints*, arXiv:1702.00786 (Feb. 2017), arXiv:1702.00786. arXiv: [1702.00786](https://arxiv.org/abs/1702.00786) [astro-ph.IM].
- [169] Sascha Husa et al. “Frequency-domain gravitational waves from nonprecessing black-hole binaries. I. New numerical waveforms and anatomy of the signal”. In: *Phys. Rev. D* 93.4 (2016), p. 044006. DOI: [10.1103/PhysRevD.93.044006](https://doi.org/10.1103/PhysRevD.93.044006). arXiv: [1508.07250](https://arxiv.org/abs/1508.07250) [gr-qc].
- [170] Kurt Hinterbichler. “Theoretical Aspects of Massive Gravity”. In: *Rev. Mod. Phys.* 84 (2012), pp. 671–710. DOI: [10.1103/RevModPhys.84.671](https://doi.org/10.1103/RevModPhys.84.671). arXiv: [1105.3735](https://arxiv.org/abs/1105.3735) [hep-th].
- [171] Angnis Schmidt-May and Mikael von Strauss. “Recent developments in bimetric theory”. In: *J. Phys.* A49.18 (2016), p. 183001. DOI: [10.1088/1751-8113/49/18/183001](https://doi.org/10.1088/1751-8113/49/18/183001). arXiv: [1512.00021](https://arxiv.org/abs/1512.00021) [hep-th].
- [172] Alexander V. Belikov and Wayne Hu. “Equivalence Principle Violation in Weakly Vainshtein-Screened Systems”. In: *Phys. Rev. D* 87.8 (2013), p. 084042. DOI: [10.1103/PhysRevD.87.084042](https://doi.org/10.1103/PhysRevD.87.084042). arXiv: [1212.0831](https://arxiv.org/abs/1212.0831) [gr-qc].
- [173] Eugeny Babichev and Gilles Esposito-Farèse. “Time-Dependent Spherically Symmetric Covariant Galileons”. In: *Phys. Rev. D* 87 (2013), p. 044032. DOI: [10.1103/PhysRevD.87.044032](https://doi.org/10.1103/PhysRevD.87.044032). arXiv: [1212.1394](https://arxiv.org/abs/1212.1394) [gr-qc].
- [174] Kazuya Koyama, Gustavo Niz, and Gianmassimo Tasinato. “Strong interactions and exact solutions in non-linear massive gravity”. In: *Phys. Rev. D* 84 (2011), p. 064033. DOI: [10.1103/PhysRevD.84.064033](https://doi.org/10.1103/PhysRevD.84.064033). arXiv: [1104.2143](https://arxiv.org/abs/1104.2143) [hep-th].
- [175] Radouane Gannouji and M. Sami. “Vainshtein mechanism in Gauss-Bonnet gravity and Galileon aether”. In: *Phys. Rev. D* 85 (2012), p. 024019. DOI: [10.1103/PhysRevD.85.024019](https://doi.org/10.1103/PhysRevD.85.024019). arXiv: [1107.1892](https://arxiv.org/abs/1107.1892) [gr-qc].

- [176] Ryotaro Kase and Shinji Tsujikawa. “Screening the fifth force in the Horndeski’s most general scalar-tensor theories”. In: *JCAP* 1308 (2013), p. 054. DOI: [10.1088/1475-7516/2013/08/054](https://doi.org/10.1088/1475-7516/2013/08/054). arXiv: [1306.6401](https://arxiv.org/abs/1306.6401) [gr-qc].
- [177] V. I. Zakharov. “Linearized gravitation theory and the graviton mass”. In: *JETP Lett.* 12 (1970). [Pisma Zh. Eksp. Teor. Fiz.12,447(1970)], p. 312.
- [178] M. Porrati. “Fully covariant van Dam-Veltman-Zakharov discontinuity, and absence thereof”. In: *Phys. Lett.* B534 (2002), pp. 209–215. DOI: [10.1016/S0370-2693\(02\)01656-8](https://doi.org/10.1016/S0370-2693(02)01656-8). arXiv: [hep-th/0203014](https://arxiv.org/abs/hep-th/0203014) [hep-th].
- [179] David G. Boulware and S. Deser. “Can Gravitation Have a Finite Range?” In: *Phys. Rev. D* 6 (12 Dec. 1972), pp. 3368–3382. DOI: [10.1103/PhysRevD.6.3368](https://doi.org/10.1103/PhysRevD.6.3368). URL: <https://link.aps.org/doi/10.1103/PhysRevD.6.3368>.
- [180] P. A. R. Ade et al. “Planck 2015 results. XIII. Cosmological parameters”. In: *Astron. Astrophys.* 594 (2016), A13. DOI: [10.1051/0004-6361/201525830](https://doi.org/10.1051/0004-6361/201525830). arXiv: [1502.01589](https://arxiv.org/abs/1502.01589) [astro-ph.CO].
- [181] Benjamin P. Abbott et al. “Tests of General Relativity with GW150914”. In: *Phys. Rev. Lett.* 116 (22 May 2016), p. 221101. DOI: [10.1103/PhysRevLett.116.221101](https://doi.org/10.1103/PhysRevLett.116.221101). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.116.221101>.
- [182] John Ellis, Nick E. Mavromatos, and Dimitri V. Nanopoulos. “Comments on Graviton Propagation in Light of GW150914”. In: *Mod. Phys. Lett.* A31.26 (2016), p. 1675001. DOI: [10.1142/S0217732316750018](https://doi.org/10.1142/S0217732316750018). arXiv: [1602.04764](https://arxiv.org/abs/1602.04764) [gr-qc].
- [183] Pedro Bicudo. “Tighter bounds on a hypothetical graviton screening mass from the gravitational wave observation GW150914 at LIGO”. 2016. arXiv: [1602.04337](https://arxiv.org/abs/1602.04337) [gr-qc].
- [184] Michele Vallisneri. “Use and abuse of the Fisher information matrix in the assessment of gravitational-wave parameter-estimation prospects”. In: *Phys. Rev. D* 77 (4 Feb. 2008), p. 042001. DOI: [10.1103/PhysRevD.77.042001](https://doi.org/10.1103/PhysRevD.77.042001). URL: <https://link.aps.org/doi/10.1103/PhysRevD.77.042001>.
- [185] Michele Vallisneri. “Beyond the Fisher-Matrix Formalism: Exact Sampling Distributions of the Maximum-Likelihood Estimator in Gravitational-Wave Parameter Estimation”. In: *Phys. Rev. Lett.* 107 (19 Nov. 2011), p. 191104. DOI: [10.1103/PhysRevLett.107.191104](https://doi.org/10.1103/PhysRevLett.107.191104). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.107.191104>.
- [186] M. Armano et al. “Sub-Femto-*g* Free Fall for Space-Based Gravitational Wave Observatories: LISA Pathfinder Results”. In: *Phys. Rev. Lett.* 116 (23 June 2016), p. 231101. DOI: [10.1103/PhysRevLett.116.231101](https://doi.org/10.1103/PhysRevLett.116.231101). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.116.231101>.

- [187] Claudia de Rham, Andrew J. Tolley, and Daniel H. Wesley. “Vainshtein mechanism in binary pulsars”. In: *Phys. Rev. D* 87 (4 Feb. 2013), p. 044025. DOI: [10.1103/PhysRevD.87.044025](https://doi.org/10.1103/PhysRevD.87.044025). URL: <https://link.aps.org/doi/10.1103/PhysRevD.87.044025>.
- [188] B. P. Abbott et al. “GW170608: Observation of a 19 Solar-mass Binary Black Hole Coalescence”. In: *The Astrophysical Journal Letters* 851.2 (2017), p. L35. URL: <http://stacks.iop.org/2041-8205/851/i=2/a=L35>.
- [189] B. P. Abbott et al. “Properties of the binary neutron star merger GW170817”. 2018. arXiv: [1805.11579](https://arxiv.org/abs/1805.11579) [gr-qc].
- [190] B. P. Abbott et al. “GW170814: A Three-Detector Observation of Gravitational Waves from a Binary Black Hole Coalescence”. In: *Phys. Rev. Lett.* 119 (14 Oct. 2017), p. 141101. DOI: [10.1103/PhysRevLett.119.141101](https://doi.org/10.1103/PhysRevLett.119.141101). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.119.141101>.
- [191] B. P. Abbott et al. “GW151226: Observation of Gravitational Waves from a 22-Solar-Mass Binary Black Hole Coalescence”. In: *Phys. Rev. Lett.* 116 (24 June 2016), p. 241103. DOI: [10.1103/PhysRevLett.116.241103](https://doi.org/10.1103/PhysRevLett.116.241103). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.116.241103>.
- [192] B. P. Abbott et al. “Properties of the Binary Black Hole Merger GW150914”. In: *Phys. Rev. Lett.* 116 (24 June 2016), p. 241102. DOI: [10.1103/PhysRevLett.116.241102](https://doi.org/10.1103/PhysRevLett.116.241102). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.116.241102>.
- [193] Jeff Crowder and Neil J. Cornish. “Beyond LISA: Exploring future gravitational wave missions”. In: *Phys. Rev. D* 72 (8 Oct. 2005), p. 083005. DOI: [10.1103/PhysRevD.72.083005](https://doi.org/10.1103/PhysRevD.72.083005). URL: <https://link.aps.org/doi/10.1103/PhysRevD.72.083005>.
- [194] Kazunori Akiyama et al. “First M87 Event Horizon Telescope Results. I. The Shadow of the Super-massive Black Hole”. In: *Astrophys. J. Lett.* 875 (2019), p. L1. DOI: [10.3847/2041-8213/ab0ec7](https://doi.org/10.3847/2041-8213/ab0ec7). arXiv: [1906.11238](https://arxiv.org/abs/1906.11238) [astro-ph.GA].
- [195] B. P. Abbott et al. “Tests of general relativity with GW150914”. In: *Phys. Rev. Lett.* 116.22 (2016). [Erratum: *Phys. Rev. Lett.* 121, no. 12, 129902 (2018)], p. 221101. DOI: [10.1103/PhysRevLett.116.221101](https://doi.org/10.1103/PhysRevLett.116.221101), [10.1103/PhysRevLett.121.129902](https://doi.org/10.1103/PhysRevLett.121.129902). arXiv: [1602.03841](https://arxiv.org/abs/1602.03841) [gr-qc].
- [196] Adam G. Riess et al. “Observational evidence from supernovae for an accelerating universe and a cosmological constant”. In: *Astron. J.* 116 (1998), pp. 1009–1038. DOI: [10.1086/300499](https://doi.org/10.1086/300499). arXiv: [astro-ph/9805201](https://arxiv.org/abs/astro-ph/9805201) [astro-ph].
- [197] S. Perlmutter et al. “Measurements of Omega and Lambda from 42 high redshift supernovae”. In: *Astrophys. J.* 517 (1999), pp. 565–586. DOI: [10.1086/307221](https://doi.org/10.1086/307221). arXiv: [astro-ph/9812133](https://arxiv.org/abs/astro-ph/9812133) [astro-ph].

- [198] J. Aasi et al. “Advanced LIGO”. In: *Class. Quant. Grav.* 32 (2015), p. 074001. DOI: [10.1088/0264-9381/32/7/074001](#). arXiv: [1411.4547 \[gr-qc\]](#).
- [199] Bala Iyer et al. *LIGO-India, Proposal of the Consortium for Indian Initiative in Gravitational-wave Observations (IndIGO)*. Tech. rep. July 2020. URL: <https://dcc.ligo.org/LIGO-M1100296/public>.
- [200] Sheila Dwyer et al. “Gravitational wave detector with cosmological reach”. In: *Phys. Rev. D* 91.8, 082001 (Apr. 2015), p. 082001. DOI: [10.1103/PhysRevD.91.082001](#). arXiv: [1410.0612 \[astro-ph.IM\]](#).
- [201] M. Punturo et al. “The third generation of gravitational wave observatories and their science reach”. In: *Class. Quant. Grav.* 27 (2010). Ed. by Zsuzsa Marka and Szabolcs Marka, p. 084007. DOI: [10.1088/0264-9381/27/8/084007](#).
- [202] Mark Hannam et al. “Simple Model of Complete Precessing Black-Hole-Binary Gravitational Waveforms”. In: *Phys. Rev. Lett.* 113.15 (2014), p. 151101. DOI: [10.1103/PhysRevLett.113.151101](#). arXiv: [1308.3271 \[gr-qc\]](#).
- [203] Shammi Tahura, Kent Yagi, and Zack Carson. “Testing Gravity with Gravitational Waves from Binary Black Hole Mergers: Contributions from Amplitude Corrections”. In: *Phys. Rev. D* 100.10 (2019), p. 104001. DOI: [10.1103/PhysRevD.100.104001](#). arXiv: [1907.10059 \[gr-qc\]](#).
- [204] Davide Gerosa et al. “Multiband gravitational-wave event rates and stellar physics”. In: *Phys. Rev. D* 99.10 (2019), p. 103004. DOI: [10.1103/PhysRevD.99.103004](#). arXiv: [1902.00021 \[astro-ph.HE\]](#).
- [205] Christopher J. Moore, Davide Gerosa, and Antoine Klein. “Are stellar-mass black-hole binaries too quiet for LISA?” In: *Mon. Not. Roy. Astron. Soc.* 488.1 (2019), pp. L94–L98. DOI: [10.1093/mnrasl/slz104](#). arXiv: [1905.11998 \[astro-ph.HE\]](#).
- [206] Enrico Barausse, Nicolás Yunes, and Katie Chamberlain. “Theory-Agnostic Constraints on Black-Hole Dipole Radiation with Multiband Gravitational-Wave Astrophysics”. In: *Phys. Rev. Lett.* 116.24 (2016), p. 241104. DOI: [10.1103/PhysRevLett.116.241104](#). arXiv: [1603.04075 \[gr-qc\]](#).
- [207] Curt Cutler et al. “What we can learn from multi-band observations of black hole binaries”. 2019. arXiv: [1903.04069 \[astro-ph.HE\]](#).
- [208] Giuseppe Gnocchi et al. “Bounding alternative theories of gravity with multiband GW observations”. In: *Phys. Rev. D* 100.6 (2019), p. 064024. DOI: [10.1103/PhysRevD.100.064024](#). arXiv: [1905.13460 \[gr-qc\]](#).
- [209] Zack Carson and Kent Yagi. “Multi-band gravitational wave tests of general relativity”. In: *Class. Quant. Grav.* 37.2 (2020), 02LT01. DOI: [10.1088/1361-6382/ab5c9a](#). arXiv: [1905.13155 \[gr-qc\]](#).

- [210] Zack Carson and Kent Yagi. “Parametrized and inspiral-merger-ringdown consistency tests of gravity with multiband gravitational wave observations”. In: *Phys. Rev. D* 101.4 (2020), p. 044047. DOI: [10.1103/PhysRevD.101.044047](https://doi.org/10.1103/PhysRevD.101.044047). arXiv: [1911.05258](https://arxiv.org/abs/1911.05258) [gr-qc].
- [211] Alexandre Toubiana et al. “Tests of general relativity with stellar-mass black hole binaries observed by LISA”. In: *Phys. Rev. D* 101.10 (2020), p. 104038. DOI: [10.1103/PhysRevD.101.104038](https://doi.org/10.1103/PhysRevD.101.104038). arXiv: [2004.03626](https://arxiv.org/abs/2004.03626) [gr-qc].
- [212] Chang Liu et al. “Multiband Observation of LIGO/Virgo Binary Black Hole Mergers in the Gravitational-wave Transient Catalog GWTC-1”. In: *Mon. Not. Roy. Astron. Soc.* 496.1 (2020), pp. 182–196. DOI: [10.1093/mnras/staa1512](https://doi.org/10.1093/mnras/staa1512). arXiv: [2004.12096](https://arxiv.org/abs/2004.12096) [astro-ph.HE].
- [213] Alvin J. K. Chua and Michele Vallisneri. “On parametric tests of relativity with false degrees of freedom”. 2020. arXiv: [2006.08918](https://arxiv.org/abs/2006.08918) [gr-qc].
- [214] Kent Yagi. “A New constraint on scalar Gauss-Bonnet gravity and a possible explanation for the excess of the orbital decay rate in a low-mass X-ray binary”. In: *Phys. Rev. D* 86 (2012), p. 081504. DOI: [10.1103/PhysRevD.86.081504](https://doi.org/10.1103/PhysRevD.86.081504). arXiv: [1204.4524](https://arxiv.org/abs/1204.4524) [gr-qc].
- [215] Cosimo Bambi, Maurizio Giannotti, and F. L. Villante. “The Response of primordial abundances to a general modification of $G(N)$ and/or of the early Universe expansion rate”. In: *Phys. Rev. D* 71 (2005), p. 123524. DOI: [10.1103/PhysRevD.71.123524](https://doi.org/10.1103/PhysRevD.71.123524). arXiv: [astro-ph/0503502](https://arxiv.org/abs/astro-ph/0503502) [astro-ph].
- [216] Craig J. Copi, Adam N. Davis, and Lawrence M. Krauss. “A New nucleosynthesis constraint on the variation of G ”. In: *Phys. Rev. Lett.* 92 (2004), p. 171301. DOI: [10.1103/PhysRevLett.92.171301](https://doi.org/10.1103/PhysRevLett.92.171301). arXiv: [astro-ph/0311334](https://arxiv.org/abs/astro-ph/0311334) [astro-ph].
- [217] R. N. Manchester. “Pulsars and Gravity”. In: *Int. J. Mod. Phys. D* 24.06 (2015), p. 1530018. DOI: [10.1142/S0218271815300189](https://doi.org/10.1142/S0218271815300189). arXiv: [1502.05474](https://arxiv.org/abs/1502.05474) [gr-qc].
- [218] Alex S. Konopliv et al. “Mars high resolution gravity fields from MRO, Mars seasonal gravity, and other dynamical parameters”. In: *Icarus* 211.1 (2011), pp. 401–428. ISSN: 0019-1035. DOI: <https://doi.org/10.1016/j.icarus.2010.10.004>. URL: <http://www.sciencedirect.com/science/article/pii/S0019103510003830>.
- [219] F. Hofmann, J. Müller, and L. Biskupek. “Lunar laser ranging test of the Nordtvedt parameter and a possible variation in the gravitational constant”. In: *A&A* 522, L5 (Nov. 2010), p. L5. DOI: [10.1051/0004-6361/201015659](https://doi.org/10.1051/0004-6361/201015659).
- [220] M. G. Hare. “Mass of the graviton”. In: *Can. J. Phys.* 51 (1973), pp. 431–433. DOI: [10.1139/p73-056](https://doi.org/10.1139/p73-056).

- [221] A. S. Goldhaber and M. M. Nieto. “Mass of the graviton”. In: *Phys. Rev.* D9 (1974), pp. 1119–1121. DOI: [10.1103/PhysRevD.9.1119](https://doi.org/10.1103/PhysRevD.9.1119).
- [222] C. Talmadge et al. “Model Independent Constraints on Possible Modifications of Newtonian Gravity”. In: *Phys. Rev. Lett.* 61 (1988), pp. 1159–1162. DOI: [10.1103/PhysRevLett.61.1159](https://doi.org/10.1103/PhysRevLett.61.1159).
- [223] Hector O. Silva et al. “Astrophysical and theoretical physics implications from multimessenger neutron star observations”. 2020. arXiv: [2004.01253 \[gr-qc\]](https://arxiv.org/abs/2004.01253).
- [224] Archil Kobakhidze, Cyril Lager, and Adrian Manning. “Constraining noncommutative spacetime from GW150914”. In: *Phys. Rev.* D94.6 (2016), p. 064033. DOI: [10.1103/PhysRevD.94.064033](https://doi.org/10.1103/PhysRevD.94.064033). arXiv: [1607.03776 \[gr-qc\]](https://arxiv.org/abs/1607.03776).
- [225] John Baker et al. “The Laser Interferometer Space Antenna: Unveiling the Millihertz Gravitational Wave Sky”. July 2019. arXiv: [1907.06482 \[astro-ph.IM\]](https://arxiv.org/abs/1907.06482).
- [226] B. P. Abbott et al. “Prospects for Observing and Localizing Gravitational-Wave Transients with Advanced LIGO, Advanced Virgo and KAGRA”. In: *Living Rev. Rel.* 21.1 (2018), p. 3. DOI: [10.1007/s41114-018-0012-9](https://doi.org/10.1007/s41114-018-0012-9), [10.1007/lrr-2016-1](https://doi.org/10.1007/lrr-2016-1). arXiv: [1304.0670 \[gr-qc\]](https://arxiv.org/abs/1304.0670).
- [227] Brian O’Reilly et al. *LIGO Document T2000012-v1*. Tech. rep. July 2020. URL: <https://dcc.ligo.org/LIGO-T2000012/public>.
- [228] David McClelland et al. *The LSC-Virgo White Paper on Instrument Science (2016-2017 edition)*. Tech. rep. July 2020. URL: <https://dcc.ligo.org/LIGO-T1600119/public>.
- [229] David Reitze et al. “The US Program in Ground-Based Gravitational Wave Science: Contribution from the LIGO Laboratory”. In: *Bull. Am. Astron. Soc.* 51 (2019), p. 141. arXiv: [1903.04615 \[astro-ph.IM\]](https://arxiv.org/abs/1903.04615).
- [230] Cosmic Explorer. <https://cosmicexplorer.org/researchers.html>. July 2020. URL: <https://cosmicexplorer.org/researchers.html>.
- [231] S. Hild et al. “Sensitivity Studies for Third-Generation Gravitational Wave Observatories”. In: *Class. Quant. Grav.* 28 (2011), p. 094013. DOI: [10.1088/0264-9381/28/9/094013](https://doi.org/10.1088/0264-9381/28/9/094013). arXiv: [1012.0908 \[gr-qc\]](https://arxiv.org/abs/1012.0908).
- [232] Travis Robson, Neil J. Cornish, and Chang Liu. “The construction and use of LISA sensitivity curves”. In: *Class. Quant. Grav.* 36.10 (2019), p. 105011. DOI: [10.1088/1361-6382/ab1101](https://doi.org/10.1088/1361-6382/ab1101). arXiv: [1803.01944 \[astro-ph.HE\]](https://arxiv.org/abs/1803.01944).

- [233] Sashwat Tanay et al. “Convergence of Fourier-domain templates for inspiraling eccentric compact binaries”. In: *Phys. Rev. D* 100.6 (2019), p. 064006. DOI: [10.1103/PhysRevD.100.064006](#). arXiv: [1905.08811 \[gr-qc\]](#).
- [234] LIGO Scientific Collaboration. *LIGO Algorithm Library - LALSuite*. free software (GPL). 2018. DOI: [10.7935/GT1W-FZ16](#).
- [235] Curt Cutler. “Angular resolution of the LISA gravitational wave detector”. In: *Phys. Rev. D* 57 (1998), pp. 7089–7102. DOI: [10.1103/PhysRevD.57.7089](#). arXiv: [gr-qc/9703068 \[gr-qc\]](#).
- [236] Emanuele Berti, Alessandra Buonanno, and Clifford M. Will. “Estimating spinning binary parameters and testing alternative theories of gravity with LISA”. In: *Phys. Rev. D* 71 (2005), p. 084025. DOI: [10.1103/PhysRevD.71.084025](#). arXiv: [gr-qc/0411129 \[gr-qc\]](#).
- [237] Davide Gerosa et al. “Spin orientations of merging black holes formed from the evolution of stellar binaries”. In: *Phys. Rev. D* 98.8 (2018), p. 084036. DOI: [10.1103/PhysRevD.98.084036](#). arXiv: [1808.02491 \[astro-ph.HE\]](#).
- [238] Antoine Klein et al. “Science with the space-based interferometer eLISA: Supermassive black hole binaries”. In: *Phys. Rev. D* 93.2 (2016), p. 024003. DOI: [10.1103/PhysRevD.93.024003](#). arXiv: [1511.05581 \[gr-qc\]](#).
- [239] Sayantani Datta et al. “Tests of general relativity using multiband observations of intermediate mass binary black hole mergers”. In: *Phys. Rev. D* 103.2 (2021), p. 024036. DOI: [10.1103/PhysRevD.103.024036](#). arXiv: [2006.12137 \[gr-qc\]](#).
- [240] Jonathan R. Gair et al. “Exploring intermediate and massive black-hole binaries with the Einstein Telescope”. In: *Gen. Rel. Grav.* 43 (2011), pp. 485–518. DOI: [10.1007/s10714-010-1104-3](#). arXiv: [0907.5450 \[astro-ph.CO\]](#).
- [241] Karan Jani, Deirdre Shoemaker, and Curt Cutler. “Detectability of Intermediate-Mass Black Holes in Multiband Gravitational Wave Astronomy”. In: *Nature Astron.* 4.3 (2019), pp. 260–265. DOI: [10.1038/s41550-019-0932-7](#). arXiv: [1908.04985 \[gr-qc\]](#).
- [242] Michal Dominik et al. “Double Compact Objects III: Gravitational Wave Detection Rates”. In: *Astrophys. J.* 806.2 (2015), p. 263. DOI: [10.1088/0004-637X/806/2/263](#). arXiv: [1405.7016 \[astro-ph.HE\]](#).
- [243] Lee Samuel Finn. “Binary inspiral, gravitational radiation, and cosmology”. In: *Phys. Rev. D* 53 (1996), pp. 2878–2894. DOI: [10.1103/PhysRevD.53.2878](#). arXiv: [gr-qc/9601048 \[gr-qc\]](#).

- [244] P. C. Peters. “Gravitational Radiation and the Motion of Two Point Masses”. In: *Phys. Rev.* 136 (1964), B1224–B1232. DOI: [10.1103/PhysRev.136.B1224](https://doi.org/10.1103/PhysRev.136.B1224).
- [245] Katie Chamberlain et al. “Frequency-domain waveform approximants capturing Doppler shifts”. In: *Phys. Rev. D* 99.2 (2019), p. 024025. DOI: [10.1103/PhysRevD.99.024025](https://doi.org/10.1103/PhysRevD.99.024025). arXiv: [1809.04799 \[gr-qc\]](https://arxiv.org/abs/1809.04799).
- [246] Neil J. Cornish and Kevin Shuman. “Black Hole Hunting with LISA”. In: *Phys. Rev. D* 101.12 (2020), p. 124008. DOI: [10.1103/PhysRevD.101.124008](https://doi.org/10.1103/PhysRevD.101.124008). arXiv: [2005.03610 \[gr-qc\]](https://arxiv.org/abs/2005.03610).
- [247] Michael Kesden et al. “Effective potentials and morphological transitions for binary black-hole spin precession”. In: *Phys. Rev. Lett.* 114.8 (2015), p. 081103. DOI: [10.1103/PhysRevLett.114.081103](https://doi.org/10.1103/PhysRevLett.114.081103). arXiv: [1411.0674 \[gr-qc\]](https://arxiv.org/abs/1411.0674).
- [248] Davide Gerosa et al. “Multi-timescale analysis of phase transitions in precessing black-hole binaries”. In: *Phys. Rev. D* 92 (2015), p. 064016. DOI: [10.1103/PhysRevD.92.064016](https://doi.org/10.1103/PhysRevD.92.064016). arXiv: [1506.03492 \[gr-qc\]](https://arxiv.org/abs/1506.03492).
- [249] Davide Gerosa and Michael Kesden. “PRECESSION: Dynamics of spinning black-hole binaries with python”. In: *Phys. Rev. D* 93.12 (2016), p. 124066. DOI: [10.1103/PhysRevD.93.124066](https://doi.org/10.1103/PhysRevD.93.124066). arXiv: [1605.01067 \[astro-ph.HE\]](https://arxiv.org/abs/1605.01067).
- [250] Enrico Barausse. “The evolution of massive black holes and their spins in their galactic hosts”. In: *Mon. Not. Roy. Astron. Soc.* 423 (2012), pp. 2533–2557. DOI: [10.1111/j.1365-2966.2012.21057.x](https://doi.org/10.1111/j.1365-2966.2012.21057.x). arXiv: [1201.5888 \[astro-ph.CO\]](https://arxiv.org/abs/1201.5888).
- [251] A. Sesana et al. “Linking the spin evolution of massive black holes to galaxy kinematics”. In: *Astrophys. J.* 794 (2014), p. 104. DOI: [10.1088/0004-637X/794/2/104](https://doi.org/10.1088/0004-637X/794/2/104). arXiv: [1402.7088 \[astro-ph.CO\]](https://arxiv.org/abs/1402.7088).
- [252] Fabio Antonini, Enrico Barausse, and Joseph Silk. “The imprint of massive black-hole mergers on the correlation between nuclear star clusters and their host galaxies”. In: *Astrophys. J. Lett.* 806 (2015), p. L8. DOI: [10.1088/2041-8205/806/1/L8](https://doi.org/10.1088/2041-8205/806/1/L8). arXiv: [1504.04033 \[astro-ph.GA\]](https://arxiv.org/abs/1504.04033).
- [253] Krzysztof Belczynski et al. “Compact Binary Merger Rates: Comparison with LIGO/Virgo Upper Limits”. In: *Astrophys. J.* 819.2 (2016), p. 108. DOI: [10.3847/0004-637X/819/2/108](https://doi.org/10.3847/0004-637X/819/2/108). arXiv: [1510.04615 \[astro-ph.HE\]](https://arxiv.org/abs/1510.04615).
- [254] Antoine Klein et al. *MBH Simulation Data Release*. <https://people.sissa.it/~barausse/catalogs/>. URL: <https://people.sissa.it/~barausse/catalogs/>.
- [255] Emanuele Berti, Jonathan Gair, and Alberto Sesana. “Graviton mass bounds from space-based gravitational-wave observations of massive black hole populations”. In: *Phys. Rev. D* 84 (2011), p. 101501. DOI: [10.1103/PhysRevD.84.101501](https://doi.org/10.1103/PhysRevD.84.101501). arXiv: [1107.3528 \[gr-qc\]](https://arxiv.org/abs/1107.3528).

- [256] Eric Poisson and Clifford M. Will. “Gravitational waves from inspiraling compact binaries: Parameter estimation using second postNewtonian wave forms”. In: *Phys. Rev. D* 52 (1995), pp. 848–855. DOI: [10.1103/PhysRevD.52.848](#). arXiv: [gr-qc/9502040 \[gr-qc\]](#).
- [257] Curt Cutler and Eanna E. Flanagan. “Gravitational waves from merging compact binaries: How accurately can one extract the binary’s parameters from the inspiral wave form?” In: *Phys. Rev. D* 49 (1994), pp. 2658–2697. DOI: [10.1103/PhysRevD.49.2658](#). arXiv: [gr-qc/9402014](#).
- [258] William H. Press et al. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. 3rd ed. USA: Cambridge University Press, 2007. ISBN: 0521880688.
- [259] Michele Maggiore. *Gravitational Waves. Vol. 1: Theory and Experiments*. Oxford Master Series in Physics. Oxford University Press, 2007. ISBN: 9780198570745, 9780198520740. URL: <http://www.oup.com/uk/catalogue/?ci=9780198570745>.
- [260] Nicolas Yunes, Frans Pretorius, and David Spergel. “Constraining the evolutionary history of Newton’s constant with gravitational wave observations”. In: *Phys. Rev. D* 81 (2010), p. 064018. DOI: [10.1103/PhysRevD.81.064018](#). arXiv: [0912.2724 \[gr-qc\]](#).
- [261] Katerina Chatziioannou et al. “Detection and Parameter Estimation of Gravitational Waves from Compact Binary Inspirals with Analytical Double-Precessing Templates”. In: *Phys. Rev. D* 89.10 (2014), p. 104023. DOI: [10.1103/PhysRevD.89.104023](#). arXiv: [1404.3180 \[gr-qc\]](#).
- [262] Sebastian Khan et al. “Phenomenological model for the gravitational-wave signal from precessing binary black holes with two-spin effects”. In: *Phys. Rev. D* 100.2 (2019), p. 024059. DOI: [10.1103/PhysRevD.100.024059](#). arXiv: [1809.10113 \[gr-qc\]](#).
- [263] Geraint Pratten et al. “Let’s twist again: computationally efficient models for the dominant and sub-dominant harmonic modes of precessing binary black holes”. 2020. arXiv: [2004.06503 \[gr-qc\]](#).
- [264] Serguei Ossokine et al. “Multipolar Effective-One-Body Waveforms for Precessing Binary Black Holes: Construction and Validation”. In: *Phys. Rev. D* 102.4 (2020), p. 044055. DOI: [10.1103/PhysRevD.102.044055](#). arXiv: [2004.09442 \[gr-qc\]](#).
- [265] Carl L. Rodriguez et al. “Illuminating Black Hole Binary Formation Channels with Spins in Advanced LIGO”. In: *Astrophys. J. Lett.* 832.1 (2016), p. L2. DOI: [10.3847/2041-8205/832/1/L2](#). arXiv: [1609.05916 \[astro-ph.HE\]](#).
- [266] Nathan Steinle and Michael Kesden. “Pathways for producing binary black holes with large misaligned spins in the isolated formation channel”. 2020. arXiv: [2010.00078 \[astro-ph.HE\]](#).

- [267] José M. M. Senovilla and David Garfinkle. “The 1965 Penrose singularity theorem”. In: *Class. Quant. Grav.* 32.12 (2015), p. 124008. DOI: [10.1088/0264-9381/32/12/124008](https://doi.org/10.1088/0264-9381/32/12/124008). arXiv: [1410.5226](https://arxiv.org/abs/1410.5226) [gr-qc].
- [268] Roger Penrose. “Gravitational collapse and space-time singularities”. In: *Phys. Rev. Lett.* 14 (1965), pp. 57–59. DOI: [10.1103/PhysRevLett.14.57](https://doi.org/10.1103/PhysRevLett.14.57).
- [269] Laurent Canetti, Marco Drewes, and Mikhail Shaposhnikov. “Matter and Antimatter in the Universe”. In: *New J. Phys.* 14 (2012), p. 095012. DOI: [10.1088/1367-2630/14/9/095012](https://doi.org/10.1088/1367-2630/14/9/095012). arXiv: [1204.4186](https://arxiv.org/abs/1204.4186) [hep-ph].
- [270] Zack Carson and Kent Yagi. *Testing General Relativity with Gravitational Waves*. Nov. 2020. DOI: [10.1007/978-981-15-4702-7_41-1](https://doi.org/10.1007/978-981-15-4702-7_41-1). arXiv: [2011.02938](https://arxiv.org/abs/2011.02938) [gr-qc].
- [271] Neil J. Cornish et al. “BayesWave analysis pipeline in the era of gravitational wave observations”. In: *Phys. Rev. D* 103.4 (2021), p. 044006. DOI: [10.1103/PhysRevD.103.044006](https://doi.org/10.1103/PhysRevD.103.044006). arXiv: [2011.09494](https://arxiv.org/abs/2011.09494) [gr-qc].
- [272] Katerina Chatziioannou et al. “Morphology-independent test of the mixed polarization content of transient gravitational wave signals”. In: *Phys. Rev. D* 104.4 (2021), p. 044005. DOI: [10.1103/PhysRevD.104.044005](https://doi.org/10.1103/PhysRevD.104.044005). arXiv: [2105.01521](https://arxiv.org/abs/2105.01521) [gr-qc].
- [273] Michele Vallisneri. “Testing general relativity with gravitational waves: a reality check”. In: *Phys. Rev. D* 86 (2012), p. 082001. DOI: [10.1103/PhysRevD.86.082001](https://doi.org/10.1103/PhysRevD.86.082001). arXiv: [1207.4759](https://arxiv.org/abs/1207.4759) [gr-qc].
- [274] Thibault Damour and Alessandro Nagar. “The Effective-One-Body Approach to the General Relativistic Two Body Problem”. In: *Astrophysical Black Holes*. Ed. by Francesco Haardt et al. Cham: Springer International Publishing, 2016, pp. 273–312. ISBN: 978-3-319-19416-5. DOI: [10.1007/978-3-319-19416-5_7](https://doi.org/10.1007/978-3-319-19416-5_7). URL: https://doi.org/10.1007/978-3-319-19416-5_7.
- [275] Banafsheh Shiralilou et al. “Post-Newtonian Gravitational and Scalar Waves in Scalar-Gauss-Bonnet Gravity”. May 2021. arXiv: [2105.13972](https://arxiv.org/abs/2105.13972) [gr-qc].
- [276] Thibault Damour and Gilles Esposito-Farese. “Nonperturbative strong field effects in tensor - scalar theories of gravitation”. In: *Phys. Rev. Lett.* 70 (1993), pp. 2220–2223. DOI: [10.1103/PhysRevLett.70.2220](https://doi.org/10.1103/PhysRevLett.70.2220).
- [277] Thibault Damour and Gilles Esposito-Farese. “Tensor - scalar gravity and binary pulsar experiments”. In: *Phys. Rev. D* 54 (1996), pp. 1474–1491. DOI: [10.1103/PhysRevD.54.1474](https://doi.org/10.1103/PhysRevD.54.1474). arXiv: [gr-qc/9602056](https://arxiv.org/abs/gr-qc/9602056).
- [278] Daniela D. Doneva and Stoytcho S. Yazadjiev. “New Gauss-Bonnet Black Holes with Curvature-Induced Scalarization in Extended Scalar-Tensor Theories”. In: *Phys. Rev. Lett.* 120.13 (2018), p. 131103. DOI: [10.1103/PhysRevLett.120.131103](https://doi.org/10.1103/PhysRevLett.120.131103). arXiv: [1711.01187](https://arxiv.org/abs/1711.01187) [gr-qc].

- [279] Hector O. Silva et al. “Spontaneous scalarization of black holes and compact stars from a Gauss-Bonnet coupling”. In: *Phys. Rev. Lett.* 120.13 (2018), p. 131104. DOI: [10.1103/PhysRevLett.120.131104](https://doi.org/10.1103/PhysRevLett.120.131104). arXiv: [1711.02080](https://arxiv.org/abs/1711.02080) [gr-qc].
- [280] Fethi M. Ramazanoğlu. “Spontaneous growth of vector fields in gravity”. In: *Phys. Rev. D* 96.6 (2017), p. 064009. DOI: [10.1103/PhysRevD.96.064009](https://doi.org/10.1103/PhysRevD.96.064009). arXiv: [1706.01056](https://arxiv.org/abs/1706.01056) [gr-qc].
- [281] Hector O. Silva et al. “The ghost of vector fields in compact stars”. Oct. 2021. arXiv: [2110.04594](https://arxiv.org/abs/2110.04594) [gr-qc].
- [282] Laura Sampson, Neil Cornish, and Nicolás Yunes. “Mismodeling in gravitational-wave astronomy: The trouble with templates”. In: *Phys. Rev. D* 89.6 (2014), p. 064037. DOI: [10.1103/PhysRevD.89.064037](https://doi.org/10.1103/PhysRevD.89.064037). arXiv: [1311.4898](https://arxiv.org/abs/1311.4898) [gr-qc].
- [283] Robert H. Swendsen and Jian-Sheng Wang. “Replica Monte Carlo Simulation of Spin-Glasses”. In: *Phys. Rev. Lett.* 57 (21 Nov. 1986), pp. 2607–2609. DOI: [10.1103/PhysRevLett.57.2607](https://doi.org/10.1103/PhysRevLett.57.2607). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.57.2607>.
- [284] David J. Earl and Michael W. Deem. “Parallel tempering: theory, applications, and new perspectives.” In: *Physical chemistry chemical physics : PCCP* 7 23 (2005), pp. 3910–6.
- [285] W. D. Vousden, W. M. Farr, and I. Mandel. “Dynamic temperature selection for parallel tempering in Markov chain Monte Carlo simulations”. In: *MNRAS* 455.2 (Jan. 2016), pp. 1919–1937. DOI: [10.1093/mnras/stv2422](https://doi.org/10.1093/mnras/stv2422). arXiv: [1501.05823](https://arxiv.org/abs/1501.05823) [astro-ph.IM].
- [286] Katerina Chatziioannou et al. “Constructing Gravitational Waves from Generic Spin-Precessing Compact Binary Inspirals”. In: *Phys. Rev. D* 95.10 (2017), p. 104004. DOI: [10.1103/PhysRevD.95.104004](https://doi.org/10.1103/PhysRevD.95.104004). arXiv: [1703.03967](https://arxiv.org/abs/1703.03967) [gr-qc].
- [287] Luc Blanchet. “On the accuracy of the postNewtonian approximation”. In: *25th Johns Hopkins Workshop on Current Problems in Particle Theory: 2001: A Relativistic Spacetime Odyssey. Experiments and Theoretical Viewpoints on General Relativity and Quantum Gravity*. July 2002, pp. 411–430. DOI: [10.1142/9789812791368_0022](https://doi.org/10.1142/9789812791368_0022). arXiv: [gr-qc/0207037](https://arxiv.org/abs/gr-qc/0207037).
- [288] Nicolas Yunes and Emanuele Berti. “Accuracy of the post-Newtonian approximation: Optimal asymptotic expansion for quasicircular, extreme-mass ratio inspirals”. In: *Phys. Rev. D* 77 (2008). [Erratum: *Phys.Rev.D* 83, 109901 (2011)], p. 124006. DOI: [10.1103/PhysRevD.77.124006](https://doi.org/10.1103/PhysRevD.77.124006). arXiv: [0803.1853](https://arxiv.org/abs/0803.1853) [gr-qc].

- [289] Thibault Damour, Bala R. Iyer, and B. S. Sathyaprakash. “A Comparison of search templates for gravitational waves from binary inspiral”. In: *Phys. Rev. D* 63 (2001). [Erratum: Phys.Rev.D 72, 029902 (2005)], p. 044023. DOI: [10.1103/PhysRevD.63.044023](https://doi.org/10.1103/PhysRevD.63.044023). arXiv: [gr-qc/0010009](https://arxiv.org/abs/gr-qc/0010009).
- [290] Thibault Damour, Bala R. Iyer, and B. S. Sathyaprakash. “A Comparison of search templates for gravitational waves from binary inspiral - 3.5PN update”. In: *Phys. Rev. D* 66 (2002), p. 027502. DOI: [10.1103/PhysRevD.66.027502](https://doi.org/10.1103/PhysRevD.66.027502). arXiv: [gr-qc/0207021](https://arxiv.org/abs/gr-qc/0207021).
- [291] K. G. Arun et al. “Parameter estimation of inspiralling compact binaries using 3.5 post-Newtonian gravitational wave phasing: The Non-spinning case”. In: *Phys. Rev. D* 71 (2005). [Erratum: Phys.Rev.D 72, 069903 (2005)], p. 084008. DOI: [10.1103/PhysRevD.71.084008](https://doi.org/10.1103/PhysRevD.71.084008). arXiv: [gr-qc/0411146](https://arxiv.org/abs/gr-qc/0411146).
- [292] F. Acernese et al. “Advanced Virgo: a second-generation interferometric gravitational wave detector”. In: *Class. Quant. Grav.* 32.2 (2015), p. 024001. DOI: [10.1088/0264-9381/32/2/024001](https://doi.org/10.1088/0264-9381/32/2/024001). arXiv: [1408.3978 \[gr-qc\]](https://arxiv.org/abs/1408.3978).
- [293] David G. Boulware and Stanley Deser. “String Generated Gravity Models”. In: *Phys. Rev. Lett.* 55 (1985), p. 2656. DOI: [10.1103/PhysRevLett.55.2656](https://doi.org/10.1103/PhysRevLett.55.2656).
- [294] David J. Gross and John H. Sloan. “The quartic effective action for the heterotic string”. In: *Nuclear Physics B* 291 (Jan. 1987), pp. 41–89. DOI: [10.1016/0550-3213\(87\)90465-2](https://doi.org/10.1016/0550-3213(87)90465-2).
- [295] Nicolas Yunes and Leo C. Stein. “Non-Spinning Black Holes in Alternative Theories of Gravity”. In: *Phys. Rev. D* 83 (2011), p. 104002. DOI: [10.1103/PhysRevD.83.104002](https://doi.org/10.1103/PhysRevD.83.104002). arXiv: [1101.2921 \[gr-qc\]](https://arxiv.org/abs/1101.2921).
- [296] Kent Yagi, Leo C. Stein, and Nicolas Yunes. “Challenging the Presence of Scalar Charge and Dipolar Radiation in Binary Pulsars”. In: *Phys. Rev. D* 93.2 (2016), p. 024010. DOI: [10.1103/PhysRevD.93.024010](https://doi.org/10.1103/PhysRevD.93.024010). arXiv: [1510.02152 \[gr-qc\]](https://arxiv.org/abs/1510.02152).
- [297] Thomas P. Sotiriou and Shuang-Yong Zhou. “Black hole hair in generalized scalar-tensor gravity: An explicit example”. In: *Phys. Rev. D* 90 (2014), p. 124063. DOI: [10.1103/PhysRevD.90.124063](https://doi.org/10.1103/PhysRevD.90.124063). arXiv: [1408.1698 \[gr-qc\]](https://arxiv.org/abs/1408.1698).
- [298] Zhenwei Lyu, Nan Jiang, and Kent Yagi. “Constraints on Einstein-dilation-Gauss-Bonnet gravity from black hole-neutron star gravitational wave events”. In: *Phys. Rev. D* 105.6 (2022), p. 064001. DOI: [10.1103/PhysRevD.105.064001](https://doi.org/10.1103/PhysRevD.105.064001). arXiv: [2201.02543 \[gr-qc\]](https://arxiv.org/abs/2201.02543).
- [299] Muhammed Saleem et al. “Parametrized tests of post-Newtonian theory using principal component analysis”. Oct. 2021. arXiv: [2110.10147 \[gr-qc\]](https://arxiv.org/abs/2110.10147).
- [300] Andrey A. Shoom et al. “Testing GR with the Gravitational Wave Inspiral Signal GW170817”. May 2021. arXiv: [2105.02191 \[gr-qc\]](https://arxiv.org/abs/2105.02191).

- [301] Archana Pai and K. G. Arun. “Singular value decomposition in parametrised tests of post-Newtonian theory”. In: *Class. Quant. Grav.* 30 (2013), p. 025011. DOI: [10.1088/0264-9381/30/2/025011](#). arXiv: [1207.1943 \[gr-qc\]](#).
- [302] Mohammed Khalil et al. “Theory-agnostic framework for dynamical scalarization of compact binaries”. In: *Phys. Rev. D* 100.12 (2019), p. 124013. DOI: [10.1103/PhysRevD.100.124013](#). arXiv: [1906.08161 \[gr-qc\]](#).
- [303] Hector O. Silva et al. “Dynamical Descalarization in Binary Black Hole Mergers”. In: *Phys. Rev. Lett.* 127.3 (2021), p. 031101. DOI: [10.1103/PhysRevLett.127.031101](#). arXiv: [2012.10436 \[gr-qc\]](#).
- [304] Timothy Clifton et al. “Modified Gravity and Cosmology”. In: *Phys. Rept.* 513 (2012), pp. 1–189. DOI: [10.1016/j.physrep.2012.01.001](#). arXiv: [1106.2476 \[astro-ph.CO\]](#).
- [305] D. N. Spergel et al. “First year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Determination of cosmological parameters”. In: *Astrophys. J. Suppl.* 148 (2003), pp. 175–194. DOI: [10.1086/377226](#). arXiv: [astro-ph/0302209](#).
- [306] Thomas P. Sotiriou and Enrico Barausse. “Post-Newtonian expansion for Gauss-Bonnet gravity”. In: *Phys. Rev. D* 75 (2007), p. 084007. DOI: [10.1103/PhysRevD.75.084007](#). arXiv: [gr-qc/0612065 \[gr-qc\]](#).
- [307] Stephon Alexander and Nicolas Yunes. “A New PPN parameter to test Chern-Simons gravity”. In: *Phys. Rev. Lett.* 99 (2007), p. 241101. DOI: [10.1103/PhysRevLett.99.241101](#). arXiv: [hep-th/0703265 \[hep-th\]](#).
- [308] I. Ciufolini and E. C. Pavlis. “A confirmation of the general relativistic prediction of the Lense-Thirring effect”. In: *Nature* 431 (Oct. 2004), pp. 958–960. DOI: [10.1038/nature03007](#).
- [309] C. W. F. Everitt et al. “Gravity Probe B: Final Results of a Space Experiment to Test General Relativity”. In: *Phys. Rev. Lett.* 106 (2011), p. 221101. DOI: [10.1103/PhysRevLett.106.221101](#). arXiv: [1105.3456 \[gr-qc\]](#).
- [310] Yacine Ali-Haïmoud and Yanbei Chen. “Slowly-rotating stars and black holes in dynamical Chern-Simons gravity”. In: *Phys. Rev. D* 84 (2011), p. 124033. DOI: [10.1103/PhysRevD.84.124033](#). arXiv: [1110.5329 \[astro-ph.HE\]](#).
- [311] Nicholas Loutrel, Takahiro Tanaka, and Nicolas Yunes. “Scalar Tops and Perturbed Quadrupoles: Probing Fundamental Physics with Spin-Precessing Binaries”. In: *Class. Quant. Grav.* 36.10 (2019), 10LT02. DOI: [10.1088/1361-6382/ab15fa](#). arXiv: [1806.07425 \[gr-qc\]](#).

- [312] Nicholas Loutrel, Takahiro Tanaka, and Nicolas Yunes. “Spin-Precessing Black Hole Binaries in Dynamical Chern-Simons Gravity”. In: *Phys. Rev. D* 98.6 (2018), p. 064020. DOI: [10.1103/PhysRevD.98.064020](https://doi.org/10.1103/PhysRevD.98.064020). arXiv: [1806.07431 \[gr-qc\]](https://arxiv.org/abs/1806.07431).
- [313] Maria Okounkova et al. “Numerical binary black hole mergers in dynamical Chern-Simons gravity: Scalar field”. In: *Phys. Rev. D* 96.4 (2017), p. 044020. DOI: [10.1103/PhysRevD.96.044020](https://doi.org/10.1103/PhysRevD.96.044020). arXiv: [1705.07924 \[gr-qc\]](https://arxiv.org/abs/1705.07924).
- [314] Maria Okounkova, Mark A. Scheel, and Saul A. Teukolsky. “Numerical black hole initial data and shadows in dynamical Chern–Simons gravity”. In: *Class. Quant. Grav.* 36.5 (2019), p. 054001. DOI: [10.1088/1361-6382/aafcdf](https://doi.org/10.1088/1361-6382/aafcdf). arXiv: [1810.05306 \[gr-qc\]](https://arxiv.org/abs/1810.05306).
- [315] Maria Okounkova, Mark A. Scheel, and Saul A. Teukolsky. “Evolving Metric Perturbations in dynamical Chern-Simons Gravity”. In: *Phys. Rev. D* 99.4 (2019), p. 044019. DOI: [10.1103/PhysRevD.99.044019](https://doi.org/10.1103/PhysRevD.99.044019). arXiv: [1811.10713 \[gr-qc\]](https://arxiv.org/abs/1811.10713).
- [316] Helvi Wittek et al. “Black holes and binary mergers in scalar Gauss-Bonnet gravity: scalar field dynamics”. In: *Phys. Rev. D* 99.6 (2019), p. 064035. DOI: [10.1103/PhysRevD.99.064035](https://doi.org/10.1103/PhysRevD.99.064035). arXiv: [1810.05177 \[gr-qc\]](https://arxiv.org/abs/1810.05177).
- [317] Vitor Cardoso and Leonardo Gualtieri. “Perturbations of Schwarzschild black holes in Dynamical Chern-Simons modified gravity”. In: *Phys. Rev. D* 80 (2009). [Erratum: *Phys. Rev. D* 81,089903(2010)], p. 064008. DOI: [10.1103/PhysRevD.81.089903](https://doi.org/10.1103/PhysRevD.81.089903), [10.1103/PhysRevD.80.064008](https://doi.org/10.1103/PhysRevD.80.064008). arXiv: [0907.5008 \[gr-qc\]](https://arxiv.org/abs/0907.5008).
- [318] Jose Luis Blázquez-Salcedo et al. “Perturbed black holes in Einstein-dilaton-Gauss-Bonnet gravity: Stability, ringdown, and gravitational-wave emission”. In: *Phys. Rev. D* 94.10 (2016), p. 104024. DOI: [10.1103/PhysRevD.94.104024](https://doi.org/10.1103/PhysRevD.94.104024). arXiv: [1609.01286 \[gr-qc\]](https://arxiv.org/abs/1609.01286).
- [319] B. P. Abbott et al. “Observation of Gravitational Waves from a Binary Black Hole Merger”. In: *Phys. Rev. Lett.* 116.6 (2016), p. 061102. DOI: [10.1103/PhysRevLett.116.061102](https://doi.org/10.1103/PhysRevLett.116.061102). arXiv: [1602.03837 \[gr-qc\]](https://arxiv.org/abs/1602.03837).
- [320] B. P. Abbott et al. “GW150914: First results from the search for binary black hole coalescence with Advanced LIGO”. In: *Phys. Rev. D* 93.12 (2016), p. 122003. DOI: [10.1103/PhysRevD.93.122003](https://doi.org/10.1103/PhysRevD.93.122003). arXiv: [1602.03839 \[gr-qc\]](https://arxiv.org/abs/1602.03839).
- [321] B. P. Abbott et al. “GW151226: Observation of Gravitational Waves from a 22-Solar-Mass Binary Black Hole Coalescence”. In: *Phys. Rev. Lett.* 116.24 (2016), p. 241103. DOI: [10.1103/PhysRevLett.116.241103](https://doi.org/10.1103/PhysRevLett.116.241103). arXiv: [1606.04855 \[gr-qc\]](https://arxiv.org/abs/1606.04855).

- [322] B. P. Abbott et al. “GW170608: Observation of a 19-solar-mass Binary Black Hole Coalescence”. In: *Astrophys. J.* 851.2 (2017), p. L35. DOI: [10.3847/2041-8213/aa9f0c](#). arXiv: [1711.05578 \[astro-ph.HE\]](#).
- [323] B. P. Abbott et al. “GW170814: A Three-Detector Observation of Gravitational Waves from a Binary Black Hole Coalescence”. In: *Phys. Rev. Lett.* 119.14 (2017), p. 141101. DOI: [10.1103/PhysRevLett.119.141101](#). arXiv: [1709.09660 \[gr-qc\]](#).
- [324] LIGO Scientific Collaboration and Virgo Collaboration. *Data release for testing GR with GWTC-1*. 2019.
- [325] Nicolas Yunes and Frans Pretorius. “Dynamical Chern-Simons Modified Gravity. I. Spinning Black Holes in the Slow-Rotation Approximation”. In: *Phys. Rev. D* 79 (2009), p. 084043. DOI: [10.1103/PhysRevD.79.084043](#). arXiv: [0902.4669 \[gr-qc\]](#).
- [326] Kent Yagi, Nicolas Yunes, and Takahiro Tanaka. “Gravitational Waves from Quasi-Circular Black Hole Binaries in Dynamical Chern-Simons Gravity”. In: *Phys. Rev. Lett.* 109 (2012). [Erratum: *Phys.Rev.Lett.* 116, 169902 (2016), Erratum: *Phys.Rev.Lett.* 124, 029901 (2020)], p. 251105. DOI: [10.1103/PhysRevLett.116.169902](#). arXiv: [1208.5102 \[gr-qc\]](#).
- [327] Emanuele Berti, Kent Yagi, and Nicolás Yunes. “Extreme Gravity Tests with Gravitational Waves from Compact Binary Coalescences: (I) Inspiral-Merger”. In: *Gen. Rel. Grav.* 50.4 (2018), p. 46. DOI: [10.1007/s10714-018-2362-8](#). arXiv: [1801.03208 \[gr-qc\]](#).
- [328] Parameswaran Ajith et al. “Phenomenological template family for black-hole coalescence waveforms”. In: *Class. Quant. Grav.* 24 (2007), S689–S700. DOI: [10.1088/0264-9381/24/19/S31](#). arXiv: [0704.3764 \[gr-qc\]](#).
- [329] P. Ajith et al. “Inspiral-merger-ringdown waveforms for black-hole binaries with non-precessing spins”. In: *Phys. Rev. Lett.* 106 (2011), p. 241101. DOI: [10.1103/PhysRevLett.106.241101](#). arXiv: [0909.2867 \[gr-qc\]](#).
- [330] L. Santamaria et al. “Matching post-Newtonian and numerical relativity waveforms: systematic errors and a new phenomenological model for non-precessing black hole binaries”. In: *Phys. Rev. D* 82 (2010), p. 064016. DOI: [10.1103/PhysRevD.82.064016](#). arXiv: [1005.3306 \[gr-qc\]](#).
- [331] Hector O. Silva et al. “Stability of scalarized black hole solutions in scalar-Gauss-Bonnet gravity”. In: *Phys. Rev. D* 99.6 (2019), p. 064011. DOI: [10.1103/PhysRevD.99.064011](#). arXiv: [1812.05590 \[gr-qc\]](#).

- [332] Caio F. B. Macedo et al. “Self-interactions and Spontaneous Black Hole Scalarization”. In: *Phys. Rev. D* 99.10 (2019), p. 104041. DOI: [10.1103/PhysRevD.99.104041](#). arXiv: [1903.06784 \[gr-qc\]](#).
- [333] Pedro V. P. Cunha, Carlos A. R. Herdeiro, and Eugen Radu. “Spontaneously Scalarized Kerr Black Holes in Extended Scalar-Tensor–Gauss-Bonnet Gravity”. In: *Phys. Rev. Lett.* 123.1 (2019), p. 011101. DOI: [10.1103/PhysRevLett.123.011101](#). arXiv: [1904.09997 \[gr-qc\]](#).
- [334] Marco Crisostomi et al. “Beyond Lovelock gravity: Higher derivative metric theories”. In: *Phys. Rev. D* 97.4 (2018), p. 044034. DOI: [10.1103/PhysRevD.97.044034](#). arXiv: [1710.04531 \[hep-th\]](#).
- [335] Atsushi Nishizawa and Tsutomu Kobayashi. “Parity-violating gravity and GW170817”. In: *Phys. Rev. D* 98.12 (2018), p. 124018. DOI: [10.1103/PhysRevD.98.124018](#). arXiv: [1809.00815 \[gr-qc\]](#).
- [336] B. P. Abbott et al. “GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral”. In: *Phys. Rev. Lett.* 119.16 (2017), p. 161101. DOI: [10.1103/PhysRevLett.119.161101](#). arXiv: [1710.05832 \[gr-qc\]](#).
- [337] B. P. Abbott et al. “Gravitational Waves and Gamma-rays from a Binary Neutron Star Merger: GW170817 and GRB 170817A”. In: *Astrophys. J.* 848.2 (2017), p. L13. DOI: [10.3847/2041-8213/aa920c](#). arXiv: [1710.05834 \[astro-ph.HE\]](#).
- [338] D. Lovelock. “The Einstein tensor and its generalizations”. In: *J. Math. Phys.* 12 (1971), pp. 498–501. DOI: [10.1063/1.1665613](#).
- [339] D. Lovelock. “The four-dimensionality of space and the einstein tensor”. In: *J. Math. Phys.* 13 (1972), pp. 874–876. DOI: [10.1063/1.1666069](#).
- [340] T rence Delsate, David Hilditch, and Helvi Witek. “Initial value formulation of dynamical Chern-Simons gravity”. In: *Phys. Rev. D* 91.2 (2015), p. 024027. DOI: [10.1103/PhysRevD.91.024027](#). arXiv: [1407.6727 \[gr-qc\]](#).
- [341] Emanuele Berti et al. “Extreme Gravity Tests with Gravitational Waves from Compact Binary Coalescences: (II) Ringdown”. In: *Gen. Rel. Grav.* 50.5 (2018), p. 49. DOI: [10.1007/s10714-018-2372-6](#). arXiv: [1801.03587 \[gr-qc\]](#).
- [342] B. P. Abbott et al. “Binary Black Hole Mergers in the first Advanced LIGO Observing Run”. In: *Phys. Rev. X* 6.4 (2016). [Erratum: *Phys. Rev. X* 8, 039903 (2018)], p. 041015. DOI: [10.1103/PhysRevX.6.041015](#). arXiv: [1606.04856 \[gr-qc\]](#).
- [343] B. P. Abbott et al. “Tests of General Relativity with GW170817”. In: *Phys. Rev. Lett.* 123.1 (2019), p. 011102. DOI: [10.1103/PhysRevLett.123.011102](#). arXiv: [1811.00364 \[gr-qc\]](#).

- [344] Richard Brito, Alessandra Buonanno, and Vivien Raymond. “Black-hole Spectroscopy by Making Full Use of Gravitational-Wave Modeling”. In: *Phys. Rev. D* 98.8 (2018), p. 084038. DOI: [10.1103/PhysRevD.98.084038](#). arXiv: [1805.00293 \[gr-qc\]](#).
- [345] Noah Sennett et al. “Gravitational-Wave Constraints on an Effective Field-Theory Extension of General Relativity”. In: *Phys. Rev. D* 102.4 (2020), p. 044056. DOI: [10.1103/PhysRevD.102.044056](#). arXiv: [1912.09917 \[gr-qc\]](#).
- [346] Kei Yamada and Takahiro Tanaka. “Parametrized test of parity-violating gravity using GWTC-1 events”. In: *PTEP* 2020.9 (2020), 093E01. DOI: [10.1093/ptep/ptaa103](#). arXiv: [2006.11086 \[gr-qc\]](#).
- [347] Abhirup Ghosh, Richard Brito, and Alessandra Buonanno. “Constraints on quasi-normal-mode frequencies with LIGO-Virgo binary-black-hole observations”. Apr. 2021. arXiv: [2104.01906 \[gr-qc\]](#).
- [348] T. G. F. Li et al. “Towards a generic test of the strong field dynamics of general relativity using compact binary coalescence: Further investigations”. In: *J. Phys. Conf. Ser.* 363 (2012). Ed. by Mark Hannam et al., p. 012028. DOI: [10.1088/1742-6596/363/1/012028](#). arXiv: [1111.5274 \[gr-qc\]](#).
- [349] Félix-Louis Julié and Emanuele Berti. “Post-Newtonian dynamics and black hole thermodynamics in Einstein-scalar-Gauss-Bonnet gravity”. In: *Phys. Rev. D* 100.10 (2019), p. 104061. DOI: [10.1103/PhysRevD.100.104061](#). arXiv: [1909.05258 \[gr-qc\]](#).
- [350] B. P. Abbott et al. “GW190425: Observation of a Compact Binary Coalescence with Total Mass $\sim 3.4M_{\odot}$ ”. In: *Astrophys. J. Lett.* 892.1 (2020), p. L3. DOI: [10.3847/2041-8213/ab75f5](#). arXiv: [2001.01761 \[astro-ph.HE\]](#).
- [351] R. Abbott et al. “GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run”. Oct. 2020. arXiv: [2010.14527 \[gr-qc\]](#).
- [352] R. Abbott et al. “GW190814: Gravitational Waves from the Coalescence of a 23 Solar Mass Black Hole with a 2.6 Solar Mass Compact Object”. In: *Astrophys. J. Lett.* 896.2 (2020), p. L44. DOI: [10.3847/2041-8213/ab960f](#). arXiv: [2006.12611 \[astro-ph.HE\]](#).
- [353] Pratik Wagle et al. “Hair loss in parity violating gravity”. In: *Class. Quant. Grav.* 36.11 (2019), p. 115004. DOI: [10.1088/1361-6382/ab0eed](#). arXiv: [1812.05646 \[gr-qc\]](#).
- [354] Alexander Saffer, Hector O. Silva, and Nico Yunes. “Exterior spacetime of relativistic stars in scalar-Gauss-Bonnet gravity”. In: *Phys. Rev. D* 100.4 (2019), p. 044030. DOI: [10.1103/PhysRevD.100.044030](#). arXiv: [1903.07779 \[gr-qc\]](#).

- [355] Kohkichi Konno, Toyoki Matsuyama, and Satoshi Tanda. “Rotating black hole in extended Chern-Simons modified gravity”. In: *Prog. Theor. Phys.* 122 (2009), pp. 561–568. DOI: [10.1143/PTP.122.561](#). arXiv: [0902.4767 \[gr-qc\]](#).
- [356] Kartik Prabhu and Leo C. Stein. “Black hole scalar charge from a topological horizon integral in Einstein-dilaton-Gauss-Bonnet gravity”. In: *Phys. Rev. D* 98.2 (2018), p. 021503. DOI: [10.1103/PhysRevD.98.021503](#). arXiv: [1805.02668 \[gr-qc\]](#).
- [357] Warren G. Anderson et al. “An Excess power statistic for detection of burst sources of gravitational radiation”. In: *Phys. Rev. D* 63 (2001), p. 042003. DOI: [10.1103/PhysRevD.63.042003](#). arXiv: [gr-qc/0008066](#).
- [358] Maximiliano Isi, Katerina Chatziioannou, and Will M. Farr. “Hierarchical test of general relativity with gravitational waves”. In: *Phys. Rev. Lett.* 123.12 (2019), p. 121101. DOI: [10.1103/PhysRevLett.123.121101](#). arXiv: [1904.08011 \[gr-qc\]](#).
- [359] Lee S. Finn. “Detection, measurement and gravitational radiation”. In: *Phys. Rev. D* 46 (1992), pp. 5236–5249. DOI: [10.1103/PhysRevD.46.5236](#). arXiv: [gr-qc/9209010](#).
- [360] Michele Vallisneri. “Use and abuse of the Fisher information matrix in the assessment of gravitational-wave parameter-estimation prospects”. In: *Phys. Rev. D* 77 (2008), p. 042001. DOI: [10.1103/PhysRevD.77.042001](#). arXiv: [gr-qc/0703086](#).
- [361] Thomas E. Riley et al. “A *NICER* View of PSR J0030+0451: Millisecond Pulsar Parameter Estimation”. In: *Astrophys. J. Lett.* 887.1 (2019), p. L21. DOI: [10.3847/2041-8213/ab481c](#). arXiv: [1912.05702 \[astro-ph.HE\]](#).
- [362] M. C. Miller et al. “PSR J0030+0451 Mass and Radius from *NICER* Data and Implications for the Properties of Neutron Star Matter”. In: *Astrophys. J. Lett.* 887.1 (2019), p. L24. DOI: [10.3847/2041-8213/ab50c5](#). arXiv: [1912.05705 \[astro-ph.HE\]](#).
- [363] B. P. Abbott et al. “GW170817: Measurements of neutron star radii and equation of state”. In: *Phys. Rev. Lett.* 121.16 (2018), p. 161101. DOI: [10.1103/PhysRevLett.121.161101](#). arXiv: [1805.11581 \[gr-qc\]](#).
- [364] Walter Del Pozzo, John Veitch, and Alberto Vecchio. “Testing General Relativity using Bayesian model selection: Applications to observations of gravitational waves from compact binary systems”. In: *Phys. Rev. D* 83 (2011), p. 082002. DOI: [10.1103/PhysRevD.83.082002](#). arXiv: [1101.1391 \[gr-qc\]](#).
- [365] Soichiro Isoyama, Riccardo Sturani, and Hiroyuki Nakano. “Post-Newtonian templates for gravitational waves from compact binary inspirals”. Dec. 2020. arXiv: [2012.01350 \[gr-qc\]](#).

- [366] Maria Okounkova et al. “Numerical binary black hole collisions in dynamical Chern-Simons gravity”. In: *Phys. Rev. D* 100.10 (2019), p. 104026. DOI: [10.1103/PhysRevD.100.104026](https://doi.org/10.1103/PhysRevD.100.104026). arXiv: [1906.08789 \[gr-qc\]](https://arxiv.org/abs/1906.08789).
- [367] Maria Okounkova et al. “Numerical relativity simulation of GW150914 beyond general relativity”. In: *Phys. Rev. D* 101.10 (2020), p. 104016. DOI: [10.1103/PhysRevD.101.104016](https://doi.org/10.1103/PhysRevD.101.104016). arXiv: [1911.02588 \[gr-qc\]](https://arxiv.org/abs/1911.02588).
- [368] Maria Okounkova. “Numerical relativity simulation of GW150914 in Einstein dilaton Gauss-Bonnet gravity”. In: *Phys. Rev. D* 102.8 (2020), p. 084046. DOI: [10.1103/PhysRevD.102.084046](https://doi.org/10.1103/PhysRevD.102.084046). arXiv: [2001.03571 \[gr-qc\]](https://arxiv.org/abs/2001.03571).
- [369] William E. East and Justin L. Ripley. “Evolution of Einstein-scalar-Gauss-Bonnet gravity using a modified harmonic formulation”. In: *Phys. Rev. D* 103.4 (2021), p. 044040. DOI: [10.1103/PhysRevD.103.044040](https://doi.org/10.1103/PhysRevD.103.044040). arXiv: [2011.03547 \[gr-qc\]](https://arxiv.org/abs/2011.03547).
- [370] Zack Carson and Kent Yagi. “Probing Einstein-dilaton Gauss-Bonnet Gravity with the inspiral and ringdown of gravitational waves”. In: *Phys. Rev. D* 101.10 (2020), p. 104030. DOI: [10.1103/PhysRevD.101.104030](https://doi.org/10.1103/PhysRevD.101.104030). arXiv: [2003.00286 \[gr-qc\]](https://arxiv.org/abs/2003.00286).
- [371] C. Molina et al. “Gravitational signature of Schwarzschild black holes in dynamical Chern-Simons gravity”. In: *Phys. Rev. D* 81 (2010), p. 124021. DOI: [10.1103/PhysRevD.81.124021](https://doi.org/10.1103/PhysRevD.81.124021). arXiv: [1004.4007 \[gr-qc\]](https://arxiv.org/abs/1004.4007).
- [372] Pablo A. Cano, Kwinten Fransen, and Thomas Hertog. “Ringing of rotating black holes in higher-derivative gravity”. In: *Phys. Rev. D* 102.4 (2020), p. 044047. DOI: [10.1103/PhysRevD.102.044047](https://doi.org/10.1103/PhysRevD.102.044047). arXiv: [2005.03671 \[gr-qc\]](https://arxiv.org/abs/2005.03671).
- [373] Pratik K. Wagle, Nicolás Yunes, and Hector O. Silva. “Quasinormal modes of slowly-rotating black holes in dynamical Chern-Simons gravity”. Mar. 2021. arXiv: [2103.09913 \[gr-qc\]](https://arxiv.org/abs/2103.09913).
- [374] Lorenzo Pierini and Leonardo Gualtieri. “Quasi-normal modes of rotating black holes in Einstein-dilaton Gauss-Bonnet gravity: the first order in rotation”. Mar. 2021. arXiv: [2103.09870 \[gr-qc\]](https://arxiv.org/abs/2103.09870).
- [375] Hai-Tian Wang et al. “Tight constraints on Einstein-dilation-Gauss-Bonnet model with the GWTC-2 events”. Apr. 2021. arXiv: [2104.07590 \[gr-qc\]](https://arxiv.org/abs/2104.07590).
- [376] *Gravitational-Wave Candidate Event Database*. <https://gracedb.ligo.org/superevents/public/03/>. Accessed: 2010-04-16.

- [377] Ming-Zhe Han et al. “Is GW190425 consistent with being a neutron star–black hole merger?” In: *Astrophys. J. Lett.* 891.1 (2020), p. L5. DOI: [10.3847/2041-8213/ab745a](#). arXiv: [2001.07882 \[astro-ph.HE\]](#).
- [378] Benjamin P Abbott et al. “A guide to LIGO–Virgo detector noise and extraction of transient gravitational-wave signals”. In: *Class. Quant. Grav.* 37.5 (2020), p. 055002. DOI: [10.1088/1361-6382/ab685e](#). arXiv: [1908.11170 \[gr-qc\]](#).
- [379] Nima Arkani-Hamed, Savas Dimopoulos, and G. R. Dvali. “The Hierarchy problem and new dimensions at a millimeter”. In: *Phys. Lett.* B429 (1998), pp. 263–272. DOI: [10.1016/S0370-2693\(98\)00466-3](#). arXiv: [hep-ph/9803315 \[hep-ph\]](#).
- [380] Nima Arkani-Hamed, Savas Dimopoulos, and G. R. Dvali. “Phenomenology, astrophysics and cosmology of theories with submillimeter dimensions and TeV scale quantum gravity”. In: *Phys. Rev.* D59 (1999), p. 086004. DOI: [10.1103/PhysRevD.59.086004](#). arXiv: [hep-ph/9807344 \[hep-ph\]](#).
- [381] Lisa Randall and Raman Sundrum. “A Large mass hierarchy from a small extra dimension”. In: *Phys. Rev. Lett.* 83 (1999), pp. 3370–3373. DOI: [10.1103/PhysRevLett.83.3370](#). arXiv: [hep-ph/9905221 \[hep-ph\]](#).
- [382] Lisa Randall and Raman Sundrum. “An Alternative to compactification”. In: *Phys. Rev. Lett.* 83 (1999), pp. 4690–4693. DOI: [10.1103/PhysRevLett.83.4690](#). arXiv: [hep-th/9906064 \[hep-th\]](#).
- [383] Roberto Emparan, Juan Garcia-Bellido, and Nemanja Kaloper. “Black hole astrophysics in AdS brane worlds”. In: *JHEP* 01 (2003), p. 079. DOI: [10.1088/1126-6708/2003/01/079](#). arXiv: [hep-th/0212132 \[hep-th\]](#).
- [384] Pau Figueras and Toby Wiseman. “Gravity and large black holes in Randall-Sundrum II braneworlds”. In: *Phys. Rev. Lett.* 107 (2011), p. 081101. DOI: [10.1103/PhysRevLett.107.081101](#). arXiv: [1105.2558 \[hep-th\]](#).
- [385] Shohreh Abdolrahimi et al. “Large Randall-Sundrum II Black Holes”. In: *Phys. Lett. B* 720 (2013), pp. 405–409. DOI: [10.1016/j.physletb.2013.02.034](#). arXiv: [1206.0708 \[hep-th\]](#).
- [386] Kent Yagi, Norihiro Tanahashi, and Takahiro Tanaka. “Probing the size of extra dimension with gravitational wave astronomy”. In: *Phys. Rev.* D83 (2011), p. 084036. DOI: [10.1103/PhysRevD.83.084036](#). arXiv: [1101.4997 \[gr-qc\]](#).
- [387] Maya Fishbach, Reed Essick, and Daniel E. Holz. “Does Matter Matter? Using the mass distribution to distinguish neutron stars and black holes”. In: *Astrophys. J. Lett.* 899 (2020), p. L8. DOI: [10.3847/2041-8213/aba7b6](#). arXiv: [2006.13178 \[astro-ph.HE\]](#).

- [388] Elias R. Most et al. “A lower bound on the maximum mass if the secondary in GW190814 was once a rapidly spinning neutron star”. In: *Mon. Not. Roy. Astron. Soc.* 499.1 (2020), pp. L82–L86. DOI: [10.1093/mnrasl/slaa168](https://doi.org/10.1093/mnrasl/slaa168). arXiv: [2006.14601](https://arxiv.org/abs/2006.14601) [[astro-ph.HE](#)].
- [389] Tom Broadhurst, Jose M. Diego, and George F. Smoot. “Interpreting LIGO/Virgo ”Mass-Gap” events as lensed Neutron Star-Black Hole binaries”. June 2020. arXiv: [2006.13219](https://arxiv.org/abs/2006.13219) [[astro-ph.CO](#)].
- [390] Christopher J. Moore et al. “Testing general relativity with gravitational-wave catalogs: the insidious nature of waveform systematics”. Mar. 2021. arXiv: [2103.16486](https://arxiv.org/abs/2103.16486) [[gr-qc](#)].