

Physics and Astrophysics with Gravitational
Waves from Compact Binary Coalescence in
Ground Based Interferometers

by

Katherine L. Grover

A thesis submitted to the University of Birmingham for the degree of
DOCTOR OF PHILOSOPHY

School of Physics and Astronomy

College of Engineering and Physical Sciences

University of Birmingham

October 2015

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

Advanced ground based laser interferometer gravitational wave detectors are due to come online in late 2015 and are expected to make the first direct detections of gravitational waves, with compact binary coalescence widely regarded as one of the most promising sources for detection.

In Chapter I I compare two techniques for predicting the uncertainty of sky localisation of these sources with full Bayesian inference. I find that timing triangulation alone tends to over-estimate the uncertainty and that average predictions can be brought to better agreement by the inclusion of phase consistency information in timing-triangulation techniques.

Gravitational wave signals will provide a testing ground for the strong field dynamics of GR. Bayesian data analysis pipelines are being developed to test GR in this new regime, as presented in Chapter 3 Appendix B. In Chapter II and Appendix C I compare the predicted from of the Bayes factor, presented by Cornish et al. [55] and Vallisneri [181], with full Bayesian inference. I find that the approximate scheme predicts exact results with good accuracy above fitting factors of ~ 0.9 .

The expected rate of detection of Compact Binary Coalescence signals has large associated uncertainties due to unknown merger rates. The tool presented in Chapter III provides a way to estimate the expected rate of specified CBC systems in a selected detector.

Acknowledgements

I would like to acknowledge and thank the following people without whom this work would not have been possible.

My co-authors and collaborators, as referenced throughout, for for letting me tag along. Alberto Vecchio and Ilya Mandel, for their guidance and patience. Will Farr and Stas Babak for their corrections to this work. Sally Schofield for her understanding during the times surrounding my hand in and viva dates. The whole of the Gravitational Wave Group at Birmingham past and present, for all their help in research and helping me find my career outside of it. David Stops, for fixing my computing issues. Andreas Freise, for his voice of reason. Ray Jones, for giving me a second chance and suggesting this Ph.D. Trevor Sidery, Carl Haster, Maggie Leiu, Charlotte Bond, Hannah Middleton and anyone else who made the trip to Starbucks, for helping me hold on to my sanity. Chiara Mingarelli, for all her encouragement, support, retail therapy and coffee. Daniel Elsby, for getting me through.

Contents

1	INTRODUCTION	1
1.1	Gravitational Waves	1
1.2	Sources	4
1.2.1	Compact Binary Coalescence	5
1.2.2	Bursts, Stochastic Background and Continuous Waves	7
1.3	Gravitational Wave Detectors	8
1.3.1	Laser Interferometers	8
1.3.2	Pulsar Timing Arrays	15
1.4	Data Analysis	16
1.4.1	Matched Filtering	17
1.4.2	Bayesian Inference	19
1.4.3	Fisher Matrix Approximation	23
1.5	Waveforms	24
1.6	Testing General Relativity	30
1.6.1	Binary Pulsars	31
1.6.2	Gravitational Waves	32
2	COMPARISON OF GRAVITATIONAL WAVE DETECTOR NETWORK SKY LOCALISATION APPROXIMATIONS	36

2.1	Introduction	38
2.2	Methods	41
2.2.1	Bayesian Parameter Estimation	42
2.3	Timing Triangulation Approximation	43
2.3.1	Fisher Information Matrix Approximation	45
2.4	Detector Network and Injections	47
2.5	Comparison of Methods	49
2.6	Timing vs Coherent Bayesian	52
2.7	Fisher Matrix vs Coherent Bayesian	56
2.8	Timing Triangulation with Phase Correction	57
2.9	Discussion	63
3	TESTING GENERAL RELATIVITY WITH COMPACT BINARY	
	COALESCENCES IN GROUND BASED DETECTORS	66
3.1	Introduction	67
3.2	Use of Bayesian Odds Ratio in Testing GR	68
3.2.1	Bayesian Model Selection	68
3.2.2	Towards a Generic Test of GR	69
3.2.3	Approximating the Odds Ratio	72
3.3	Detectability of non-GR Signals	73
3.4	Example Signal	76
3.5	Results	78
3.6	Discussion and Conclusions	79
4	ONLINE CALCULATOR FOR GRAVITATIONAL WAVE DETEC-	
	TION RATES	85
4.1	Introduction	86

4.2	Merger Rate	88
4.2.1	Redshifted merger rate	89
4.3	Detection Rate	90
4.3.1	Redshift	92
4.4	The Calculator	95
4.4.1	Inputs	95
4.4.2	Calculation	98
4.4.3	Include Redshift Effects	98
4.4.4	Outputs	99
4.5	Comparison to Published Results	108
5	FUTURE WORK	109
5.1	Sky Localisation	109
5.2	Testing General Relativity	110
5.3	Rates Calculator	110
5.3.1	Astrophysical Parameters	110
5.3.2	Detection Parameters	111
5.3.3	Calculation	111
5.3.4	Outputs	112
A	OBSERVING THE DYNAMICS OF SUPERMASSIVE BLACK HOLE BINARIES WITH PULSAR TIMING ARRAYS	113
A.1	Introduction	114
A.2	Signals from SMBHBs	117
A.3	Observations using the Earth-term only	119
A.4	Measuring SMBHB evolution using the Earth and pulsar term	122
A.5	Conclusions	124

B	TOWARDS A GENERIC TEST OF THE STRONG FIELD DYNAMICS OF GENERAL RELATIVITY USING COMPACT BINARY COALESCENCE	126
B.1	Introduction	128
B.2	Waveform model and its sensitivity to changes in phase coefficients . . .	134
B.2.1	Model waveform(s) and detector configuration	135
B.2.2	Changes in phase coefficients and detectability	137
B.3	Method	139
B.3.1	Bayesian inference and nested sampling	140
B.3.2	Basic method for a single source	143
B.3.3	The general case	151
B.3.4	Combining information from multiple sources	154
B.4	Results	156
B.4.1	Measurability of a deviation in a post-Newtonian coefficient . . .	158
B.4.2	Effect of number of hypotheses used	168
B.4.3	Measurability of deviations with non-PN frequency dependences	170
B.4.4	Parameter estimation	180
B.4.5	A note on parameter estimation and multiple sources	195
B.5	Conclusions and future directions	196
C	TESTING GENERAL RELATIVITY WITH COMPACT COALESCING BINARIES: COMPARING EXACT AND PREDICTIVE METHODS TO COMPUTE THE BAYES FACTOR	201
C.1	Introduction	202
C.2	Bayesian Inference for gravitational wave signals	205
C.2.1	Analytical Approximation	206

C.3	Comparison between the exact calculation and the analytical approximation	208
C.3.1	Scaling with the signal-to-noise ratio	211
C.3.2	Scaling with the fitting factor	212
C.3.3	Correcting the analytical expression for lower fitting factors	214
C.4	Discussion	218
C.5	Appendix: Computing Fitting Factors from logLikelihoods	221

List of Figures

1.1	Effect of h_+ and h_\times on a ring of free test masses	3
1.2	Michelson Interferometer	9
1.3	Advanced LIGO noise contributions	10
1.4	Broadband vs Burst Signal	13
1.5	Initial LIGO Sensitivities	14
1.6	Agreement of Binary Pulsar Data with GR	33
1.7	Parametrised tests of General Relativity using the double pulsar	34
2.1	Noise Power Spectral Density for LIGO S6	48
2.2	The chirp mass and coherent network signal-to-noise ratio of the set of sources being used to compare the sky localization methods.	50
2.3	Histograms of the estimated 50% credible-interval sky areas	51
2.4	Comparisons of 50% credible-interval sky areas	53
2.5	50% credible intervals for the a typical injection and for patchy Bayesian analysis	54
2.6	Histograms of Bayesian sky areas	55
2.7	Time and Phase Correlation	59
2.8	Fractional reduction in timing triangulation prediction for the sky localization area after incorporating phase consistency	61

3.1	Fitting factors for examples of generic deviation from GR	80
3.2	Shifting of mass parameters within the prior to absorb the mismatch caused by a deviation from the General Relativity template bank. . . .	81
3.3	Loss in detection rate due to mismatched template bank.	82
3.4	Bayes Factors calculated via lalinference and via approximation for a fixed fitting factor and varying signal to noise ratio.	83
3.5	Bayes Factors calculated via lalinference and via approximation for a fixed signal to noise ratio and varying fitting factor.	84
4.1	Rates Calculator Astrophysical Parameters Input	94
4.2	Rates Calculator Detection Parameters Input	96
4.3	Θ CDF	99
4.4	Rates Calculator Default Outputs	100
4.5	Rates Calculator Default Noise Amplitude Spectral Density	101
4.6	Rates Calculator Default Frequency Domain Waveforms	102
4.7	Rates Calculator Default Time Domain Waveforms	103
4.8	Rates Calculator Default Detections per Redshift	104
4.9	Rates Calculator Default SNR as a Function of Redshift	105
4.10	Distribution of detected SNRs	106
4.11	Merger rate as a function of redshift	107
A.1	The fraction of parameter space in $\hat{\Omega}$ and $\hat{\mathbf{p}}$ for which the maximum excursion of φ_p over the time $L_p(1 + \hat{\Omega} \cdot \hat{\mathbf{p}})$ for $L_p = 1$ kpc exceeds a certain value	125
B.1	The high-power, zero-detuning noise curve for Advanced LIGO, and the BNS-optimized Advanced Virgo noise curve.	137

B.2	The fitting factors for a range of \mathcal{M} , once the other parameters are maximized over	139
B.3	The fitting factors for a range of η , once the other parameters are maximized over	140
B.4	The log odds ratios for individual sources.	159
B.5	The normalized distribution $P(\ln O_{\text{GR}}^{\text{modGR}})$ of log odds ratios for individual and combined sources.	160
B.6	For a given SNR, the cumulative number of times that the Bayes factor against noise for a particular component hypothesis is the largest for injections with that SNR or below, for $\delta\chi_3 = 0.1$	162
B.7	The log odds ratios for individual sources.	163
B.8	The normalized distribution $P(\ln O_{\text{GR}}^{\text{modGR}})$ of log odds ratios for individual and combined sources	165
B.9	The build-up of cumulative Bayes factors against GR for individual hypotheses, and the odds ratio, for a typical catalog with $\delta\chi_3 = 0.025$. . .	166
B.10	The efficiency of detecting a GR violation for sources with $\delta\chi_3 = 0.025$, as a function of catalog size for false alarm probabilities $\beta \in \{0.05, 0.01\}$. . .	168
B.11	The change in false alarm probabilities (FAPs) in going from two to three testing parameters and the change in efficiencies when keeping the false alarm probabilities fixed.	171
B.12	The difference of false alarm probabilities and efficiencies, ${}^{(3)}\beta - {}^{(\text{target})}\beta$, for fixed log odds ratio thresholds and signals having $\delta\chi_3 = 0.025$, between our 3-parameter test and a ‘targeted search’ which only looks for a deviation in ψ_3	172
B.13	The log odds ratios for individual sources for injections with an anomalous frequency dependence in the phase.	174

B.14	The normalized distribution $P(\ln O_{\text{GR}}^{\text{modGR}})$ of log odds ratios for individual and combined sources, where the injections are either GR or have the anomalous frequency dependence.	175
B.15	The cumulative number of times that the Bayes factor against noise for a particular component hypothesis is the largest	177
B.16	A couple of examples of how cumulative Bayes factors against GR for individual hypotheses, and the odds ratio, grow as sources with increasing SNR are being added within three different catalogs of 15 sources in total.	178
B.17	A few examples of how cumulative Bayes factors against GR for individual hypotheses, and the odds ratio, grow as sources with increasing SNR are being added within three different catalogs of 15 sources in total.	179
B.18	Posterior PDFs for a single GR injection with network SNR of 23.0	183
B.19	Posterior PDFs for a single GR injection with network SNR of 23.0	184
B.20	Posterior PDFs for a single injection with $\delta\chi_3 = 0.1$ and network SNR 23.2.	186
B.21	Posterior PDFs for a single injection with $\delta\chi_3 = 0.1$ and network SNR 23.2.	187
B.22	The posterior PDF for the injection with $\delta\chi_3 = 0.1$ as in the previous figure, but recovered with waveforms where $\{\vec{\theta}, \delta\chi_1, \delta\chi_2, \delta\chi_3\}$ are all free.	187
B.23	The 68% and 95% confidence contours of the two-dimensional PDF for $\{\delta\chi_2, \delta\chi_3\}$, still for an injection with $\delta\chi_3 = 0.1$, and with a waveform that has $\{\vec{\theta}, \delta\chi_1, \delta\chi_2, \delta\chi_3\}$ as free parameters.	188
B.24	Posterior PDFs for a single injection with $\delta\chi_3 = 0.025$ and network SNR 20.6	190

B.25	The posterior PDFs for $\delta\chi_1$ (top), $\delta\chi_2$ (middle), and $\delta\chi_3$ (bottom) for a single injection with $\delta\chi_3 = 0.025$ and network SNR 20.6, recovered with waveforms where $\{\vec{\theta}, \delta\chi_3\}$ is free.	191
B.26	The posterior PDFs for $\delta\chi_1$ (top) and $\delta\chi_2$ (bottom) for a single injection with $\delta\chi_A = -2.2$ and network SNR 22.4, recovered with waveforms where, respectively, $\{\vec{\theta}, \delta\chi_1\}$ and $\{\vec{\theta}, \delta\chi_2\}$ are free.	193
B.27	The posterior PDFs for $\delta\chi_3$ (bottom) for a single injection with $\delta\chi_A = -2.2$ and network SNR 22.4, recovered with waveforms where, , $\{\vec{\theta}, \delta\chi_3\}$ are free.	194
B.28	The 68% and 95% confidence contours of the two-dimensional PDF for $\{\delta\chi_2, \delta\chi_3\}$, for an injection with $\delta\chi_A = -2.2$, and with a waveform that has $\{\vec{\theta}, \delta\chi_1, \delta\chi_2, \delta\chi_3\}$ as free parameters.	194
C.1	Two-dimensional logarithmic prior probability distribution in the \mathcal{M} and η space.	209
C.2	Logarithmic Bayes factor from <code>lalInference</code> (errorbars) and from Eq. (C.6) (empty circles) as a function of the SNR.	211
C.3	Logarithmic Bayes factor from <code>lalInference</code> (error bars) and from Eq. (C.6) (empty circles) as a function of the FF.	213
C.4	Logarithmic Bayes factor from <code>lalInference</code> (error bars), from Eq. (C.6) (empty circles), and from Eq. (C.21) (red dots) as a function of the FF.	217

List of Tables

1.1	The likely estimated rate of compact binary coalescences per Milky Way equivalent galaxy per mega-year and per mega-parsec per mega-year.	7
2.1	Credible-interval sky areas	50
2.2	Median areas of 50% Bayesian credible intervals for coherent Bayesian method, timing triangulation and time-phase localization	61
4.1	Comparison of rates calculator outputs to published results	108
A.1	Number of Gravitational Wave cycles for supermassive black hole binary signal	121

Conventions

Throughout this work geometric units are used where $G = c = 1$. G is Newton's Gravitational Constant and c is the speed of light. M_{\odot} is used to denote the mass of the sun $\approx 4.93 \times 10^{-6}$ s.

Unless stated otherwise, Einstein summation convention is used.

The Fourier transform and its inverse are used in the form,

$$\tilde{a}(f) = \int_{-\infty}^{\infty} a(t) e^{-2\pi i f t} dt, \quad (1)$$

$$a(t) = \int_{-\infty}^{\infty} \tilde{a}(f) e^{+2\pi i f t} df. \quad (2)$$

Chapter 1

INTRODUCTION

1.1 Gravitational Waves

Gravitational waves (GWs) – time varying oscillations in the space time metric – are a direct consequence of Einstein’s theory of General Relativity, see for example [132]. Far from a GW source one can use a linearised approximation, considering only the leading order perturbations to the flat Minkowski metric,

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \quad |h_{\mu\nu}| \ll 1, \quad (1.1)$$

where μ and $\nu = 0, 1, 2, 3$, $g_{\mu\nu}$ is the metric and $\eta_{\mu\nu}$ is the flat Minkowski metric,

$$\eta_{\mu\nu} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (1.2)$$

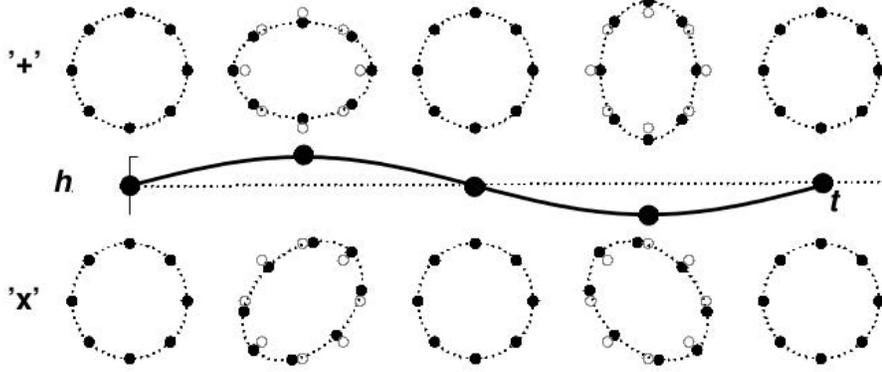
$h_{\mu\nu}$ is the metric perturbation, whose typical non-zero components are assumed to be much less than 1 in order to apply an approximation which is valid to the order of linear terms in h . The Einstein equations can then be solved in this limit, the derivation is not reproduced here but can be found for example in [78]. The transverse-traceless (TT) gauge is chosen reducing the 10 degrees of freedom of the symmetric matrix $h_{\mu\nu}$. The solution is a wave equation with just two independent polarisations known as the plus (h_+) and cross (h_\times) polarisations,

$$h_{\mu\nu}(\mathbf{x}) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & h_+ & h_\times & 0 \\ 0 & h_\times & -h_+ & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \exp[ik(z - t)], \quad (1.3)$$

where the direction of propagation is in the z-direction. Imagining a ring of free particles in the x-y plane demonstrates the effects of the two polarizations as the wave progresses. The ring is tidally deformed as the GW passes, seen in Figure 1.1, the angle between the h_+ and h_\times polarisations is $\frac{\pi}{4}$, with the shape of the tidal deformation of the rings giving the polarisations their names.

At the lowest order the amplitude of a GW can be approximated by the quadrupole

Figure 1.1 The effect of h_+ and h_\times on a ring of free particles. (Source: [155])



formula,

$$h_{ij} = \frac{2}{D} \ddot{Q}_{ij}, \quad (1.4)$$

where i, j and $k = 1, 2, 3$. D is the distance to the source and dots represent derivatives with respect to time. Q_{ij} is the traceless mass quadrupole moment which is dependant on the time varying mass-energy density of the source $\rho(\mathbf{x}, t)$,

$$Q_{ij} \equiv I_{ij} - \frac{1}{3} \delta_{ij} I_{kk}, \quad (1.5)$$

$$I_{ij} \equiv \int \rho(\mathbf{x}, t) x_i x_j d\mathbf{x}. \quad (1.6)$$

At lowest order GWs are produced from asymmetric accelerated masses with luminosity given by,

$$L = \frac{1}{5} \langle \ddot{Q}_{jk} \ddot{Q}_{jk} \rangle, \quad (1.7)$$

where angled brackets denote the average over several wavelengths. The energy flux at large distances from the source is proportional to $\dot{h}_{ij} \dot{h}_{ij}$ and the factor of a fifth is

a consequence of integrating the power over the full solid angle in the TT gauge. To estimate the GW luminosity we see that,

$$L \sim \frac{1}{5} \dot{Q}^2 \sim \frac{1}{5} \left(\frac{MR^2}{T^3} \right)^2 \sim \frac{1}{5} \left(\frac{Mv^3}{R} \right)^2 . \quad (1.8)$$

Where the characteristic mass, size, timescale and velocity of the source are M , R , T and v respectively. Or in terms of the schwartzchild radius $R_s = 2M$,

$$L \sim \frac{1}{20} \left(\frac{R_s}{R} \right)^2 (v)^6 , \quad (1.9)$$

which is dimensionless in our chosen unit convention, but for comparison can be converted to standard units by multiplying by,

$$L_0 = \frac{c^5}{G} = 3.6 \times 10^{52} W . \quad (1.10)$$

The maximum luminosity is produced when $R_s \sim R$ and $v \sim c$, i.e. compact, fast moving objects. Higher mass systems evolve slower than low mass ones, producing more integrated luminosity over their life time.

1.2 Sources

As gravitational waves (GWs) interact very weakly with matter we can only hope to detect the most strongly emitting sources, massive objects accelerating in a strong gravitational field. This excludes detecting man-made or local sources of GWs and

limits possible detections to astrophysical sources. In this work I concentrate solely on compact binary coalescence (CBC) sources, primarily in the stellar mass range, as detected in ground based laser interferometers, in particular in the Laser Interferometer Gravitational-wave Observatory (LIGO) [11] and Virgo [13] detectors, see Section 1.3. In Appendix A I also consider GWs from super massive black hole binaries (SMBHBs) detected in pulsar timing arrays (PTAs).

1.2.1 Compact Binary Coalescence

GW signals from CBCs are one of the most promising sources for GW detection. CBC signals are the GWs emitted during the inspiral, merger and ring-down of either a binary black hole (BBH), binary neutron star (BNS) or neutron star - black hole (NSBH) pair. During the inspiral the binary loses energy and angular momentum via GWs causing the orbit to shrink, the objects to spiral in towards each other and eventually merge. After the merger of a black hole (BH) it will oscillate in its ring-down period to reach stability.

From Equation 1.8 we can estimate the GW luminosity from a binary system,

$$L_{CBC} \sim \frac{1}{5} \left(\frac{M}{R} \right)^5, \quad (1.11)$$

Where M is the total mass and R is the Keplerian orbital separation. This means that binary systems will emit at a fraction of L_0 and they will emit more strongly as the orbit shrinks. The minimum separation of the objects during the inspiral phase of the coalescence is known as the innermost stable circular orbit (ISCO). For a binary with non-spinning components the ISCO is approximated by assuming a point mass in a

Schwarzschild metric,

$$R_{ISCO} = 6M . \tag{1.12}$$

At leading order the GW frequency will be twice the orbital frequency. Using Kepler's third law gives an estimate for the maximum GW frequency of the inspiral at the ISCO,

$$f_{ISCO} \sim 4400 \text{ Hz} \frac{M_{\odot}}{M} . \tag{1.13}$$

Where as usual M_{\odot} denotes the solar mass.

BNSs are the only type of binary compact objects we have direct evidence for so far. A handful of BNS systems have been detected in which one or both of the neutron stars (NSs) are radio pulsars, e.g. [96] and [120]. By extrapolating the distribution and coalescence timescales from these objects an estimate of the rate of BNS coalescences can be made [102]. For the other types of binaries (BBH and NSBH binaries) population synthesis methods must be used to estimate the rates. These involve inputting a number of observed astrophysical parameters into models for stellar formation, binary formation and compact object evolution, e.g. [36]. The input parameters are mostly not well constrained and the models themselves may not account for all the relevant physics, leading to a large range of predicted merger rates. The most simple models for coalescence rates assume that the merger rates are proportional to the stellar birth rate which is in turn assumed to be proportional to the blue light luminosity of a galaxy (which is thought to be incorrect for older elliptical galaxies). A list of likely predicted merger rates is shown in Table, 1.1 and the details of the calculations that generated them can be found in [4].

Table 1.1 The likely estimated rate of compact binary coalescences per Milky Way equivalent galaxy per mega-year and per mega-parsec per mega-year (Source [4]). More predictions and details of the assumptions behind these rates can be found in [4].

Source	Rate MWEG ⁻¹ Myr ⁻¹	Rate Mpc ⁻³ Myr ⁻¹
BNS	100	1
NSBH	3	0.03
BBH	0.4	0.005

1.2.2 Bursts, Stochastic Background and Continuous Waves

Other promising sources of gravitational waves for early detections include bursts, stochastic backgrounds and continuous wave sources.

Burst sources are those which produce strong short-lived signals with sufficient power in the sensitive band of the interferometer to be detectable. One such predicted source is a stellar core collapse as found in a type II supernova [135]. The strength of the GW produced in such a collapse would depend strongly on its geometry, a spherically symmetric collapse would produce no GWs at all. Supernova simulations are currently unable to include all the relevant physics to model the phenomena and therefore accurate GW waveform models from this source do not exist. There may be many other possible burst sources caused by unpredicted violent astrophysical events, see [6] and references there in for details of sources and the LIGO search for them.

Stochastic backgrounds of GWs are predicted to be generated from a large number of discrete uncorrelated sources such as the superposition of many astrophysical emitters, e.g. spinning Neutron Stars [143] or from cosmological backgrounds like relic gravitational waves from the early universe [121]. Multiple interferometers are needed to differentiate a stochastic background from noise, and such searches have been carried out by LIGO and Virgo placing upper limits on this GW background [8]. Searches

are also underway for imprints of relic GWs on the Cosmic microwave background via electromagnetic observations, e.g. [43].

GW pulsars are a possible source of continuous waves. A mass deformation on a spherically symmetric spinning NS would produce a continuous source of GWs at a frequency twice that of the spin frequency for an extended period of time. The amplitude of these GWs would depend on the size of the mass deformation and is limited by the size which could be stably supported by the NS [136]. Details of recent LIGO searches for continuous waves can be found in [144].

1.3 Gravitational Wave Detectors

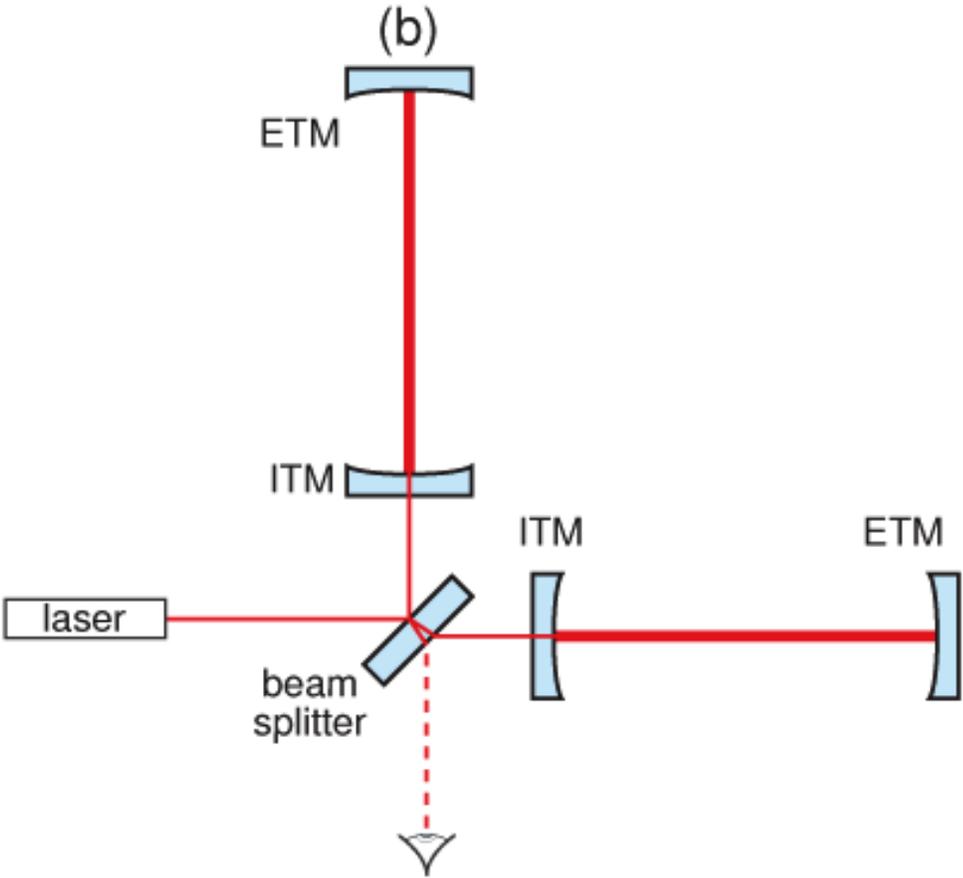
Gravitational waves (GWs) interact very weakly with matter making them difficult to detect. There are various methods of detection that have been proposed, some of which have been active for a number of years. These detection methods are characterised by their sensitive frequency range, which in turn sets the possible sources of GWs they can hope to detect. In this work I focus primarily on stellar mass compact binary coalescence (CBC) signals detected in ground based laser interferometers. In Appendix A I also consider GWs from super massive black hole binaries (SMBHBs) as detected by pulsar timing arrays (PTAs).

1.3.1 Laser Interferometers

Various ground based interferometers have completed their initial science runs and are now being upgraded for improved sensitivity alongside more interferometers in planning stages across the world. These detectors use laser interferometry to detect the

relative motion of freely suspended test masses. The ground based detectors are modified Michelson interferometers, with two perpendicular arms, with many additional features to increase sensitivity to GW signals while minimising noise. For example many detectors include Fabry-Perot cavities in both arms [79], seen in Figure 1.2.

Figure 1.2 Schematic of a basic Michelson Interferometer with Fabry-Perot cavities for GW detection (Source: [139]). The cavities are formed by the initial test mass (ITM) and end test mass (ETM) in each perpendicular arm.

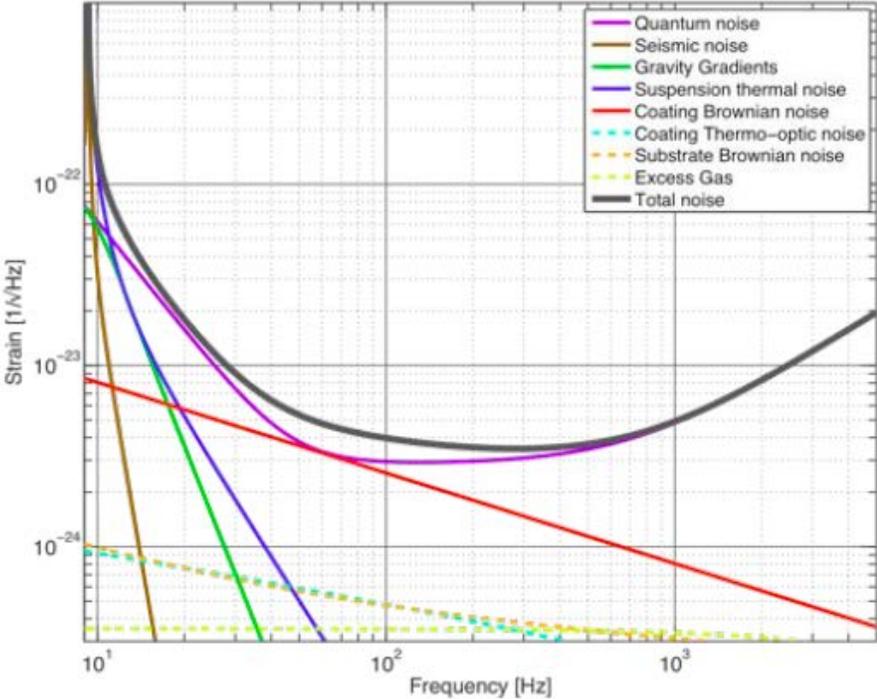


Noise in Ground Based Interferometers

Interferometers have a sensitive frequency band, the frequency range in which they can hope to detect GWs, set by various sources of noise in the detector. The noise can

never be fully eradicated but via various design and operational techniques the noise in a specific frequency band can be lowered to the point of allowing GWs detection. The main sources of noise in advanced detectors can be seen in Figure 1.3.

Figure 1.3 The anticipated main noise sources and the total noise curve for the advanced Laser Interferometer Gravitational-wave Observatory (aLIGO) detectors. (Source: [177])



Seismic noise in advanced ground based detectors limits sensitivity below about 10 Hz. Seismic noise is due to the seismic motion of the Earth which causes motion in the test masses of the interferometer. To minimise seismic noise the test masses can be suspended from a series of pendulums to dampen any horizontal motion and springs can be used to dampen vertical motion [139].

Gravity gradient noise is caused by local fluctuations in gravitational field strength effecting the motion of the test masses and is limiting around 10 Hz [70]. It is impossible

to screen the test masses from the local changes in gravity as they have the same physical effect as GWs. As the largest effect is from surface waves [95] possible future solutions to gravity gradient noise would be to build detectors underground, such as the KAGRA detector currently under construction in Japan [29]. Another possible solution is to use a monitor and subtract method in order to remove the noise electronically rather than mechanically [139].

Thermal noise is caused by the thermal vibrations of the test masses and their suspension, to combat this the test masses have minimised thermal coupling to the suspension, with resonances chosen outside of the sensitive band. It may also be possible to cryogenically cool future interferometers to reduce thermal noise, [67].

Radiation pressure noise is due to uncertainty in the amplitude of the laser field which can be reduced by using heavier mirrors or decreasing the laser power. However this would increase the shot noise, or photon counting noise, which dominates the high end of the frequency band. In order to reduce shot noise the light power in the interferometer can be increased by either increasing the power in the interferometer via various methods, such as multipass arms, or power recycling [208].

The levels of noise or the sensitivity of a given detector is usually represented by a power spectral density (PSD). The output of the detector, $d(t)$ is the linear sum of any GW signal, $h(t)$ and the noise, $n(t)$. Assuming the noise is stationary, let κ be defined as,

$$\kappa(\tau) \equiv \langle n(t_1)n(t_2) \rangle , \tag{1.14}$$

where $\tau = | t_1 - t_2 |$ and $\langle \dots \rangle$ denotes the ensemble average. Then the one-sided noise

PSD $S_n(f)$ is defined as zero for negative frequencies and for positive frequencies as,

$$S_n(f) = \frac{1}{2} \int_{-\infty}^{\infty} \kappa(\tau) e^{2\pi i f t} d\tau, \quad (1.15)$$

$$\langle \tilde{n}(f) \tilde{n}^*(f) \rangle = \frac{1}{2} S_n(f) \delta(f - f'), \quad (1.16)$$

where $S_n(f)$ has units of Hz^{-1} . Sometimes the noise amplitude spectral density is reported, as in Figure 1.5, which is simply the square root of the PSD. Both these functions measure the noise per linear frequency bin but to compare the noise to a broadband GW signal (e.g. figure 1.4) the power per logarithmic frequency bin is often used via the effective gravitational wave noise,

$$h_n(f) \equiv \sqrt{f S_n(f)}, \quad (1.17)$$

(note $S_n(f)$ has units of time) as this quantity can then be directly compared to some characteristic signal amplitude,

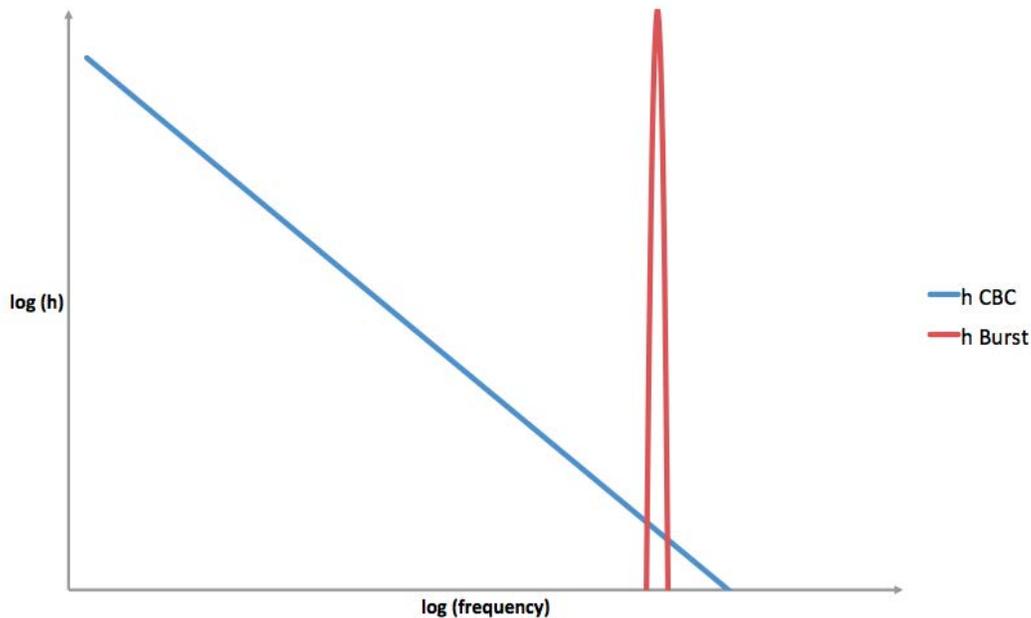
$$h_c \equiv f | \tilde{h}(f) |, \quad (1.18)$$

where $\tilde{h}(f)$ is the Fourier transform of $h(t)$.

The Existing Network of Interferometers

In contrast to electromagnetic telescopes, GW detectors are nearly omni-directional, they are sensitive to signals from across the whole sky. Therefore it is necessary to use a network of detectors in order to locate short lived GW signals via triangulation, while long lived (\sim months) signals may be triangulated via the Earth's motion. A network of

Figure 1.4 A CBC produces a broadband signal which sweeps the sensitive frequencies of the detector, meaning the integrated signal is of interest. Contrastingly a burst signal may be narrow band and focused in a small frequency band.

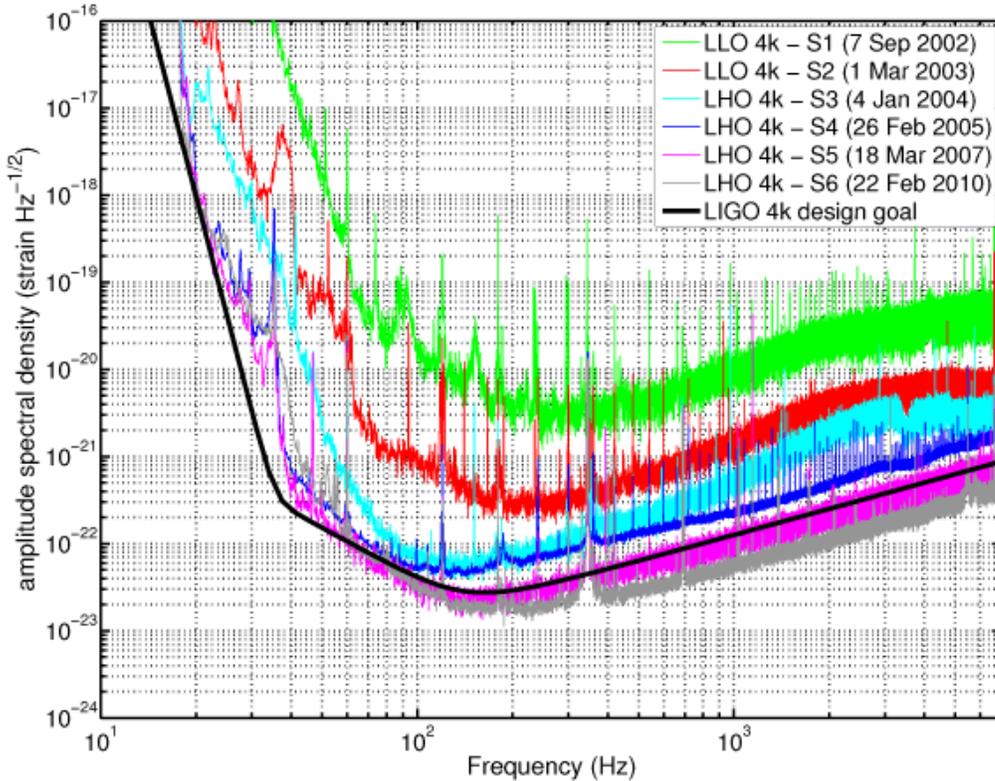


detectors also allows identification of false signals in the output of the detectors caused by local noise, as these “glitches” will not be coincident in multiple locations.

The LIGO Virgo Collaboration (LVC) is a worldwide collaboration of detectors and scientists aiming to build and utilise such a network of detectors. The LVC detectors are the Laser Interferometer Gravitational-wave Observatory (LIGO), GEO and Virgo instruments. LIGO [11] is a set of three detectors in the United States comprised of two detectors (H1 and H2) located at Hanford, Washington and a third (L1) at Livingston, Louisiana. H1 and L1 have arm lengths of 4 km while H2 has 2 km long arms. The first 6 periods of scientific data taking, science runs, using the LIGO detectors are known as initial and enhanced LIGO and took place between 2002 and 2010, see e.g. [10]. The sensitivities achieved in this time can be seen in Figure 1.5. Since 2010 LIGO has been undergoing upgrades to improve its sensitivity [1]. Science runs of this advanced LIGO

are scheduled to begin in September 2015. Additionally there are plans to relocate the H2 detector to India [150] to increase the baseline of the world wide network of detectors.

Figure 1.5 Sensitivities achieved by initial LIGO Hanford 4km (LHO 4k) and LIGO Livingston 4km (LLO 4k) interferometers. (Source: [139])



Virgo [13] is a similar interferometer located near Pisa, Italy with arm lengths of 3 km. Initial Virgo had three science runs between 2007 and 2010. Like LIGO, Virgo is also implementing advanced technologies to increase sensitivities [179]. GEO600 [85] is a smaller detector located near Hannover, Germany with 600 m arms. GEO has a different geometry to the other detectors, using folded arms instead of Fabry-Perot cavities [208]. GEO has been a testing ground for the new technologies now being implemented in the advanced LIGO upgrades. Finally, KAGRA is an advanced detector currently under

construction in Japan. KAGRA has additional features such as cryogenic cooled mirrors and a seismically quiet location inside a large mountain. KAGRA construction began in 2010 and is scheduled to begin taking data in 2017 [29].

Looking to the future, planning has already been under way for the next generation of laser interferometers for a number of years. The proposed underground Einstein Telescope (ET) and space based Laser Interferometer Space Antenna (LISA) [60] could provide greatly improved sensitivities as well extending the frequency range of laser interferometer GW detectors.

1.3.2 Pulsar Timing Arrays

Complementary to interferometer detectors are PTAs, which aim to utilise the extremely stable time of arrivals of radio wave pulses emitted by millisecond pulsars. Any GW travelling between the observed pulsar and the Earth will cause a time varying shift in the time of arrival of the radio signal. By monitoring an array of such pulsars and cross correlating the shift in arrival times it is theoretically possible to detect a GW signal. The sensitive frequency band is $\sim 10^{-7} - 10^{-9}$ Hz for PTAs is determined by the length of the data set and the cadence of observations, i.e. how long the pulsars have been observed for and the frequency of such observations. In order to search for a signal in pulsar timing data, the other variations in timing which are not caused by GWs must be taken into account. This noise is modelled as a number of different contributions [111]. the uncertainty in the pulsar's astronomical and timing properties, such as it's spin period and it's possible spin noise. There is white noise caused by the radiometer and also jitter noise caused by averaging multiple pulses. The pulsar signal can be delayed via dispersion in the interstellar medium. PTAs are also insensitive to

signal at a frequency of 1/year due to solar system ephemeris errors.

The International Pulsar Timing Array (IPTA) [91] and constituting PTAs (European Pulsar Timing Array (EPTA) [77], North American Nanohertz Observatory for Gravitational Waves (NANOgrav) [99] and Parkes Pulsar Timing Array (PPTA) [90]) have published several upper limits of stochastic backgrounds of GWs in this band, see for example [162], [187], [65].

1.4 Data Analysis

Gravitational waves (GWs) interact very weakly with matter and therefore are extremely difficult to detect. Advanced instruments should be on the limit of sensitivity needed to detect these signals and the output of the interferometer will include any GW signals buried in detector noise. As such, sophisticated data analysis techniques must be employed in order to firstly detect possible GWs in the data and then to analyse the precise form of the signal in order to study the physics and astrophysics of the source. In this work I focus on the data analysis of compact binary coalescence (CBC) signals in ground based interferometers. CBC signals are promising sources in these detectors due to their expected rates and also because of the accuracy at which the form of the GW signal can be predicted, allowing the data analysis techniques described below to be performed.

The output of an interferometer, known as the data $d(t)$, is a linear combination of the detector noise $n(t)$ and any GW signal $h(t)$,

$$d(t) = n(t) + h(t) . \tag{1.19}$$

The data from a network of detectors may be combined using coincident or coherent methods. Coincident analysis processes data from each detector individually and then looks for coincident results with physically consistent parameters in multiple detectors. Whereas coherent analysis combines outputs from each detector into a single data set before analysis. In principle, coherent analysis is the correct method but is more computationally expensive [133] while coincident analysis can provide quick checks of false signals which are present only in one detector due to local noise. Therefore usually searches for GWs have used coincident analysis, e.g. [7], however where computationally feasible coherent searches have also been performed, e.g. [6].

For data analysis is it often useful to work in the frequency domain, for wide band signals as it is assumed that the noise is uncorrelated in frequency.

$$\tilde{d}(f) = \tilde{n}(f) + \tilde{h}(f). \quad (1.20)$$

For the purpose of data analysis the noise is usually assumed to be stationary and Gaussian with zero mean and variance described by the single sided noise power spectral density (PSD) S_n as defined in Equation 1.15. It is useful to define a noise weighted inner product,

$$\langle a | b \rangle \equiv 2 \int_0^\infty \frac{\tilde{a}(f)\tilde{b}^*(f) + \tilde{a}^*(f)\tilde{b}(f)}{S_n(f)} df = 4\Re \left[\int_0^\infty \frac{\tilde{a}(f)\tilde{b}^*(f)}{S_n(f)} df \right]. \quad (1.21)$$

1.4.1 Matched Filtering

The general form of a CBC signal has been approximated via various method (see Section 1.5, however the precise form will depend on the unknown source parameters.

To find CBC signals in the noisy output of the detector a matched, or Wiener, filter technique is used, which is well known to be the optimal filter to maximise the signal to noise ratio for a known form of signal buried in noise, see e.g. [89]. As the precise form is unknown a range of filters is created to span the source parameter space, known as a template bank. Filtered potential signals that pass some threshold signal to noise ratio (SNR) can then be analysed further in order to ensure they are true signals and not noise, see [22] for details.

In order to optimally filter the data the correct form of the GW signal must be used, which is not known a priori. If a filter proportional to the exact form of h is used an optimal SNR, ρ_{opt} , is recovered,

$$\rho_{\text{opt}} = \langle h | h \rangle^{1/2} . \quad (1.22)$$

While if some suboptimal filter u is used the recovered SNR ρ would be,

$$\rho = \frac{\langle u | h \rangle}{\sqrt{\langle u | u \rangle}} . \quad (1.23)$$

This method relies on creating a template bank of possible signals to use as filters, which is densely distributed enough in parameter space and created from accurate enough waveforms in order to ensure that there is not significant drop in optimal SNR while still being computationally effective. This potential loss of SNR due to inaccurate waveforms or sparsely populated parameter space can be quantified via the maximised match or fitting factor.

Let h_T be the true waveform which depends on a set of parameters, $\underline{\theta}$. The output of the detector is $d(t) = n(t) + h_T(\underline{\theta}_T)$, where again the subscript T denotes the true

values or function. The template bank is constructed from an approximate waveform h_A which is evaluated at some value of the parameters $\underline{\theta}$. The match M is defined as the fractional loss in SNR from using the filter of the form $h_A(\underline{\theta})$,

$$M = \frac{\rho}{\rho_{opt}} = \frac{\langle h_A(\underline{\theta}) | h_T(\underline{\theta}_T) \rangle}{\sqrt{\langle h_A(\underline{\theta}) | h_A(\underline{\theta}) \rangle \langle h_T(\underline{\theta}_T) | h_T(\underline{\theta}_T) \rangle}}. \quad (1.24)$$

The maximised match or fitting factor FF is the match maximised over all the values of $\underline{\theta}$ that are included in the template bank.

$$FF = \left[\frac{\langle h_A(\underline{\theta}) | h_T(\underline{\theta}_T) \rangle}{\sqrt{\langle h_A(\underline{\theta}) | h_A(\underline{\theta}) \rangle \langle h_T(\underline{\theta}_T) | h_T(\underline{\theta}_T) \rangle}} \right]_{\text{maximised over } \underline{\theta}}. \quad (1.25)$$

Note this maximum will not necessarily occur at $\underline{\theta} = \underline{\theta}_T$ due to the use of an approximate waveform.

1.4.2 Bayesian Inference

Bayesian analysis provides a robust framework for hypothesis testing and parameter estimation and as such is used extensively in GW data analysis. Bayes theorem allows us to calculate the probability of some hypothesis \mathcal{H}_i given some data D and any relevant prior information I regarding the hypothesis and the data,

$$P(\mathcal{H}_i | D, I) = \frac{P(\mathcal{H}_i | I) P(D | \mathcal{H}_i, I)}{P(D | I)}. \quad (1.26)$$

$P(\mathcal{H}_i | \text{I})$ is the prior probability of the hypothesis, $P(D | \mathcal{H}_i, \text{I})$ is the likelihood of the hypothesis and $P(D | \text{I})$ is a normalisation such that,

$$P(D | \text{I}) = \sum_i P(\mathcal{H}_i | \text{I}) P(D | \mathcal{H}_i, \text{I}) . \quad (1.27)$$

Model Selection

To use Bayes theorem to compare two models, each of which could possibly describe the data, we assign the hypothesis \mathcal{H}_i to some data model M_i which depends on some parameters $\underline{\theta}$. As the hypothesis depends on some free parameters we must marginalise over these to provide the marginalised likelihood or evidence Z ,

$$Z = P(D | M_i) = \int p(\underline{\theta} | M_i, \text{I}) p(D | M_i, \underline{\theta}, \text{I}) d\underline{\theta} . \quad (1.28)$$

The normalisation factor in Equation 1.26 is only calculable if there is a complete set of known models for which $\sum_i P(M_i | D, \text{I}) = 1$, otherwise we use the odds ratio $O_{i,j}$ to compare models where this quantity vanishes.

$$O_{i,j} = \frac{P(M_i | D, \text{I})}{P(M_j | D, \text{I})} = \frac{P(M_i | \text{I})}{P(M_j | \text{I})} B_{i,j} , \quad (1.29)$$

$$B_{i,j} = \frac{P(D | M_i, \text{I})}{P(D | M_j, \text{I})} , \quad (1.30)$$

where $B_{i,j}$ is the ratio of the evidences and is known as the Bayes factor. The odds ratio can be used to compare models such as $d = n$ vs. $d = n + h$. In this case both models are valid even if there is no signal present (in the second case the model parameters would be such that $h = 0$). However Bayes theorem has a built in ‘Occam’s

razor', penalising extra parameters in the hypothesis which aren't justified by the data, meaning the noise only model would be preferred in the absence of a GW signal.

Parameter Estimation

In order to use detected CBC signals to study astrophysics, Bayes theorem is used as a parameter estimation technique. Although in principle there is no distinction between the two idealised results of matched filtering and parameter estimation (i.e. the best filter uses the true parameters of system) the computational techniques used in one are currently not efficient for the other. Additionally in the presence of noise the optimal filter may be shifted from the source parameters.

In this case the model describing the data is assumed to be true (i.e. is contained within I) and the hypothesis space is the value of the parameters $\underline{\theta}$ that describe that model. Equation 1.26 becomes,

$$p(\underline{\theta} | D, I) = \frac{p(\underline{\theta} | I)p(D | \underline{\theta}, I)}{p(D | I)}, \quad (1.31)$$

where $p(\underline{\theta} | D, I)$ is the posterior density function (PDF) of the parameters given the data. $p(\underline{\theta} | I)$ is the prior, $p(D | \underline{\theta}, I)$ is the likelihood and $p(D | I)$ is the evidence, as in Equation 1.28 where in this case M_i is included in I . The marginalised PDF on a subset of parameters such that $\underline{\theta} = \{\underline{\theta}_A, \underline{\theta}_B\}$ is,

$$p(\underline{\theta}_A | D, \mathcal{H}, I) = \int p(\underline{\theta} | D, \mathcal{H}, I) d\underline{\theta}_B. \quad (1.32)$$

This allows the PDF of a single parameter (or any combination of parameters) to be computed.

In the case of a CBC GW signal in Gaussian stationary noise the likelihood is:

$$\mathcal{L} = \mathcal{N} \exp \left[-\frac{\langle n | n \rangle}{2} \right] = \mathcal{N} \exp \left[-\frac{\langle d - h(\boldsymbol{\theta}) | d - h(\boldsymbol{\theta}) \rangle}{2} \right], \quad (1.33)$$

where \mathcal{N} is the normalisation constant.

Bayesian Codes

Bayesian Inference for CBC model selection or parameter estimation relies on accurately exploring a highly dimensional parameter space, 9 dimensions in the case of CBC signals with non-spinning components and 15 for spinning ones. This integral is very computationally expensive so statistical sampling techniques are used. Multi-dimensional integrals are not unique to the field of GW data analysis so existing algorithms are used which are then tailored and further developed specifically for this problem in order to optimise the computation. Even so, full parameter estimation computations can take days, weeks or even months to complete.

There are two main methods used by the CBC Parameter Estimation group of the LIGO Virgo Collaboration (LVC) and are implemented in the LIGO Scientific Collaboration (LSC) algorithm library (LAL) [119], they are called Nested Sampling and Markov chain Monte Carlo (MCMC). Both codes use sampling methods to concentrate samples in the part of the parameter space which has the highest contribution to the evidence. Nested Sampling is an algorithm specifically designed to compute the Bayesian Evidence, with the by-product of producing parameter PDFs. It was created by Skilling [167] [168] and developed for use in GW data analysis by Veitch and Vecchio [192] [193]. Nested Sampling works by evolving a set of N samples (or live points), which are initially randomly distributed in parameter space, to higher likelihood values at each iteration,

zooming in on the peak(s) of the PDF and performing the sum of the evidence as it goes. The PDF can then be reconstructed via the density of samples at each iteration of the algorithm, see [193] for details. In contrast the MCMC algorithm is designed to produce the PDF and can produce the evidence as a by-product of the computation. MCMC methods have a long history of being used in a wide variety of fields including a number of Bayesian applications , e.g. [82], and has been used in GW parameter estimation for a number of years, e.g. [147] [186]. The MCMC algorithm uses a weighted random walk to draw samples from the parameter space such that the probability of being in a region of parameter space is proportional to the posterior density for that region. Hence, after some termination condition is met, the density of samples is directly proportional to the PDF.

1.4.3 Fisher Matrix Approximation

Full Bayesian parameter estimation is computationally costly and often a faster approximation for the accuracy to which one can recover the parameters of the GW source is desirable. The fisher matrix has been used extensively for this purpose, see for example [56] although it has many limitations some which are only recently being investigated [145] [125].

The Fisher Information Matrix Γ_{ij} is defined as,

$$\Gamma_{ij} = \left\langle -\frac{\partial}{\partial\theta_i} \frac{\partial}{\partial\theta_j} \log \mathcal{L}(d | \boldsymbol{\theta}) \right\rangle \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_T}. \quad (1.34)$$

Where $\langle \dots \rangle$ denotes averaging over noise realisations. In the limit of high SNR, equivalently the linear signal approximation (LSA) see [180], the inverse of the Fisher matrix

approximates the covariance matrix of the PDF such that,

$$\sigma_i = \sqrt{\Sigma^{ij}} = \sqrt{(\Gamma^{-1})^{ij}}, \quad (1.35)$$

$$\text{covar}(\theta^i, \theta^j) = \frac{\Sigma^{ij}}{\sqrt{\Sigma^{ii}\Sigma^{jj}}}. \quad (1.36)$$

For this likelihood,

$$\Gamma_{ij} = \langle h_i | h_j \rangle. \quad (1.37)$$

1.5 Waveforms

As discussed above, compact binary coalescence (CBC) data analysis depends on using accurate waveforms of the gravitational wave (GW) signal that can be produced across a range of source parameters at a rate which allows computation of a template bank or sample of likelihoods. These are possible to produce as CBC systems are relatively clean sources, meaning the signal depends on a small number of parameters. The parameters which control the form of the signal received at the Earth can be classified as intrinsic or extrinsic to the source. The intrinsic parameters are the spins and the masses of the objects, often expressed as the Chirp Mass \mathcal{M} and symmetric mass ratio η ,

$$\eta = \frac{m_1 m_2}{M^2}, \quad (1.38)$$

$$\mathcal{M} = \eta^{3/5} M, \quad (1.39)$$

where m_1 and m_2 , are the binary's component masses and M is the total mass of the system. The extrinsic parameters are those which depend on the detector location and orientation relative to the source: sky position (right ascension α and declination δ), the luminosity distance of the source (D_L), the polarisation phase (ψ), the binary's in-

clination (ι) and some reference time (t_c) and phase (ϕ_c). This is a total of 9 parameters for binaries composed of non-spinning compact objects and 15 parameters for spinning objects. The eccentricity of the orbit has a negligible effect on the GW signal in the Laser Interferometer Gravitational-wave Observatory (LIGO) band. This is because the radiation reaction causes any eccentricity in the orbit to decrease [138] so that by the time the signal is in the sensitive band of the detector the orbit it can be assumed circular.

The waveforms produced by the complex two-body gravitationally bound system may be approximated using a variety of methods. Various numerical relativity simulations have been able to generate waveforms in recent years over \sim tens of orbits for a selected choice of extrinsic parameters as well as the merger of the objects, see [19] for details of a selection of these waveforms. As many of the data analysis techniques rely on using a large number of waveforms across the parameter space these numerical methods do not provide all the waveforms needed and other approximate techniques are used.

The post-Newtonian (PN) expansion is commonly used to approximate the inspiral signal generated by CBC [49]. It is a low velocity expansion whose expansion parameter is the velocity v of the reduced mass μ of the system. By using expansions in this parameter for the GW luminosity and energy of the system the energy balance equation can be used to compute the adiabatic expressions for orbital velocity and phase. These equations can then be solved in a number of ways leading to different families of waveforms, see [53] for a comparison. TaylorF2 is one such waveform family which is used extensively in this work, it uses a Taylor expansion and stationary phase approximation to generate a frequency domain waveform. Here I outline the derivation of the orbital parameters and TaylorF2 waveform at the leading (Newtonian) order based on the derivations described in [53], [56] and [122].

Let the orbit of the binary lie in the (x, y) plane such that,

$$x(t) = r \cos\left(\pi ft + \frac{\pi}{2}\right), \quad (1.40)$$

$$y(t) = r \sin\left(\pi ft + \frac{\pi}{2}\right), \quad (1.41)$$

$$z(t) = 0, \quad (1.42)$$

where r is the orbital separation and f is the gravitational wave frequency, which (at this lowest order) is twice the orbital frequency.

Then using the expressions for Q_{ij} given in Equations 1.5 and 1.6, in the centre of mass frame, gives the mass quadrupole and its traceless second and third time derivatives,

$$I_{ij} = \mu x^i x^j, \quad (1.43)$$

$$I_{11} = \mu r^2 \left(\frac{1 - \cos(2\pi ft)}{2} \right), \quad (1.44)$$

$$I_{22} = \mu r^2 \left(\frac{1 + \cos(2\pi ft)}{2} \right), \quad (1.45)$$

$$I_{12} = I_{21} = -\mu r^2 \frac{\sin(2\pi ft)}{2}, \quad (1.46)$$

$$I_{i3} = I_{3i} = 0, \quad (1.47)$$

where $i = 1, 2, 3$.

$$\ddot{Q}_{11} = -\ddot{Q}_{22} = 2\mu r^2 \pi^2 f^2 \cos(2\pi ft), \quad (1.48)$$

$$\ddot{Q}_{12} = \ddot{Q}_{21} = 2\mu r^2 \pi^2 f^2 \sin(2\pi ft). \quad (1.49)$$

$$\dddot{Q}_{11} = -\dddot{Q}_{22} = -4\mu r^2 \pi^3 f^3 \sin(2\pi ft), \quad (1.50)$$

$$\dddot{Q}_{12} = \dddot{Q}_{21} = 4\mu r^2 \pi^3 f^3 \cos(2\pi ft). \quad (1.51)$$

Using the equation for quadrupole luminosity given in Equation 1.7,

$$\mathcal{L} = \frac{32}{5} \mu^2 r^4 \pi^6 f^6 . \quad (1.52)$$

Using the quadrupole formula, Equation 1.4,

$$h = \frac{4\mu r^2 \pi^2 f^2}{D_L} \begin{bmatrix} \cos(2\pi ft) & \sin(2\pi ft) & 0 \\ \sin(2\pi ft) & -\cos(2\pi ft) & 0 \\ 0 & 0 & 0 \end{bmatrix} . \quad (1.53)$$

In order to express this as a simple time-domain waveform we assume an adiabatic inspiral such that,

$$\mathcal{L} = -\frac{dE}{dt} , \quad (1.54)$$

where E is, at leading order, the sum of potential and kinetic energies,

$$E = -\frac{m_1 m_2}{r} + \frac{1}{2} \mu v^2 . \quad (1.55)$$

Assuming a circular Keplerian orbit,

$$v = \sqrt{\frac{M}{r}} = (\pi M f)^{1/3} . \quad (1.56)$$

We can now write the equation of motion of the binary as,

$$\frac{dr}{dt} = \frac{dE}{dt} \frac{dr}{dE} = -\mathcal{L} \frac{dr}{dE} . \quad (1.57)$$

$$\frac{dr}{dt} = -\frac{64 m_1 m_2 M}{5 r^3} . \quad (1.58)$$

Letting $r \rightarrow 0$ as $t \rightarrow t_c$,

$$r(t) = \left(\frac{256}{5} m_1 m_2 M \right)^{1/4} (t_c - t)^{1/4} . \quad (1.59)$$

$$(1.60)$$

And equivalently,

$$f(t) = \frac{\mathcal{M}^{-5/8}}{\pi} \left[\frac{256}{5} (t_c - t) \right]^{-3/8} . \quad (1.61)$$

The phase of the wave is,

$$\phi = \int 2\pi f dt . \quad (1.62)$$

Integrating and using constant of integration such that as $t \rightarrow t_c$ $\phi \rightarrow \phi_c$ gives,

$$\phi(t) = -2 \left[\frac{1}{5\mathcal{M}} (t - t_c) \right]^{5/8} + \phi_c . \quad (1.63)$$

Leading to,

$$h_+ = \frac{4\eta M^2}{D_L r} \cos(\phi(t)) , \quad (1.64)$$

$$h_\times = \frac{4\eta M^2}{D_L r} \sin(\phi(t)) . \quad (1.65)$$

To approximate the Fourier transform we use the well known stationary phase approx-

imation which states that given a function of the form,

$$B(t) = A(t) \cos \phi(t), \quad (1.66)$$

the estimate of the fourier transform is given by

$$\tilde{B}(f) \approx \frac{1}{2} A(t) \left(\frac{df}{dt} \right)^{-1/2} e^{i(2\pi ft - \phi(f) - \pi/4)} \quad (1.67)$$

where t is defined as the time when $\frac{d\phi(t)}{dt} = 2\pi f$ and $\phi(f) = \phi(f(t))$. The following conditions must be met,

$$\frac{1}{A} \frac{dA}{dt} \ll 2\pi f, \quad (1.68)$$

$$\frac{1}{f} \frac{df}{dt} \ll 2\pi f. \quad (1.69)$$

I.e. this is valid for CBC systems within the LIGO band which have many cycles at a slowly changing amplitude and frequency. Therefore we can approximate $\tilde{h}(f)$ as,

$$\tilde{h}(f) = \frac{1}{D_L} \mathcal{M}^{5/6} f^{-7/6} e^{i\Psi(f)}, \quad (1.70)$$

$$\Psi(f) = 2\pi f t_c - \phi_c - \frac{\pi}{4} + \frac{3}{128} (\pi \mathcal{M} f)^{-5/3}. \quad (1.71)$$

This is the quadrupole Newtonian approximation to the waveform, PN corrections and higher harmonics add corrections to both the amplitude and phase of the wave, which can be expanded in terms of the classical velocity of the reduced mass.

The corrections to the phase, which have been calculated to 3.5 PN order ($\mathcal{O}(v^7)$), are vital for detection and parameter estimation, see section 1.4. The 3.5 PN TaylorF2

phase for a non-spinning objects is given by (e.g. [26])

$$\Psi(f) = 2\pi f t_c - \phi_c - \frac{\pi}{4} + \frac{3}{128} (\pi \mathcal{M} f)^{-5/3} \sum_{k=0}^N \alpha_k (\pi M f)^{k/3}, \quad (1.72)$$

$$\alpha_0 = 1, \quad (1.73)$$

$$\alpha_1 = 0, \quad (1.74)$$

$$\alpha_2 = \frac{20}{9} \left(\frac{743}{336} + \frac{11}{4} \eta \right), \quad (1.75)$$

$$\alpha_3 = -16\pi, \quad (1.76)$$

$$\alpha_4 = 10 \left(\frac{3058673}{1016064} + \frac{5429}{1008} \eta + \frac{617}{144} \eta^2 \right), \quad (1.77)$$

$$\alpha_5 = \pi \left(\frac{38645}{756} + \frac{38645}{252} \log \left(\frac{v}{v_{ISCO}} \right) - \frac{65}{9} \eta \left[1 + 3 \log \left(\frac{v}{v_{ISCO}} \right) \right] \right), \quad (1.78)$$

$$\begin{aligned} \alpha_6 = & \left(\frac{11583231236531}{4694215680} - \frac{640 \pi^2}{3} - \frac{6848 \gamma}{21} \right) \\ & + \eta \left(-\frac{15335597827}{3048192} + \frac{2255 \pi^2}{12} - \frac{1760 \theta}{3} + \frac{12320 \lambda}{9} \right) \\ & + \frac{76055}{1728} \eta^2 - \frac{127825}{1296} \eta^3 - \frac{6848}{21} \log(4v), \end{aligned} \quad (1.79)$$

$$\alpha_7 = \pi \left(\frac{77096675}{254016} + \frac{378515}{1512} \eta - \frac{74045}{756} \eta^2 \right). \quad (1.80)$$

Where $\lambda = -\frac{1987}{3080}$, $\theta = -\frac{11831}{9240}$ and γ is the Euler-Mascheroni constant = 0.577... The presence of spins, as expected to be common if a black hole (BH) is one of the compact objects, causes the orbit to precess, introducing a phase and amplitude modulation.

1.6 Testing General Relativity

Einstein's theory of General Relativity (GR) has passed every test it has undergone so far, see [207] and references therein for comprehensive review of these tests. The direct detection of gravitational waves (GWs) from compact binary coalescence (CBC)

sources will provide a new testing ground for GR. Tests of GR can be classified as weak or strong field tests, the strength of gravity can be classified via the Newtonian gravitational field strength calculated from the total mass M of a system and its linear size r ,

$$\phi = \frac{M}{r}. \quad (1.81)$$

For example for a gravitationally bound system with reduced mass μ and typical velocity v , the kinetic forces must be of order of the gravitational forces,

$$\frac{1}{2} \frac{\mu M}{r} \sim \frac{1}{2} \mu v^2, \quad (1.82)$$

$$\phi \sim v^2. \quad (1.83)$$

GR is relatively well tested in the weak field regime within our solar system, e.g. [42], where $\phi \ll 1$, for example the Newtonian gravitational field strength on the surface of the Sun $\phi_{\odot} = M_{\odot}/r_{\odot} \approx (4.93 \times 10^{-6} s) / (2.32 s) \approx 2.12 \times 10^{-6}$. In order to more fully test GR it must be tested against observations from strong field systems also.

1.6.1 Binary Pulsars

The most stringent tests of GR so far have come from the observations of binary pulsars which are much more sensitive to relativistic effects than any solar system tests, $v \sim 10^{-3}$, they have a short orbital period allowing effects to rapidly build up and they are relatively clean systems allowing relativistic effects to be accurately determined [207]. The first pulsar in a binary system to be discovered was PSR B1913+16, also known as the Hulse-Taylor pulsar after its discoverers [96]. The binary is composed of a radio pulsar in orbit around a quiet companion with an orbital period of ~ 7.75 hours and

binary separation $\sim 1R_{\odot}$ leading to the conclusion that the companion must be a compact object. The pulsar is very stable with a pulse period of $P \sim 59ms$ and no glitches. GR predicts the orbit of the binary to be decaying in a predictable way due to loss of energy and angular momentum via GWs. This decay is dependant on the masses and eccentricity of the system which can be determined via other observables. Therefore the predicted shift in periastron due to the orbits decay can be calculated and compared to the observed shift as a test of GR. This agreement in observation and prediction over 30 years of observations of the pulsar is seen in Figure 1.6.

Since the discovery of the Hulse Taylor pulsar a handful of other pulsars in binary systems have been discovered, see [118] and references therein. The most stringent tests of GR come from the only known double pulsar system PSR J0737-3039, in which both objects have been detected via pulsar emissions [120]. As each of the relativistic observational parameters of the binary have a precise dependence on the masses of the system predicted by GR these parameters are used as a self consistency check of the theory. seen in graphical form in Figure 1.7.

1.6.2 Gravitational Waves

When CBC GW signals are detected in advanced interferometers they will be the strongest field systems to be observed, with $\phi \sim \frac{1}{6}$. If GR is the incorrect theory of gravity there are various effects which could possibly be observed in the inspiral signal, especially if those theories effect the orbital evolution of the binary as ground based interferometers are very sensitive to phase evolution of the signal. Rather than investigate the exact form of GWs from all possible theories of gravity, which would each involve complex analytical solutions which are not even complete for the most

Figure 1.6 The Periastron advance of the Binary Pulsar PSR B1913+16. The points are data (with error bars) and the curve is the prediction of GR (Source: [201])

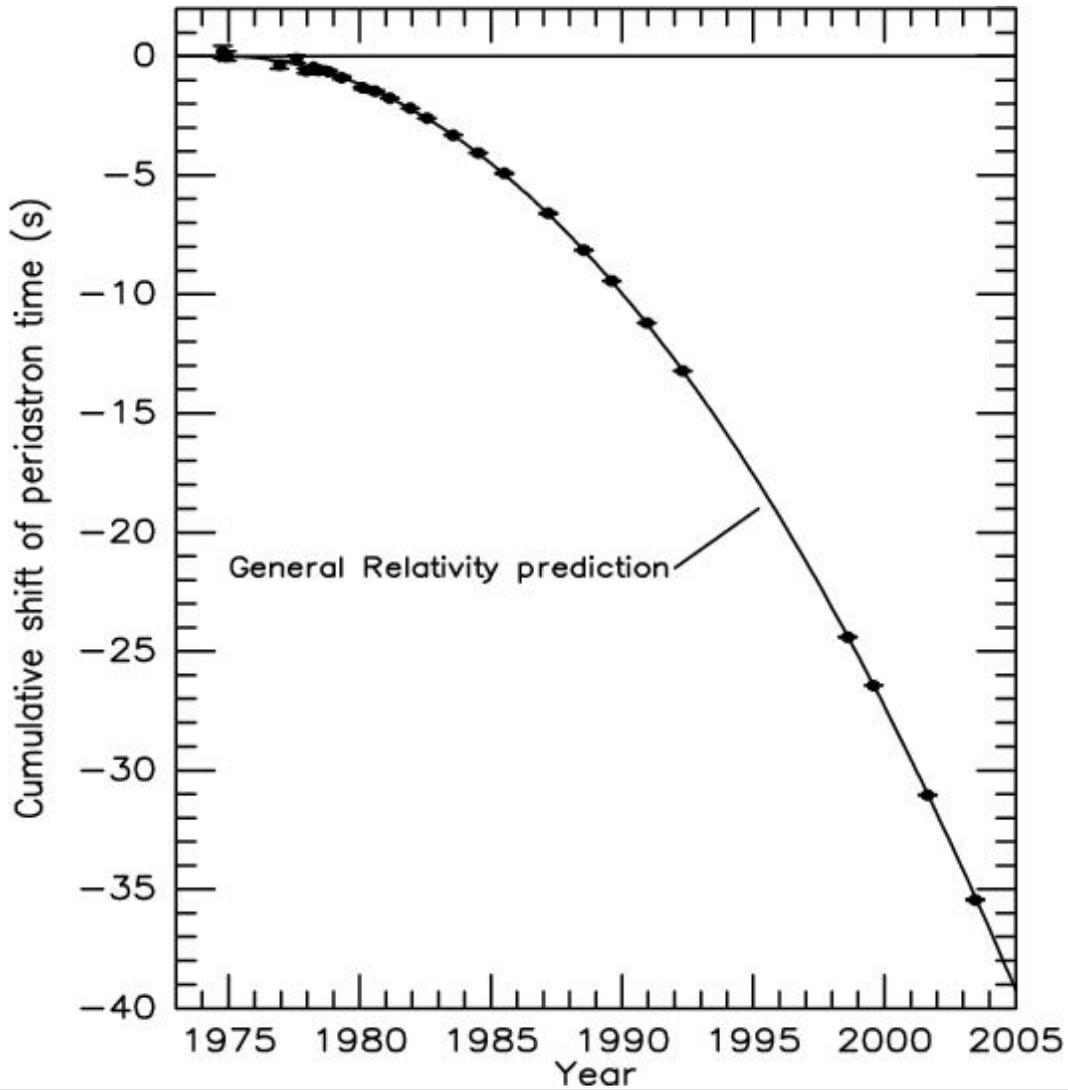
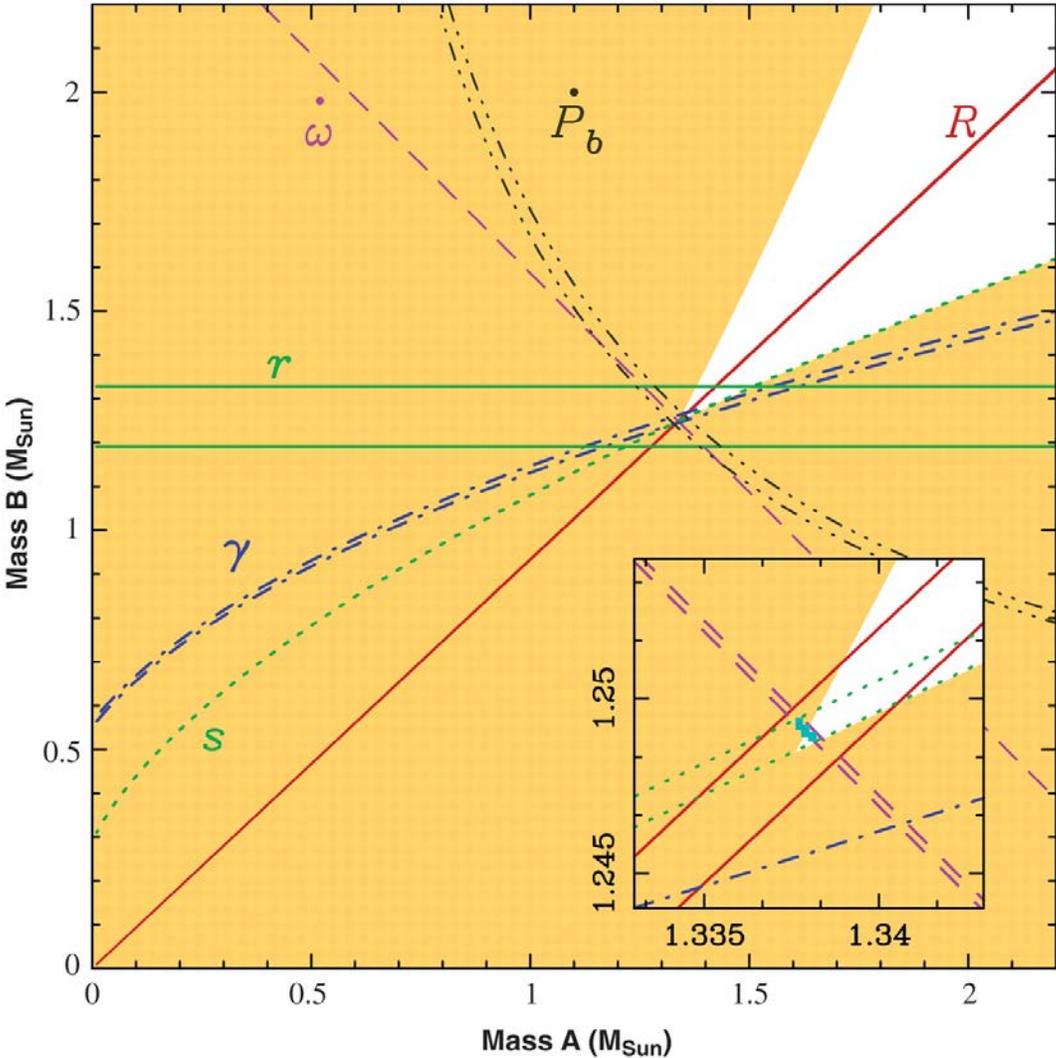


Figure 1.7 Parametrised tests of General Relativity using the observed post-Keplerian parameters of the double pulsar. The shaded area is the parameter space forbidden via mass functions which ensure \sin of the inclination of the orbit is less than 1. Other post-Keplerian parameters and their respective errors are represented as pairs of lines (some only distinguishable in the zoomed panel). All of the mass functions are seen to agree within a small error area, marked in blue. R is the mass ratio, $\dot{\omega}$ describes the relativistic precession of the orbit, γ describes the relativistic redshift and time dilation, \dot{P}_b describes the decrease in orbital period dues to GW emission, s and r describe the Shapiro time delay. (Source: [109])



respected theory of gravity, it is possible instead to focus on general phenomenological changes to GW signals and parameterisation of signals in such a way to test a range of alternative theories. This topic is discussed further in Chapter 3, Appendix B and Appendix C.

The rest of this work is arranged as follows. Chapter 2 is a study on the predictive methods for the accuracy of sky localisation of compact binary coalescence (CBC) signals in an advanced detector network, as published in [86]. Chapter 3 presents the first steps in studying data analysis of non-General Relativity (GR) gravitational wave (GW) signal, the work being included in published works [62] and [112], both found in the appendix. Chapter 4 presents an online tool for estimating the rates of CBC signals in advanced detectors. The appendices include the papers mentioned above as well as the paper published in [130] which studies the accuracy to which the dynamics of a super massive black hole binary (SMBHB) could be recovered using a pulsar timing array (PTA).

Chapter 2

COMPARISON OF GRAVITATIONAL WAVE DETECTOR NETWORK SKY LOCALISATION APPROXIMATIONS

The work in this chapter is presented in the form of a paper published in Physical Review D [86].

Authorship of paper: K. Grover, S. Fairhurst, B. F. Farr, I. Mandel, C. Rodriguez, T. Sidery, and A. Vecchio. K. Grover wrote sections 2.1 - 2.7 (with edits by I. Mandel and S. Fairhurst) including all plots. The full Bayesian PDFs and Fisher Matrices were produced by B. F. Farr and C. Rodriguez which K. Grover selected and post processed

to obtain comparable sky areas, the timing triangulation areas were calculated using a code written by S. Fairhurst and modified by K. Grover. K. Grover was part of discussions for sections 2.8 and 2.9 but did not produce the plots and K. Grover edited the text which was written by I. Mandel and S. Fairhurst.

Gravitational waves emitted during compact binary coalescences are a promising source for gravitational-wave detector networks. The accuracy with which the location of the source on the sky can be inferred from gravitational wave data is a limiting factor for several potential scientific goals of gravitational-wave astronomy, including multi-messenger observations. Various methods have been used to estimate the ability of a proposed network to localize sources. Here we compare two techniques for predicting the uncertainty of sky localization – timing triangulation and the Fisher information matrix approximations – with Bayesian inference on the full, coherent data set.

We find that timing triangulation alone tends to over-estimate the uncertainty in sky localization by a median factor of 4 for a set of signals from non-spinning compact object binaries ranging up to a total mass of $20M_{\odot}$, and the over-estimation increases with the mass of the system. We find that average predictions can be brought to better agreement by the inclusion of phase consistency information in timing-triangulation techniques. However, even after corrections, these techniques can yield significantly different results to the full analysis on specific mock signals. Thus, while the approximate techniques may be useful in providing rapid, large scale estimates of network localization capability, the fully coherent Bayesian analysis gives more robust results for individual signals, particularly in the presence of detector noise.

2.1 Introduction

Neutron star (NS) and black hole (BH) binaries will be an important source of gravitational waves (GWs) detectable by advanced ground-based laser interferometers that are expected to come online in 2015, the two LIGO detectors in the US and the Virgo detector in Italy [87,179]. These instruments are sensitive to the final minutes to seconds of a binary coalescence and the detection rate is anticipated to be between 0.4 yr^{-1} and 400 yr^{-1} at final design sensitivity [4]. This new class of observations is anticipated to provide us with new insights into the formation and evolution of this class of relativistic objects and their environments e.g. [102,112,126].

An important element in any astronomical observation is the ability to locate a source in the sky. Compact binaries that are detected via GWs provide a radically new sample of compact object binaries in a highly relativistic regime, which is otherwise difficult to identify. If some of the binding energy is released in the electromagnetic band as a prompt burst of radiation during the merger or as an afterglow, as is expected in the case of NS-NS and NS-BH coalescences, signatures in the gamma-ray, X-ray, optical, and radio spectra may be identified by telescopes, and will provide a multi-wavelength view of these phenomena e.g. [129]. If, on the other hand, the merger is electro-magnetically silent, as expected for a binary black hole merger in vacuum, constraining the source location in the sky may provide unprecedented clues about the environments that harbor such exotic objects.

Locating a GW source as precisely, and rapidly, as is technically possible is therefore an important element of GW observations. GW laser interferometers are not pointing telescopes, but the sky location can be reconstructed through the time of arrival of

GW radiation at the different detector sites, i.e. *timing triangulation*, as well as the relative amplitude and phase of the GWs in different detectors, that carry additional information about the source geometry. During the final observational run (S6/VSR2-3) of the LIGO and Virgo instruments, in their so-called “initial” configuration, a rapid sky localization algorithm based in part on timing triangulation was implemented and used to provide alerts on the time-scale of minutes to telescopes for follow-ups of detection candidates (none of which represented a confident GW detection) [9]. In parallel, a Bayesian inference analysis designed to provide accurate estimates of all the source parameters, including location in the sky, was deployed on a range of software and hardware injections, including the “blind injection” [5]. Now that the instruments are undergoing upgrades [87, 179] a major effort is taking place to refine these sky localization techniques in preparation for observations at advanced sensitivity, and to estimate the pointing performance of the GW network in order to put in place an electro-magnetic (EM) follow-up observational strategy.

When computing the typical size of the error-box in the sky, one needs to draw a careful distinction between actual parameter estimation on noisy data from a detector network, and predictions of the expected parameter-recovery accuracy in principle. Ultimately, one would like to know the probability that the source is enclosed within a given sky area in a real analysis. This information is provided by the posterior probability density function of the sky location given the data and all the assumptions of the model. For a multi-dimensional and possibly multi-modal parameter space, as is the case of models that describe GWs from coalescing binaries, this function can be efficiently obtained using a variety of stochastic sampling techniques. Markov-chain Monte Carlo [186] and nested sampling [193] have both been applied to inference on GW data; in fact, a sophisticated software library specifically developed for this goal is now in place and actively developed [119]. Several assessments of the localization

capability of the trans-continental network of GW instruments in different observational scenarios have recently been made using full, coherent parameter estimation techniques e.g. [104, 134, 189].

However, the exploration of a large-dimensional parameter space can be computationally intensive and lengthy. Other methods have therefore been devised to estimate theoretically the uncertainty of the sky location recovery without going through the actual analysis process.

One technique that has been used in countless studies is based on the evaluation of the Fisher information matrix (FIM), whose inverse provides an estimate of the covariance matrix on the parameter space [56]. However, the FIM provides robust estimates only for a sufficiently high signal-to-noise ratio to validate the approximation of the likelihood as a Gaussian. It is now clear that it is of limited use to describe the actual performance of the network of ground-based laser interferometers in realistic conditions, see [180] and most recently [145].

A different approach that has been considered specifically for predicting the sky localization error-box is based on using time delays between the arrival of GWs at different detectors in the network and theoretical estimates of time-of-arrival measurement uncertainty in individual detectors [74, 202]. A predictive method using solely timing accuracy is therefore based on the assumption of independent rather than coherent analyses of data streams from different detectors, and potentially misses out on additional information from a coherent analysis, such as the relative phase and amplitude of the GW strain between detectors, that could improve localization accuracy. This timing triangulation approach [75] was used as the basis of the recent LIGO Scientific Collaboration and Virgo Collaboration document on prospects for localization of GW transients with advanced observatories [115].

The main question that we want to address in this paper is the following: how accurate are the predictions based solely on triangulation estimates with respect to what could be actually achieved in a real coherent analysis of a stretch of data in which a GW candidate has been identified? We answer this question by considering a set of synthetic gravitational-wave signals from coalescing compact binaries added to Gaussian and stationary noise representative of the sensitivity achieved in the last LIGO – Virgo science run. We then examine and compare the predicted sky localization uncertainty estimates obtained with different techniques for this common set of simulated sources. Our study measures the accuracy of estimates from timing triangulation and Fisher information matrix analyses relative to the optimal coherent Bayesian inference, which defines what is achievable in practice with a given data set and model.

The paper is organized as follows. Section 2.2 outlines the methods used to calculate the sky areas. Section III applies these methods to a set of simulated GW signals. We compare the three methods in Section IV, and conclude with a discussion in Section V.

2.2 Methods

The data, $d_j(t)$, from each GW detector $j = 1, \dots, N$, where N is the number of instruments, is a sum of the noise $n_j(t)$ and any signal $h(t; \boldsymbol{\theta})$,

$$d_j(t) = n_j(t) + h_j(t; \boldsymbol{\theta}) , \tag{2.1}$$

where $\boldsymbol{\theta}$ is a vector that describes the set of unknown parameters that characterize the emitting source. The waveforms for coalescing binary systems $h(\boldsymbol{\theta})$ used in this study are discussed in Section 2.4. The noise in each detector is assumed to be a

zero-mean, stationary, Gaussian process characterized by a one-sided noise spectral density $S_{n_j}(|f|)$. The signal, $h(t; \boldsymbol{\theta})$, from compact binary coalescences (CBCs) with non-spinning components can be described with a nine-dimensional parameter vector $\boldsymbol{\theta}$: the two component masses m_1 and m_2 (or, alternatively, the symmetric mass ratio $\eta = m_1 m_2 / (m_1 + m_2)^2$ and the chirp mass $\mathcal{M}_c = \eta^{3/5} (m_1 + m_2)$), the distance to the source D , the source location in the sky – right ascension α and declination δ , the orientation of the binary – polarization ψ and inclination ι , and the reference phase ϕ_0 and time t_0 . It is useful to define the noise-weighted inner product between two functions a and b as

$$\langle a|b \rangle = 2 \int_0^\infty df \frac{\tilde{a}^*(f)\tilde{b}(f) + \tilde{a}(f)\tilde{b}^*(f)}{S_n(f)}, \quad (2.2)$$

where the integral is formally over all positive frequencies, but in practice is restricted, through $S_n(f)$, to the finite bandwidth over which the instrument is sensitive. Through the rest of the paper we use the notation $\langle a|b \rangle = \sum_j \langle a_j|b_j \rangle$ in order to describe the data analysis of N detectors with uncorrelated noise.

2.2.1 Bayesian Parameter Estimation

The parameter-estimation pipeline of the LIGO – Virgo collaboration uses a Bayesian framework to coherently analyze data from all detectors in the network [5]. Bayes' law gives the posterior probability density function $p(\boldsymbol{\theta} | d)$ in terms of the likelihood $p(d | \boldsymbol{\theta})$ and prior $p(\boldsymbol{\theta})$:

$$p(\boldsymbol{\theta} | d) = \frac{p(\boldsymbol{\theta})p(d | \boldsymbol{\theta})}{p(d)}, \quad (2.3)$$

where the denominator $p(d)$ is the evidence, which we treat as a normalization factor. In stationary, Gaussian noise, the likelihood is

$$p(d | \boldsymbol{\theta}) = \exp \left[-\frac{1}{2} \langle d - h(\boldsymbol{\theta}) | d - h(\boldsymbol{\theta}) \rangle \right] . \quad (2.4)$$

The Bayesian pipeline maps out the posterior probability density function $p(\boldsymbol{\theta} | d)$ of the signal parameters given the data and model, and thus directly provides the statistical measurement uncertainty on parameter estimates. The sampling algorithm used in this paper is the LAL [119] implementation of MCMC [186], known as `LALinference_mcmc`. It explores the parameter space using a random walk based on Metropolis-Hastings sampling methods [82]. The code returns samples from the posterior which are binned using a k D-tree algorithm [165] to compute the relevant two-dimensional probability intervals on the sky marginalized over all other parameters. For the comparisons reported below, we considered the smallest area which encloses 50% of the total posterior probability A_{Bayes} , also known as the 50% credible interval. The performance of this pipeline has been demonstrated on a number of simulated events [5]. Sky localization with this pipeline, among other parameters, was investigated for binary neutron star sources at high signal-to-noise ratios (SNRs) in [146], and the sky localization performance was validated on the same set of simulated sources as here in [164], where the self-consistency of credible intervals was tested.

2.3 Timing Triangulation Approximation

The Bayesian methods described above are computationally expensive; therefore, approximate techniques capable of analyzing a large set of synthetic GW sources and

predicting their localization accuracies are desirable. In particular, a timing triangulation (TT) approximation, outlined in [74] and [75], has been used to predict the sky localization ability of the upcoming advanced LIGO-Virgo network [115]. TT assumes that the bulk of information that allows for source localization on the sky is contained in the relative time delays of the arrival of the GW train at different detectors in the network, and so uses estimates of time-of-arrival uncertainties in individual detectors to predict the overall localization uncertainty [74]. The timing uncertainty in each detector σ_t is estimated to be:

$$\sigma_t = \frac{1}{2\pi\rho\sigma_f}, \quad (2.5)$$

where ρ is the expectation value of the SNR in the detector $\rho^2 = \langle h | h \rangle$ and the effective bandwidth of the signal in the detector σ_f is

$$\sigma_f^2 = \overline{f^2} - (\overline{f})^2, \quad (2.6)$$

where

$$\overline{f^n} = \frac{1}{\rho^2} \langle f^n h | h \rangle. \quad (2.7)$$

Considering only timing information, the likelihood for the sky location in the high-SNR regime is given by

$$p(\mathbf{r}|\mathbf{R}) = \exp \left[-\frac{1}{2}(\mathbf{r} - \mathbf{R})^T \mathbf{M}(\mathbf{r} - \mathbf{R}) \right] \quad (2.8)$$

where \mathbf{R} and \mathbf{r} are unit vectors describing the true location and recovered locations of the source, respectively. The matrix \mathbf{M} describes the network localization accuracy (whose inverse is the covariance matrix on the sky) and is given by

$$\mathbf{M} = \frac{1}{\sum_k \sigma_{t_k}^{-2}} \sum_{i,j} \frac{\mathbf{D}_{ij} \mathbf{D}_{ij}^T}{2\sigma_{t_i}^2 \sigma_{t_j}^2}, \quad (2.9)$$

where i, j, k label the detectors. \mathbf{M} is a function of the timing uncertainty σ_{t_i} in each detector and the pairwise separation vectors of the detectors in the network \mathbf{D}_{ij} expressed in light seconds.

We use the sky area as defined for the three-detector network in equation (33) of [74] to calculate the $P = 50\%$ credible interval,

$$A_{\text{TT}}(P) = -2\pi\sigma_x\sigma_y \frac{\ln(1-P)}{\cos\gamma}, \quad (2.10)$$

where γ is the angle between the direction to the source and the normal to the plane containing the three detectors and σ_x and σ_y are the inverse square roots of the two non-zero eigenvalues of the matrix \mathbf{M} .

2.3.1 Fisher Information Matrix Approximation

The calculation of the timing uncertainty, Eq. (2.5), in the TT method described above is in essence a two-dimensional (time and phase) application of the Fisher information Matrix (FIM) technique. The FIM is an approximation with a long history of use in GW data analysis e.g. [56]. In the high SNR limit, the likelihood over the desired credible region can be approximated as a Gaussian centered on the true parameters:

$$p(d|\boldsymbol{\theta}) \propto \exp \left[-\frac{1}{2} \langle h_a | h_b \rangle \Delta\theta^a \Delta\theta^b \right], \quad (2.11)$$

where $\Delta\theta_a = \theta_a - \theta_a^{\text{true}}$, $h_a = \frac{\partial h}{\partial \theta^a}$ and repeated indices are summed over.

In this case the Fisher information matrix is

$$\Gamma_{ab} \equiv \left\langle \frac{\partial h}{\partial \theta^a} \middle| \frac{\partial h}{\partial \theta^b} \right\rangle, \quad (2.12)$$

where h is the gravitational-wave signal and the inner product is defined in Eq. (2.2). The FIM approximation has been found to be useful for large-scale investigations into sky localization of detector networks, although individual sky areas estimated via FIM can be orders of magnitude different from full coherent analyses for some particular cases [145]. Assuming flat priors, the covariance matrix is then

$$\langle \Delta\theta^a \Delta\theta^b \rangle = (\Gamma^{-1})^{ab} \equiv \Sigma^{ab}. \quad (2.13)$$

We use the solid angle sky area as defined in equation (43) of [34] to calculate the $P = 50\%$ credible interval,

$$A_F(P) = -2\pi \cos(\delta^{\text{true}}) [\ln(1 - P)] \sqrt{\Sigma^{\alpha\alpha} \Sigma^{\delta\delta} - (\Sigma^{\alpha\delta})^2}, \quad (2.14)$$

where α and δ label the right ascension and declination respectively, δ^{true} is the declination of the source and here indices are not summed over. Below, we describe two different FIM analyses. In the first, the full 9-dimensional parameter space described at the beginning of section 2.2 is considered. The 9-dimensional FIM and its inverse, the 9-dimensional covariance matrix, are used to estimate the area A_{F9} via Eq. (2.14).

As an interesting comparison we also include FIM results based on a reduced four-dimensional analysis incorporating only the two sky location parameters of the source (the right ascension and declination) as well as the reference phase and time of merger, with the remaining parameters held fixed, i.e., assumed to be known perfectly. Such

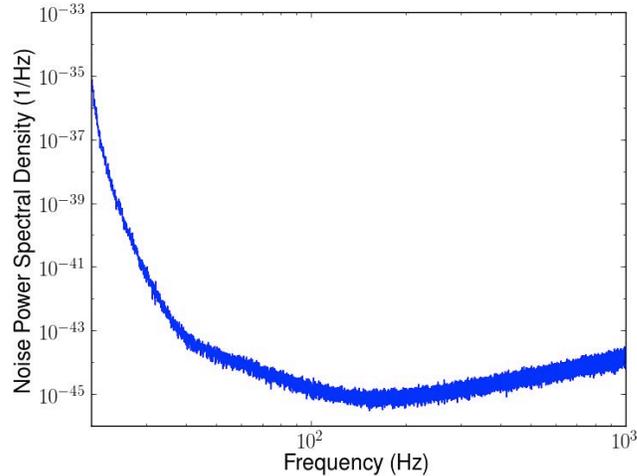
an analysis artificially restricts the parameter space and assumes perfect knowledge of parameters which are correlated with sky location. The 50% credible level sky localization areas computed from the four-dimensional FIM approximation are reported as A_{F4} .

2.4 Detector Network and Injections

To compare the methods described in the previous section we consider a set of mock observations made with the LIGO – Virgo detector network. The network is composed of three detectors located at the LIGO Hanford (Washington state, USA) LIGO Livingston (Louisiana, USA) and Virgo (near Pisa, Italy) sites. We model the noise in each instrument as zero-mean, Gaussian and stationary with the same one-sided power spectral density $S_{n_j} = S_n(|f|)$ for each instrument, $j = 1, 2, 3$. We use the noise power spectral density typical of the sensitivity achieved during the last science run (LIGO S6) [178]; Figure 2.1 shows the noise power spectral density.

The signal model $h(t; \boldsymbol{\theta})$ describes the inspiral phase of a low-mass compact-object binary in circular orbit. The waveform is generated in the time domain and is modeled using the restricted post-Newtonian (pN) approximation, with amplitude at the leading Newtonian order and phase at the 3.5 pN order. The compact objects are assumed to be non-spinning and the specific waveform approximant used for injection and Bayesian recovery is TaylorT4 [53]. The waveform used in the FIM analysis is the frequency-domain TaylorF2 approximant, which uses the restricted post-Newtonian approximation with amplitude at the leading Newtonian order and phase at the 2pN order [53]. The difference in waveform family will not have a significant effect on sky localization for the sources considered here, i.e., binaries with non-spinning components detected by

Figure 2.1 The one-sided noise power spectral density $S_n(|f|)$ as estimated from the mock data. The same noise spectral density, which is representative of the noise during the LIGO S6 operational period, was used to generate noise realizations in each of the three instruments (LIGO Hanford, LIGO Livingston and Virgo) that make up the detector network used in the study.



a network of instruments with identical noise spectra. Given the noise spectral density adopted for this study, the waveforms are generated from a starting frequency of 40 Hz and are terminated at the usual condition for these approximants, when the waveform reaches the innermost stable circular orbit frequency, $f_{\text{ISCO}} = 1/[\mathfrak{G}^{3/2}\pi(m_1 + m_2)]$. In all cases, we neglect the merger and ringdown parts of the waveform even though, for the higher mass systems, they would contribute to the signal power and to the source localization.

The set of signals added to the Gaussian noise was generated with parameters drawn randomly from the following distribution, which was chosen to test the techniques and does not represent expected astrophysical populations: the distance is distributed linearly in $\log(D)$ in the range 10 – 40 Mpc, the sky location and orientation are isotropic on the respective two-spheres, and the component masses (m_1, m_2) are drawn uniformly from $1 M_\odot \leq m_{1,2} \leq 15 M_\odot$ with an additional cut on the total mass at

$m_1 + m_2 \leq 20 M_\odot$. These injection distributions are also used as the priors in the Bayesian analysis. A total of 200 injections were generated. However, in this study we consider only 166 of the 200, based on the following two cuts. We crudely emulate the detection pipeline by considering only signals with $\rho \geq 5$ in at least two detectors. We also consider only signals whose sky position was at least 5 degrees away from the plane defined by the detector network in order to use the simple TT approximation in Eq. (2.10). The combination of distance and masses of the sources selected for the analysis leads to a distribution of chirp mass and total network SNR $\rho_{\text{network}} = \sqrt{\sum_j \rho_j^2}$ that is shown in Figure 2.2.

Sky localization is largely determined by the frequency content of the signals. For this set of injections, the mean frequency \bar{f} varies from 150 Hz for the lowest mass system to 120 Hz for the highest, with a median value of 130 Hz, while the bandwidth σ_f varies from 90 Hz at low mass to 40 Hz at high mass, with a median of 60 Hz.

2.5 Comparison of Methods

For each of the 166 signals described in the previous section, we computed the associated sky area at the 50% credible level (CL) using each of the four methods: coherent Bayesian analysis (A_{Bayes}), timing triangulation (A_{TT}), 9-dimensional Fisher information matrix (A_{F9}) and 4-dimensional Fisher information matrix (A_{F4}), as described in Section 2.2. Figures 2.3(a)-2.3(d) and Table 2.1 summarize the results.

Figure 2.2 The chirp mass and coherent network signal-to-noise ratio of the set of sources being used to compare the sky localization methods.

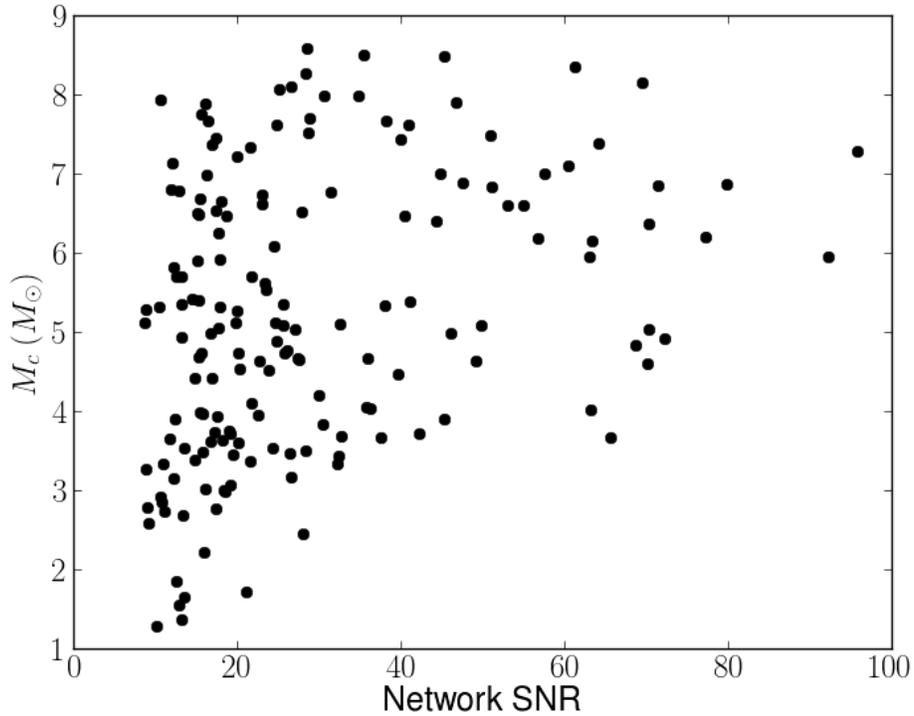
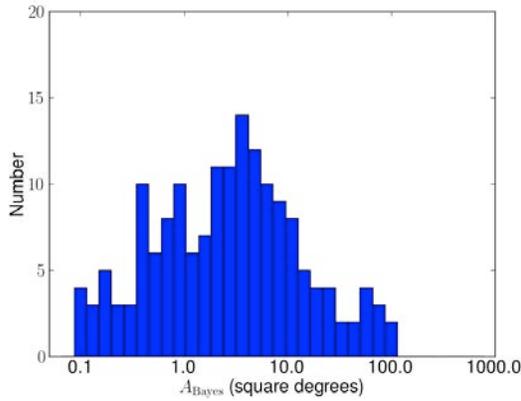


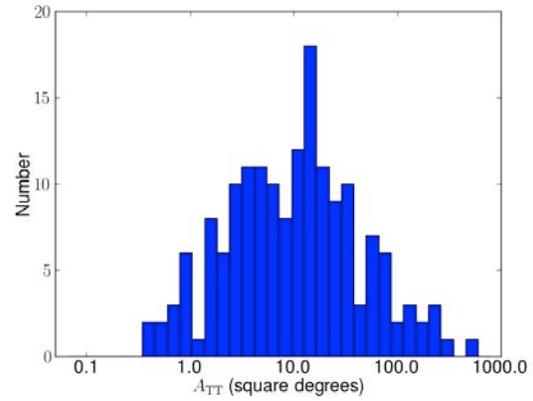
Table 2.1 Average 50% credible-interval sky areas, in square degrees, computed using the four methods in Section 2.2.

Method	Median (sq. deg.)	Mean (sq. deg.)	Standard Dev. (sq. deg.)
A_{Bayes}	2.9	8.9	17.1
A_{TT}	10.6	29.3	59.3
A_{F9}	4.0	8.7	16.0
A_{F4}	1.6	3.2	6.1

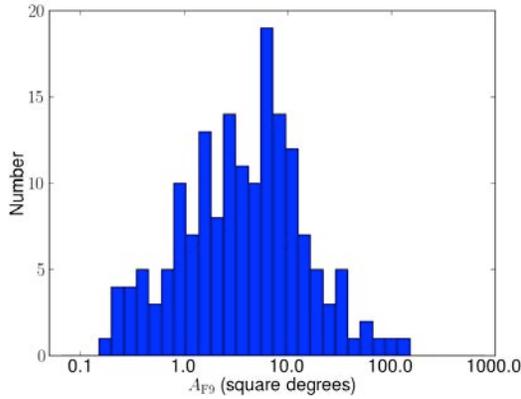
Figure 2.3 Histograms of the estimated 50% credible-interval sky areas, in square degrees, calculated using the Bayesian coherent analysis A_{Bayes} (panel (a)), timing triangulation A_{TT} (panel (b)), 9-dimensional Fisher information matrix A_{F9} (panel (c)) and 4-dimensional Fisher information matrix, A_{F4} (panel (d)).



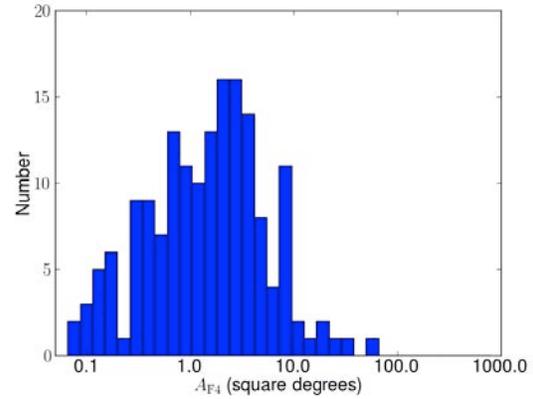
(a) Full coherent Bayesian



(b) Timing triangulation approximation



(c) Fisher information matrix approximation 9D



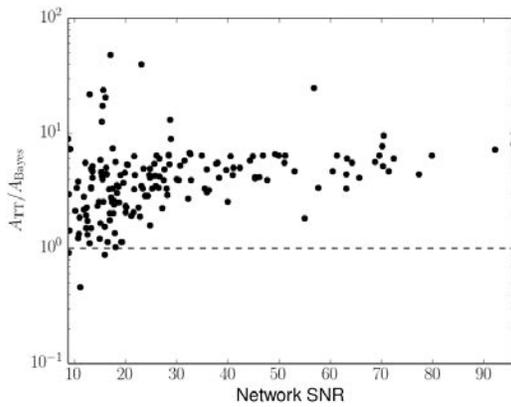
(d) Fisher information matrix approximation 4D

2.6 Timing vs Coherent Bayesian

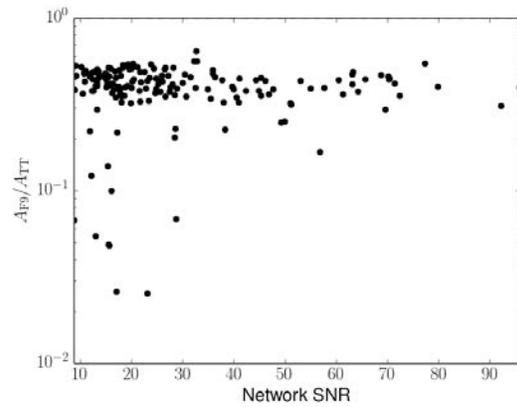
The comparison of the timing triangulation (TT) approximate sky areas and those from the full Bayesian method provides an indication of the ability of the TT approximation to predict the sky localization uncertainty for GW transients with advanced networks. It was expected that the extra information used in the full Bayesian coherent analysis should provide better localization and so smaller sky areas. However, prior to this study the size of this improvement was unknown. As seen in Table 2.1, the sky areas are on average significantly smaller when measured using the coherent Bayesian method than when estimated with TT. When comparing injections individually, the median of the ratio $A_{\text{TT}}/A_{\text{Bayes}}$ is 4.1, see Figure 2.4(a). However, there is a large scatter in individual results. The distribution of the ratio $A_{\text{TT}}/A_{\text{Bayes}}$ is strongly skewed, but the distribution of the logarithm of the ratio appears relatively Gaussian, so we use the standard deviation of the ratio as a proxy for the scatter. The standard deviation of the natural logarithm of this ratio is 0.7, meaning that individual sky localization areas typically differ by a factor of ~ 2 .

The Bayesian 50% credible interval for a typical mock event is shown in Figure 2.5(a), along with timing triangulation and FIM predictions for localization ellipses. This event, at a network SNR of 23.6, has $A_{\text{TT}}/A_{\text{Bayes}} = 4.01$. The Bayesian posteriors are slightly offset from the true value because of the impact of noise in the data, while the predicted TT and FIM uncertainties that do not consider the actual data realization are centred on the true value. However, the qualitative shape of the Bayesian posterior here is ellipsoidal (the KD binning method tends to produce elongated bins at the edges of distributions, exaggerating the “jagged” edges of these ellipses) and the orientation and eccentricity of this ellipse match the analytical predictions, although the size of

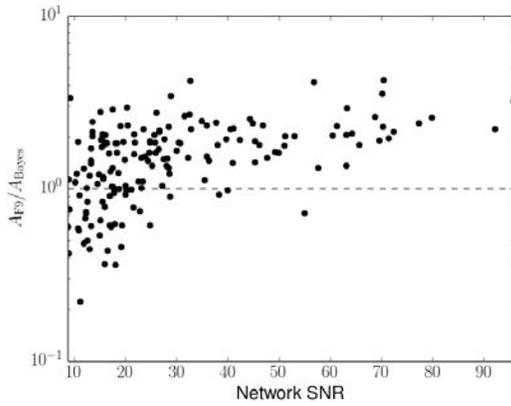
Figure 2.4 Comparisons of 50% credible-interval sky areas found using the timing triangulation approximation A_{TT} , the 4d Fisher information matrix approximation A_{F4} , the 9d Fisher information matrix approximation A_{F9} and the coherent Bayesian method A_{Bayes} , as a function of the network SNR of the signal. Panel (a) compares A_{TT} and A_{Bayes} , panel (b) A_{F9} and A_{TT} , panel (c) A_{F9} and A_{Bayes} , and panel (d) A_{F4} and A_{Bayes} .



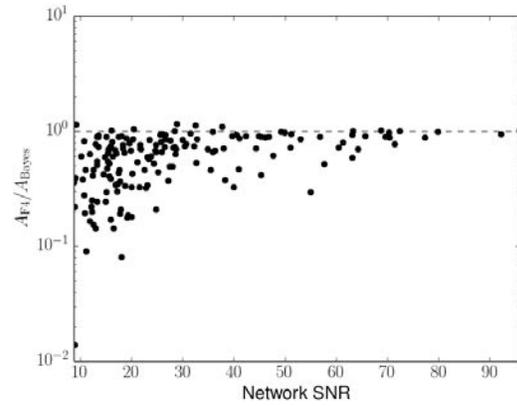
(a) Timing triangulation and full coherent Bayesian comparison



(b) 9d Fisher information matrix and timing triangulation comparison



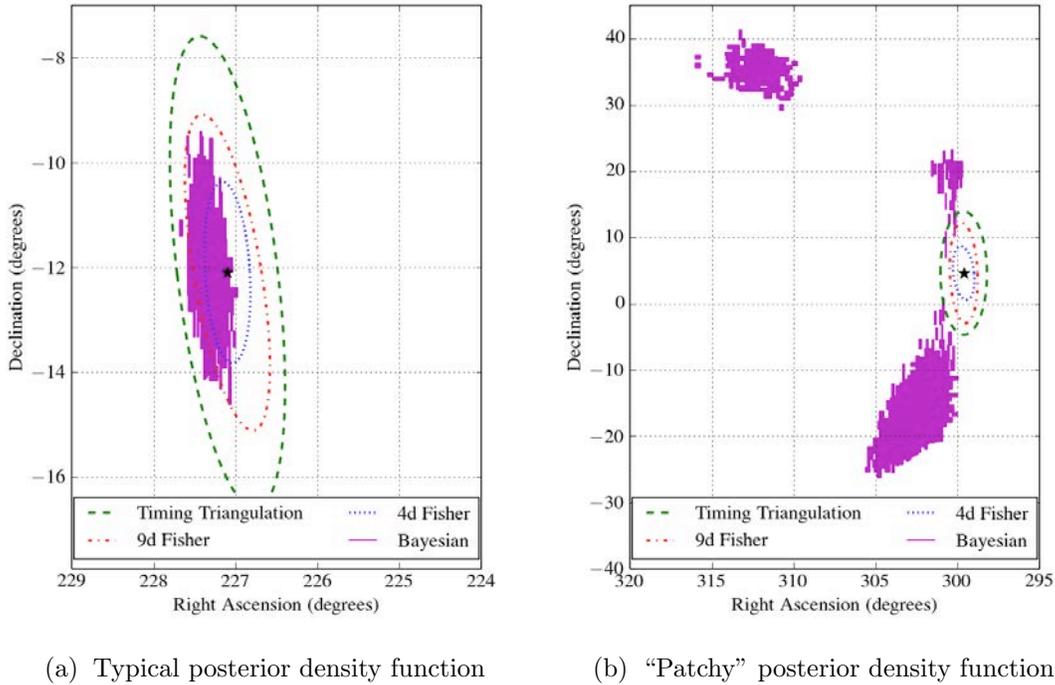
(c) 9d Fisher information matrix and full coherent Bayesian comparison



(d) 4d Fisher information matrix and full coherent Bayesian comparison

the ellipses vary because of the omission of key phase and amplitude information when constructing the TT uncertainty region.

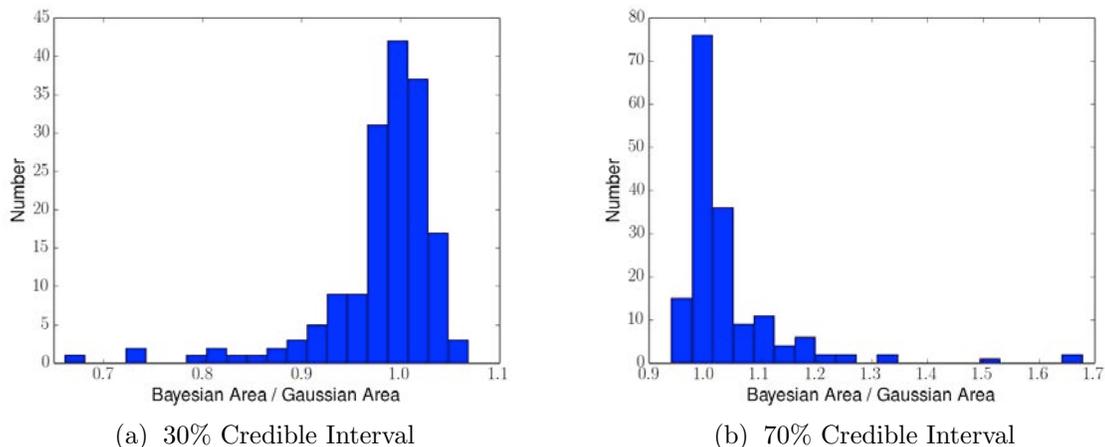
Figure 2.5 50% credible intervals for the a typical injection (panel (a), $A_{\text{TT}}/A_{\text{Bayes}} = 4.01$, $A_{\text{F4}}/A = 0.84$, $A_{\text{F9}}/A_{\text{Bayes}} = 2.23$, $\rho_{\text{Hanford}} = 7.48$, $\rho_{\text{Livingston}} = 14.3$, $\rho_{\text{Virgo}} = 17.1$ and $\rho_{\text{Network}} = 23.6$) and for a injection where the posterior density function found via the Bayesian analysis is patchy (panel (b), $A_{\text{TT}}/A_{\text{Bayes}} = 0.38$, $A_{\text{F4}}/A_{\text{Bayes}} = 0.09$, $A_{\text{F9}}/A_{\text{Bayes}} = 0.22$, $\rho_{\text{Hanford}} = 6.62$, $\rho_{\text{Livingston}} = 4.54$, $\rho_{\text{Virgo}} = 7.77$ and $\rho_{\text{Network}} = 11.2$).



Meanwhile, Figure 2.5(b) shows the Bayesian credible interval and the analytical localization predictions for the outlier seen in Figure 2.4(a) at a network SNR of 11.2 and $\frac{A_{\text{TT}}}{A_{\text{Bayes}}} = 0.38$. In this case, the partial degeneracy between multiple sky locations cannot be broken, and the Bayesian 50% credible interval is composed of multiple patches in the sky. The TT and FIM technique, which assume that such degeneracies can always be broken using amplitude information, do not account for this degeneracy and significantly under-predict the sky localisation uncertainty. Note the MCMC code may have not completely converged on the peak here, as the code was terminated when set

number of independent samples were drawn from the posterior, as such it is possible the code could be run for longer to more accurately estimate the sky location.

Figure 2.6 Histograms of Bayesian sky areas at the 30% (panel (a)) and 70% (panel (b)) credible levels to the area of a Gaussian posterior at the same credible level, scaled to match the Bayesian posterior at the 50% credible level.



As the TT and FIM approximations assume a Gaussian posterior distribution, any deviations from this, as seen in an extreme case in Figure 2.5(b), will affect the validity of these approximations. In order to investigate the Gaussianity of the Bayesian posteriors we compare the two-dimensional 30% credible intervals and 70% credible intervals to those expected for a Gaussian posterior distribution scaled to match the Bayesian posterior at the 50% CL, Figure 2.6. The tails extending to lower ratios of Bayesian to Gaussian ratios at the 30% credible intervals and to higher ratios at the 70% credible intervals CI indicate that some of the Bayesian posteriors are not Gaussian, but have greater peakedness with long tails.

2.7 Fisher Matrix vs Coherent Bayesian

The fractional difference in the sky location angles, right ascension and declination, between FIM and Bayesian for this set of simulated sources have been reported in [145], where it was shown that the FIM tends to overestimate the error in individual sky angles at low SNRs and underestimate it at high SNRs. In Figure 2.4(b) we show the ratio of sky areas for the FIM and TT approximations, while Figure 2.4(c) shows the ratio of FIM sky areas to the coherent Bayesian 50% credible intervals. With the cuts used in this study (see Section 2.4), the median of ratios A_{F9}/A_{Bayes} is 1.6 with standard deviation in the log of the ratio 0.6, while the median of the ratio A_{TT}/A_{F9} is 2.4 and standard deviation in the log of the ratio of 0.5. The inclusion of the full 9-dimensional parameter space in the FIM results provides a closer estimation of the full coherent analysis than the areas estimated via timing triangulation alone, but still overestimates the sky areas. While the FIM does include all the signal parameters, it is still limited to considering the leading order, quadratic terms to determine localization, and it makes no use of physically motivated priors on the signal parameters. Both of these are included in the Bayesian approach and are likely to explain the difference in results.

The restricted 4-dimensional FIM analysis over-constrains the sky areas by artificially assuming perfect knowledge of the remaining 5 parameters. The ratio of 4-dimensional FIM 50% confidence-interval sky areas to corresponding areas from the coherent Bayesian analysis A_{F4}/A_{Bayes} , shown in Figure 2.4(d), have a median 0.68 with standard deviation in the log of the ratio 0.6. It is not surprising that the 4-dimensional FIM gives the smallest localization regions, as it is imposing physically unrealistic restrictions on the parameters.

2.8 Timing Triangulation with Phase Correction

We have seen that signal parameters not considered in basic timing triangulation still affect the sky localization, yielding an area that is a factor of 2.7 larger than predicted by the 9-dimensional FIM. Ignoring these parameters, particularly the phase consistency between detectors, leads to pessimistic predictions of the ability of a detector network to localize GW sources. We can attempt to correct the predicted sky localization based on timing triangulation alone [74, 75] to account for phase information.

To understand the impact of this, we can consider a simplified case where we ignore the polarization of the waveform and keep the relative phase of the signal between detectors fixed. Generically, this is overly restrictive, although it would be appropriate for a face-on, circularly polarized binary.

Consider a waveform of the form $\tilde{h}(f) = A(f) \exp[i\phi_0 + i2\pi ft_0 + i\Psi(f)]$ where ϕ_0 and t_0 are the phase and time at a particular frequency (e.g., at the fiducial moment of coalescence) and $A(f)$ and $\Psi(f)$ are the amplitude and phase, respectively, which depend on the other seven parameters of the signal. For a signal of this form, it is possible to measure both ϕ_0 and t_0 from each detector's data. The timing accuracy in each detector σ_t , considered independently, can be obtained by applying a 2-dimensional FIM calculation in $\{\phi_0, t_0\}$ to each detector's data set. This FIM is:

$$\begin{aligned}\Gamma_{tt} &= 4\pi^2 \rho^2 \bar{f}^2, \\ \Gamma_{t\phi} &= 2\pi \rho^2 \bar{f}, \\ \Gamma_{\phi\phi} &= \rho^2,\end{aligned}\tag{2.15}$$

where \bar{f} and \bar{f}^2 are defined in Eq. (2.7). In the high SNR limit, the covariance matrix

can be approximated by the inverse of the FIM, Eq. (2.13). The timing uncertainty in an individual detector is given by $\sigma_t^2 = (\Gamma^{-1})_{tt} = \left[4\pi^2\rho^2(\bar{f}^2 - \bar{f}^2)\right]^{-1}$, in agreement with the expression for σ_t given in Eq. (2.5).

This calculation, a variant of which was used for the original TT prediction [74], ignores the requirement of phase consistency between detectors. Although we don't know the actual value of ϕ_0 in the detectors, its value must be the same for all detectors regardless of detector location. Fixing the phase, rather than marginalizing over it, corresponds to using a 1-dimensional FIM Γ_{tt} in place of the two-dimensional matrix considered above. Thus, with the phase consistency requirement, the timing uncertainty in each detector is just the inverse of the tt component of the 2-dimensional FIM, $\sigma_{t,\text{new}}^2 = (\Gamma_{tt})^{-1}$. We can use this timing uncertainty in place of σ_t in Eq. (2.9) to compute a prediction for the sky localization uncertainty with the phase consistency requirement.

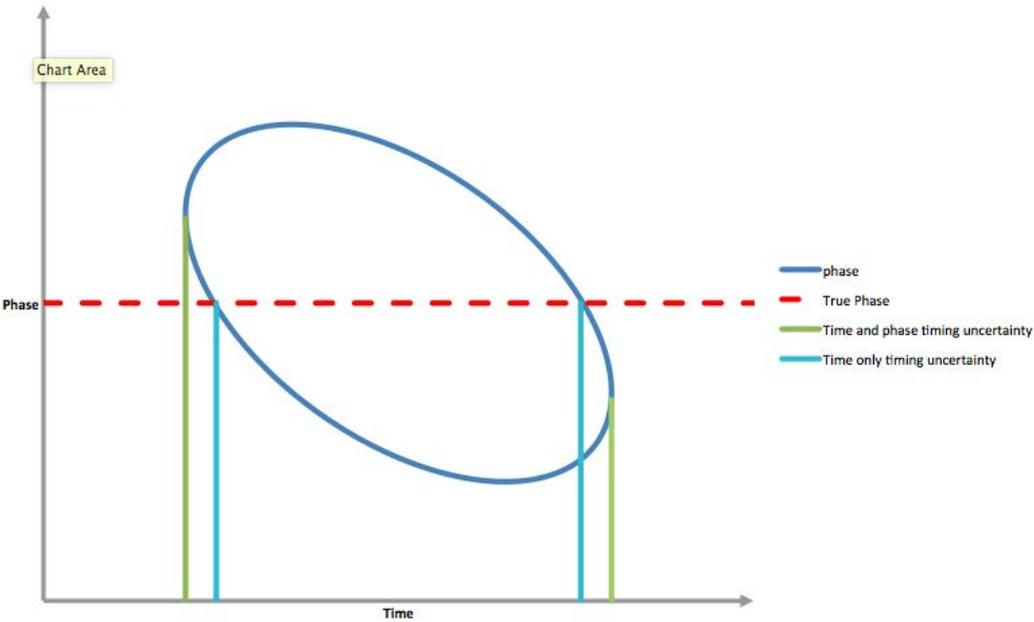
If there is a correlation between ϕ_0 and t_0 , the uncertainty ellipse in time–phase space will be inclined. In this case, the ellipse's projection onto the time axis, measured by σ_t , which was the measure of localization used in the original TT approach, can be much wider than the width of the ellipse at a fixed value of phase. Thus, the largest improvements from including phase constraints in timing triangulation arise when there are large correlations between time and phase measurements, as explained pictorially in Figure 2.7.

For the simple case of detectors with equal noise power spectral densities, such as those considered in this study, sky localization area uncertainties will decrease by a factor of

$$\frac{\sigma_{t,1\text{-phase}}^2}{\sigma_t^2} = \frac{\bar{f}^2 - \bar{f}^2}{\bar{f}^2}. \quad (2.16)$$

The precise value added depends on both the detector bandwidth and the system

Figure 2.7 An example of a correlated error ellipse in time and phase space. No assumption or other information to measure phase gives a wide timing error when the parameters are highly correlated as here (green lines). Alternately assuming perfect knowledge of the phase (red dashed line) constricts the ellipse to smaller timing error (blue lines).



masses, with larger contributions for higher masses because of greater correlation. The effect is illustrated in Figure 2.8. For instance, for a signal from a typical binary neutron star system observed by the network considered here, we expect approximately a factor of 3 improvement in sky localization by including phasing information. However, for the most massive systems in the set of injections considered above, the improvement is as large as a factor of 8, while the median improvement is a factor of 6.

The additional freedom contained in the choice of polarization lowers the phase consistency improvement to the sky localization area to only a square root of the value predicted above, i.e.,

$$\frac{\sigma_{t, 2\text{-phase}}^2}{\sigma_t^2} = \sqrt{\frac{f^2 - \bar{f}^2}{f^2}}. \quad (2.17)$$

Heuristically, this can be understood by noting that the GW has two free phases, while TT (with a three-site network) assumes three. Consider the case where two detectors see the $+$ gravitational wave polarization while the third detector is sensitive to the \times . In this case, the observations in the first two detectors must be phase consistent, while the phase in the third detector is not restricted. Timing triangulation with a three-site network (TT) allows for three free phases while the calculation above allows a single overall phase. Thus, we should expect the actual result to be midway between TT and fixed phase approximations. The improvement in predicted sky localization relative to TT alone varies from a factor of 1.8 for a signal from a typical binary neutron star system, to a factor of three for the most massive system considered here (i.e. the signal with the shortest bandwidth), with a median value of 2.5 for the set of injections considered above (see Figure 2.8 for the distribution with mass).

We see in Table 2.2 that the inclusion of phase information improves the TT predictions and brings them closer in-line with the observed sky localization ability of

Figure 2.8 Fractional reduction in timing triangulation prediction for the sky localization area after incorporating phase consistency versus using only timing information; crosses correspond to the single-free-phase approximation described here, while circles correspond to the two-free-phase approximation from [76].

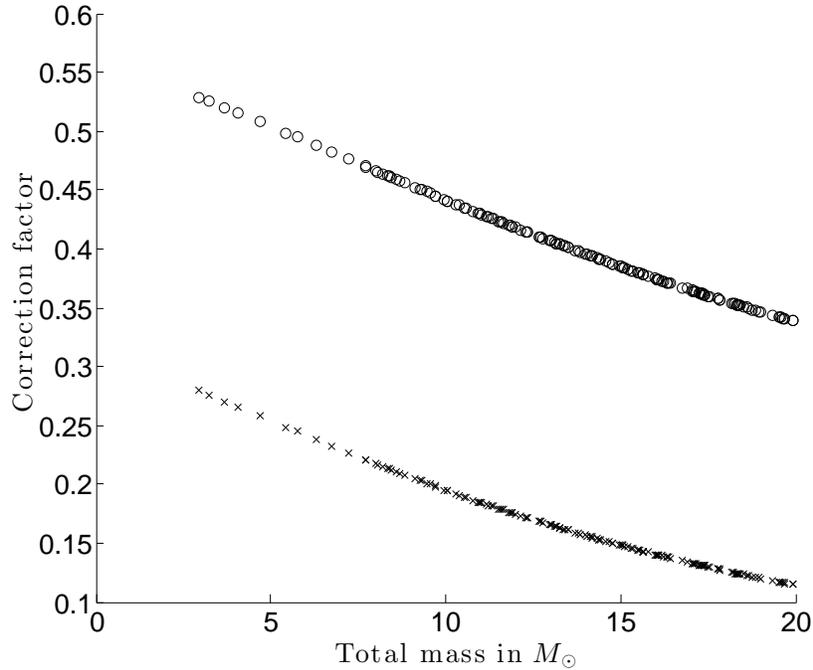


Table 2.2 The median areas of 50% Bayesian credible intervals for coherent Bayesian method, timing triangulation and time-phase localization. Median ratios of the areas relative to the coherent Bayesian method are also given.

Method	Median Area (square degrees)	Median Ratio with A_{Bayes}
Bayesian	2.9	-
standard TT	10.6	4.1
TT + two phases from [76]	4.4	1.6
TT + one free phase	1.7	0.6

coherent Bayesian analysis. In the case where we ignore the freedom of choosing the polarization angle, we effectively over-constrained the system with the phase consistency requirement, which leads to predicted credible intervals that are smaller than what can be achieved even with coherent Bayesian analysis. In fact, the median 50% credible-interval area for the TT prediction with this phase constraint, 1.7 square degrees for our injection set (see Table 2.2), matches with the 4-dimensional FIM median interval of 1.6 square degrees (see Table 2.1). This is to be expected as both methods effectively consider three free angles: two sky angles and phase.

Predictions from the TT calculation with two free phases are similar to the full 9-dimensional FIM, and, indeed, the median sky localization areas for the two methods are 4.4 and 3.9 square degrees, respectively. This indicates that by requiring a coherent signal across the detectors, we have incorporated the most significant additional parameter after arrival time. However, the median area predicted with this method is a factor of ~ 1.6 larger than the median Bayesian area, which indicates that additional inputs can further improve sky localization ability.

This importance of coherence, and particularly phase consistency, is relevant for the development of actual (rather than merely predictive) rapid methods to localize sources based on TT, e.g., [9]. We expect these rapid localization methods to benefit from including phase consistency information along with timing triangulation. As discussed above, the largest benefits accrue for more massive systems.

Although we see that the inclusion of phase information brings average predictions of timing-triangulation based estimates closer to the full coherent Bayesian localization, there is very significant scatter in results on individual mock data sets when compared on an injection-by-injection basis. The standard deviation of the natural logarithm of the ratio of the 50% MCMC credible intervals to 50% TT predicted areas is ~ 0.7 both

without phase information and for both correction schemes.

2.9 Discussion

We have compared the predicted accuracy of source localization with various techniques with the performance obtained by a fully coherent Bayesian parameter estimation method. In general, the full exploration of the parameter space yields the most accurate results, and also the smallest confidence regions. For the systems considered here, the median areas are a factor of four smaller than those obtained with the timing triangulation approximation. We have seen that requiring a consistent phase between the sites can significantly improve the localization of sources. The agreement between this result and the 9-dimensional FIM result suggests that this is the most significant additional parameter to include. For the set of simulations considered here, incorporating this correlation reduces the median area by a factor of 2.5. However, for binary neutron star systems, the correction will be smaller, a factor of 1.8. The effect could be less significant in advanced detectors as their sensitive band starts at a lower frequency. Additionally, the inclusion of merger and ringdown waveform phases will significantly increase the frequency content of the higher mass signals and consequently may reduce the impact of phase consistency for stellar-mass black hole binaries.

The best of the approximation methods still differ, on average, by at least 50% from the full results and, more significantly, there is a large variation in the results on an event-by-event basis. The difference between the full analysis results and the best approximations indicate that there are still significant factors which are being overlooked in the approximation methods. These include: correlations with other parameters such as the binary's component masses, priors on signal parameters, particularly distance

and inclination, and the specific noise realization in the data set being analyzed. Furthermore, all of the approximation techniques consider only the leading-order effects. At low SNR, expected for the first gravitational-wave detections, these approximations are not sufficient and a full examination of the parameter space is required to accurately extract the signal parameters. While the approximation methods can be used to give a sense of the localization capabilities of various detector networks, the fully coherent Bayesian analysis gives the most accurate results both on a population of sources and, most significantly, on individual events. It also provides a critical cross-check on the self-consistency of purely predictive methods techniques for parameter estimation.

In this study, we only analyzed the case of non-spinning compact objects. The results should extend straightforwardly to systems with aligned spins. However, for binaries in which at least one of the components is a black hole, and therefore could carry significant spin, precession of the orbital plane through spin-orbit and spin-spin coupling carries information about the system location and orientation parameters and therefore affects our ability to recover the sky location of the source. It is likely that these effects will produce a systematic offset in a parameter estimation approach based on intermediate data products from single-detector analyses. A fully coherent Bayesian parameter recovery can straightforwardly incorporate such effects. It would be interesting to investigate the effect of precession on source localization, but we leave that for the future.

Acknowledgements

We would like to acknowledge the useful discussions with our colleagues from the LIGO – Virgo Collaboration, especially the helpful comments from Leo Singer. This work has

been supported in part by the UK Science and Technology Facilities Council. SF would like to thank the Royal Society for support.

Chapter 3

TESTING GENERAL RELATIVITY WITH COMPACT BINARY COALESCENCES IN GROUND BASED DETECTORS

The work in this chapter is a combination of the work presented in its published form in Appendix B and C. K. Grover wrote all text and produced all plots apart from figures and , which were produced by W. Del Pozzo, based on work done by the both of us. Equation 3.8 was a joint work by the authors of Appendix B.

The advanced generation of gravitational wave interferometer detectors are expected to start making regular detections in late 2015. Gravitational wave signals from compact binary coalescences as detected in these instruments will provide a new regime in which to test General Relativity, with signals providing detailed information on strong field

dynamics of the theory. A bayesian pipeline which tests GR using these detections is described here. The pipeline works by combining a number of tests of the consistency of phase coefficients with those predicted via GR, without assuming any particular specific deviation from GR. As this is a computationally intensive process a means to cheaply estimate the Odds ratio has been suggested by various sources, including [181]. Here we compare the output of a full Bayesian inference calculation with that calculated via this approximation. We find that the two agree for high Fitting Factors, around 0.9, but progressively disagree for larger differences between the General Relativity (GR) and alternative gravity (AG) waveforms.

3.1 Introduction

Gravitational waves (GWs) from compact binary coalescence (CBC) sources are one of the most promising signals predicted in the advanced Laser Interferometer Gravitational-wave Observatory (aLIGO) and advanced Virgo ground based interferometers [87, 179], which are expected to start taking science data in 2015. These signals will provide access to a new regime of strong field tests of GR. As such, it is vital to investigate at what level these advanced detectors can be expected to test GR and prepare data analysis techniques to utilise these observations to their full potential.

Various methods have been suggested to test GR using CBC signals either searching directly for specific alternate theories or for some generic deviation from GR. For example using the CBC signal to test GR using the post-Newtonian (PN) coefficients as test parameters [28] following the example of the binary pulsar tests of GR. Alternatively Bayesian model selection can be used e.g. [63] to compare alternative theories or a generic parametrised deviations from GR [55]. Such a pipeline, the test infrastructure

for general relativity (TIGER), has been implemented in `lalInference` [190].

Before CBC signals become routine it is useful to investigate at what level particular deviations from GR can be detected. Full pipelines such as TIGER are very computationally expensive, as a background of simulated sources must be calculated in order to account for unmodelled effects. Therefore an approximation to the Odds ratio, like the one described by Vallisneri in [181] could prove useful as a predictive method.

Here we briefly describe the TIGER pipeline, where full testing of the pipeline is shown in Appendix B. We then investigate the validity of the approximation to the Odds ratio, and find it matches full Bayesian inference at fitting factors (FFs) less than approximately 0.9. The rest of this chapter is arranged as follows, in Section 3.2.1 we recap the Bayesian framework for model selection, before outlining the TIGER pipeline in Section 3.2.3. The approximation to the Odds ratio is described in Section 3.9 and the Fitting Factor and detectability are described in Section 3.3. An example signal is used to test the validity of the Odds approximation in Sections 3.4 and 3.5.

3.2 Use of Bayesian Odds Ratio in Testing GR

3.2.1 Bayesian Model Selection

Bayesian model selection provides a comprehensive framework in which it is possible to compare two competing models. As Bayesian infrastructure is already in place to perform GW data analysis, an obvious method to test GR is using model selection. The odds ratio $O_{i,j}$ between models M_i and M_j is,

$$O_{i,j} = \frac{P(M_i | D, I)}{P(M_j | D, I)} = \frac{P(M_i | I)}{P(M_j | I)} B_{i,j}, \quad (3.1)$$

$$B_{i,j} = \frac{P(D | M_i, I)}{P(D | M_j, I)}. \quad (3.2)$$

Where D is the data and I is any prior information. In this case the two models are a detection of a GW as generated by GR and as generated by some AG theory, $M_i = AG$ and $M_j = GR$. The odds ratio can be strongly influenced by the prior probabilities assigned to each model $P(M_i | I)$ and in this case there is no strong reasoning for a particular quantified value of these prior odds. Therefore it seems prudent to give each model equal prior probabilities or equivalently to report the Bayes factor $B_{i,j}$ which is the ratio of evidences, Z , between the models, allowing the reader to apply their own priors as a simple weighting. In order to calculate the evidence for a particular model we must parametrise the model such that it is possible to calculate the likelihood, $\mathcal{L} = p(D | M_i, \underline{\theta}, I)$,

$$Z = P(D | M_i) = \int p(\underline{\theta} | M_i, I) p(D | M_i, \underline{\theta}, I) d\underline{\theta}. \quad (3.3)$$

Where $\underline{\theta}$ are the free parameters of the model.

3.2.2 Towards a Generic Test of GR

It is impossible to calculate the odds ratio between a GR GW and a completely general non-GR GW as we must parametrise the non-GR GW in some way. This can be done by either selecting a specific non-GR model to compare (e.g. [63]) or by using a parametrised framework which encompasses a range of theories (e.g. the ppE framework [55]). A

Bayesian pipeline to test GR with parametrised deformations, the TIGER, has been implemented in `lalInference` [190].

TIGER aims to test generic deviations from GR by testing the consistency of free parameters of the waveform, such that when the parameters are zero the waveform corresponds to GR. In principle this could be applied to any parametrised modified theory of gravity, however in this first case we consider the free parameters to be the PN coefficients of the phase. The phase of the GW directly encodes the orbital dynamics of the compact binary and as such any difference from GR will have a direct imprint on the GW phase. This pipeline is designed to be useful in the early detection era, when SNRs are expected to be low, making parameter estimation and model selection with large number of free parameters useless. Instead the odds ratio of subsets of parameters are calculated and then combined. Bayesian inference provides a simple way to combine such tests as well as combining results from multiple detections.

For example, consider a waveform with two phase coefficients ψ_1 and ψ_2 . We wish to calculate the odds of either of these coefficients being varied from the expected function in GR such that $\psi_1 \neq \psi_1^{GR}(\mathcal{M}, \eta)$ or $\psi_2 \neq \psi_2^{GR}(\mathcal{M}, \eta)$. I.e. the models we are comparing are,

$$M_i = AG = AG_1 \vee AG_2 \vee AG_{12}, \quad (3.4)$$

$$M_j = GR. \quad (3.5)$$

Where $AG_{i,j}$ denotes a model where ψ_i and ψ_j differ from GR while the other coefficients, if any, take the form as prescribed by GR. The Odds ratio we compute is

then,

$$O_{AG,GR} = \frac{P(AG | D, I)}{P(GR | D, I)} = \frac{P(AG_1 \vee AG_2 \vee AG_{12} | I)}{P(GR | I)} B_{AG,GR}. \quad (3.6)$$

As these are logically disjoint we can write,

$$O_{AG,GR} = \frac{P(AG_1 | D, I)}{P(GR | D, I)} + \frac{P(AG_2 | D, I)}{P(GR | D, I)} + \frac{P(AG_{12} | D, I)}{P(GR | D, I)}. \quad (3.7)$$

In practice we set the priors on the free phase coefficients such that the hypothesis AG_1 and AG_2 are embedded within hypothesis AG_{12} , which in principle could invalidate the assumption that these hypotheses are logically disjoint. However the $\psi_1 = \psi_1^{GR}(\mathcal{M}, \eta)$ (i.e. AG_1) is a space of zero measure within the full AG_{12} integration space, meaning it does not contribute to B_{12} and we are in no way “double counting” the contribution from this hypothesis.

We currently have no reason for preferring one of these types of deviation over the others so can assign equal prior weighting,

$$O_{AG,GR} = \alpha [B_{AG_1,GR} + B_{AG_2,GR} + B_{AG_{12},GR}]. \quad (3.8)$$

This may seem like some averaging factor which could win over a high Bayes factor for the case of many parameters, remember that the “correct” Bayes factor grows exponentially with the log Likelihood and will overcome this averaging factor, see B for specific examples. \mathcal{N} independent detections can be combined simply via the product of Bayes factors, allowing the odds ratio for a catalog of sources to be calculated. GR signals may still have a resulting Odds ratio which favours AG due to noise data or other

unmodelled effects. As such a background of Odds ratios for simulated GR signals is created and used as a comparison. The pipeline is tested in Appendix B, where a deviation at the few percent level in the 1.5 PN phase coefficient is discernible. Additionally it is tested on a deviation in the phase which does not fall within the parametrised deviations, demonstrating that a generic phase deviation may be detectable via this pipeline.

3.2.3 Approximating the Odds Ratio

Using such a pipeline is computationally intensive process even for this simple case and including spinning waveforms, which must be done when testing true signals, can cause Bayesian codes to take days or weeks to calculate. This makes the type of systematic study that it would be wise to carryout prior to the first tests of GR, or to construct a background distribution, extremely time consuming. As such, an approximate method to calculate the odds ratio has been suggested in [181]. This approximation uses a linear signal approximation (LSA), one additional parameter to describe an alternative gravity waveform, uniform priors on all parameters, and assumes a FF close to one.

Using these approximations the evidence can be calculated analytically and in the zero noise realisation the Odds Ratio is given by,

$$O_{AG,GR} \approx \frac{p(AG)}{p(GR)} \frac{\sqrt{2\pi} \Delta\theta_{\text{Fisher}}^a}{\Delta\theta_{\text{prior}}^a} \exp [\rho^2 (1 - FF)] . \quad (3.9)$$

ρ is the optimal signal to noise ratio (SNR),

$$\rho^2 = \langle h(\underline{\theta}_T) | h(\underline{\theta}_T) \rangle , \quad (3.10)$$

and $\underline{\theta}_T$ are the true parameters of the waveform. Where $\langle a | b \rangle$ denotes the noise weighted inner product such that,

$$\langle a | b \rangle \equiv 2 \int_0^\infty \frac{\tilde{a}(f)\tilde{b}^*(f) + \tilde{a}^*(f)\tilde{b}(f)}{S_n(f)} df = 4\Re \left[\int_0^\infty \frac{\tilde{a}(f)\tilde{b}^*(f)}{S_n(f)} df \right]. \quad (3.11)$$

The FF measures the fraction of optimal SNR recovered, due to mismatch between the model used for recover (GR) and the true form of the GW (AG).

$$FF = \left[\frac{\langle h_{GR}(\underline{\theta}) | h_{AG}(\underline{\mu}) \rangle}{\sqrt{\langle h_{GR}(\underline{\theta}) | h_{GR}(\underline{\theta}) \rangle \langle h_{AG}(\underline{\mu}) | h_{AG}(\underline{\mu}) \rangle}} \right]_{\text{maximised over } \underline{\theta}}, \quad (3.12)$$

where $\underline{\theta}$ and $\underline{\mu}$ are the parameters of the GR and AG waveform respectively. $\Delta\theta_{\text{Fisher}}^a$ and $\Delta\theta_{\text{prior}}^a$ denote the width of the prior and Fisher matrix $1 - \sigma$ uncertainty estimate of the additional AG parameter. Using this approximation it would be possible to investigate which AG theories would be distinguishable from GR via Bayesian model selection with out using lengthy full code runs. It is important to ensure that such an approximation gives accurate results, and in which regime by comparisons to full Bayesian codes, before using in any systematic study.

3.3 Detectability of non-GR Signals

Detection of CBC signals depends on the construction of a template bank of accurate waveforms in order to perform detection via matched filtering. By definition the type of generic deviation from GR as discussed in the previous section will not be represented in the standard detection template bank. Any mismatch between the template bank and the true signals will result in a loss of effective SNR and so a loss in the number

of detected signals, i.e. signals with near threshold optimal SNR become undetectable. This loss in SNR due to using a GR template bank to detect a AG signal is quantified by the FF as defined above in Equation 3.12. This can be converted to a fractional decrease in detection rate,

$$\frac{N_{opt} - N}{N_{opt}} \sim 1 - FF^3, \quad (3.13)$$

where the subscript *opt*, denotes the optimal number of detections per time, which is recovered by using a template of exactly the same form as the signal.

In order to calculate the FF a maximisation over the whole template bank must be performed. Consider a GW signal of the form,

$$\tilde{h}(f) = \mathcal{A}^T f^{-7/6} e^{i[2\pi f t_c^T + 7\phi_c^T + \Psi^T]}, \quad (3.14)$$

where \mathcal{A} is the frequency independent amplitude, f is the GW frequency, t_c and ϕ_c are a reference time and phase respectively. And a template bank constructed from the same waveform family, where the superscript T denotes the true values of the parameter. It can be seen that the FF is independent of any maximisation over the frequency independent amplitude. Let $\mathcal{A}\bar{h} = \tilde{h}$ and $\{\underline{\theta}', \mathcal{A}\} = \underline{\theta}$,

$$\begin{aligned} FF &= \left[\frac{\langle \mathcal{A}\bar{h}(\underline{\theta}') | \mathcal{A}^T \bar{h}(\underline{\theta}^T) \rangle}{\sqrt{\mathcal{A}^2 \langle \bar{h}(\underline{\theta}') | \bar{h}(\underline{\theta}') \rangle \mathcal{A}^{T^2} \langle \bar{h}(\underline{\theta}^T) | \bar{h}(\underline{\theta}^T) \rangle}} \right]_{\text{maximised over } \{\underline{\theta}', \mathcal{A}\}} \\ &= \left[\frac{\langle \bar{h}(\underline{\theta}') | \bar{h}(\underline{\theta}^T) \rangle}{\sqrt{\langle \bar{h}(\underline{\theta}') | \bar{h}(\underline{\theta}') \rangle \langle \bar{h}(\underline{\theta}^T) | \bar{h}(\underline{\theta}^T) \rangle}} \right]_{\text{maximised over } \underline{\theta}'} . \end{aligned} \quad (3.15)$$

To maximise over the reference time t_c , notice that the denominator in Equation 3.15

is independent from the value of t_c . The numerator, when maximising over t_c may be re-expressed as

$$\left[4\mathcal{R} \int_0^\infty \frac{\tilde{h}(f, \underline{\theta}, t_c = 0) \tilde{h}(f, \underline{\theta}^T, t_c = t_c^T)}{S_n(f)} e^{i2\pi f t_c} df \right]_{\text{maximised over } t_c}, \quad (3.16)$$

$$= 4\mathcal{R} \text{FT} \left[\frac{\tilde{h}(f, \underline{\theta}, t_c = 0) \tilde{h}(f, \underline{\theta}^T, t_c = t_c^T)}{S_n(f)} \right]_{\text{maximised over } t_c}. \quad (3.17)$$

where FT [...] is the inverse Fourier transform, i.e. in order to maximise over reference time it is possible to perform a Fourier transform and find the maximum of the resulting time series.

To maximise over phase notice that again the denominator does not depend on this parameter and the numerator in Equation 3.15 can be reexpressed as,

$$\left[\sqrt{x^2 + y^2} \cos \left(\phi_c - \arctan \frac{y}{x} \right) \right]_{\text{maximised over } \phi_c}. \quad (3.18)$$

where $x = \langle \bar{h}(\phi_c = 0) | \bar{h}_T \rangle$ and $y = \langle \bar{h}(\phi_c = \frac{\pi}{2}) | \bar{h}_T \rangle$. This is simply maximised at $\phi_c = \arctan(\frac{y}{x})$, allowing the phase maximisation to be computed via just two templates.

The remaining parameters, in the case of non-spinning inspiral only waveforms the two mass parameters, must be searched over to find the maximum. Here we look at a generic deviation to ensure it is detectable before using it to further investigate the odds ratio approximation.

3.4 Example Signal

In what follows we use a simple arbitrary deviation from GR based on the TaylorF2 waveform family but with a deviation in the 1.5PN phase term. For an optimally oriented binary the TaylorF2 waveform is given by,

$$\tilde{h}(f) = \frac{\pi^{-2/3}}{D_L} \sqrt{\frac{5}{24}} \mathcal{M}^{5/6} f^{-7/6} e^{i\Psi(f)}, \quad (3.19)$$

$$\Psi(f) = 2\pi f t_c + \phi_c - \frac{\pi}{4} + \frac{3}{128} (\pi \mathcal{M} f)^{-5/3} \sum_{k=0}^N \alpha_k (\pi M f)^{k/3}. \quad (3.20)$$

Where,

$$\eta = \frac{m_1 m_2}{M^2}, \quad (3.21)$$

$$\mathcal{M} = \eta^{3/5} M, \quad (3.22)$$

and m_1 and m_2 , are the binary's component masses and M is the total mass of the system. D_L is the luminosity distance to the source, t_c and ϕ_c are a reference time and phase respectively. f is the GW frequency and α_k is the phase coefficient as found in Equation 1.72. The AG waveform is modified such that,

$$\alpha_3 \rightarrow \alpha_3(1 + \delta\chi_3) \quad (3.23)$$

Such a signal is not designed to be physically motivated and rather demonstrates some

generic deviation from GR. However it should be noted such a shift in one of the phase coefficients is not impossible, for example in the massive graviton model a deviation from GR occurs just in the 1 PN coefficient, as explored in [63].

Here we consider an example CBC system to understand the effects of a AG waveform on the Odds Ratio approximation. The Odds ratio is calculated via the lalinference [190] implementation of Nested Sampling [168]. The predicted noise power spectral density (PSD) of a single advanced LIGO detector at full sensitivity in the high power zero detuning mode is used [163]. The mass combination investigated here is an unequal mass binary of masses $4.5 M_{\odot}$ and $1.4 M_{\odot}$. This combination was chosen as to be central to the mass parameter space relative to the prior boundaries. The prior is set to be between $2 M_{\odot}$ and $8 M_{\odot}$ in total mass with a minimum individual mass of $1 M_{\odot}$, see Figure 3.2. A range of injected $\delta\chi$ s were used from -0.5 to 0.5 , in order to generate a range of FFs, and the prior set to this range accordingly. The sky location and orientation prior is uniform on the two-sphere and the distance prior is set to be uniform in volume.

The signal waveform, known as the injection, is generated as a discrete frequency series, $\tilde{h}(f_k)$ from an inspiral only frequency domain waveform family. As such the inner product integrals are approximated as discrete Riemann sums.

$$\langle a | b \rangle \equiv 4\Re \left[\int_0^{\infty} \frac{\tilde{a}(f)\tilde{b}^*(f)}{S_n(f)} df \right] \approx 4\Re \left[\sum_{f_k=f_{min}}^{f_{max}} \frac{\tilde{a}(f_k)\tilde{b}^*(f_k)}{S_n(f_k)} \Delta f \right] \quad (3.24)$$

f_{min} is set by sensitive band of the detector, for a signal in advanced Laser Interferometer Gravitational-wave Observatory (LIGO) there is negligible contribution to the sum below $10 Hz$. f_{max} is set to $512 Hz$ to avoid complications to the Fisher matrix approximation caused when signals terminate in band, as reported in [124].

Here we use a simple algorithm to calculate the FF suited for our example cases for a number of test cases in a (reasonably) short time scale. For the event rate estimation we use a simple heirarchical grid based search used to maximise over the masses. To investigate the odds ratio approximation we use the FF calculated from the lalinference maximum likelihoods,

$$FF = \sqrt{\frac{2\Delta L}{\rho^2} + 1}, \quad (3.25)$$

where ΔL is the difference in maximised log likelihoods, see Appendix C.5.

The Fisher Matrix uncertainty $\Delta\theta_{\text{Fisher}}^a$ is calculated from a six-dimensional covariance matrix calculation (masses, time, phase, mass independent amplitude and AG parameter), calculated at the parameter values corresponding to the injected values for GR parameters and $d\chi = 0$, as specified in the derivation by the Nested Sampling algorithm [167] as implemented in `lalinference` [190].

3.5 Results

The fitting factors for the example mass combination “injected” with an SNR of 20 were found to lie between 0.7 and 1, seen in Figure 3.1. As expected when the signals can be included in the GR template bank, i.e. with $d\chi = 0$, the signal has maximum overlap, giving a FF of unity. As the deviation from GR increases the difference in the waveforms can be partially absorbed by a shift from the true parameters, however as the deviation increases the mismatch between the waveforms can no longer be completely absorbed, decreasing the FF. Due to the particular form of the deviation chosen in this study the shift in the symmetric mass ratio is more extreme than the shift in chirp mass. Eventually the parameter shift is halted by the prior boundary, which are set by

computational or physical reasoning, meaning the parameters cannot shift any further to counteract the mismatch, and will rail against the priors as seen in Figure 3.2. In this case this happens for deviations greater than -0.2 or 0.4, hence the sharp decrease in FFs seen in Figure 3.1. In this case these prior bounds come from two distinct effects. The negative deviations hit the physically motivated prior on the symmetric mass ratio which has a maximum of 0.25, a reparameterization of the mass parameters could allow greater freedom within the mass parameter space to absorb these deviations. The positive $d\chi$ hit the combined chirp mass, symmetric mass ratio prior which has been set to allow relative speed of the Bayesian codes, which could be extended if needed. For interest the corresponding fractional loss in detection rate can be seen in Figure 3.3.

Next we compare the Bayes factor from the LALInference codes with that estimated via Equation 3.9. We investigate the dependence of the Bayes Factor on the SNR and FF, as the approximation in Equation 3.9 essentially depends on these two quantities. For a fixed FF of 0.992 we compare the Bayes factors at SNRs of 10, 20, 30 and 40. As shown in Figure 3.4 the approximation gives the expected scaling of the log Bayes factor with square of the SNR at this high FF. Fixing the SNR to 20, we then vary the FF, seen in Figure 3.5. The two methods agree for small variations from GR, i.e. small FFs. However as the FF decreases the approximation begins to progressively over estimate the Bayes factor.

3.6 Discussion and Conclusions

Calculating FFs using a GR template bank to recover an arbitrary AG signal shows the importance of priors in detection. If the deviation shifts the best fit parameters within

Figure 3.1 The plot shows the fitting factors as calculated by a hierarchical grid based search for two compact binary coalescence systems. As expected the fitting factors are unity when the signals are included within the General Relativity model ($d\chi=0$) and then decrease with the deviation from GR. The sharp decrease occurs when the recovered parameters can long shift to absorb the deviation due to the prior bounds.

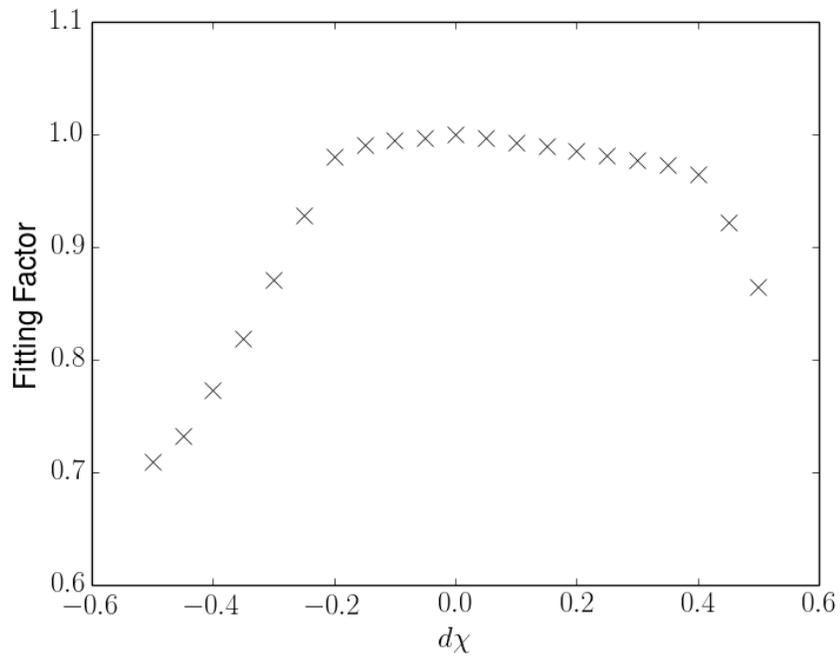
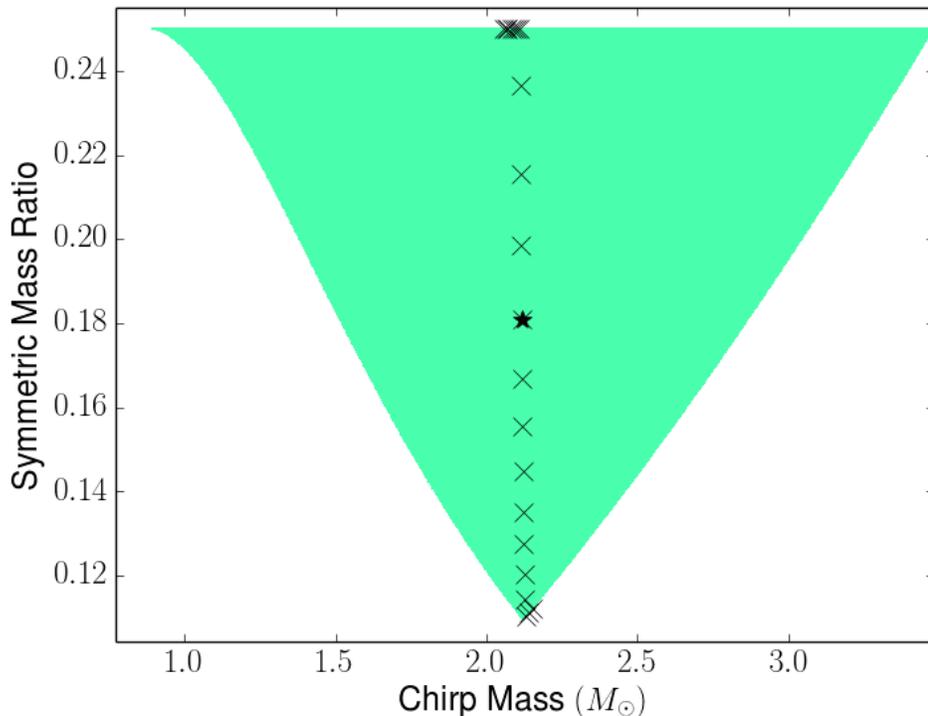
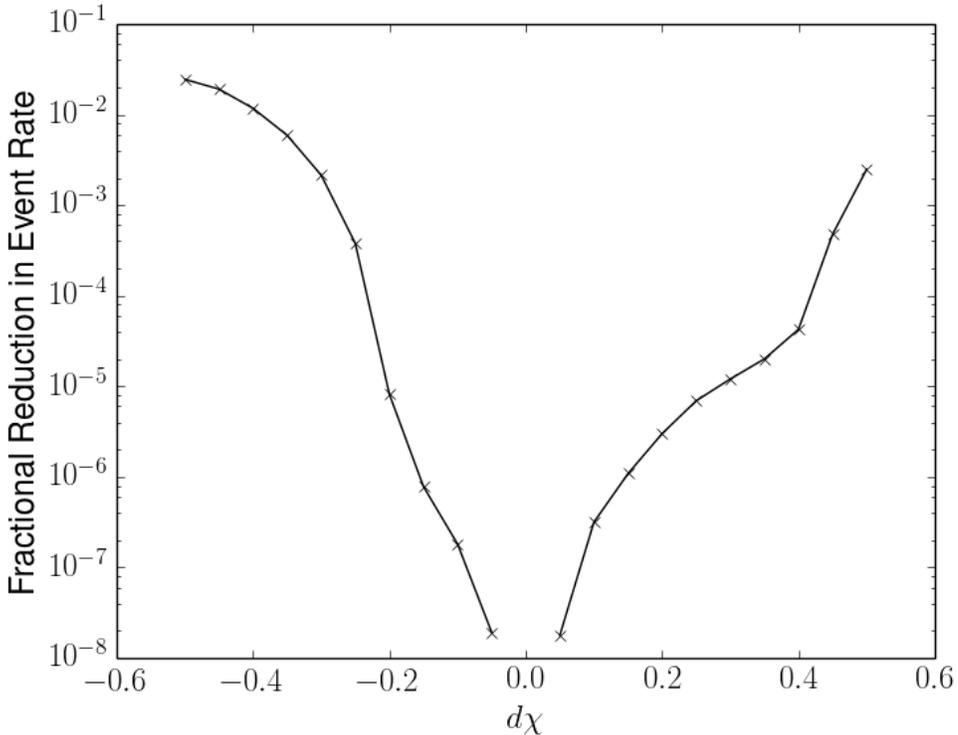


Figure 3.2 The points in mass parameter space at which the maximum match between General Relativity and alternative gravity waveform is found. This shift in masses away from the injected value, starred, allows some of the mismatch caused by the non-zero $d\chi$ to be absorbed. However the prior bounds allow templates only in the shaded region of mass space meaning the mass parameters can't shift indefinitely.



the prior boundary high fitting factors can still be recovered, suggesting the signals are still recoverable with minimal loss in event rate, albeit with false best fit parameters. Comparing the Bayes factor calculated via Lalinference and via an approximate analytical prediction shows agreement when FFs > 0.9 . This is unsurprising as the approximation, as described in [181] assumed a small deviation from GR. Via investigating the scaling of the odds ratio with SNR additional corrections to the approximation are suggested in appendix C. Using this approximation may provide a quick indication as to the possibility of differentiating between GR and a GR signal with out running computationally intensive simulations. When real detections are made the full TIGER

Figure 3.3 The fractional reduction in detection rate ($1 - FF^3$) for the injected AG waveforms which are detected using GR waveforms. The $d\chi = 0$ case is omitted to allow a log scale plotting.



pipeline will be used to analyse these signals.

Figure 3.4 The Bayes factors as calculated via lalinference (bars) and via approximation (circles) as a function of SNR for a fixed fitting factor of 0.992.

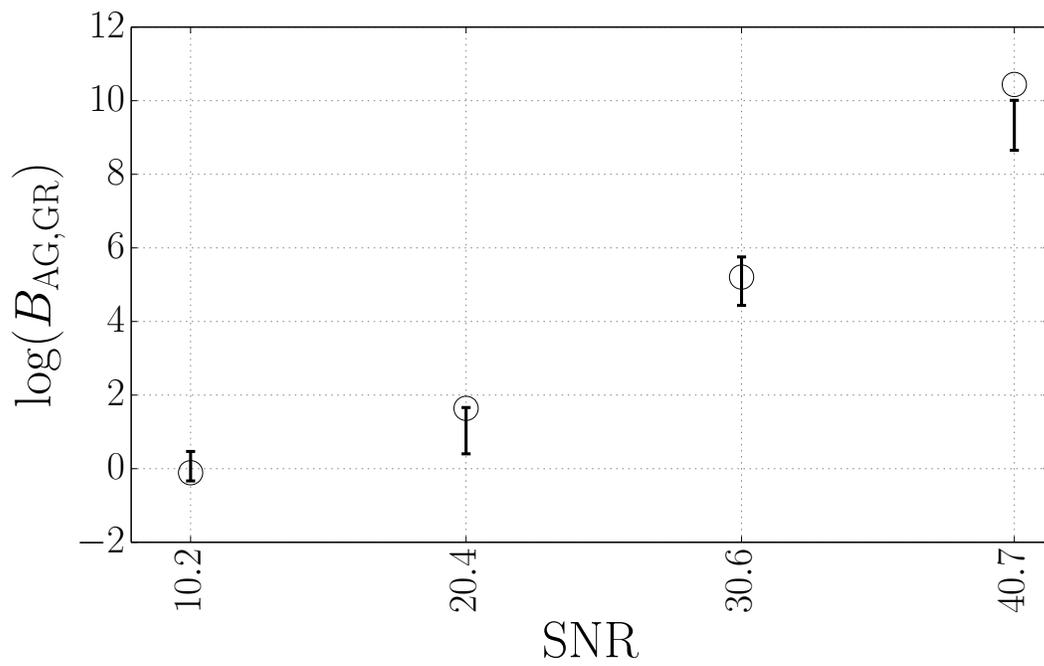
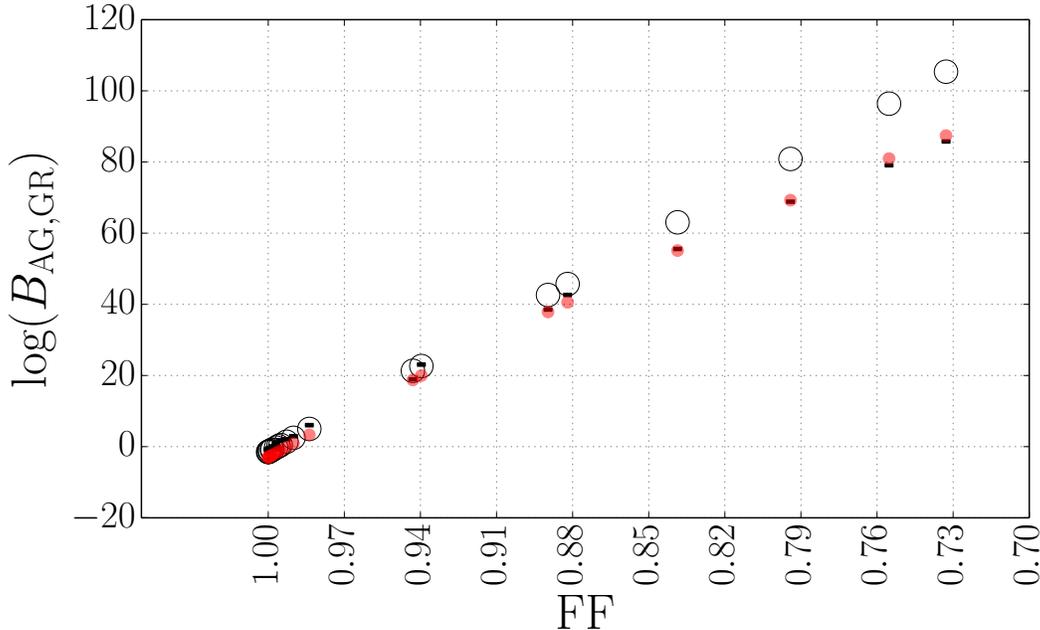


Figure 3.5 The Bayes factors as calculated via lalinference (bars) and via approximation (circles) as a function of Fitting Factor for a fixed SNR of 20. The red dots show the edited version of the approximation, described in Appendix C



Chapter 4

ONLINE CALCULATOR FOR GRAVITATIONAL WAVE DETECTION RATES

The work in this chapter is presented in the form of documentation to support the online tools available at www.sr.bham.ac.uk/gwastro. All text was written by K. Grover. The calculator was jointly created by K Grover, J Klemkowsky, I Mandel, J Tye, W Vousden and T Wantock. J Klemkowsky, J Tye and T Wanock were undergraduate summer students who were supervised by I. Mandel and K. Grover.

The first of the advanced generation of ground based laser interferometer gravitational wave (GW) detectors are schedule to begin taking data in the later half of 2015. One of the most promising sources of GWs in these detectors are those from compact binary coalescence (CBC) sources. However the expected rate of detection of such binaries has large associated uncertainties due to the unknown merger rates. This online tool

provides a way to calculate the expected rate in a selected detector given certain input parameters which control the merger rate of compact binaries and the conditions of detection. The tool is aimed at the interested public as well being useful for those within the scientific community.

4.1 Introduction

General Relativity has passed every test it has undergone and is widely regarded as an accurate description of gravity. One consequence of General Relativity (GR) yet to be directly observed are ripples in space time known as Gravitational waves (GWs). GWs are caused by accelerating asymmetric masses which travel out at the speed of light in every direction from the source, causing distances to stretch and squash as the wave passes. The direct detection of these waves by measuring this distance change has not yet been achieved. However indirect evidence of gravitational waves has been observed, by measuring the energy loss from a binary pulsar system [201].

Ground based laser interferometers are advanced versions of a Michelson interferometers with arms that are kilometres long. A world wide network of advanced detectors hope to make the first direct detections of gravitational waves. These GW detectors are currently being upgraded and will start taking data in late 2015. As GWs interact very weakly with matter we can only hope to detect the most strongly emitting sources of GWs. One category of these potential sources is known as compact binary coalescence (CBC); the merger of black holes (BHs) and/or neutron stars (NSs). If two of these compact objects are in a binary system the binary will lose energy via gravitational wave emission, causing the objects to spiral in towards each other until they eventually merge. The last stages of the CBC inspiral would be “loud” enough to detect directly

in an advanced interferometer detector.

The data output from the detector will be a combination of any GW signal, h , and some unwanted output called noise, n , which comes from a variety of sources, see for example [139]. The output of the detector is simply the sum of any GW signal and the noise,

$$d(t) = h(t) + n(t) . \quad (4.1)$$

For the purposes of data analysis we often assume the noise is stationary and gaussian with zero mean.

CBC signals are “clean”, meaning the form of the GW can be described using relatively few parameters and the form of the wave can be predicted with relative accuracy. This is vital to data analysis techniques which allow the GW signal to be detected and analysed, even when the noise has a higher amplitude than the GW signal. CBC signals are found in the output by filtering $d(t)$ with a range of filters known as a template bank. Templates are simply the predicted waveform (or approximation to it) at different possible parameter combinations. The optimal filter is an accurate waveform generated using the true parameters of the source. The optimally filtered data provides us with an optimal signal to noise ratio (SNR), ρ , which is calculated via the noise weighted inner product,

$$\rho^2 = 4\Re \int_0^\infty \frac{\tilde{h}(f)\tilde{h}^*(f)}{S_n(f)} df . \quad (4.2)$$

Here $\tilde{h}(f)$ is the fourier transform of $h(t)$ and $S_n(f)$ is the single sided noise power spectral density which describes the variance of the noise in the detector.

The criteria for detection in the Laser Interferometer Gravitational-wave Observatory

(LIGO) detector is in reality a function of the data pipeline and real noise in the data, rather than the gaussian stationary idealised scenario we use here. The data analysis pipeline will also vary depending on the detector network and precise detection algorithm, e.g. [22]. We approximate this process by assuming the optimal SNR must be above some threshold, $\rho_{threshold}$, in a single detector network in order to be called a detection. The rate at which such SNRs will be reached depends on the noise level in the detector, the precise form of the GW and the merger rate of binaries in the universe. An example threshold is usually taken to be 8 for CBC signals in a ground based detector, this means we are only accepting high sigma events. This is due to the fact that CBC signals are expected with rates of up to a few per year, while the noise could generate many false alarms per year for a lower threshold after being matched with 10^{13} filters for a template bank of 10^3 filters for a year of data sampled at $1kHz$ [152]. For a false alarm rate less than 10^{-2} per year in aLIGO-AdV a 2 detector network has a SNR threshold of 8.5 and a 3 detector network has a detector threshold of 7 in each detector [115].

4.2 Merger Rate

The only compact binary systems which have been directly observed are galactic binary neutron stars (BNSs) where at least one of the neutron stars is a pulsar. The merger rates of BNS are then extrapolated from the number of these observations that will merge in the Hubble time including the selection effects of the observations. Due to the small sample size and uncertainties in the selection effects, the predicted merger rates of BNS systems varies by several orders of magnitude. We rely on population synthesis models for predicted merger rates where one or more of the objects is a black hole.

These models depend on a number of free parameters, which again results in predicted merger rates which can vary by several orders of magnitude. For a review of these rates see [127].

As a first approximation we can assume merger rates, R_m , in other galaxies scales as the star formation rate, R_s . Star formation rate is often estimated via the blue light luminosity, meaning it is often expressed per L_{10} , which is 10^{10} times the solar blue light luminosity. Here we use a simple direct proportionality between R_s and R_m with constant of proportionality C such that $R_m = CR_s$. However it should be noted that BH mergers can be significantly delayed from the formation of the binary where blue light luminosity becomes a less accurate measure of star formation.

4.2.1 Redshifted merger rate

Star formation rate is not constant throughout the history of the universe. The rates presented in [4] do not account for this fact, as it is a negligible factor for mergers at low redshift. We use, as an example, a star formation rate density as a function of redshift that is presented in [171], for redshifts $z < 5$ this can be approximated to,

$$R_s = \frac{\partial^2 M}{\partial V_c \partial t_{source}} \approx 0.014(1+z)^{1.3} \frac{M_\odot}{\text{Mpc}^3 \text{ yr}}, \quad (4.3)$$

where M is stellar mass, V_c is comoving distance and t_{source} is time at the source.

4.3 Detection Rate

Let us initially ignore redshift effects, given a constant merger rate per volume the detection rate R_D is simply ,

$$R_D = V R_m , \tag{4.4}$$

where V is the detectable volume of space within which a given mass combination has $\rho > \rho_{threshold}$. V is not a sphere as the detector is not equally sensitive to all sky positions and orientations. The GW signal depends on relatively few parameters, some intrinsic and some extrinsic to the source. The intrinsic parameters are the masses and spins of the two compact objects. Here we sometimes use the Chirp Mass \mathcal{M} and symmetric mass ratio η as the mass parameters,

$$\eta = \frac{m_1 m_2}{M^2} , \tag{4.5}$$

$$\mathcal{M} = \eta^{3/5} M , \tag{4.6}$$

where m_1 and m_2 , are the binary's component masses and M is the total mass of the system. We assume non-spinning components. The extrinsic parameters are the four angles which define the position and orientation of the binary relative to the detector, the inclination ι , polarisation phase ψ , polar angle θ and azimuthal angle ϕ . As well as the luminosity Distance to the source, D_L and a reference time and phase, t_c and ϕ_c .

The GW signal is a combination of the two GW polarisations, h_+ and h_\times , weighted by

the detector response functions, F_+ and F_\times .

$$h = \frac{4\mathcal{M}}{D_L} (\pi\mathcal{M}f)^{2/3} \cos(\Phi + \phi_c) \left[\left(\frac{1 + \cos^2\iota}{2} \right) F_+^2 + \cos^2\iota F_\times^2 \right]^{1/2}. \quad (4.7)$$

Φ is the GW phase.

$$F_+ = \frac{1}{2} (1 + \cos^2\theta) \cos 2\phi \cos 2\psi - \cos\theta \sin 2\phi \sin 2\psi, \quad (4.8)$$

$$F_\times = \frac{1}{2} (1 + \cos^2\theta) \cos 2\phi \sin 2\psi + \cos\theta \sin 2\phi \cos 2\psi. \quad (4.9)$$

Let the function Θ encode all angular dependence in the amplitude of the GW signal such that $h \propto \Theta$ and $\Theta = 1$ for an optimally orientated binary.

$$\Theta = \left[\left(\frac{1 + \cos^2\iota}{2} \right) F_+^2 + \cos^2\iota F_\times^2 \right]^{1/2}. \quad (4.10)$$

Let,

$$\rho_1 = \rho(\Theta = 1, D_L = 1Mpc), \quad (4.11)$$

then the maximum distance which is detectable for a given mass combination and Θ is,

$$D_L^{max} = \frac{\rho_1}{\rho_{threshold}} \Theta = D_L^{horizon} \Theta \text{ Mpc}, \quad (4.12)$$

where the horizon distance, $D_L^{horizon}$, is defined as the maximum distance which is detectable for an optimally orientated binary. If we assume the merger rate, R_m , is

a constant in space and time, we can imagine how a sphere with radius $D_L^{horizon}$ is fractionally reduced due to random binary positions and orientations by averaging Θ across the whole sky and possible orientations.

$$V = \overline{\frac{4}{3}\pi D_L^{max^3}} = \frac{4}{3}\pi \left(\frac{\rho_1}{\rho_{threshold}} \right)^3 \overline{\Theta^3} \quad (4.13)$$

4.3.1 Redshift

The detection rate calculation described above is a fair estimation of detection rate for distances at which redshifts can be neglected, however advanced detectors are expected to make detections out to cosmological distances, i.e. where $z \sim 1$. The effect this has on star formation rate is discussed above and redshift also has an effect on the detected GW signal. The propagation of the GW over cosmological distances acts like a factor of $(1+z)$ on the masses, so although redshift has no physical effect on the masses of the system the observed wave has “redshifted masses”, such that $\mathcal{M}_{obs} = (1+z)\mathcal{M}_{source}$. Additionally the observed frequency is redshifted with respect to the source frequency, $f_{obs} = \frac{f_{source}}{1+z}$.

When discussing distances and volumes at cosmological distances the distinction between distance measures must be made. Here we are interested in the comoving volume, V_c , i.e. a volume calculated using the comoving distance,

$$D_c = \int_0^z \frac{D_H}{E(z')} dz' , \quad (4.14)$$

where D_H is the Hubble distance, or the inverse of the Hubble parameter, taken here

to be $H_0 = 70.4 \text{ km s}^{-1} \text{ Mpc}^{-1}$. $E(z) \equiv \sqrt{\Omega_M(1+z)^3 + \Omega_\Lambda}$, where Ω_M and Ω_Λ are dimensionless matter and vacuum density respectively, here taken as $\Omega_M = 0.272$ and $\Omega_\Lambda = 0.728$ (we assume a flat universe). The luminosity distance is then

$$D_L = D_c(1+z) . \quad (4.15)$$

Finally when calculating the rates we must not forget that $dt_{obs} = (1+z)dt_{source}$. For detailed derivations of the effect of a CBC signal at cosmological distances see Chapter 4.1.4 of [122] and for more information on cosmological distance scales see [92].

Therefore the detection rate is now,

$$R_D = \int_0^{z_{max}} R_s(z) \frac{dV_c}{dz} f_d(z) \frac{1}{1+z} dz , \quad (4.16)$$

where z_{max} is the redshift at which a redshifted optimally orientated binary has an SNR at threshold, f_d is the fraction of binaries that are above threshold due to non-optimal orientation and the factor $\frac{1}{1+z}$ is the difference in rate at the source and observer due to cosmic time dilation.

Figure 4.1 The user interface for the online rates calculator. There are five astrophysical options, redshift effects can be turned on or off, the star formation rate and conversion to merger rate, the maximum redshift to which to count mergers and the example masses to be used for detection rates.

Parameters

Astrophysical parameters

Redshift Effects	<input checked="" type="radio"/> Include Redshift Effects <input type="radio"/> No Redshift Effects
Star Formation Rate (SFR)	<input checked="" type="radio"/> Constant SFR <input type="text" value="3.5e6"/> Solar Mass/MWEG/Myr ↕ <input type="radio"/> Redshift Dependent SFR <input type="text" value="1.2e6"/> (1 + z) ^{<input type="text" value="1.3"/>} Solar Mass/MWEG/Myr ↕
Conversion Factor from SFR to Merger Rate	Binary Neutron Stars <input type="text" value="3e-5"/> /Solar Masses Binary Black Holes <input type="text" value="1e-7"/> /Solar Masses Neutron Star-Black Hole Binaries <input type="text" value="1e-6"/> /Solar Masses
Max. redshift	<input type="text" value="5"/>
Component masses	<input checked="" type="checkbox"/> Binary Neutron Stars <input type="text" value="1.4"/> Solar Masses and <input type="text" value="1.4"/> Solar Masses <input checked="" type="checkbox"/> Binary Black Holes <input type="text" value="10.0"/> Solar Masses and <input type="text" value="10.0"/> Solar Masses <input checked="" type="checkbox"/> Neutron Star-Black Hole <input type="text" value="1.4"/> Solar Masses and <input type="text" value="10.0"/> Solar Masses

4.4 The Calculator

4.4.1 Inputs

Astrophysical Parameters

Redshift Effects

Selecting “No Redshift Effects” performs a simple calculation which does not take into account the effect of redshift on star formation or the difference between luminosity and co-moving distances.

Star Formation Rate

The user can select a constant or redshift dependent star formation rate, each can be specified in the following units,

- Solar Mass per Milky Way Equivalent Galaxy per Mega-year,
- Solar Mass per L_{10} per Mega-year,
- Solar Mass per Mega-parsec cubed per mega year.

The default star formation rate for the constant option is that used in [4]. If choosing a redshift dependent star formation rate it must be in the form, $\alpha(1+z)^\beta$ with defaults set as described in Equation 4.3, $\alpha = 0.014 \frac{M_s}{Mpc^3 yr}$ and $\beta = 1.3$.

Conversion Factor from SFR to Merger Rate

The constant of proportionality between star formation rate and merger rate controls the number of mergers of each type of binary. This number is physically set by the delay from formation of the compact objects to their merger and is estimated via population synthesis models. The default values are set to those used in [4], $C_{BNS} = 3 \times 10^{-5}$,

$$C_{BBH} = 1 \times 10^{-7} \text{ and } C_{NSBH} = 1 \times 10^{-6}.$$

Max. Redshift

If redshift effects are turned on the maximum redshift sets the maximum distance at which mergers are counted, this simply stops the calculation from taking too long, especially if large masses are chosen.

Component Masses

The user may choose up to three types of system for which to estimate the detection rate. The NSs must have masses between $1M_{\odot}$ and $2.5M_{\odot}$ (default $1.4M_{\odot}$) and the BHs must have masses between $2.5M_{\odot}$ and $100M_{\odot}$ (default $10M_{\odot}$).

Detection Parameters

Figure 4.2 The user interface for the online rates calculator. There are four detection options, the detector and its operating mode, the waveform family with which to calculate the SNR, the minimum frequency to begin the SNR and the threshold SNR above which a detection is counted.

Detection parameters	
Detector/operating mode	AdvLIGO: Zero Detuning High Power ▾
Waveform approximant	<input checked="" type="radio"/> Time Domain Waveforms TaylorT1 ▾ <input type="radio"/> Frequency Domain Waveforms TaylorF2 ▾
Starting frequency	20.0 Hz
Threshold SNR	8.0

Detector/Operating Mode

The user can select from a choice of detectors and operating modes. This choice sets the level of the noise, the stronger the noise the lower the signal to noise ratio and fewer detections will be made. Users can view the exact form of the noise in the detector noise PSD plot in the results section. The current choices are a variety of operating modes of Advanced LIGO [87] or variants of the Einstein Telescope [141].

Waveform Approximant

The SNR of the signal is calculated using a waveform approximant. These can be generated in the time or frequency domain. If the waveform is generated in the time domain a fast Fourier transform (FFT) is used to transform to the frequency domain for the computation.

Starting Frequency

Although in principle the SNR is calculated from $f = 0$ to $f = \infty$ this is not suitable for the discrete computation used here. The maximum frequency is set to the innermost stable circular orbit (ISCO) frequency and the starting frequency is set here by the user. This should be around the frequency at which the noise power spectral density (PSD) drops in order to not significantly under estimate the SNR. As lowering the minimum frequency can dramatically increase the length of the calculation the calculator may change the users choice if the waveform will be over 120s in length, a message will inform the user if this is the case. The default value is 20 Hz .

Threshold SNR

The threshold SNR is the way we approximate the detection criteria, which in reality can only be calculated empirically depending on the real noise, detector network and data analysis algorithm. The default value is 8, as commonly used to calculate horizon distances, e.g. [4].

4.4.2 Calculation

Waveform and SNR

The waveforms are generated using the LIGO Scientific Collaboration (LSC) algorithm library (LAL), if a time domain waveform is chosen a FFT is used to transform it to the frequency domain. The waveforms are generated at a distance of $1 Mpc$ as the simple $1/D_L$ scaling allows each distance to be calculated without the computationally intensive process of waveform generation being repeated. The SNR is calculated using a discrete form of Equation 4.2 summed between the minimum frequency set by the user and the ISCO frequency.

No Redshift Effects

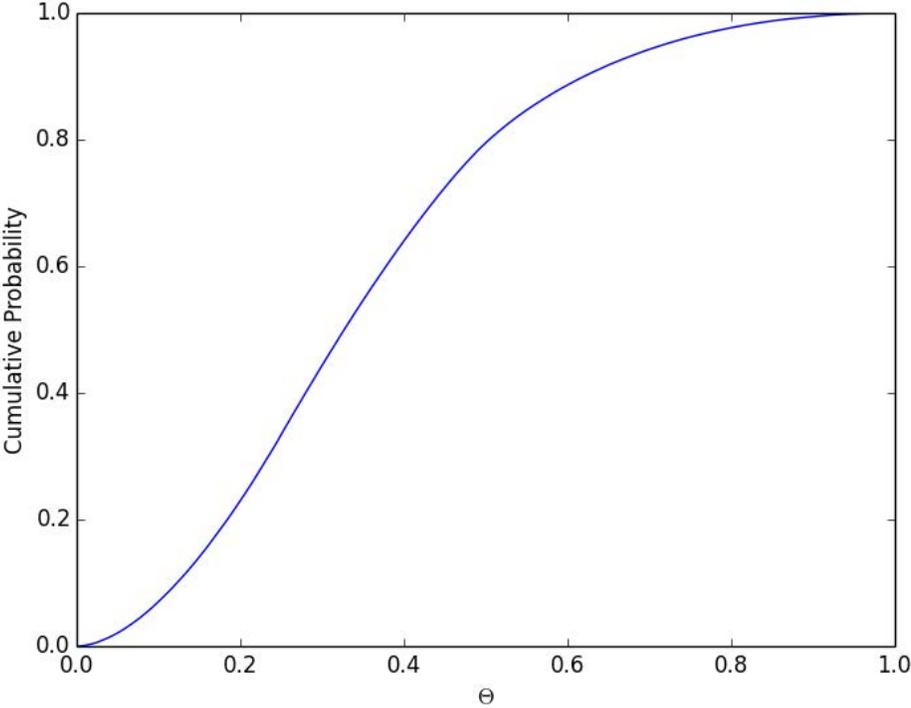
If no redshift effects are included the calculation is a simple one, the total detection rate is the merger rate times the detectable volume. The merger rate is the star formation rate multiplied by the conversion factor both supplied by the user. The volume is calculated as shown in Equation 4.13, using the threshold SNR set by the user. $\overline{\Theta^3}$ has been calculated via a Monte Carlo to be 0.086, assuming the following are uniformly distributed $\cos \iota[-1, 1]$, $\cos \theta[-1, 1]$, $\phi[0, 2\pi]$ and $\psi[0, 2\pi]$.

4.4.3 Include Redshift Effects

As described in Equation 4.16 the detection rate is a integration over a complex function of redshift. Therefore the calculator works by computing the detection count per year at discrete redshift shells, summing the total until an optimally oriented binary no

longer reaches the threshold SNR. The fraction, f_d , that will be above threshold due to sky position and orientation is determined via a precalculated, Monte Carlo generated, cumulative distribution function, seen in Figure 4.3.

Figure 4.3 Θ cumulative distribution function generated via a Monte carlo of 1000000 randomly distributed sky locations and orientations.



4.4.4 Outputs

Numerical Outputs

Detection Rate

The result of the calculation as described in the previous sections are shown for each of the selected mass combinations as well as the total sum.

Distance

The horizon distance is that at which an optimally orientated binary is at threshold SNR, i.e the furthest distance this detector can detect a binary of a given mass combination. This is shown for each of the selected masses.

Figure 4.4 The default outputs of the rates calculator including the number of detections and horizon distance for each of the mass systems.

Results

Total Detection rate	54 per year
Detection rate BBH	8 per year
Detection rate BNS	38 per year
Detection rate NSBH	8 per year
Horizon luminosity distance BNS	488 Mpc
Horizon luminosity distance BBH	2.59e+03 Mpc
Horizon luminosity distance NSBH	1e+03 Mpc

Plots

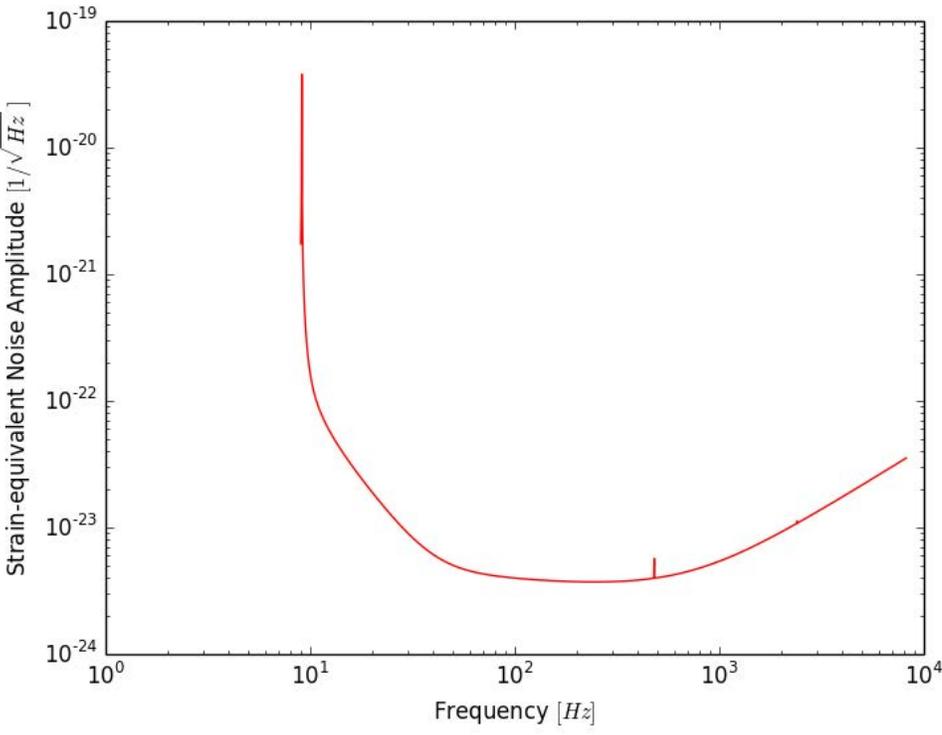
Noise amplitude spectral density

This plot shows the strain equivalent noise amplitude for the detector and operating mode that the user has selected, the default is shown as an example in Figure 4.5.

Frequency Domain Waveforms

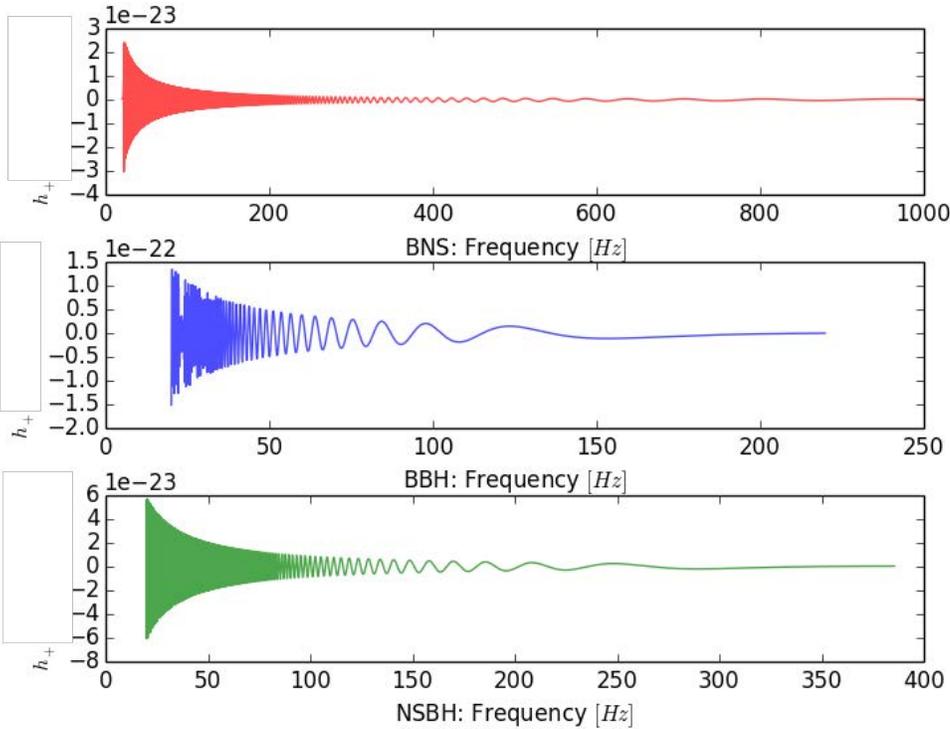
The frequency domain waveforms are plotted for the selected mass systems at a dis-

Figure 4.5 The default noise amplitude spectral density used in the rates calculator. This is the predicted noise curve for the high power zero detuning mode of the advanced LIGO detectors.



tance of 1 Mpc with an optimal orientation and sky position. If a frequency domain waveform is requested the waveform is plotted as generated by the LAL. If a time domain waveform is requested the frequency domain waveform is generated via a FFT of the waveform generated by the LAL. The default is shown as an example in Figure 4.6.

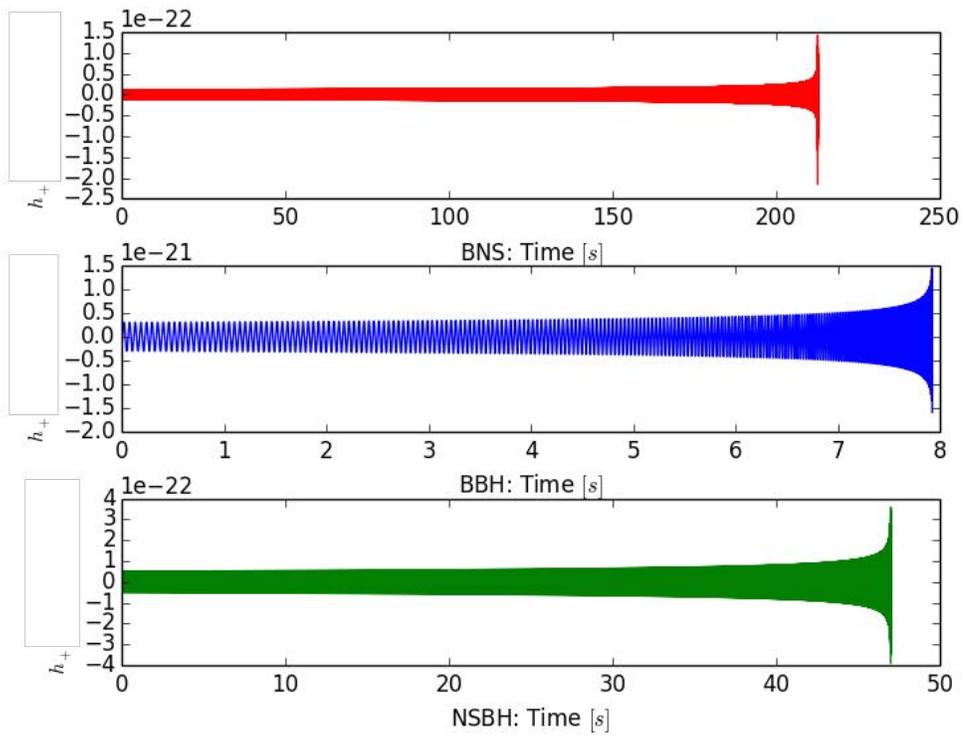
Figure 4.6 The default frequency domain waveforms generated by a fast fourier transform of a TaylorT1 waveform generated by LAL for the three mass systems. In each case the real part of the waveform is plotted.



Time Domain Waveforms

The time domain waveforms are plotted for the selected mass systems at a distance of 1 Mpc with an optimal orientation and sky position. If a frequency domain waveform is requested the time domain waveform plot is not generated as it is not needed for the rates calculation. The default is shown as an example in Figure 4.7

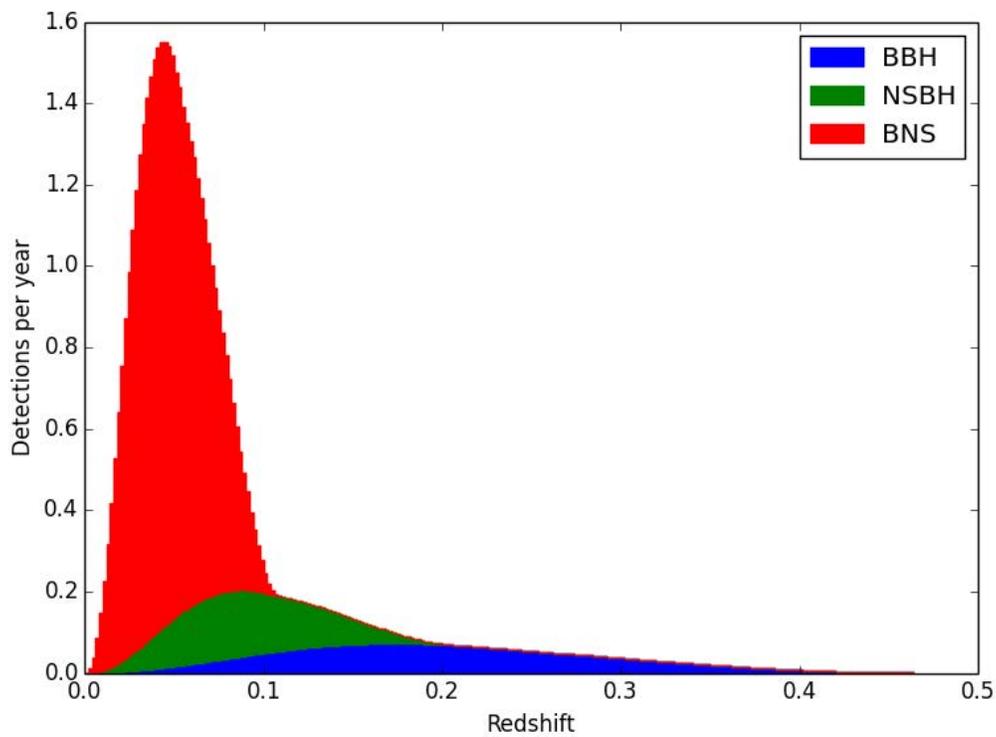
Figure 4.7 The default time domain waveforms, a TaylorT1 waveform generated by LAL for the three mass systems.



Detections per Redshift Shell

If redshift effects are turned on the detection rate per year at each discrete redshift shell is shown as a stacked histogram for all the mass systems that are selected. The total number of detections is the sum of these rates. The default is shown as an example in Figure 4.8

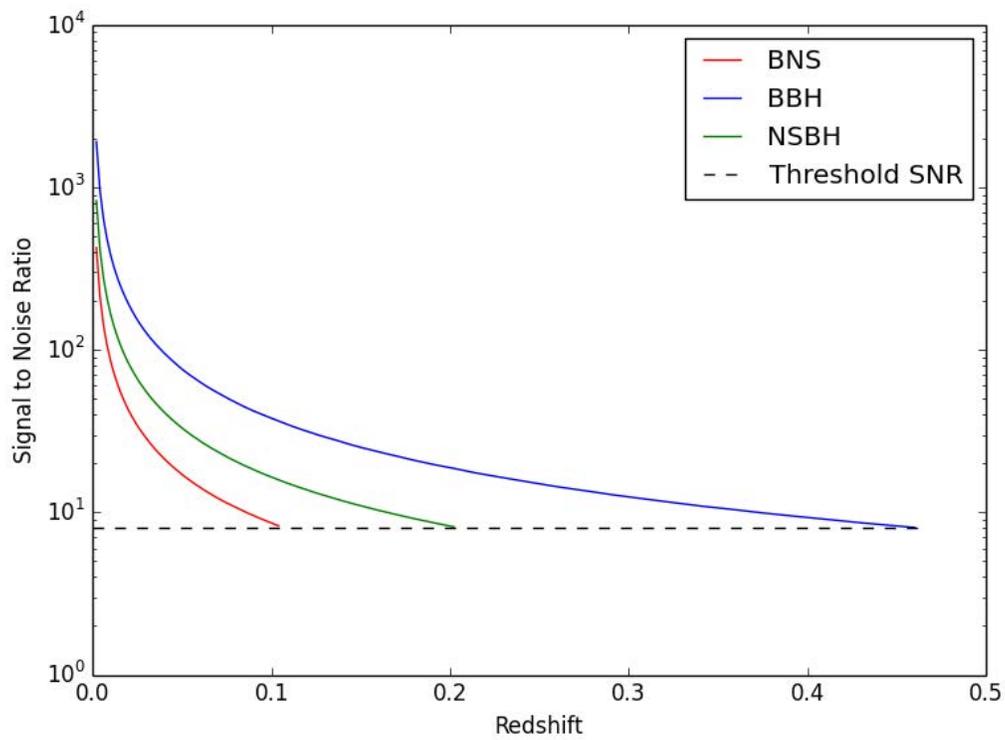
Figure 4.8 The default detection rate per discrete redshift shell for each of the three mass systems.



SNR as a Function of Redshift

If redshift effects are turned on the signal to noise ratio for an optimally oriented binary is calculated at each discrete redshift shell and plotted here for each of the selected mass systems, the default is shown as an example in Figure 4.9.

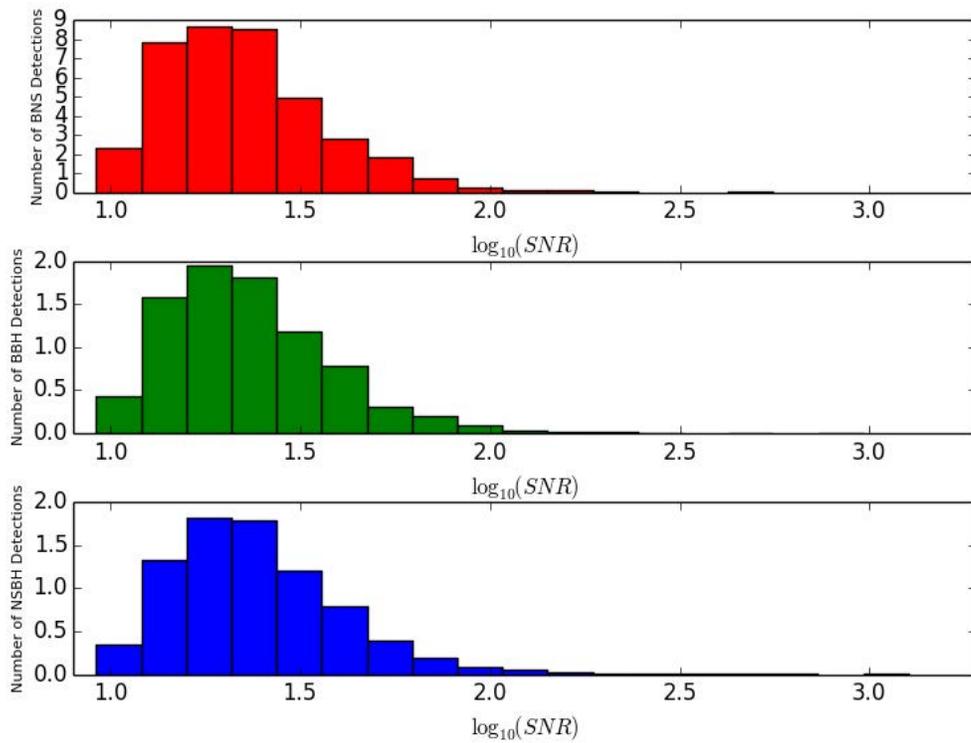
Figure 4.9 The default Signal to Noise Ratio for an optimally orientated binary as a function of redshift for each of the three mass systems.



Distribution of SNRs

If redshift effects are turned on the number of detections per year is shown as a distribution of SNR for each of the selected systems, the default is shown as an example in Figure 4.10.

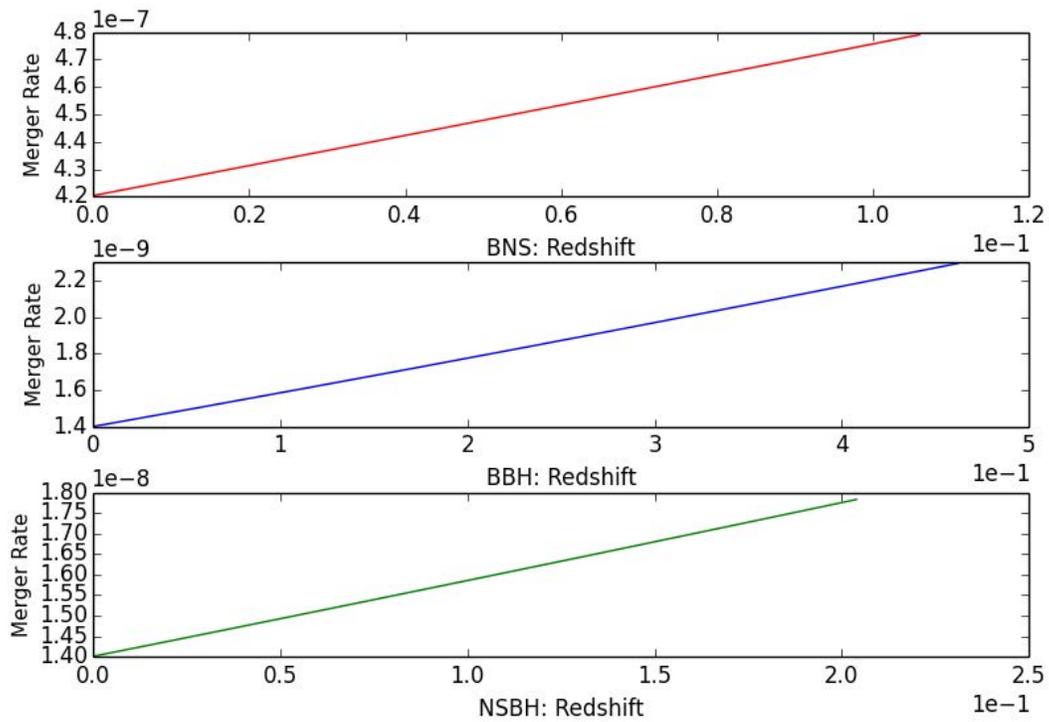
Figure 4.10 The default distribution of Signal to Noise Ratios that are detected for the three mass systems.



Merger Rate as a Function of Redshift

If a redshift dependant star formation rate is selected the merger rate is plotted as a function of redshift, the default is shown as an example in Figure 4.11.

Figure 4.11 The default merger rate as a function of redshift for the three mass systems.



4.5 Comparison to Published Results

Many of the default input parameters are taken from the calculations and results presented in [4], so a comparison of outputs of the calculator to these numbers is shown in table 4.1 . The outputs of the rates calculator match fairly closely with the published results, although the binary black hole (BBH) rate seems to be slightly low when using the same merger rate. However it should be noted that this calculation varies to the order of a few depending on the waveform family selected, as some include the merger and ringdown portion of the signal which can significantly contribute to SNR. It should be noted however that redshifting the waveforms in this case does not significantly affect the results, and this effect is negligible compared to errors on Merger Rates. Using star formation rate with the redshift dependence estimated from results published [171] does cut rates by around two-thirds, but again this falls within the uncertainty from merger rates.

Table 4.1 The table shows the plausible pessimistic R_D^{low} , most likely, R_D^{likely} and plausible optimistic R_D^{high} rates of detection as published in [4], which do not include redshift effects, along side some example outputs of the rates calculator. We use a TaylorF2 waveform with no redshift effects, R_D , including redshifted waveforms but constant star formation rate R_D^z and finally redshifted waveforms and redshift dependent star formation rate, $R_D^{z,SFR}$.

Source	R_D^{low}	R_D^{likely}	R_D^{high}	R_D	R_D^z	$R_D^{z,SFR}$
BNS	0.4	40	400	37	34	12
Neutron star - black hole (NSBH)	0.2	10	300	11	9	3
BBH	0.4	20	1000	15	10	4

Chapter 5

FUTURE WORK

5.1 Sky Localisation

As described in Chapter 2 approximate methods to characterise the sky localisation ability of detector networks can provide useful upper limits. However approximate methods still differ by at least 50% from full results and there is a large variation in individual results. Any studies going forward are best suited to use Bayesian inference techniques with which to characterise the network. A recent work investigates the sky localisation capabilities of the advanced detector network using a rapid sky localisation technique, BAYESTAR, which is shown to be as accurate as full parameter estimation for Binary Neutron Star signals [166].

5.2 Testing General Relativity

Testing GR using gravitational wave signals looks to be a challenging and complex endeavour as the advanced detection era approaches. The TIGER pipeline has been tested on modelled and unmodeled deviations from GR as described in Appendix B and has been shown to be robust against a number of unmodeled effects for BNS signals in [17]. However there is still large uncertainties as the expected level of unmodeled effects that could effect signals, both within GR signals and deviations from it, for example testing the pipeline in real noise from the advanced detectors.

5.3 Rates Calculator

The online rates calculator as described in Chapter 4 is a useful tool for the general public and scientific community to quickly check the impact of various inputs to the expected rates of CBC signals in ground based interferometers. There are various improvements that could be made to improve the accuracy and functionality of the calculator, described below.

5.3.1 Astrophysical Parameters

The current calculator only allows a redshift dependent star formation rate (SFR) in the form, $\alpha(1+z)^\beta$, where z is the redshift, α and β are free parameters set by the user. It is likely that some users may want to enter functions with other redshift dependencies, especially with more sensitive detector that has a higher redshift horizon, where it is known that this approximation no longer holds.

Currently users can select up to three example masses for the CBC systems they are interested in, the rates are then calculated using the SNR calculated from waveforms with exactly these mass parameters. A more accurate scenario would be to allow the user to input a mass distribution from which to calculate the rates. Similarly the sources in the current calculator are assumed to be non-spinning compact objects. As a first step it could be possible for the user to enter an example spin, like the current options for the masses. However going forward there should again be a spin distribution input option.

5.3.2 Detection Parameters

The current “detector/operating mode” option allows the user to select from a variety of predicted advanced LIGO and Einstein Telescope (ET) operating modes. As advanced LIGO comes online and more accurate noise curves become available it may be possible to implement them in the calculator as well including those from other detectors such as the Virgo detector, GEO600 and Kagra. It may be useful for the user to see this options graphically before selection. The waveform approximant section should be kept up to date with the most accurate waveform family implemented within the LAL.

5.3.3 Calculation

The calculation can be slow for time domain waveforms, which is not ideal for a web based application as casual users are unlikely to stick with the page for calculations even as long as a minute. This is due to the computational cost of performing the Fourier transform on the time domain waveform (the reverse is not carried out for frequency domain waveforms). To mitigate this problem currently long waveforms are truncated,

via increasing the minimum frequency, however this technique underestimates the SNR of the signal and therefore underestimates the rates. A better solution would be to implement a faster FFT routine (the current code uses standard Python libraries which are known to be slow).

5.3.4 Outputs

The current outputs of the calculator are fairly comprehensive, however they could be improved by outputting the horizon as a redshift and comoving volume, which would be of interest to astrophysicists.

Appendix A

OBSERVING THE DYNAMICS OF SUPERMASSIVE BLACK HOLE BINARIES WITH PULSAR TIMING ARRAYS

The work in this Appendix is presented in the form of the paper Published in Physical Review Letters [130]

Text was written by C. M. F. Minagrelli, K. Grover, T. Sidery, R. Smith, A. Vecchio. K. Grover was involved in discussions for all parts of this paper but did not produced any of the key figures or tables in the final paper. K. Grover performed the calculation to ensure correct mass and frequencies were chosen for example systems to ensure separation of the Earth and pulsar terms in frequency space.

Pulsar timing arrays are a prime tool to study unexplored astrophysical regimes with

gravitational waves. Here, we show that the detection of gravitational radiation from individually resolvable supermassive black hole binary systems can yield direct information about the masses and spins of the black holes, provided that the gravitational-wave-induced timing fluctuations both at the pulsar and at Earth are detected. This in turn provides a map of the nonlinear dynamics of the gravitational field and a new avenue to tackle open problems in astrophysics connected to the formation and evolution of supermassive black holes. We discuss the potential, the challenges, and the limitations of these observations.

A.1 Introduction

Gravitational waves (GWs) provide a new means for studying black holes and addressing open questions in astrophysics and fundamental physics: from their formation, evolution and demographics, to the assembly history of galactic structures and the dynamical behaviour of gravitational fields in the strong non-linear regime. Specifically, GW observations through a network of radio pulsars used as ultra-stable clocks – Pulsar Timing Arrays (PTAs) [68, 73, 153] – represent the only direct observational avenue for the study of supermassive black hole binary (SMBHB) systems in the $\sim 10^8 - 10^9 M_\odot$ mass range, with orbital periods between ~ 1 month and a few years, see e.g. [156, 198] and references therein. Ongoing observations [77, 91, 99, 194] and future instruments, e.g. the Square Kilometre Array [3], are expected to yield the necessary timing precision [116, 195] to observe the diffuse GW background. This background is likely dominated by the incoherent superposition of radiation from the cosmic population of massive black holes [66, 88, 98, 100, 142, 157, 159, 188, 209] and within it, we expect a handful of sources that are sufficiently close, massive and high-frequency to be individually

resolvable [32, 71, 72, 101, 110, 160, 161, 203, 210].

Massive black hole formation and evolution scenarios [108, 123, 199, 211] predict the existence of a large number of SMBHBs. Furthermore, SMBHBs are expected to be (possibly rapidly) spinning [94, 128]. In fact the dynamics of such systems – which according to general relativity are entirely determined by the masses and spins of the black holes [132] – leave a direct imprint on the emitted gravitational waveforms. From these, one could measure SMBHB masses and their distribution, yielding new insights into the assembly of galaxies and the dynamical processes in galactic nuclei [203]. Moreover, measuring the magnitude and/or orientation of spins in SMBHBs would provide new information on the role of accretion processes [41, 69, 81, 137, 200]. Finally, detections of SMBHBs could allow us to probe general relativistic effects in the non-linear regime in an astrophysical context not directly accessible by other means, see [140, 172, 207], and references therein.

The observation of GWs with PTAs relies on the detection of the small deviation induced by gravitational radiation in the times of arrival (TOAs) of radio pulses from millisecond pulsars that function as ultra-stable reference clocks. This deviation, called the residual, is the difference between the expected (without GW contribution) and actual TOAs once all the other physical effects are taken into account. The imprint of GWs on the timing residuals is the result of how the propagation of radio waves is affected by the GW-induced space-time perturbations along the travel path. It is a linear combination of the GW perturbation at the time when the radiation transits at a pulsar, the so-called “pulsar term”, and then when it passes at the radio receiver, the “Earth term” [68, 73, 153]. The two terms reflect the state of a GW source at two different times of its evolution separated by $\tau \equiv (1 + \hat{\Omega} \cdot \hat{\mathbf{p}}) L_p \sim 3.3 \times 10^3 (1 + \hat{\Omega} \cdot \hat{\mathbf{p}}) (L_p/1 \text{ kpc}) \text{ yr}$ where $\hat{\Omega}$ and $\hat{\mathbf{p}}$ are the unit vectors that identify the GW propagation

direction and the pulsar sky location at a distance L_p from the Earth, respectively, see e.g. [161]. [We use geometrical units in which $G = c = 1$.] In a network (array) of pulsars all the perturbations at the Earth add coherently and therefore boost the signal-to-noise ratio (S/N) of the signal. Each pulsar term is at a slightly different frequency since the orbital period of the binary evolves over the time τ .

Measuring the key physics of SMBHBs is hampered by the short (typically $T = 10$ yr) observation time compared to the typical orbital evolution timescale $f/\dot{f} = 1.6 \times 10^3 (\mathcal{M}/10^9 M_\odot)^{-5/3} (f/50 \text{ nHz})^{-8/3}$ yr of binaries that are still in the weak field adiabatic inspiral regime, with an orbital velocity $v = 0.12 (M/10^9 M_\odot)^{1/3} (f/50 \text{ nHz})^{1/3}$ [138]. Here $M = m_1 + m_2$, $\mu = m_1 m_2 / M$ and $\mathcal{M} = M^{2/5} \mu^{3/5}$ are the total, reduced and chirp mass, respectively, of a binary with component masses $m_{1,2}$, and f is the GW emission frequency at the leading quadrupole order. The chirp mass determines the frequency evolution at the leading Newtonian order. In the post-Newtonian (pN) expansion of the binary evolution [49] in terms of $v \ll 1$, the second mass parameter enters at p¹N order ($\mathcal{O}(v^2)$ correction); spins contribute at p^{1.5}N order and above ($\mathcal{O}(v^3)$) causing the orbital plane to precess through spin-orbit coupling, at leading order. These contributions are therefore seemingly out of observational reach.

The GW effect at the pulsar – the pulsar term – may be detectable in future surveys, and for selected pulsars their distance could be determined to sub-parsec precision [64, 110, 169]. If this is indeed the case, it opens the opportunity to coherently connect the signal observed at the Earth and at pulsars, therefore providing snapshots of the binary evolution over $\sim 10^3$ yr. These observations would drastically change the ability to infer SMBHB dynamics, and study the relevant astrophysical process and fundamental physics.

In this Letter we show that for SMBHBs at the high end of the mass and frequency

spectrum observable by PTAs, say $m_{1,2} = 10^9 M_\odot$ and $f = 10^{-7}$ Hz, the observations of a source still in the weak-field regime become sensitive to post-Newtonian contributions up to p^{1.5}N, including spin-orbit effects, if both the pulsar and Earth term can be detected. This in principle enables the measurement of the two mass parameters and a combination of the spin’s magnitude and relative orientation. We also show that the Earth-term only can still be sensitive to spin-orbit coupling due to geometrical effects produced by precession. We discuss the key factors that enable these measurements, and future observational prospects and limitations.

A.2 Signals from SMBHBs

Consider a radio pulsar emitting radio pulses at frequency ν_0 in the source rest-frame. GWs modify the rate at which the radio signals are received at the Earth [68, 73, 153], inducing a relative frequency shift $\delta\nu(t)/\nu_0 = h(t - \tau) - h(t)$, where $h(t)$ is the GW strain. The quantities that are actually produced at the end of the data reduction process of a PTA are the timing residuals, $\int dt' \delta\nu(t')/\nu_0$, although without loss of generality, we will base the discussion on $h(t)$. The perturbation induced by GWs is repeated twice, and carries information about the source at time t , the “Earth term”, and at past time $t - \tau$, the “pulsar term”.

We model the radiation from a SMBHB using the so-called restricted pN approximation, in which pN corrections are included only in the phase and the amplitude is retained at the leading Newtonian order, but we include the leading order modulation effects produced by spin-orbit coupling. The strain is given by

$$h(t) = -A_{\text{gw}}(t)A_{\text{p}}(t) \cos[\Phi(t) + \varphi_{\text{p}}(t) + \varphi_{\text{T}}(t)], \quad (\text{A.1})$$

where $A_{\text{gw}}(t) = 2[\pi f(t)]^{2/3} \mathcal{M}^{5/3}/D$ is the Newtonian order GW amplitude, $\Phi(t)$ is the GW phase, see e.g. Eq. (232, 234) in [49] and Eq. (8.4) in [44], and D is the distance to the GW source. $A_p(t)$ and $\varphi_p(t)$ are the time-dependent polarisation amplitude and phase and $\varphi_{\text{T}}(t)$ is an additional phase term, analogous to Thomas precession, see Eq. (29) in [23].

The physical parameters leave different observational signatures in the GW strain $h(t)$ and are therefore found in the TOA residuals. At the leading Newtonian order, \mathcal{M} drives the frequency and therefore the phase $\Phi(t)$ evolution, with the second independent mass parameter entering from the p¹N onwards. SMBHs are believed to be rapidly spinning, and the spins are responsible for three distinctive imprints in the waveform: (i) they alter the phase evolution through spin-orbit coupling and spin-spin coupling at p^{1.5}N and p²N order, respectively [107], (ii) they cause the orbital plane to precess due to (at lowest order) spin-orbit coupling and therefore induce amplitude and phase modulations in the waveform through $A_p(t)$ and $\varphi_p(t)$; and (iii) through orbital precession they introduce an additional secular contribution $\varphi_{\text{T}}(t)$ to the waveform phase. Astrophysically we expect PTAs to detect SMBHBs of comparable component masses [161]. We therefore model the spin-orbit precession using the simple precession approximation [23], which formally applies when $m_1 = m_2$, or when one of the two spins is negligible with respect to the other. Let $\mathbf{S}_{1,2}$ and \mathbf{L} be the black holes' spins and the orbital angular momentum, respectively. Then both $\mathbf{S} = \mathbf{S}_1 + \mathbf{S}_2$ and \mathbf{L} precess around the (essentially) constant direction of the total angular momentum, $\mathbf{J} = \mathbf{S} + \mathbf{L}$, at the same rate $d\alpha/dt = \pi^2 (2 + 3m_2/(2m_1)) (|\mathbf{L} + \mathbf{S}|)f^2(t)/M$ [23], where α is the precession angle, while preserving the angle of the precession cone, λ_L , see Fig. 4 of Ref. [23]. This approximation is adequate to conceptually explore these effects, however in the case of real observations, one will need to consider the exact expressions [106]. We also assume the binaries are circularised and have zero eccentricity. For this simple precession case

the waveforms are derived in [23], the equations are lengthy and not repeated here but the interested reader is pointed to eqns (64a-c) in [23].

The detection and particularly the measurement of the aforementioned parameters relies on coherently matching the signal with a template that faithfully reproduces its amplitude and, importantly, its phase evolution. We therefore consider the contribution to the total number of wave cycles a proxy for the significance of a specific parameter. Individual terms that contribute ~ 1 GW cycle or more mean that the effect is in principle detectable, hence one can infer information about the associated parameter(s). We show that information about the parameters can only be inferred for SMBHBs at the high end of the mass spectrum and PTA observational frequency range. Having a sufficiently high-mass and high-frequency GW source is also essential to ensure sufficient frequency evolution over the time τ , so that the Earth and pulsar term are clearly separated in frequency space cf. Table A.1. We therefore take fiducial source parameters of $m_1 = m_2 = 10^9 M_\odot$, frequency at the Earth at the beginning of the observation $f_E = 10^{-7}$ Hz and an observational time $T = 10$ years to illustrate the main results. We provide scaling relations as a function of the relevant quantities, allowing the reader to rescale the results for different astrophysical and/or observational values.

A.3 Observations using the Earth-term only

We start by considering analyses that rely only on the Earth-term contribution to the residuals, as done in Ref. [117, 210]. The case of a coherent analysis based both on the Earth- and pulsar-term, introduced in Ref. [101], is discussed later in this Letter. Table A.1 shows that, in general, the frequency change over 10 yr is small compared to the frequency bin width, $3.2(10 \text{ yr}/T)$ nHz [110, 161]. The observed signal is effectively

monochromatic, making the dynamics of the system impossible to infer. However, the presence of spins affects the waveform not only through the GW phase evolution, but also via the modulations of $A_p(t)$ and $\varphi_p(t)$ that are periodic over the precession period, and also introduces the secular contribution $\varphi_T(t)$. For $m_{1,2} = 10^9 M_\odot$ and $f_E = 10^{-7}$ Hz the orbital angular momentum precesses by $\Delta\alpha = 2$ rad (for dimensionless spin parameter $a \equiv S/M^2 = 0.1$) and $\Delta\alpha = 3$ rad (for $a = 0.98$), and therefore the additional modulation effect on $A_p(t)$ and $\varphi_p(t)$ is small, and likely undetectable. However, the overall change of $\varphi_T(t)$ over 10 yrs could be appreciable: the average contribution for each precession cycle of this additional phase term is $\langle\Delta\varphi_T\rangle = 4\pi$ or $4\pi(1 - \cos\lambda_L)$, depending on whether $\hat{\Omega}$ lies inside or outside the precession cone, respectively [23]. If $\hat{\Omega}$ lies inside the precession cone, and given that the observation will cover between a third and a half of a full precession cycle, then $\langle\Delta\varphi_T\rangle \sim \pi$, which could surely indicate the presence of spins. On the other hand, the precession cone will be small in general since $|S/L| \sim av(M/\mu) \simeq 0.1 a(M/\mu)(M/10^9 M_\odot)^{1/3}(f/100 \text{ nHz})^{1/3}$, therefore the likelihood of $\hat{\Omega}$ lying inside the precession cone is small, assuming an isotropic distribution and orientation of sources. In this case the Thomas precession contribution (per precession cycle) is suppressed by a factor $(1 - \cos\lambda_L) \simeq \lambda_L^2/2 \sim 5 \times 10^{-3} a^2 (M/\mu)^2 (M/10^9 M_\odot)^{2/3} (f/100 \text{ nHz})^{2/3}$, which will produce a negligible contribution $\Delta\varphi_T(t) \ll 1$. However unlikely, spins may still introduce observable effects that need to be taken into account.

Table A.1 Frequency change Δf , total number of Gravitational Wave cycles and individual contributions from the leading order terms in the pN expansion over the two relevant time scales – a 10 yr period starting at the Earth and the time period $L_p(1 + \hat{\Omega} \cdot \hat{\mathbf{p}})$ between the Earth and pulsar term (hence the negative sign) – for selected values of $m_{1,2}$ and f_E .

$m_1(M_\odot)$	$m_2(M_\odot)$	$f_E(\text{nHz})$	$(v/c \times 10^{-2})$	timespan	Δf (nHz)	Total	Newtonian	p ¹ N	p ^{1.5} N	spin-orbit/ β	p ² N
10^9	10^9	100	14.6	10 yr	3.22	32.1	31.7	0.9	-0.7	0.06	0.04
			9.6	-1 kpc	71.2	4305.1	4267.8	77.3	-45.8	3.6	2.2
		50	11.6	10 yr	0.24	15.8	15.7	0.3	-0.2	0.01	< 0.01
			9.4	-1 kpc	23.1	3533.1	3504.8	53.5	-28.7	2.3	1.2
10^8	10^8	100	6.8	10 yr	0.07	31.6	31.4	0.2	-0.07	< 0.01	< 0.01
			6.4	-1 kpc	15.8	9396.3	9355.7	58.3	-19.9	1.6	0.5
		50	5.4	10 yr	0.005	15.8	15.7	0.06	-0.02	< 0.01	< 0.01
			5.3	-1 kpc	1.62	5061.4	5045.8	20.8	-5.8	0.5	0.1

A.4 Measuring SMBHB evolution using the Earth and pulsar term

With more sensitive observations and the increasing possibility of precisely determining L_p see e.g. [169], the prospect of also observing the contribution from the pulsar-term from one or more pulsars becomes more realistic. We show below that if at least one of the pulsar terms can be observed together with the Earth-term, this opens opportunities to study the dynamical evolution of SMBHBs and, in principle, to measure their masses and spins. This is a straightforward consequence of the fact that PTAs become sensitive to $\sim 10^3$ yr of SMBHB evolutionary history, in “snippets” of length $T \ll L_p$ that can be coherently concatenated.

The signal from each pulsar term will be at a S/N which is significantly smaller than the Earth-term by a factor $\sim \sqrt{N_p}$, where N_p is the number of pulsars that effectively contribute to the S/N of the array. For example, if the Earth-term yields an S/N of $\sim 36\sqrt{N_p/20}$, then each individual pulsar term would give an S/N ~ 8 . The possibility of coherently connecting the Earth-term signal with each pulsar term becomes therefore a question of S/N, prior information about the pulsar-Earth baseline and how accurately the SMBHB location in the sky can be reconstructed, as part of a “global fit”, e.g. [110]. Assuming for simplicity that the uncertainties on L_p and $\hat{\Omega}$ are uncorrelated, this requires that the distance to the pulsar and the location of the GW source are known with errors $\lesssim 0.01(100 \text{ nHz}/f)$ pc and $\lesssim 3(100 \text{ nHz}/f)(1 \text{ kpc}/L_p)$ arcsec, respectively. These are very stringent constraints [32, 161, 169], and a detailed analysis is needed in order to assess the feasibility of reaching this precision. Clearly if an electromagnetic counterpart to the GW source were to be found [158, 176], it would

enable the identification of the source location in the sky, making the latter constraint unnecessary.

We can now consider the contribution from the different terms in the pN expansion to the total number of cycles in observations that cover the GW source evolution over the time τ that are encoded in the simultaneous analysis of the Earth and pulsar terms. The results are summarised in Table A.1, for selected values of $m_{1,2}$, and f_E and for a fiducial value $\tau = 1$ kpc. The wavecycle contributions from the spin-orbit parameter are normalised to $\beta = (1/12) \sum_{i=1}^2 [113(m_i/M)^2 + 75\eta] \hat{\mathbf{L}} \cdot \hat{\mathbf{S}}_i$, which has a maximum value of 7.8. Contributions from the p²N order spin-spin terms are negligible. The results clearly show that despite the fact that the source is in the weak field regime the extended Earth-pulsar baseline requires the p^{1.5}N, and in some rare cases the p²N contribution, to accurately (i.e. within ~ 1 GW cycle) reproduce the full phase evolution.

For $m_{1,2} = 10^9 M_\odot$ and $f_E = 10^{-7}$ Hz there is a total of 4305 GW cycles over a 1 kpc light travel time evolution, with the majority (4267) accounted for by the leading order Newtonian term, providing information about the chirp mass, and tens of cycles due to the p¹N and p^{1.5}N terms (77 and 45, respectively), that provide information about a second independent mass parameter. Spins contribute to phasing at p^{1.5}N with $\sim 3\beta$ cycles. Therefore their total contribution is smaller than the p^{1.5}N mass contribution by a factor between a few and ~ 10 . The additional Thomas precession phase contribution may become comparable to the p¹N mass contribution in some cases. In fact, for $a = 0.1(0.98)$ the binary undergoes 24(34) precession cycles. This corresponds to a total Thomas precession phase contribution of 306(426) rad if $\hat{\mathbf{\Omega}}$ lies outside the precession cone.

The modulations of $A_p(t)$ and $\varphi_p(t)$ are characterised by a small λ_L , because for most of the inspiral $S \ll L$, and are likely to leave a smaller imprint on the waveform than

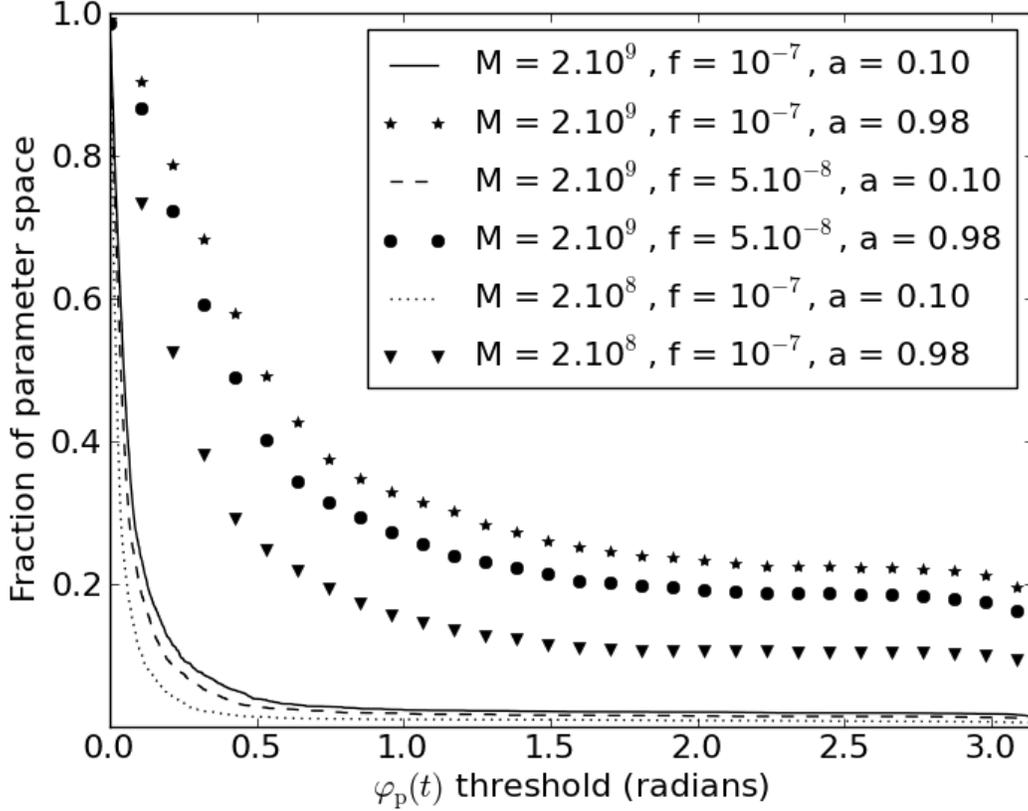
those discussed so far. We can indeed estimate the importance of this effect for the most favourable parameter combinations. The value of $\varphi_p(t)$ oscillates over time with an amplitude which depends on the time to coalescence, \mathbf{S} , \mathbf{L} , $\hat{\boldsymbol{\Omega}}$ and $\hat{\mathbf{p}}$. We choose the orientation of $\hat{\mathbf{S}}$ such that λ_L is maximised, and we vary $\hat{\boldsymbol{\Omega}}$ and $\hat{\mathbf{p}}$, each of which is drawn from a uniform distribution on the two-sphere.

In Figure 1 we show that for rapidly spinning ($a = 0.98$) SMBHBs this effect could introduce modulations larger than $\pi/2$ in $\varphi_p(t)$ over 30% of the parameter space of possible $\hat{\boldsymbol{\Omega}}$ and $\hat{\mathbf{p}}$ geometries. The amplitude would correspondingly change over the same portion of the parameter space by at most 60% with respect to its unmodulated value. Since these effects are modulated, they will not be easily identifiable.

A.5 Conclusions

We have established that the coherent observation of both the Earth and pulsar term provides information about the dynamical evolution of a GW source. The question now is whether they can be unambiguously identified. A rigorous analysis would require extensive simulations based on the actual analysis of synthetic data sets. We can however gain the key information with a much simpler order of magnitude calculation. The phase (or number of cycles) error scales as $\sim 1/(\text{S/N})$. Assuming $\text{S/N} \sim 40$ means that the total number of wave cycles over the Earth-pulsar baseline can be determined with an error $\sim 4300/40 \sim 100$ wave cycles. This is comparable to the p^1N contribution to the GW phase and, in very favourable circumstances, to the Thomas precession phase contribution, and larger by a factor of a few or more than all the other contributions. It may therefore be possible to measure the chirp mass and, say, the symmetric mass ratio of a SMBHB, and possibly a combination of the spin parameters. Effects due to the

Figure A.1 The fraction of parameter space in $\hat{\Omega}$ and $\hat{\mathbf{p}}$ for which the maximum excursion of φ_p over the time $L_p(1 + \hat{\Omega} \cdot \hat{\mathbf{p}})$ for $L_p = 1 \text{ kpc}$ exceeds a certain value, shown on the horizontal axis. Several values of $m_{1,2}$, a and f_E are considered (see legend).



$p^{1.5N}$ and higher phase terms are likely to remain unobservable, as well as amplitude and phase modulations. Correlations between the parameters, in particular masses and spins, will further degrade the measurements. The details will depend on the actual S/N of the observations, the GW source parameters, and the accuracy with which the source location and the pulsar distance can be determined. We plan to explore these issues in detail in a future study.

Appendix B

TOWARDS A GENERIC TEST OF THE STRONG FIELD DYNAMICS OF GENERAL RELATIVITY USING COMPACT BINARY COALESCENCE

The work in this Appendix is presented in the form of the paper Published in Physical Review D [112].

Authorship of paper: T. G. F. Li, W. Del Pozzo, S. Vitale, C. Van Den Broeck, M. Agathos, J. Veitch, K. Grover, T. Sidery, R. Sturani, A. Vecchio. K. Grover was involved in development of the method and discussions of the key results. K. Grover wrote the text and produced the plots for section B.2.2.

Coalescences of binary neutron stars and/or black holes are amongst the most likely gravitational-wave signals to be observed in ground-based interferometric detectors. Apart from the astrophysical importance of their detection, they will also provide us with our very first empirical access to the genuine strong-field dynamics of general relativity (GR). We present a new framework based on Bayesian model selection aimed at detecting deviations from GR, subject to the constraints of the Advanced Virgo and LIGO detectors. The method tests the consistency of coefficients appearing in the waveform with the predictions made by GR, without relying on any specific alternative theory of gravity. The framework is suitable for low signal-to-noise ratio events through the construction of multiple subtests, most of which involve only a limited number of coefficients. It also naturally allows for the combination of information from multiple sources to increase one's confidence in GR or a violation thereof. We expect it to be capable of finding a wide range of possible deviations from GR, including ones which, in principle, cannot be accommodated by the model waveforms, on the condition that the induced change in phase at frequencies where the detectors are the most sensitive is comparable to the effect of a few percent change in one or more of the low-order post-Newtonian phase coefficients. In principle, the framework can be used with any GR waveform approximant, with arbitrary parametrized deformations, to serve as model waveforms. In order to illustrate the workings of the method, we perform a range of numerical experiments in which simulated gravitational waves modelled in the restricted post-Newtonian, stationary phase approximation are added to Gaussian and stationary noise that follows the expected Advanced LIGO/Virgo noise curves.

B.1 Introduction

General Relativity (GR) is a non-linear, dynamical theory of gravity. Until the 1970s, all of its tests involved the weak-field, stationary regime; these are the standard Solar System tests that are discussed in most textbooks, e.g. [132]. GR passed them with impressive accuracy. Nevertheless, the more interesting part of any field theory resides in its dynamics, and this is especially true of GR [152, 207]. A first test of the latter came from the Hulse-Taylor binary and a handful of similar tight neutron star binaries [54, 96, 109], whose orbital elements are changing in close agreement with GR under the assumption that energy and angular momentum are carried away by gravitational waves (GW). Thus, these discoveries led to the very first, albeit indirect, evidence for GW. However, even the most relativistic of these binaries, PSR J0737-3039 [54, 109], is still in the relatively slowly varying, weak-field regime from a GR point of view, with a compactness of $GM/(c^2R) \simeq 4.4 \times 10^{-6}$, with M the total mass, R the orbital separation, and a typical orbital speed $v/c \simeq 2 \times 10^{-3}$. By contrast, for an inspiraling compact binary, in the limit of a test particle around a non-spinning black hole, the last stable orbit occurs at a separation of $R = 6GM/c^2$, where $GM/(c^2R) = 1/6$ and $v/c = 1/\sqrt{6}$. This constitutes the genuinely strong-field, dynamical regime of General Relativity, which in the foreseeable future will only be empirically accessible by means of gravitational-wave detectors.

Several large gravitational wave observatories have been operational for some years now: the two 4 km arm length LIGO interferometers in the US [11], the 3 km arm length Virgo in Italy [13–15], and the 600 m arm length GEO600 [84, 85]. By around 2015, LIGO and Virgo will have been upgraded to their so-called advanced configurations [1, 2, 87, 179], and shortly afterwards up to tens of detections per year are expected [4]. Another

planned GW observatory is the Japanese detector KAGRA [29], and the construction of a further large interferometer in India is under consideration [150]. Among the most promising sources are the inspiral and merger of compact binaries composed of two neutron stars (BNS), a neutron star and a black hole (NSBH), or two black holes (BBH).

Within GR, especially the inspiral part of the coalescence process has been modeled in great detail using the post-Newtonian (PN) formalism (see [49] and references therein), in which quantities such as the conserved energy and flux are found as expansions in v/c , where v is a characteristic speed. During inspiral, the GW signals will carry a detailed imprint of the orbital motion. Indeed, the contribution at leading order in amplitude has a phase that is simply $2\Phi(t)$, with $\Phi(t)$ being the orbital phase. Thus, the angular motion of the binary is directly encoded in the waveform’s phase, and assuming quasi-circular inspiral, the radial motion follows from the instantaneous angular frequency $\omega(t) = \dot{\Phi}(t)$ through the relativistic version of Kepler’s Third Law. If there are deviations from GR, the different emission mechanism and/or differences in the orbital motion will be encoded in the phase of the signal waveform, allowing us to probe the strong-field dynamics of gravity. In this regard we note that with binary pulsars, one can only constrain the conservative sector of the dynamics to 1PN order (i.e. $(v/c)^2$ beyond leading order), and the dissipative sector to leading order; see, e.g., the discussion in [122] and references therein. Hence, when it comes to $\Phi(t)$, these observations do not fully constrain the 1PN contribution. Yet several of the more interesting dynamical effects occur starting from 1.5PN order; this includes ‘tail effects’ [47,48] and spin-orbit interactions; spin-spin effects first appear at 2PN [45]. As indicated by the Fisher matrix results of [131], Advanced LIGO/Virgo should be able to put significant constraints on the 1.5PN contribution to the phase, and possibly also higher-order contributions.

Possible deviations from GR that have been considered in the past in the context of compact binary coalescence include scalar-tensor theories [37, 38, 58, 154, 205, 206]; a varying Newton constant [215]; modified dispersion relation theories, usually referred to in literature as ‘massive graviton’ models¹ [27, 37, 38, 40, 63, 105, 173, 204]; violations of the No Hair Theorem [33, 39, 80, 93, 103]; violations of Cosmic Censorship [103, 184]; and parity violating theories [21, 170, 212, 214]. The (rather few) specific alternative theories of gravity that have been considered in the context of ground-based gravitational wave detectors – essentially scalar-tensor and ‘massive graviton’ theories – happen to be hard to constrain much further with GW observations, and we will not consider them in this paper. However, General Relativity may be violated in some other manner, including a way that is yet to be envisaged. This makes it imperative to develop methods that can search for generic deviations from GR.

In the past several years, several proposals have been put forward to test GR using coalescing compact binary coalescence:

1. One can search directly for the imprint of specific alternative theories, such as the so-called ‘massive graviton’ models and scalar-tensor theories [27, 37, 38, 40, 58, 105, 154, 173, 204–206]. For the ‘massive graviton’ case, a full Bayesian analysis was recently performed by Del Pozzo, Veitch, and Vecchio [63].
2. A method due to Arun et al. exploits the fact that, at least for binaries where neither component has spin, all coefficients ψ_i in the PN expansion (see Eq. (B.2) below for their definition) of the inspiral phase depend only on the two component

¹The designations ‘massive gravity’ and ‘massive graviton’ originate from [45] where actually the effect of a modified dispersion relation, or a wavelength dependent propagation velocity has been taken into account. While it is attractive to ascribe such a modification to a graviton mass, a modification of the dispersion relation can be a more general effect, and moreover endowing the graviton with a mass introduces additional deviations from GR than a mere modified dispersion relation. See e.g. the original [183] and the recent [61] for a thorough discussion of the issues related to massive gravity models.

masses, m_1 and m_2 [25, 28, 131]. In that case only two of the ψ_i are independent, so that a comparison of any three of them allows for a test of GR. Such a method would be very general, in that one does not have to look for any particular way in which gravity might deviate from GR; instead it allows generic failures of GR to be searched for. However, so far its viability was only explored using Fisher matrix calculations.

3. In the so-called parameterized post-Einsteinian (ppE) formalism of Yunes and Pretorius, gravitational waveforms are parameterized so as to include effects from a variety of alternative theories of gravity [213, 216]. A Bayesian analysis was performed in [55].
4. Recently a test of the No Hair Theorem was presented, also in a Bayesian setting [83].

The first method presupposes that GR will be violated in a particular way. As for the third and fourth points, the particular Bayesian implementation that was used involves comparing the GR waveform with a waveform model that includes parameterized deformations away from GR, thus introducing further free parameters.

Now, imagine one introduces free parameters p_i , $i = 1, 2, \dots, N_T$ in such a way that $p_i = 0$ for all i corresponds to GR being correct. Then one can compare a waveform model in which all the p_i are allowed to vary with the GR waveform model in which all the p_i are zero. As we will explain in this paper, this amounts to asking the question: “Do all the p_i differ from zero at the same time?” Let us call the associated hypothesis $H_{12\dots N_T}$, which is to be compared with the GR hypothesis \mathcal{H}_{GR} .

A more interesting (because much more general) question would be: “Do one or more of the p_i differ from zero”, without specifying which. As we shall see, this question

is more difficult to cast into the language of model selection. Below we will call the associated hypothesis $\mathcal{H}_{\text{modGR}}$, to be compared with the GR hypothesis \mathcal{H}_{GR} . What we will show is that, although there is no single waveform model that corresponds to $\mathcal{H}_{\text{modGR}}$, testing the latter amounts to testing $2^{N_T} - 1$ hypotheses $H_{i_1 i_2 \dots i_k}$ corresponding to all subsets $\{p_{i_1}, p_{i_2}, \dots, p_{i_k}\}$ of the full set $\{p_1, p_2, \dots, p_{N_T}\}$. Each of the hypotheses $H_{i_1 i_2 \dots i_k}$ is tested by a waveform model in which $p_{i_1}, p_{i_2}, \dots, p_{i_k}$ are free, but all the other p_j are fixed to zero. The Bayes factors against GR for all of these tests can then be combined into a single odds ratio which compares the full hypothesis $\mathcal{H}_{\text{modGR}}$ with \mathcal{H}_{GR} .

In a scenario with low signal-to-noise ratio (SNR), as will be the case with advanced ground-based detectors, parameter estimation will degrade significantly when trying to estimate too many parameters at once, and so will model selection if the alternative model to GR has too many additional degrees of freedom [63]. This could be problematic if one only tests the ‘all-inclusive’ hypothesis $H_{12\dots N_T}$ against GR, i.e., if the question one asks is “Do all the p_i differ from zero at the same time”. The question we want to ask instead, namely “Do one or more of the p_i differ from zero”, is not only more general; most of the sub-hypotheses $H_{i_1 i_2 \dots i_k}$ involve a smaller number of free parameters, making the corresponding test more powerful in a low-SNR situation. And, as in most Bayesian frameworks, information from multiple sources can be combined in a straightforward way.

Our framework can be used with any family of waveforms with parameterized deformations, including the ppE family. In order to illustrate the method, following [25, 28, 131] we make the simplest choice by adopting model waveforms in which the deviations away from GR take the form of shifts in a subset of the post-Newtonian inspiral phase coefficients. To establish the validity of the framework, we perform simulations with

simple analytic frequency domain waveforms in stationary, Gaussian noise that follows the expected Advanced LIGO/Virgo noise curves. In the future, for actual tests of GR, one may want to use time domain waveforms and introduce deviations in e.g. a Hamiltonian used to numerically evolve the motion of the binary.

As noted in [63,216] and also illustrated here, if one is only interested in detection, then it might suffice to only search with template waveforms predicted by GR, but parameter estimation can be badly off; this is what is called ‘fundamental bias’. Indeed, even if there is a deviation from GR in the signal, then a model waveform with completely different values for the masses and other parameters could still be a good fit to the data, with minimal loss of signal-to-noise ratio. We note that given model waveforms that feature a certain family of deformations away from GR, ‘fundamental bias’ can still occur if the signal has a deviation that does not belong to the particular class of deformations allowed for by the models. However, in that case one expects the non-GR model waveforms to still be preferred over GR ones in the sense of model selection, even if parameter estimation may be deceptive in interpreting the nature of the deviation. We will show an explicit example of this, and will argue that generic deviations from GR can be picked up, of course subject to the limitations imposed by the detectors, as is the case with any kind of measurement. In particular, we expect that a GR violation will generally be visible, on condition that its effect on the phase at frequencies where the detector is the most sensitive is comparable to the effect of a few percent shift in one of the lower-order phase coefficients.

This paper is structured as follows. We first introduce our waveform model for compact binary inspiral, and discuss its sensitivity to changes in phase parameters (Section B.2). In Section B.3, we explain the basic method for single and multiple sources. In Section B.4, we construct different simulated catalogs of sources and evaluate the level at which

deviations from GR can be found. We end with a summary and a discussion of future steps to be taken.

Unless stated otherwise, we will take $G = c = 1$.

B.2 Waveform model and its sensitivity to changes in phase coefficients

We now introduce our waveform model. Since we are concerned with testing the strong-field dynamics of gravity, eventually all of the effects which we expect to see with compact binary coalescence should be represented in the waveform. This includes, but is not limited to, precession due to spin-orbit and spin-spin interactions [45], sub-dominant signal harmonics [46], and merger/ringdown [35]. In due time these will indeed need to be taken into account. However, in this paper we first and foremost wish to demonstrate the validity of a particular method, for which it will not be necessary to use very sophisticated waveforms. Also, as suggested by Fisher matrix calculations such as those of Mishra *et al.* [131], methods based on measuring phase coefficients will be the most accurate at low total mass. In this paper we limit ourselves to BNS sources, for which spin will be negligible, as well as sub-dominant signal harmonics [184, 185]. Since we will assume a network of Advanced LIGO and Virgo detectors, the merger and ringdown signals will also not have a large impact [197]. Thus, for a first analysis we will focus on the inspiral part of the coalescence process, modeling the waveform in the frequency domain using the stationary phase approximation (SPA) [97, 151]. In particular, we use the so-called restricted TaylorF2 waveforms [53, 119] up to 2PN in phase.

Since the way we illustrate our method here is based on allowing for deviations in phase coefficients, we will need to know how sensitive our waveform model is to minor changes in the values of these coefficients. To get a sense of this, one could use the results of [131] as a guide, but since these are based on the Fisher matrix they necessarily assume that signal and template are from the same waveform family. Before explaining our method for testing GR, we will first look at what happens both to detectability and parameter estimation when the signal contains a deviation from GR but is being searched for with a bank of GR templates.

B.2.1 Model waveform(s) and detector configuration

We start from the way TaylorF2 is implemented in the LIGO Algorithms Library [119]:

$$h(f) = \frac{1}{D} \frac{\mathcal{A}(\theta, \phi, \iota, \psi, \mathcal{M}, \eta)}{\sqrt{\dot{F}(\mathcal{M}, \eta; f)}} f^{2/3} e^{i\Psi(t_c, \phi_c, \mathcal{M}, \eta; f)}, \quad (\text{B.1})$$

where D is the luminosity distance to the source, (θ, ϕ) specify the sky position, (ι, ψ) give the orientation of the inspiral plane with respect to the line of sight, \mathcal{M} is the chirp mass, and η is the symmetric mass ratio. In terms of the component masses (m_1, m_2) , one has $\eta = m_1 m_2 / (m_1 + m_2)^2$ and $\mathcal{M} = (m_1 + m_2) \eta^{3/5}$. t_c and ϕ_c are the time and phase at coalescence, respectively. The ‘frequency sweep’ $\dot{F}(\mathcal{M}, \eta; f)$ is an expansion in powers of the frequency f with mass-dependent coefficients, and

$$\begin{aligned} & \Psi(t_c, \phi_c, \mathcal{M}, \eta; f) \\ &= 2\pi f t_c - \phi_c - \pi/4 \\ & \quad + \sum_{i=0}^7 \left[\psi_i + \psi_i^{(l)} \ln f \right] f^{(i-5)/3}. \end{aligned} \quad (\text{B.2})$$

In the case of GR, the functional dependence of the coefficients ψ_i and $\psi_i^{(l)}$ on (\mathcal{M}, η) can be found in [131]. However, here we will not assume that those relationships necessarily hold, except in the case of ψ_0 , which has been tested using binary pulsars. With minor abuse of notation, let us re-label the remaining coefficients as ψ_i , $i = 1, \dots, M$.

We note that \dot{F} is related to the phase Ψ , and in principle we should also leave open the possibility that its expansion coefficients deviate from their GR values. However, with the Advanced LIGO and Virgo network and for stellar mass binaries we do not expect to be very sensitive to sub-dominant contributions to the amplitude [184, 185].

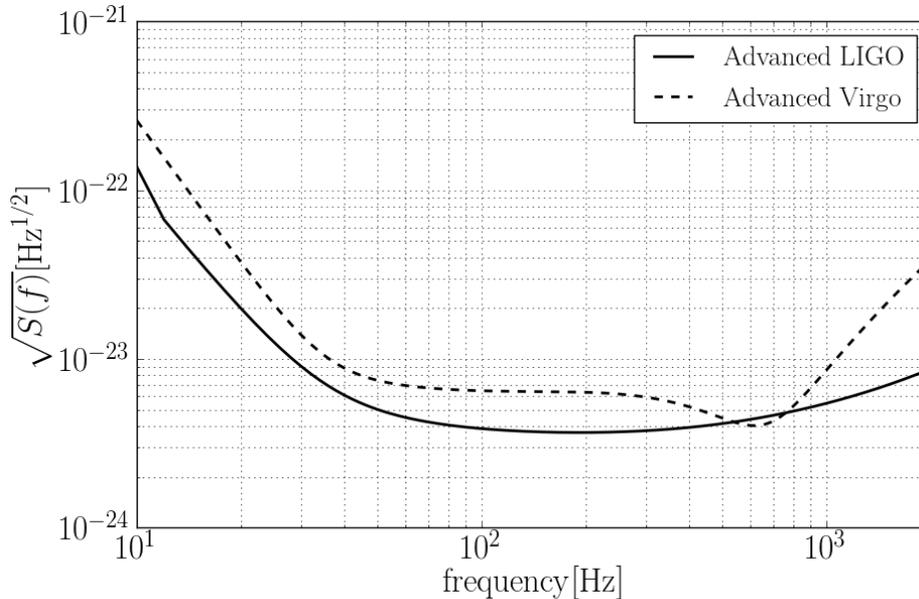
Let us focus on the phase (B.2). One way of testing GR would be to use a model waveform in which all the ψ_i are considered free parameters, measure these together with \mathcal{M} and η , and check whether one obtains agreement with the functional relations $\psi_i(\mathcal{M}, \eta)$ predicted by GR. However, the events we expect in Advanced LIGO/Virgo will probably not have sufficient SNR for this to be directly feasible [131].

Below, we instead suggest a scheme where a large number of tests are done, in each of which a specific, limited subset of the phase coefficients is left free while the others have the dependence on masses as in GR. The results from all of these tests can then be combined into the odds ratio for a general deviation from GR versus GR being correct.

In this study we will assume a network of two Advanced LIGO detectors, one in Hanford, WA, and the other one in Livingston, LA, together with the Advanced Virgo detector in Cascina, Italy. We take the Advanced LIGO noise curve to be the one with zero-detuning of the signal recycling mirror and high laser power [163]. With these assumptions, the curves in Fig. B.1 represent the incoherent sums of the principal noise sources as they are currently understood; however, there may be unexpected, additional sources of noise. The high-power, zero-detuning option gives most of the desired sensi-

tivity with the fewest technical difficulties. Advanced Virgo can also be optimized for BNS sources by an appropriate choice of the signal recycling detuning and the signal recycling mirror transmittance [179], and this is what we assume here.

Figure B.1 The high-power, zero-detuning noise curve for Advanced LIGO, and the BNS-optimized Advanced Virgo noise curve.



B.2.2 Changes in phase coefficients and detectability

It is important to investigate to what extent signals that may deviate from General Relativity could be detected in the first place. The fitting factor (FF) is a measure of the adequateness of a template family to fit the signal; $1 - FF$ is the reduction in signal-to-noise ratio that occurs from using a model waveform which differs from the exact signal waveform when searching the data. Let $h_e(\vec{\lambda})$ be the ‘exact’ waveform of the signal and $h_m(\vec{\theta})$ the model used for detection; the exact and model waveforms are dependent on sets of parameters $\vec{\lambda}$ and $\vec{\theta}$, respectively. The fitting factor is then defined

as [24]

$$FF \equiv \max_{\vec{\theta}} \left(\frac{\langle h_e(\vec{\lambda}) | h_m(\vec{\theta}) \rangle}{\sqrt{\langle h_m(\vec{\theta}) | h_m(\vec{\theta}) \rangle \langle h_e(\vec{\lambda}) | h_e(\vec{\lambda}) \rangle}} \right), \quad (\text{B.3})$$

where $\langle a | b \rangle$ denotes the usual noise weighted inner product,

$$\langle a | b \rangle = 2 \int_{f_{\min}}^{f_{\max}} \frac{a^*(f)b(f) + a(f)b^*(f)}{S_n(f)} df, \quad (\text{B.4})$$

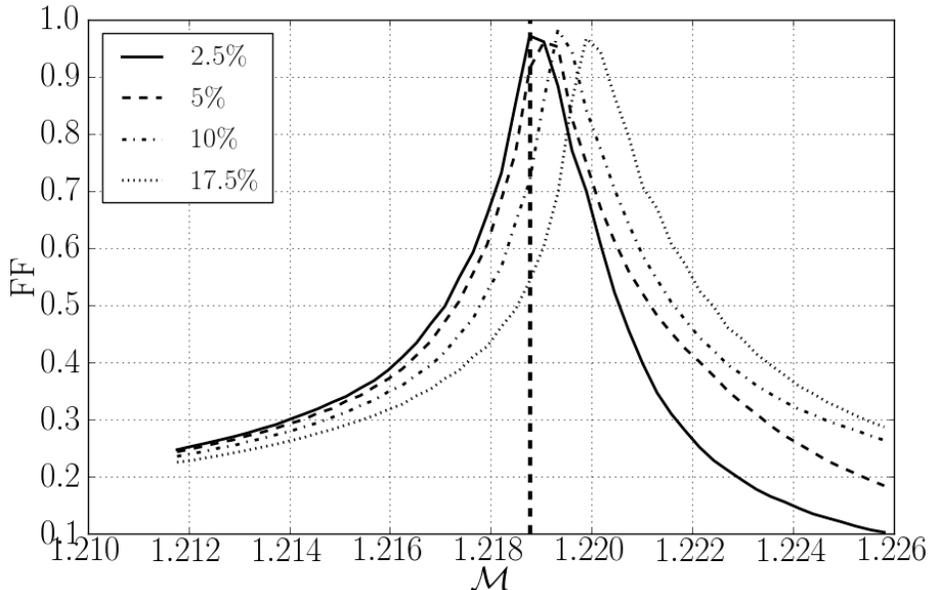
and $S_n(f)$ is the one-sided noise spectral density [122]. f_{\min} is the detectors' lower cut-off frequency, and in our case, f_{\max} is the frequency at last stable orbit, $f_{\max} = (6^{3/2}\pi\mathcal{M}\eta^{-3/5})^{-1}$. We note that the detection rate scales like the cube of the signal-to-noise ratio, so that the fractional reduction in event rate is $1 - FF^3$ [24]. In this case the waveform is not 'exact' in the sense of numerical relativity; instead we use the fitting factor as a measure of how similar a modified TaylorF2 waveform and a GR version are.

A sample deviation is tested using a modified waveform different only in the 1.5PN order phase coefficient: $\psi_3^{\text{GR}}(\mathcal{M}, \eta) \rightarrow \psi_3^{\text{GR}}(\mathcal{M}, \eta) [1 + \delta\chi_3]$. The Advanced LIGO noise curve is used. The signal is 'detected' with a template bank of standard GR waveforms that is regularly spaced in the parameters present in the phase, ϕ_c , t_c , \mathcal{M} , and η . We study a (1.4, 1.4) M_\odot binary with $\delta\chi_3$ ranging from 0.025 to 0.175. As seen in figures B.2 and B.3, the mass parameters can absorb the change in the phase due to the modified phase coefficient while providing a fitting factor of over 95%.

Thus, with a template bank of GR waveforms it is possible to detect a signal containing a large deviation from GR without significant loss in signal-to-noise ratio, but recovered with intrinsic parameters that deviate significantly from the true values. We now describe a method which will be able to nevertheless recognize a deviation from GR

when one is present.

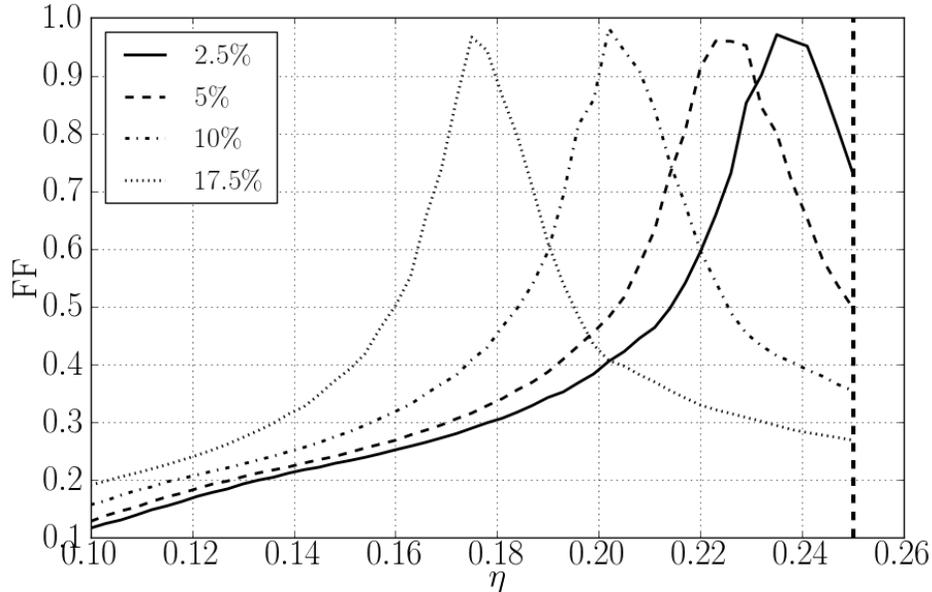
Figure B.2 The fitting factors for a range of \mathcal{M} , once the other parameters are maximized over. Here a deviation of ψ_3 from 2.5% to 17.5% is used for a $(1.4, 1.4) M_\odot$ system. The vertical dashed line represents the true value of \mathcal{M} , while the maximum is offset to compensate for the modification in the phase.



B.3 Method

We will first recall some basic facts about Bayesian inference. Next we outline our method for finding a possible violation of GR using inspiral signals by allowing for deviations in phase coefficients. For simplicity, we start with an example where only two phase coefficients are taken into account; then we go on to the general case. Finally we explain how to combine information from a catalog of sources.

Figure B.3 The fitting factors for a range of η , once the other parameters are maximized over. Here a deviation of ψ_3 from 2.5% to 17.5% is used for a $(1.4, 1.4) M_\odot$ system. The vertical dashed line represents the true value of η , while the maximum is offset to compensate for the modification in the phase.



B.3.1 Bayesian inference and nested sampling

We now give a brief overview of Bayesian inference, as well as the method of nested sampling due to Skilling [167], which was first introduced into ground-based GW data analysis by Veitch and Vecchio [191–193] and which we will adopt here as well.

Let us consider hypotheses $\mathcal{H}_i, \mathcal{H}_j$. Here \mathcal{H}_i could be the hypothesis that there is a deviation from GR while \mathcal{H}_j is the hypothesis that GR is correct; or \mathcal{H}_i could simply be the hypothesis that a signal of a particular form is present in the data while \mathcal{H}_j is the hypothesis that there is only noise. The statements that we can make about any hypothesis are based on a data set d (observations) and all the relevant prior information I that we hold.

Within the framework of Bayesian inference, the key quantity that one needs to compute

is the posterior probability of a hypothesis \mathcal{H}_i . Applying Bayes' theorem we obtain

$$P(\mathcal{H}_i|d, \mathbb{I}) = \frac{P(\mathcal{H}_i|\mathbb{I}) P(d|\mathcal{H}_i, \mathbb{I})}{P(d|\mathbb{I})}, \quad (\text{B.5})$$

where $P(\mathcal{H}_i|d, \mathbb{I})$ is the posterior probability of the hypothesis \mathcal{H}_i given the data, $P(\mathcal{H}_i|\mathbb{I})$ is the prior probability of the hypothesis, and $P(d|\mathcal{H}_i, \mathbb{I})$ is the marginal likelihood or evidence for \mathcal{H}_i , which can be written as:

$$\begin{aligned} P(d|\mathcal{H}_i, \mathbb{I}) &= \mathcal{L}(\mathcal{H}_i) \\ &= \int d\vec{\theta} p(\vec{\theta}|\mathcal{H}_i, \mathbb{I}) p(d|\vec{\theta}, \mathcal{H}_i, \mathbb{I}). \end{aligned} \quad (\text{B.6})$$

In the above expression, $p(\vec{\theta}|\mathcal{H}_i, \mathbb{I})$ is the prior probability density of the unknown parameter vector $\vec{\theta}$ within the model corresponding to \mathcal{H}_i , and $p(d|\vec{\theta}, \mathcal{H}_i, \mathbb{I})$ is the likelihood function of the observation d , assuming a given value of the parameters $\vec{\theta}$ and the model \mathcal{H}_i .

If we want to compare different hypotheses, \mathcal{H}_i and \mathcal{H}_j , in light of the observations made, we can compute the ratio of posterior probabilities, which is known as the odds ratio:

$$\begin{aligned} O_j^i &= \frac{P(\mathcal{H}_i|d, \mathbb{I})}{P(\mathcal{H}_j|d, \mathbb{I})} \\ &= \frac{P(\mathcal{H}_i|\mathbb{I}) P(d|\mathcal{H}_i, \mathbb{I})}{P(\mathcal{H}_j|\mathbb{I}) P(d|\mathcal{H}_j, \mathbb{I})} \\ &= \frac{P(\mathcal{H}_i|\mathbb{I})}{P(\mathcal{H}_j|\mathbb{I})} B_j^i, \end{aligned} \quad (\text{B.7})$$

where $P(\mathcal{H}_j|\mathbb{I})/P(\mathcal{H}_i|\mathbb{I})$ is the prior odds of the two hypotheses, the relative confidence we assign to the models before any observation, and B_j^i is the Bayes factor.

In addition to computing the relative probabilities of different models, one usually wants to make inference on the unknown parameters, and therefore one needs to compute the joint posterior probability density function (PDF)

$$p(\vec{\theta}|d, \mathcal{H}_i, \mathbb{I}) = \frac{p(\vec{\theta}|\mathcal{H}_i, \mathbb{I})p(d|\vec{\theta}, \mathcal{H}_i, \mathbb{I})}{p(d|\mathcal{H}_i, \mathbb{I})}. \quad (\text{B.8})$$

From the previous expression it is simple to compute the marginalized PDF on any given parameter, say θ_1 within a given model \mathcal{H}_i :

$$p(\theta_1|d, \mathcal{H}_i, \mathbb{I}) = \int d\theta_2 \dots \int d\theta_N p(\vec{\theta}|d, \mathcal{H}_i, \mathbb{I}). \quad (\text{B.9})$$

The key quantities for Bayesian inference in Eq. (B.7), (B.8) and (B.9) can be efficiently computed using e.g. a nested sampling algorithm [167]. The basic idea of nested sampling is to use a collection of n objects, called live points, randomly sampled from the prior distributions, but subject to a constraint over the value of their likelihood. A live point $\vec{\xi}$ is a point in the multidimensional parameter space. At each iteration, the live point $\vec{\xi}^*$ having the lowest likelihood $\mathcal{L}(\vec{\xi}^*)$ is replaced with a new point $\vec{\xi}$ sampled from the prior distribution. To be accepted, the new point must obey to the condition

$$\mathcal{L}(\vec{\xi}) > \mathcal{L}(\vec{\xi}^*). \quad (\text{B.10})$$

The above condition ensures that regions of progressively increasing likelihood are explored, and the evidence integral, Eq. (B.6), is calculated using those points as the computation progresses.

In this paper we use a specific implementation of this technique that was developed for ground-based observations of coalescing binaries by Veitch and Vecchio; we point the

interested reader to [30, 31, 191–193] for technical details. To select a new live point, a Metropolis-Hastings Markov chain Monte Carlo (MCMC) is used with p steps. The uncertainty in the evidence computation for a given number of live points n and MCMC steps p was quantified in [193]. For the calculations in this paper we took $n = 1000$ and $p = 100$, in which case the standard deviation on log Bayes factors is $\mathcal{O}(1)$.

B.3.2 Basic method for a single source

Given that we have no knowledge of which coefficient(s) might deviate from the GR values, we want to test the hypothesis that at least one of the known coefficients $\{\psi_0, \psi_1, \psi_2, \dots, \psi_M\}$ is different. Computational limitations and possible lack of sensitivity to changes in the higher-order coefficients will induce us to only look for deviations in a set of testing coefficients $\{\psi_1, \dots, \psi_{N_T}\} \subset \{\psi_0, \psi_1, \dots, \psi_M\}$, with $N_T < M$. For the examples of Sec. B.4, we will choose $N_T = 3$ due to computational constraints, but a larger number could be used.

Let us introduce some notation. We define hypotheses $H_{i_1 i_2 \dots i_k}$ as follows:

$H_{i_1 i_2 \dots i_k}$ is the hypothesis that the phasing coefficients $\psi_{i_1}, \dots, \psi_{i_k}$ do not have the functional dependence on (\mathcal{M}, η) as predicted by General Relativity, but all other coefficients $\psi_j, j \notin \{i_1, i_2, \dots, i_k\}$ do have the dependence as in GR.

Thus, for example, H_{12} is the hypothesis that ψ_1 and ψ_2 deviate from their GR values, with all other coefficients being as in GR. With each of the hypotheses above, we can associate a waveform model that can be used to test it. Let $\vec{\theta} = \{\mathcal{M}, \eta, \dots\}$ be the parameters occurring in the GR waveform. Then $H_{i_1 i_2 \dots i_k}$ is tested by a waveform in

which the independent parameters are

$$\{\vec{\theta}, \psi_{i_1}, \psi_{i_2}, \dots, \psi_{i_k}\}, \quad (\text{B.11})$$

i.e. the coefficients $\{\psi_{i_1}, \psi_{i_2}, \dots, \psi_{i_k}\}$ are allowed to vary freely. In practice, these will need to be subsets of a limited set of coefficients; in Section B.4, where we present results, we will consider all subsets of the set $\{\psi_1, \psi_2, \psi_3\}$. This choice will already suffice to illustrate the method without being overly computationally costly, but in the future one may want to use a larger set.

Now, the hypothesis we would really like to test is that one or more of the ψ_i differ from GR, without specifying which. This corresponds to a logical ‘or’ of the above hypotheses:

$$\mathcal{H}_{\text{modGR}} = \bigvee_{i_1 < i_2 < \dots < i_k} H_{i_1 i_2 \dots i_k}. \quad (\text{B.12})$$

Our aim is to compute the following odds ratio:

$$O_{\text{GR}}^{\text{modGR}} \equiv \frac{P(\mathcal{H}_{\text{modGR}}|d, \text{I})}{P(\mathcal{H}_{\text{GR}}|d, \text{I})}. \quad (\text{B.13})$$

However, the hypothesis (B.12) is not what model waveforms with one or more free coefficients ψ_i will test; rather, such waveforms test the hypotheses $H_{i_1 i_2 \dots i_k}$ themselves. What we will need to do is to break up the logical ‘or’ in $\mathcal{H}_{\text{modGR}}$ into the component hypotheses $H_{i_1 i_2 \dots i_k}$. Fortunately, this is trivial.

Before treating the problem more generally, let us consider a simple example. Imagine that only two coefficients ψ_1 and ψ_2 are being used for the testing of GR. Then

$$\mathcal{H}_{\text{modGR}} = H_1 \vee H_2 \vee H_{12}. \quad (\text{B.14})$$

In this example, the odds ratio of interest can then be written as

$${}^{(2)}O_{\text{GR}}^{\text{modGR}} \equiv \frac{P(H_1 \vee H_2 \vee H_{12}|d, \mathbf{I})}{P(\mathcal{H}_{\text{GR}}|d, \mathbf{I})}, \quad (\text{B.15})$$

where the superscript (2) reminds us that only two of the parameters are being used for testing.

An important observation is that the hypotheses H_1 , H_2 , H_{12} are logically disjoint: the ‘and’ of any two of them is always false. Indeed, in H_1 , ψ_2 takes the value predicted by GR, but in H_2 it differs from the GR value, as it does in H_{12} . Similarly, in H_2 , ψ_1 takes the GR value, but in H_1 it differs from the GR value, and the same in H_{12} . More generally, any two hypotheses $H_{i_1 i_2 \dots i_k}$ and $H_{j_1 j_2 \dots j_l}$ with $\{i_1, i_2, \dots, i_k\} \neq \{j_1, j_2, \dots, j_l\}$ are logically disjoint. This means that the odds ratio is simply

$${}^{(2)}O_{\text{GR}}^{\text{modGR}} = \frac{P(H_1|d, \mathbf{I})}{P(\mathcal{H}_{\text{GR}}|d, \mathbf{I})} + \frac{P(H_2|d, \mathbf{I})}{P(\mathcal{H}_{\text{GR}}|d, \mathbf{I})} + \frac{P(H_{12}|d, \mathbf{I})}{P(\mathcal{H}_{\text{GR}}|d, \mathbf{I})}. \quad (\text{B.16})$$

Using Bayes’ theorem, this can be written as

$${}^{(2)}O_{\text{GR}}^{\text{modGR}} = \frac{P(H_1|\mathbf{I})}{P(\mathcal{H}_{\text{GR}}|\mathbf{I})} B_{\text{GR}}^1 + \frac{P(H_2|\mathbf{I})}{P(\mathcal{H}_{\text{GR}}|\mathbf{I})} B_{\text{GR}}^2 + \frac{P(H_{12}|\mathbf{I})}{P(\mathcal{H}_{\text{GR}}|\mathbf{I})} B_{\text{GR}}^{12}. \quad (\text{B.17})$$

Here B_{GR}^1 , B_{GR}^2 , B_{GR}^{12} are the Bayes factors

$$\begin{aligned} B_{\text{GR}}^1 &= \frac{P(d|H_1, \mathbf{I})}{P(d|\mathcal{H}_{\text{GR}}, \mathbf{I})}, \\ B_{\text{GR}}^2 &= \frac{P(d|H_2, \mathbf{I})}{P(d|\mathcal{H}_{\text{GR}}, \mathbf{I})}, \\ B_{\text{GR}}^{12} &= \frac{P(d|H_{12}, \mathbf{I})}{P(d|\mathcal{H}_{\text{GR}}, \mathbf{I})}, \end{aligned} \quad (\text{B.18})$$

and $P(H_1|\mathbf{I})/P(\mathcal{H}_{\text{GR}}|\mathbf{I})$, $P(H_2|\mathbf{I})/P(\mathcal{H}_{\text{GR}}|\mathbf{I})$, $P(H_{12}|\mathbf{I})/P(\mathcal{H}_{\text{GR}}|\mathbf{I})$ are ratios of prior odds.

In practice, we will write the testing coefficients as

$$\psi_i = \psi_i^{\text{GR}}(\mathcal{M}, \eta) [1 + \delta\chi_i], \quad (\text{B.19})$$

with $\psi_i^{\text{GR}}(\mathcal{M}, \eta)$ the functional form of the dependence of ψ_i on (\mathcal{M}, η) according to GR, and the dimensionless $\delta\chi_i$ is a fractional shift in ψ_i . Note that in GR, the 0.5PN contribution is identically zero; it will be treated separately, as explained in section B.4. In the example of this subsection, one can assume that ψ_1 , ψ_2 are any of the PN coefficients other than the 0.5PN one. With the above notation, the Bayes factors (B.18) are

$$B_{\text{GR}}^1 = \frac{\int d\vec{\theta} d\delta\chi_1 \{^1\}\pi(\delta\chi_1) \pi(\vec{\theta}) p(d|\vec{\theta}, \delta\chi_1, H_1, \mathbf{I})}{\int d\vec{\theta} \pi(\vec{\theta}) p(d|\vec{\theta}, \mathcal{H}_{\text{GR}}, \mathbf{I})}, \quad (\text{B.20})$$

$$B_{\text{GR}}^2 = \frac{\int d\vec{\theta} d\delta\chi_2 \{^2\}\pi(\delta\chi_2) \pi(\vec{\theta}) p(d|\vec{\theta}, \delta\chi_2, H_2, \mathbf{I})}{\int d\vec{\theta} \pi(\vec{\theta}) p(d|\vec{\theta}, \mathcal{H}_{\text{GR}}, \mathbf{I})}, \quad (\text{B.21})$$

$$B_{\text{GR}}^{12} = \frac{\int d\vec{\theta} d\delta\chi_1 d\delta\chi_2 \{^{12}\}\pi(\delta\chi_1, \delta\chi_2) \pi(\vec{\theta}) p(d|\vec{\theta}, \delta\chi_1, \delta\chi_2, H_{12}, \mathbf{I})}{\int d\vec{\theta} \pi(\vec{\theta}) p(d|\vec{\theta}, \mathcal{H}_{\text{GR}}, \mathbf{I})}. \quad (\text{B.22})$$

Here $\{^1\}\pi(\delta\chi_1)$, $\{^2\}\pi(\delta\chi_2)$, $\{^{12}\}\pi(\delta\chi_1, \delta\chi_2)$ are priors for, respectively, $\delta\chi_1$, $\delta\chi_2$, and the pair $(\delta\chi_1, \delta\chi_2)$. We choose these to be constant functions in the relevant parameter or pair of parameters, having support within a large interval or square centered on the origin, and normalized to one. For the other parameters, $\vec{\theta}$, we use the same functional form and limits as [193], with the exception of the distance being allowed to vary between 1 and 1000 Mpc. Specifically, for the sky location and the orientation of the orbital plane we choose uniform priors on the corresponding unit spheres. For the phase

at coalescence ϕ_c we choose a flat prior with $\phi_c \in [0, 2\pi]$, and the time of coalescence t_c is in a time interval of 100 ms. The prior on η is flat on the interval $[0, 0.25]$. For chirp mass we use an approximation to the Jeffreys prior which gives $p(\mathcal{M}|\mathbf{I}) \propto \mathcal{M}^{-11/6}$; see [193] for motivation. In addition, component masses are restricted to the interval $m_1, m_2 \in [1, 34] M_\odot$.

At this point it is worth commenting on the mutual relationships of the hypotheses $H_{i_1 i_2 \dots i_k}$ amongst each other and with \mathcal{H}_{GR} , and on the waveform models used to test them. As an example, let us discuss the case of H_1 and \mathcal{H}_{GR} . Consider the numerator of the Bayes factor B_{GR}^1 in Eq. (B.20):

$$\int d\vec{\theta} d\delta\chi_1 \quad {}^{\{1\}}\pi(\delta\chi_1) \pi(\vec{\theta}) p(d|\vec{\theta}, \delta\chi_1, H_1, \mathbf{I}). \quad (\text{B.23})$$

The parameter space of the GR waveforms, $\{\vec{\theta}\}$, has a natural embedding into the parameter space $\{\vec{\theta}, \delta\chi_1\}$ of the waveforms used to test H_1 : it can be identified with the hypersurface $\delta\chi_1 = 0$. We could have explicitly excluded this hypersurface from $\{\vec{\theta}, \delta\chi_1\}$ by setting a prior on $\delta\chi_1$ of the form ${}^{\{1\}}\pi_0(\delta\chi_1) = 0$ if $\delta\chi_1 = 0$ and ${}^{\{1\}}\pi_0(\delta\chi_1) = \text{const}$ otherwise. However, this would not have made a difference in the integral above; indeed, with respect to the integration measure induced by the prior probability density on $\{\vec{\theta}, \delta\chi_1\}$, the surface $\delta\chi_1 = 0$ constitutes a set of measure zero anyway. Now look at the denominator in the expression for B_{GR}^1 , which is the evidence for the GR hypothesis:

$$\int d\vec{\theta} \pi(\vec{\theta}) p(d|\vec{\theta}, \mathcal{H}_{\text{GR}}, \mathbf{I}). \quad (\text{B.24})$$

Despite the fact that the GR waveforms form a set of measure zero within the set of waveforms used for testing H_1 , the above integral is clearly not zero. It is the evidence for a qualitatively different hypothesis, whose parameter space $\{\vec{\theta}\}$ carries a different

integration measure with respect to which the marginalization of the likelihood is carried out. In particular, \mathcal{H}_{GR} is not ‘included’ in the H_1 model in any sense that is meaningful for model selection.

Next let us consider H_{12} . The numerator of the Bayes factor B_{GR}^{12} in Eq. (B.22) is

$$\int d\vec{\theta} d\delta\chi_1 d\delta\chi_2 \text{}^{\{12\}}\pi(\delta\chi_1, \delta\chi_2) \pi(\vec{\theta}) p(d|\vec{\theta}, \delta\chi_1, \delta\chi_2, H_{12}, \mathbf{I}). \quad (\text{B.25})$$

The parameter space of the GR waveforms has a natural embedding into the parameter space $\{\vec{\theta}, \delta\chi_1, \delta\chi_2\}$ of the waveforms used to test H_{12} , by identification with the hypersurface $\delta\chi_1 = \delta\chi_2 = 0$. Similarly, the parameter spaces of the waveforms we use to test H_1 and H_2 can be identified with the hypersurface $\delta\chi_2 = 0$ and the hypersurface $\delta\chi_1 = 0$, respectively. With respect to the integration measure induced by the prior on $\{\vec{\theta}, \delta\chi_1, \delta\chi_2\}$, these hypersurfaces have measure zero. Despite this, the numerators of B_{GR}^1 and B_{GR}^2 in Eqns. (B.20) and (B.21) are not zero, because of the different integration measures. With our choices of prior probability densities on the different parameter spaces, and the associated waveform models, there is no meaningful sense in which H_1 or H_2 are ‘included’ in H_{12} . Thus, we are indeed testing the disjoint hypotheses H_1 , H_2 , and H_{12} . The latter is the hypothesis that both $\delta\chi_1$ and $\delta\chi_2$ differ from zero: the prior density $\text{}^{\{12\}}\pi(\delta\chi_1, \delta\chi_2)$ gives the surfaces $\delta\chi_1 = 0$ and $\delta\chi_2 = 0$ zero prior mass.

The waveform model that leaves $\{\vec{\theta}, \delta\chi_1, \delta\chi_2\}$ free corresponds to the hypothesis H_{12} . When computing the Bayes factor B_{GR}^{12} , the question one addresses is “Do ψ_1 and ψ_2 both differ from what GR predicts?” This is analogous to what has been done in recent Bayesian work on testing GR [55, 63, 83]. As we are in the process of showing, it is possible to address a more general question, namely “Do ψ_1 , or ψ_2 , or both at the same time, differ from their GR values?”

For completeness, we note that what nested sampling will give us directly are not the Bayes factors of Eqns. (B.20)–(B.22), but rather the Bayes factors for the various hypotheses against the noise-only hypothesis $\mathcal{H}_{\text{noise}}$:

$$\begin{aligned}
B_{\text{noise}}^1 &= \frac{P(d|H_1, \mathbb{I})}{P(d|\mathcal{H}_{\text{noise}}, \mathbb{I})}, \\
B_{\text{noise}}^2 &= \frac{P(d|H_2, \mathbb{I})}{P(d|\mathcal{H}_{\text{noise}}, \mathbb{I})}, \\
B_{\text{noise}}^{12} &= \frac{P(d|H_{12}, \mathbb{I})}{P(d|\mathcal{H}_{\text{noise}}, \mathbb{I})}, \\
B_{\text{noise}}^{\text{GR}} &= \frac{P(d|\mathcal{H}_{\text{GR}}, \mathbb{I})}{P(d|\mathcal{H}_{\text{noise}}, \mathbb{I})}.
\end{aligned} \tag{B.26}$$

These can trivially be combined to obtain the Bayes factors of Eqns. (B.20)–(B.22):

$$B_{\text{GR}}^1 = \frac{B_{\text{noise}}^1}{B_{\text{noise}}^{\text{GR}}}, \quad B_{\text{GR}}^2 = \frac{B_{\text{noise}}^2}{B_{\text{noise}}^{\text{GR}}}, \quad B_{\text{GR}}^{12} = \frac{B_{\text{noise}}^{12}}{B_{\text{noise}}^{\text{GR}}}. \tag{B.27}$$

Now, upon calculating the Bayes factors for each model, we would like to combine these measurements into an overall odds ratio between the GR model and any of the competing hypotheses (Eq. (B.17)). In order to do this, we must specify the prior odds for each model against GR, $P(H_1|\mathbb{I})/P(\mathcal{H}_{\text{GR}}|\mathbb{I})$, $P(H_2|\mathbb{I})/P(\mathcal{H}_{\text{GR}}|\mathbb{I})$, etc. Here one might want to let oneself be guided by, e.g., the expectation that a violation of GR will likely occur at higher post-Newtonian order, and give more weight to H_2 and H_{12} . Or, if one expects a deviation to happen only in a particular phase coefficient (such as ψ_2 in the case of ‘massive gravity’), one may want to down weight the most inclusive hypothesis, in this example H_{12} . In reality, we will not know beforehand what form a violation will take; in particular, it could affect all of the PN coefficients. For the purposes of this analysis, we invoke the principle of indifference among the alternative hypotheses, taking no one to be preferable to any other. This imposes the condition that the prior odds of each against GR are equal. We explicitly note that this is a

choice of the authors for computing final results. This choice results in our effectively taking the average of the Bayes factors for the alternative hypotheses when we compute the odds ratio versus GR.

When combining the Bayes factors into the odds ratio, we therefore assume

$$\frac{P(H_1|\mathbf{I})}{P(\mathcal{H}_{\text{GR}}|\mathbf{I})} = \frac{P(H_2|\mathbf{I})}{P(\mathcal{H}_{\text{GR}}|\mathbf{I})} = \frac{P(H_{12}|\mathbf{I})}{P(\mathcal{H}_{\text{GR}}|\mathbf{I})}. \quad (\text{B.28})$$

Furthermore, we let

$$\frac{P(\mathcal{H}_{\text{modGR}}|\mathbf{I})}{P(\mathcal{H}_{\text{GR}}|\mathbf{I})} = \frac{P(H_1 \vee H_2 \vee H_{12}|\mathbf{I})}{P(\mathcal{H}_{\text{GR}}|\mathbf{I})} = \alpha, \quad (\text{B.29})$$

where we do not specify α ; it will end up being an overall scaling of the odds ratio. This, together with (B.28) and the logical disjointness of the hypotheses H_1 , H_2 , H_{12} implies

$$\frac{P(H_1|\mathbf{I})}{P(\mathcal{H}_{\text{GR}}|\mathbf{I})} = \frac{P(H_2|\mathbf{I})}{P(\mathcal{H}_{\text{GR}}|\mathbf{I})} = \frac{P(H_{12}|\mathbf{I})}{P(\mathcal{H}_{\text{GR}}|\mathbf{I})} = \frac{\alpha}{3}. \quad (\text{B.30})$$

The final expression for the odds ratio for a modification of GR versus GR, in the case where up to two coefficients are used for testing, is then

$${}^{(2)}O_{\text{GR}}^{\text{modGR}} = \frac{\alpha}{3} [B_{\text{GR}}^1 + B_{\text{GR}}^2 + B_{\text{GR}}^{12}]. \quad (\text{B.31})$$

Before continuing to the case of more than two testing parameters, let us compare and contrast what is proposed here with what was done in previous Bayesian work, e.g. [55, 63, 83]. There, one introduced free parameters p_i in the waveform (not necessarily corresponding to shifts in phase coefficients; they could, e.g., be shifts in ringdown

frequencies and damping times), which are zero in GR. Next, one constructed a Bayes factor comparing a model waveform in which all of the p_i are allowed to vary freely with the GR model in which all of the p_i are fixed to zero. In our language and in the case of two testing parameters, this corresponds to only comparing the hypothesis H_{12} with the GR hypothesis \mathcal{H}_{GR} . Hence, what was effectively done in previous work was to address the question: “Do all of the additional free parameters differ from zero at the same time?” A more interesting question to ask is “Do one or more of the extra parameters differ from zero?” This corresponds to testing our hypothesis $\mathcal{H}_{\text{modGR}}$. There is no waveform model that can be used to test the latter hypothesis directly, but as we were able to show, $\mathcal{H}_{\text{modGR}}$ can be broken up into sub-hypotheses, H_1 , H_2 , and H_{12} in the case of two testing parameters. With each of these, a waveform model can be associated, hence they can be tested against the GR hypothesis, \mathcal{H}_{GR} . The resulting Bayes factors can be combined into an odds ratio as in Eq. (B.31), which does compare the more general hypothesis $\mathcal{H}_{\text{modGR}}$ with \mathcal{H}_{GR} .

B.3.3 The general case

So far we have assumed just two testing coefficients, but we may want to use more. In practice it makes sense to pick $\{\psi_1, \dots, \psi_{N_T}\}$, $N_T \leq M$. The number of coefficients used, N_T , will be dictated mostly by computational cost; in Sec. B.4 we will pick $N_T = 3$ but a larger number could be chosen. We then define

$$\mathcal{H}_{\text{modGR}} = \bigvee_{i_1 < i_2 < \dots < i_k; k \leq N_T} H_{i_1 i_2 \dots i_k}. \quad (\text{B.32})$$

When using this set of testing coefficients, the odds ratio for ‘modification to GR’ versus GR becomes:

$$\begin{aligned}
& {}^{(N_T)}O_{\text{GR}}^{\text{modGR}} \\
&= \frac{P(\mathcal{H}_{\text{modGR}}|d, \mathbb{I})}{P(\mathcal{H}_{\text{GR}}|d, \mathbb{I})} \\
&= \frac{P(\bigvee_{i_1 < i_2 < \dots < i_k; k \leq N_T} H_{i_1 i_2 \dots i_k} | d, \mathbb{I})}{P(\mathcal{H}_{\text{GR}}|d, \mathbb{I})},
\end{aligned} \tag{B.33}$$

where as before, $H_{i_1 i_2 \dots i_k}$ is the hypothesis that $\{\psi_{i_1}, \psi_{i_2}, \dots, \psi_{i_k}\}$ do not have the functional dependence on (\mathcal{M}, η) as predicted by GR, but all of the remaining coefficients do. Thus, we are considering the odds ratio for one or more of the phase coefficients $\psi_1, \dots, \psi_{N_T}$ deviating from GR, versus all of them having the functional dependence on masses as in GR.

Using the logical disjointness of the $H_{i_1 i_2 \dots i_k}$ for different subsets $\{i_1, i_2, \dots, i_k\}$ as well as Bayes’ theorem, one can write

$${}^{(N_T)}O_{\text{GR}}^{\text{modGR}} = \sum_{k=1}^{N_T} \sum_{i_1 < i_2 < \dots < i_k} \frac{P(H_{i_1 i_2 \dots i_k} | \mathbb{I})}{P(\mathcal{H}_{\text{GR}} | \mathbb{I})} B_{\text{GR}}^{i_1 i_2 \dots i_k}, \tag{B.34}$$

where

$$B_{\text{GR}}^{i_1 i_2 \dots i_k} = \frac{P(d | H_{i_1 i_2 \dots i_k}, \mathbb{I})}{P(d | \mathcal{H}_{\text{GR}}, \mathbb{I})}. \tag{B.35}$$

Again, one computes the $2^{N_T} - 1$ individual Bayes factors $B_{\text{GR}}^{i_1 i_2 \dots i_k}$ of each of the alternative hypotheses versus GR. The evaluation of the odds ratio requires that we use specific values for the prior odds ratios. We will set them equal to each other, as we

did in Eq. (B.28):

$$\frac{P(H_{i_1 i_2 \dots i_k} | \mathbb{I})}{P(\mathcal{H}_{\text{GR}} | \mathbb{I})} = \frac{P(H_{j_1 j_2 \dots j_l} | \mathbb{I})}{P(\mathcal{H}_{\text{GR}} | \mathbb{I})} \quad \text{for any } k, l \leq N_T, \quad (\text{B.36})$$

in which case the odds ratio $^{(N_T)}O_{\text{GR}}^{\text{modGR}}$ will be proportional to a straightforward average of the Bayes factors. We note once again that other choices could in principle be made. If one expects a violation to be mainly visible in a particular phase coefficient (see, e.g., Table I in [55]), then one might want to down weight the more inclusive hypotheses. Or, one may argue that a deviation will most likely affect all of the coefficients starting from some PN order, but then it would not necessarily be sensible to give more inclusive hypotheses a lower weight. Hence we assign equal prior odds to all of the sub-hypotheses $H_{i_1 i_2 \dots i_k}$.

Also as before, we let

$$\frac{P(\mathcal{H}_{\text{modGR}} | \mathbb{I})}{P(\mathcal{H}_{\text{GR}} | \mathbb{I})} = \alpha, \quad (\text{B.37})$$

where we do not specify α ; it will end up being an overall prefactor in the odds ratio. The equality (B.37), together with (B.36) and the logical disjointness of the $2^{N_T} - 1$ hypotheses $H_{i_1 \dots i_k}$ implies

$$\frac{P(H_{i_1 i_2 \dots i_k} | \mathbb{I})}{P(\mathcal{H}_{\text{GR}} | \mathbb{I})} = \frac{\alpha}{2^{N_T} - 1}. \quad (\text{B.38})$$

In terms of the $H_{i_1 i_2 \dots i_k}$, the odds ratio can then be written as

$$^{(N_T)}O_{\text{GR}}^{\text{modGR}} = \frac{\alpha}{2^{N_T} - 1} \sum_{k=1}^{N_T} \sum_{i_1 < i_2 < \dots < i_k} B_{\text{GR}}^{i_1 i_2 \dots i_k}. \quad (\text{B.39})$$

Analogously to (B.20)–(B.22), we will use priors on the parameters $\delta\chi_i$

$$\pi_{\{i_1 i_2 \dots i_k\}}(\delta\chi_{i_1}, \delta\chi_{i_2}, \dots, \delta\chi_{i_k}) \quad (\text{B.40})$$

which are constant within some large box centered on the origin.

B.3.4 Combining information from multiple sources

Although the detection rate for binary neutron stars is still rather uncertain, we expect advanced instruments to detect several events per year [4]. It is therefore important to take advantage of multiple detections to provide tighter constraints on the validity of GR. Consider a set of \mathcal{N} independent GW events, corresponding to \mathcal{N} independent data sets d_A . We do not assume that deviations from GR are necessarily consistent between events, but rather that they can vary from source to source. We assume there is a common underlying theory of gravity that describes emission of gravitational waves from the sources that are observed, but that shifts in the values of the parameters can vary from one source to another, over and above the dependence of the parameters on the masses. One can write down a combined odds ratio for the catalog of sources:

$$\begin{aligned} & {}^{(N_T)}\mathcal{O}_{\text{GR}}^{\text{modGR}} \\ &= \frac{P(\mathcal{H}_{\text{modGR}}|d_1, \dots, d_{\mathcal{N}}, \mathbb{I})}{P(\mathcal{H}_{\text{GR}}|d_1, \dots, d_{\mathcal{N}}, \mathbb{I})} \\ &= \frac{\sum_{k=1}^{N_T} \sum_{i_1 < i_2 < \dots < i_k} P(H_{i_1 i_2 \dots i_k} | d_1, \dots, d_{\mathcal{N}}, \mathbb{I})}{P(\mathcal{H}_{\text{GR}} | d_1, \dots, d_{\mathcal{N}}, \mathbb{I})} \\ &= \sum_{k=1}^{N_T} \sum_{i_1 < i_2 < \dots < i_k} \frac{P(H_{i_1 i_2 \dots i_k} | \mathbb{I})}{P(\mathcal{H}_{\text{GR}} | \mathbb{I})} {}^{(\text{cat})}B_{\text{GR}}^{i_1 i_2 \dots i_k}, \end{aligned} \quad (\text{B.41})$$

where

$${}^{(\text{cat})}B_{\text{GR}}^{i_1 i_2 \dots i_k} = \frac{P(d_1, \dots, d_{\mathcal{N}} | H_{i_1 i_2 \dots i_k}, \mathbf{I})}{P(d_1, \dots, d_{\mathcal{N}} | \mathcal{H}_{\text{GR}}, \mathbf{I})}. \quad (\text{B.42})$$

Since the events $d_1, \dots, d_{\mathcal{N}}$ are all independent, one has

$$\begin{aligned} P(d_1, \dots, d_{\mathcal{N}} | H_{i_1 i_2 \dots i_k}, \mathbf{I}) &= \prod_{A=1}^{\mathcal{N}} P(d_A | H_{i_1 i_2 \dots i_k}, \mathbf{I}), \\ P(d_1, \dots, d_{\mathcal{N}} | \mathcal{H}_{\text{GR}}, \mathbf{I}) &= \prod_{A=1}^{\mathcal{N}} P(d_A | \mathcal{H}_{\text{GR}}, \mathbf{I}). \end{aligned} \quad (\text{B.43})$$

Thus,

$${}^{(\text{cat})}B_{\text{GR}}^{i_1 i_2 \dots i_k} = \prod_{A=1}^{\mathcal{N}} {}^{(A)}B_{\text{GR}}^{i_1 i_2 \dots i_k}, \quad (\text{B.44})$$

with

$${}^{(A)}B_{\text{GR}}^{i_1 i_2 \dots i_k} = \frac{P(d_A | H_{i_1 i_2 \dots i_k}, \mathbf{I})}{P(d_A | \mathcal{H}_{\text{GR}}, \mathbf{I})}. \quad (\text{B.45})$$

To evaluate the combined odds ratio of the catalog we choose to invoke indifference (as in Eqs. (B.28) and (B.36)) and set the individual prior odds ratios equal to each other, so that

$$\frac{P(H_{i_1 i_2 \dots i_k} | \mathbf{I})}{P(\mathcal{H}_{\text{GR}} | \mathbf{I})} = \frac{\alpha}{2^{N_T} - 1}. \quad (\text{B.46})$$

Together with Eqns. (B.41), (B.44), this leads to

$${}^{(N_T)}\mathcal{O}_{\text{GR}}^{\text{modGR}} = \frac{\alpha}{2^{N_T} - 1} \sum_{k=1}^{N_T} \sum_{i_1 < i_2 < \dots < i_k} \prod_{A=1}^{\mathcal{N}} {}^{(A)}B_{\text{GR}}^{i_1 i_2 \dots i_k}, \quad (\text{B.47})$$

which, up to an overall prefactor, amounts to taking the average of the cumulative Bayes factors (B.44).

Alternatively, one may prefer not to make any assumptions for the prior odds ratios $P(H_{i_1 i_2 \dots i_k} | \mathbf{I})/P(\mathcal{H}_{\text{GR}} | \mathbf{I})$ at all, and focus on the cumulative Bayes factors ${}^{(\text{cat})}B_{\text{GR}}^{i_1 i_2 \dots i_k}$

separately and individually. It will also be of interest to look at the cumulative Bayes factors for the various hypotheses $H_{i_1 i_2 \dots i_k}$ and \mathcal{H}_{GR} against the noise-only hypothesis $\mathcal{H}_{\text{noise}}$:

$$\begin{aligned} \prod_A {}^{(A)}B_{\text{noise}}^{i_1 i_2 \dots i_k} &= \prod_A \frac{P(d_A | H_{i_1 i_2 \dots i_k}, \mathbb{I})}{P(d_A | \mathcal{H}_{\text{noise}}, \mathbb{I})}, \\ \prod_A {}^{(A)}B_{\text{noise}}^{\text{GR}} &= \prod_A \frac{P(d_A | \mathcal{H}_{\text{GR}}, \mathbb{I})}{P(d_A | \mathcal{H}_{\text{noise}}, \mathbb{I})}, \end{aligned} \quad (\text{B.48})$$

and we note that

$${}^{(\text{cat})}B_{\text{GR}}^{i_1 i_2 \dots i_k} = \frac{\prod_A {}^{(A)}B_{\text{noise}}^{i_1 i_2 \dots i_k}}{\prod_{A'} {}^{(A')}B_{\text{noise}}^{\text{GR}}}. \quad (\text{B.49})$$

Finally, we will also look at the individual contributions ${}^{(A)}B_{\text{noise}}^{i_1 i_2 \dots i_k}$ and ${}^{(A)}B_{\text{noise}}^{\text{GR}}$, for $A = 1, 2, \dots, \mathcal{N}$.

B.4 Results

To illustrate the method, we construct catalogs of 15 binary neutron star sources, distributed uniformly in volume, with random sky positions and orientations, and whose total number is taken to be on the conservative side of the ‘realistic’ estimates in [4] for the number of detectable sources in a one-year time span. We take the individual neutron star masses to lie between 1 and $2 M_{\odot}$. The distance interval is between 100 Mpc and 400 Mpc; the former number is the radius within which one would expect ~ 0.5 BNS inspirals per year, and 400 Mpc is the approximate horizon distance in Advanced LIGO. The corresponding signals are added coherently to stationary, Gaussian simulated data for the Advanced Virgo interferometer and the two Advanced LIGOs. A lower cut-off of 8 is implemented on the optimal network SNR, defined as the quadrature sum of individual detector SNRs. The analysis of the surviving signals is performed with an

appropriately modified version of the nested sampling code available in LAL [119].

For simplicity, we took the phase up to only 2PN order ($M = 4$), both in the signal and in the model waveforms. As an example, we considered the case of $N_T = 3$ testing coefficients ψ_1, ψ_2, ψ_3 . Nested sampling then gives $2^3 - 1 = 7$ Bayes factors $B_{\text{GR}}^1, B_{\text{GR}}^2, B_{\text{GR}}^3, B_{\text{GR}}^{12}, B_{\text{GR}}^{13}, B_{\text{GR}}^{23}$, and B_{GR}^{123} , which for a single source can be combined to form the odds ratio (B.39). Note that up to and including 2PN order, the post-Newtonian coefficients take the general form [131]

$$\psi_i(\mathcal{M}, \eta) = \frac{3}{128\eta} g_i(\eta) (\pi\mathcal{M}\eta^{-3/5})^{(i-5)/3}, \quad (\text{B.50})$$

where the $g_i(\eta)$ are polynomials in η ; for the lowest three sub-leading PN orders one has

$$g_1(\eta) \equiv 0, \quad g_2(\eta) = \frac{20}{9} \left(\frac{743}{336} + \frac{11}{4}\eta \right), \quad g_3(\eta) = -16\pi. \quad (\text{B.51})$$

Accordingly, in the model waveforms we allow for deformations

$$\begin{aligned} \psi_1^{\text{GR}}(\mathcal{M}, \eta) = 0 &\rightarrow \frac{3}{128\eta} (\pi\mathcal{M}\eta^{-3/5})^{-4/3} \delta\chi_1, \\ \psi_2^{\text{GR}}(\mathcal{M}, \eta) &\rightarrow \frac{3}{128\eta} g_2(\eta) (\pi\mathcal{M}\eta^{-3/5})^{-1} [1 + \delta\chi_2], \\ \psi_3^{\text{GR}}(\mathcal{M}, \eta) &\rightarrow \frac{3}{128\eta} g_3(\eta) (\pi\mathcal{M}\eta^{-3/5})^{-2/3} [1 + \delta\chi_3]. \end{aligned} \quad (\text{B.52})$$

We take the prior on $\delta\chi_i$ to be flat and centered on zero, with a total width of 0.5. This will be much larger than the deviations we will use in simulated signals and hence suffices to illustrate the method; for real measurements one may want to choose a still wider prior.

For the purposes of showing results, the factor α in Eqns. (B.39) and (B.47) will be set to one.

B.4.1 Measurability of a deviation in a post-Newtonian coefficient

In order to gauge the sensitivity of our method to deviations from GR, we first consider a large number of simulated signals with a constant relative offset $\delta\chi_3$. If GR is violated, we do not expect the deviation to be this simple, but these examples will serve to illustrate both the workings and the effectiveness of the technique for low-SNR sources of the kind we expect in Advanced LIGO and Virgo. We note that ψ_3 is the lowest-order coefficient which incorporates the non-linearity of General Relativity through so-called tail effects [47, 48], and is therefore of particular interest.

Signals with constant relative deviation $\delta\chi_3 = 0.1$

We start with signals that have $\delta\chi_3 = 0.1$. We first compute the odds ratios for individual sources, ${}^{(N_T)}O_{\text{GR}}^{\text{modGR}}$, according to (B.39), with $N_T = 3$. Next we divide these up randomly into catalogs of 15 sources each and compute the combined odds ratios ${}^{(N_T)}O_{\text{GR}}^{\text{modGR}}$ as in (B.47). We do the same thing for injections that are pure GR, i.e. $\delta\chi_i = 0$ for all i , and again compute the quantities ${}^{(N_T)}O_{\text{GR}}^{\text{modGR}}$ for individual sources, and ${}^{(N_T)}O_{\text{GR}}^{\text{modGR}}$ for catalogs of 15 sources each.

Before considering catalogs, let us look at the (log) odds ratios for individual sources as a function of SNR. This is shown in Fig. B.4. The overwhelming majority of signals have SNR between 8 and 15, consistent with our SNR cut and the placement of sources uniformly in volume up to 400 Mpc. Even for low SNR sources, there is separation

between the GR injections and the injections with modified ψ_3 . As one would expect, the separation becomes much clearer with increasing SNR.

Figure B.4 The log odds ratios for individual sources. The blue crosses represent signals with standard GR waveforms, the red circles signals with a constant 10% relative offset in ψ_3 . A separation between the two is visible for SNR greater than ~ 10 and becomes more pronounced as the SNR increases.

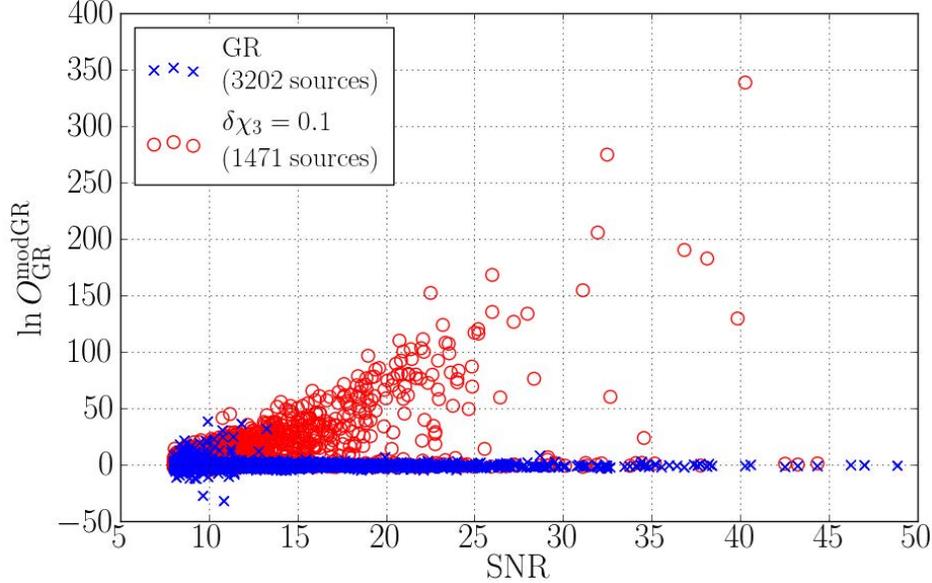
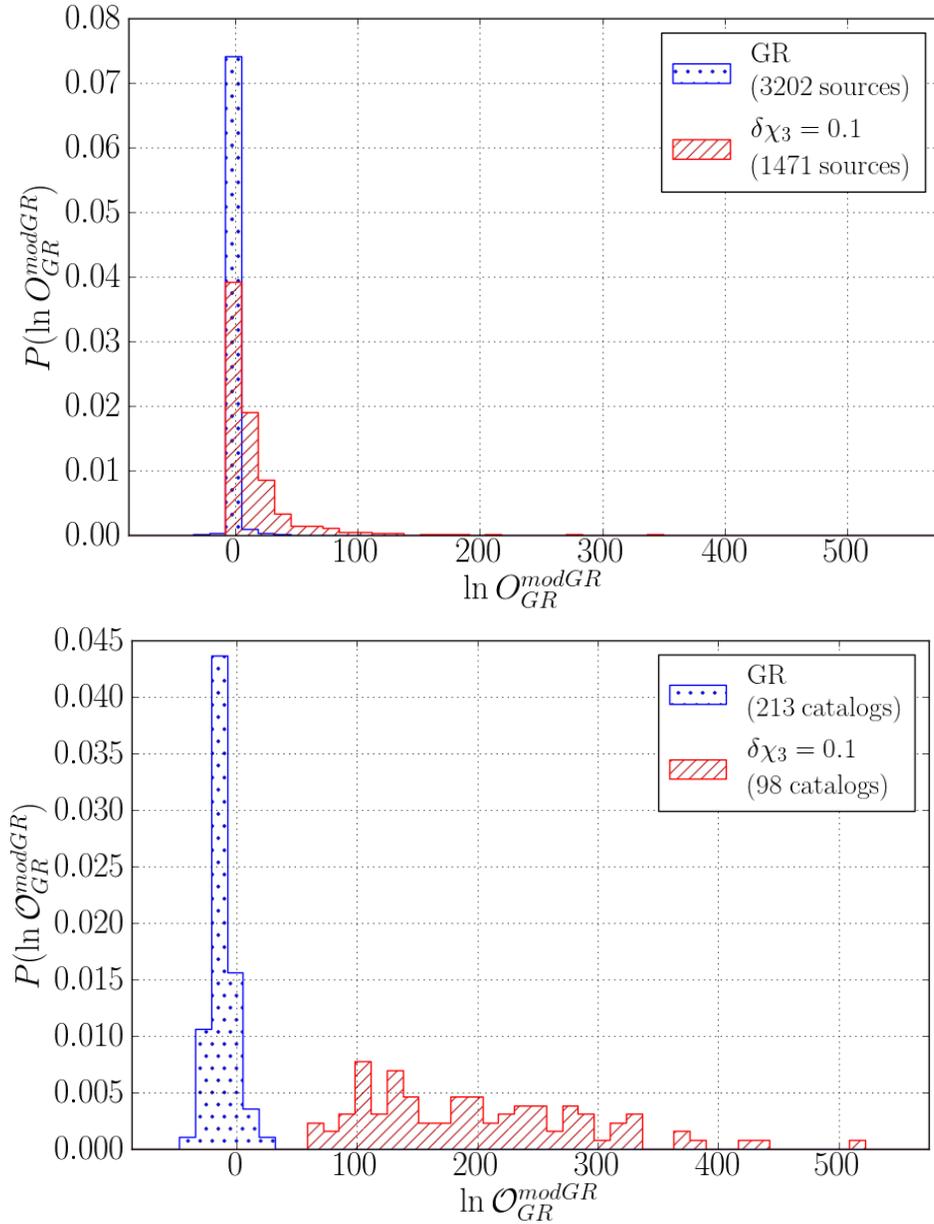


Fig. B.5 shows normalized distributions of the log odds ratios, both for individual sources and for catalogs of 15 sources each. Combining the odds ratios for sources within a catalog will strongly boost our confidence in a violation of GR if one is present at the given level. For this particular choice of $\delta\chi_3$ in the injected non-GR signals, the separation from the GR injections is complete.

It is useful to look at which of the Bayes factors of the component hypotheses tend to give the largest contribution to the odds ratio. What is computed directly by the nested sampling code is not $B_{\text{GR}}^1, \dots, B_{\text{GR}}^3, B_{\text{GR}}^{12}, \dots, B_{\text{GR}}^{23}, B_{\text{GR}}^{123}$, but rather the Bayes factors for each of the hypotheses against the noise-only hypothesis $\mathcal{H}_{\text{noise}}$:

$$B_{\text{noise}}^{i_1 \dots i_k} = \frac{P(d|H_{i_1 \dots i_k}, \mathbf{I})}{P(d|\mathcal{H}_{\text{noise}}, \mathbf{I})}, \quad B_{\text{noise}}^{\text{GR}} = \frac{P(d|\mathcal{H}_{\text{GR}}, \mathbf{I})}{P(d|\mathcal{H}_{\text{noise}}, \mathbf{I})}, \quad (\text{B.53})$$

Figure B.5 Top: The normalized distribution $P(\ln O_{GR}^{modGR})$ of log odds ratios for individual sources, where the injections are either GR or have $\delta\chi_3 = 0.1$. Bottom: The normalized distribution $P(\ln \mathcal{O}_{GR}^{modGR})$ of logs of the combined odds ratios for GR injections and injections with $\delta\chi_3 = 0.1$, for catalogs of 15 sources each. The effectiveness of the catalog approach to testing for deviations from GR comes from the combination of multiple sources, each source contributing to the overall result in proportion to its own Bayes factors.



and one has

$$B_{\text{GR}}^{i_1 \dots i_k} = \frac{B_{\text{noise}}^{i_1 \dots i_k}}{B_{\text{noise}}^{\text{GR}}}. \quad (\text{B.54})$$

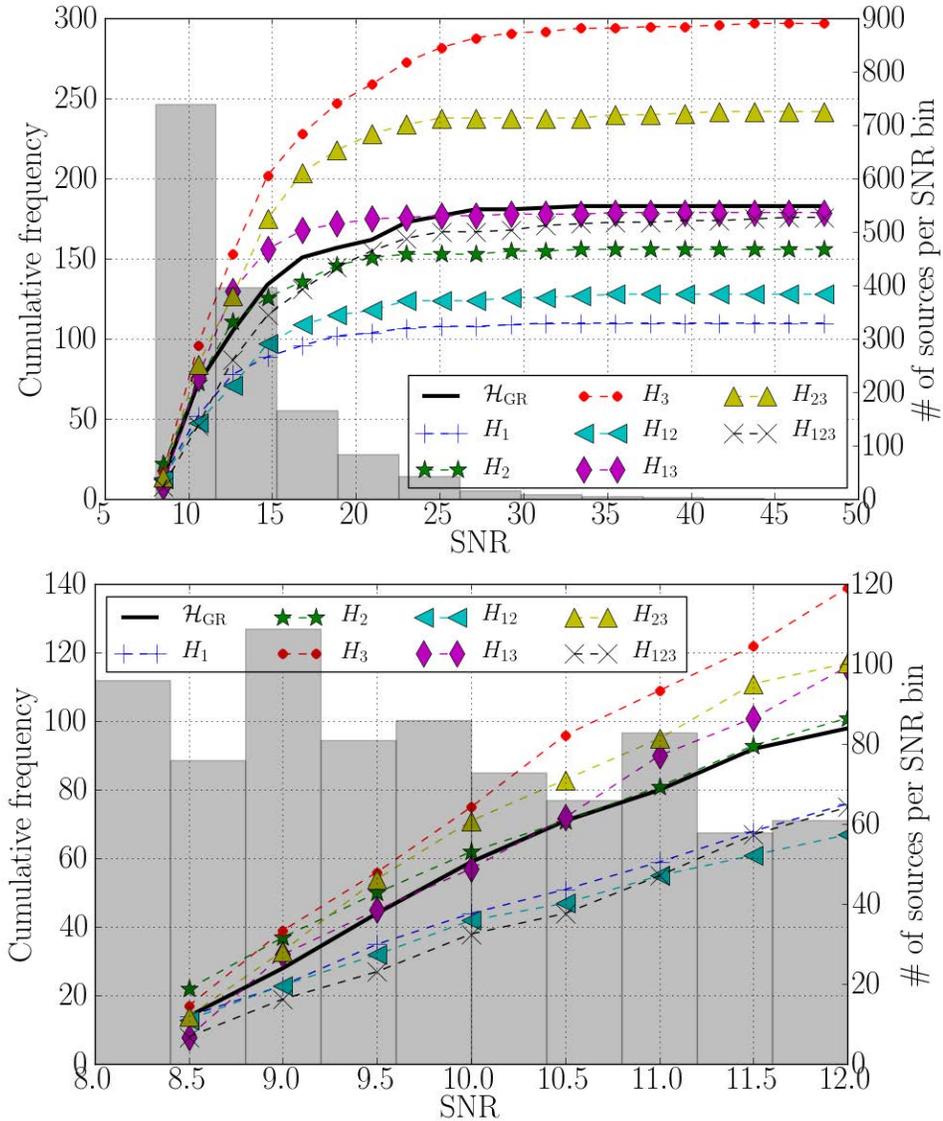
In Fig. B.6, we show the cumulative number of times that a particular $B_{\text{noise}}^{i_1 \dots i_k}$ is the largest, against SNR, for the case where the injections have $\delta\chi_3 = 0.1$. The results are entirely as expected, considering that the injected waveform has a shift in ψ_3 only:

- The Bayes factor B_{noise}^3 corresponding to the hypothesis H_3 dominates;
- The Bayes factors $B_{\text{noise}}^{i_1 \dots i_k}$ corresponding to hypotheses that involve ψ_3 being non-GR tend to outperform those that do not;
- The Bayes factors for the non-GR hypotheses deviate from the GR one already at low SNR, showing that our method will perform well in the low-SNR scenario.

Because of the first two points, one may be tempted to assign different prior odds to the various hypotheses instead of setting them all equal to each other. For instance one might consider downweighting the most inclusive hypothesis, H_{123} , by invoking Occam's razor. However, the violation of GR we assume here is of a rather special form. In reality one will not know beforehand what the nature of the deviation will be; in particular, its effect may not be restricted to a single phase coefficient. It is possible that all coefficients are affected, in which case one would not want to a priori deprecate H_{123} . As explained in Sec. B.3.3, our hypothesis $\mathcal{H}_{\text{modGR}}$ corresponds to the question whether one or more of the phasing coefficients $\{\psi_1, \psi_2, \psi_3\}$ differ from their GR values; one may want to ask a different question, but this is the one that is the most general within our framework. To retain full generality, all sub-hypotheses $H_{i_1 i_2 \dots i_k}$ need to be taken into account and given equal weight.

Signals with constant relative deviation $\delta\chi_3 = 0.025$

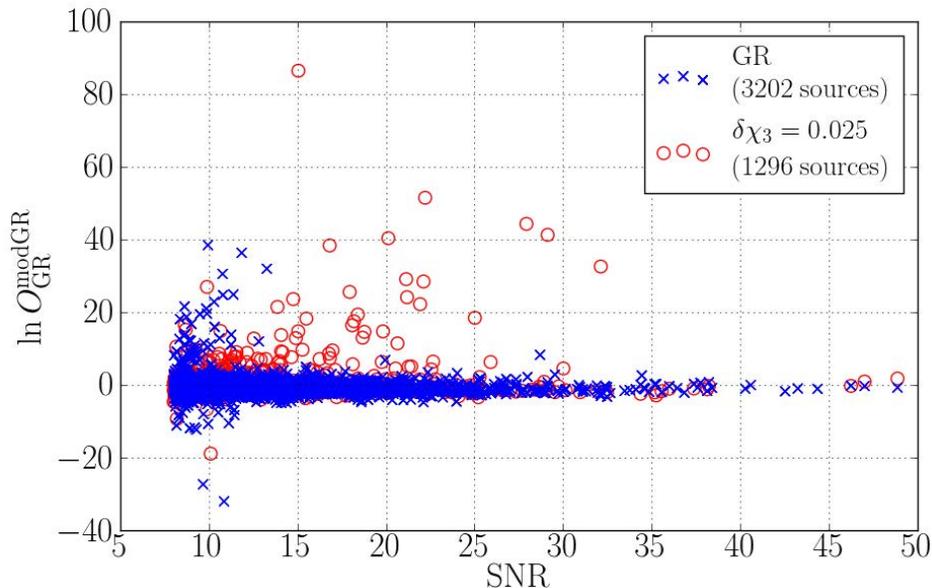
Figure B.6 Top; curves and left vertical axis: For a given SNR, the cumulative number of times that the Bayes factor against noise for a particular hypothesis is the largest for injections with that SNR or below, for $\delta\chi_3 = 0.1$. All 1471 simulated sources were used. As expected, B_{noise}^3 dominates, and Bayes factors for hypotheses that let ψ_3 be non-GR tend to outperform those that do not. The GR model outperforms the one with the largest number of free parameters. Histogram and right vertical axis: The number of sources per SNR bin. Bottom: The same as above, but restricting to sources with SNR < 12 . Similar behavior as for the full set of 1471 sources is observed. Note that already at SNR close to threshold, the GR hypothesis is more likely to be disfavored.



It is clear that, if signals arriving at the Advanced Virgo-LIGO network would have a (constant) fractional deviation in ψ_3 as large as 10%, then at least under the assumption of Gaussian noise, we would have no trouble in discerning this violation of GR even if only 15 events were ever recorded. Now let us look at a smaller deviation in ψ_3 ; say, 2.5%.

In Fig. B.7 we plot the log odds ratios for individual sources against SNR, both for signals with GR waveforms and signals with $\delta\chi_3 = 0.025$. This time the two distributions largely coincide, although there are some outliers which could boost the combined odds ratio when they are present in a catalog of sources.

Figure B.7 The log odds ratios for individual sources. The blue crosses are for signals with standard GR waveforms, the red circles for signals with a constant 2.5% relative offset in ψ_3 . This time there is little separation between the two, although there are outliers for SNR greater than ~ 15 .



In Fig. B.8 we show normalized distributions of the log odds ratio for individual sources, as well as for catalogs with 15 sources each. For individual sources, the distributions are more or less on top of each other. The picture is somewhat different for the catalogs. If a catalog with $\delta\chi_3 = 0.025$ happens to contain one of the outliers visible in Fig. B.7,

then it can boost the combined odds ratio for the catalog.

It is instructive to look at a representative catalog with $\ln^{(3)}\mathcal{O}_{\text{GR}}^{\text{modGR}} > 0$. In Fig. B.9 we show the build-up of the log Bayes factors for the various sub-hypotheses against GR, as well as the odds ratio itself. In a scenario where the evidence for a GR violation is marginal, it is imperative to include as many hypotheses as possible in the analysis. Indeed, in this example, the hypothesis H_3 is not the most favored one; instead, it is H_1 . Note also that if we had only tested the most inclusive hypothesis H_{123} against GR (as one might do if one expects a deviation in all of the PN parameters), we would have concluded that the GR hypothesis is the favored one. The same is true if we had only tested H_2 , as one would do when specifically looking for a ‘massive graviton’. Even the log Bayes factor for H_{23} ends up being negative. We also remind the reader that we will not know beforehand what the precise nature of the GR violation will be.

By using the distribution of log odds ratios for simulated catalogs of GR sources, one can establish a threshold which the odds ratio of a given catalog must overcome in order that a violation of GR becomes credible. This would be the analog of what was done in [193] (see Fig. 7 of that paper), where the distribution of the log Bayes factor $\ln B_{\text{S,N}}$ for the presence of a signal versus noise-only was computed for many realizations of the noise, in the absence of a signal. Consider the distribution $P(\ln^{(N_T)}\mathcal{O}_{\text{GR}}^{\text{modGR}}|\kappa, \mathcal{H}_{\text{GR}}, \text{I})$ of log odds ratio for the collection κ of simulated catalogs of signals that are in accordance with GR. Given a ‘false alarm probability’ β , a threshold $\ln \mathcal{O}_\beta$ for the odds ratio can be set as follows:

$$\beta = \int_{\ln \mathcal{O}_\beta}^{\infty} P(\ln \mathcal{O}|\kappa, \mathcal{H}_{\text{GR}}, \text{I}) d \ln \mathcal{O}. \quad (\text{B.55})$$

Now suppose we also have a distribution $P(\ln^{(N_T)}\mathcal{O}_{\text{GR}}^{\text{modGR}}|\kappa', \mathcal{H}_{\text{alt}}, \text{I})$ of log odds ratio for a collection κ' of simulated catalogs of signals which follow some alternative theory (in this example, one which leads to a shift $\delta\chi_3 = 0.025$). Then we can quantify the

Figure B.8 Top: The normalized distribution $P(\ln O_{GR}^{modGR})$ of log odds ratios for individual sources, where the injections are either GR or have $\delta\chi_3 = 0.025$. Bottom: The normalized distribution $P(\ln \mathcal{O}_{GR}^{modGR})$ of logs of the combined odds ratios for GR injections and injections with $\delta\chi_3 = 0.025$, for catalogs of 15 sources each. For individual sources, the two distributions essentially lie on top of each other. However, when sources are combined into catalogs, it is possible for an outlier to boost the odds ratio of the entire catalog.

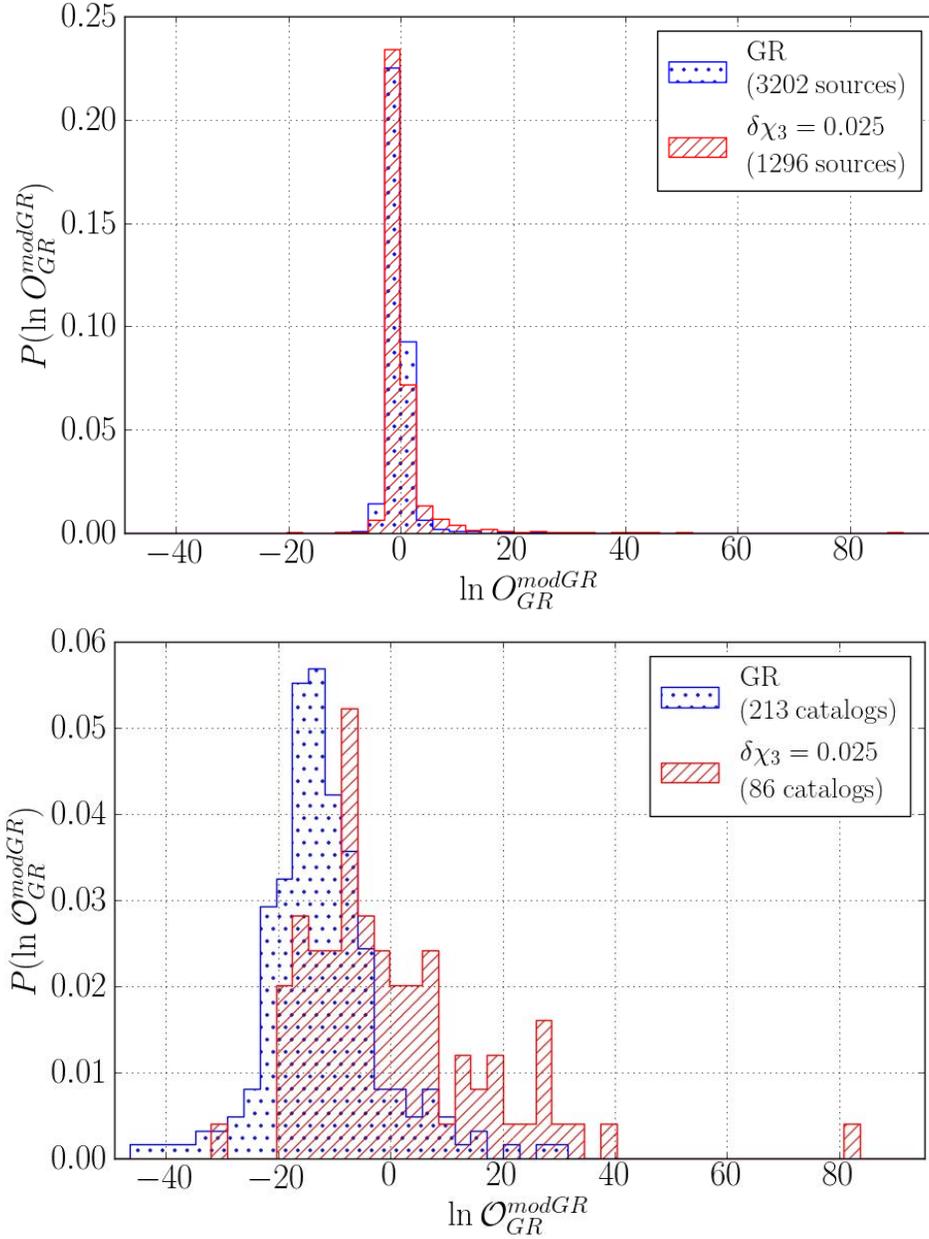
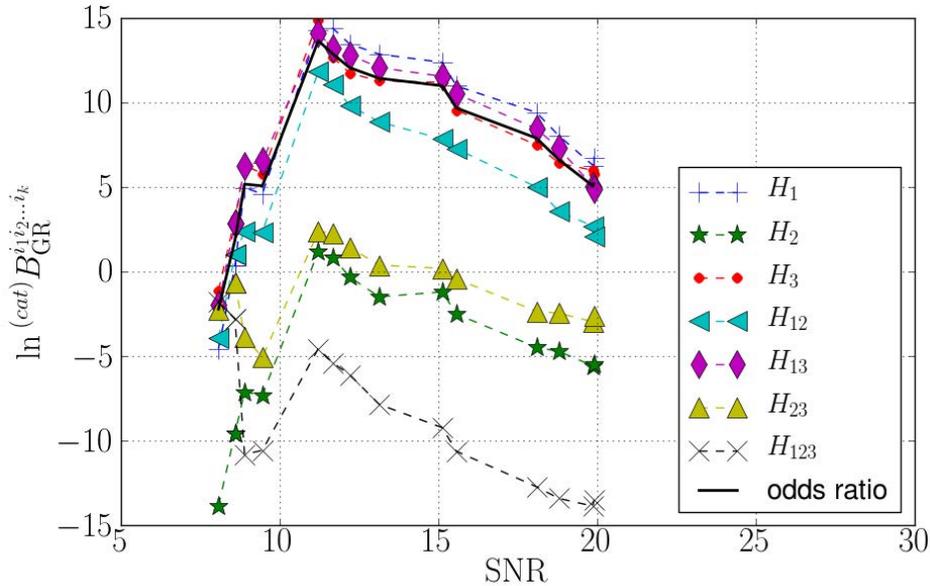


Figure B.9 The build-up of cumulative Bayes factors against GR for individual hypotheses, and the odds ratio, for a typical catalog with $\delta\chi_3 = 0.025$. Note that on the basis of the Bayes factor against GR of the most inclusive hypothesis H_{123} alone, one would have concluded that the GR model is in fact the favored one. Even the log Bayes factor for H_{23} ends up being negative. Additionally, the hypothesis with the largest Bayes factor is not H_3 but H_1 . This illustrates that it is necessary to include as many hypotheses as possible in the analysis.



chance that a deviation from GR of this particular kind will be detected with a false alarm probability smaller than the given β , by means of an efficiency ζ , defined as

$$\zeta = \int_{\ln \mathcal{O}_\beta}^{\infty} P(\ln \mathcal{O} | \kappa', \mathcal{H}_{\text{alt}}, \mathbb{I}) d \ln \mathcal{O}. \quad (\text{B.56})$$

We note that with these definitions, the efficiency is independent of the overall prior odds of $\mathcal{H}_{\text{modGR}}$ versus \mathcal{H}_{GR} , as the factor α in Eqns. (B.37) and (B.47) will just cause both the distributions $P(\ln^{(N_T)} \mathcal{O}_{\text{GR}}^{\text{modGR}} | \kappa, \mathcal{H}_{\text{GR}}, \mathbb{I})$ and $P(\ln^{(N_T)} \mathcal{O}_{\text{GR}}^{\text{modGR}} | \kappa', \mathcal{H}_{\text{alt}}, \mathbb{I})$, and the threshold $\ln \mathcal{O}_\beta$, to be shifted by $\ln \alpha$.

In the present example, with $\delta\chi_3 = 0.025$ and catalogs of 15 sources each, for $\beta = 0.05$ one has $\zeta = 0.22$. Hence, by this standard and for the given number of sources in a catalog, a GR violation of this kind is just borderline detectable.

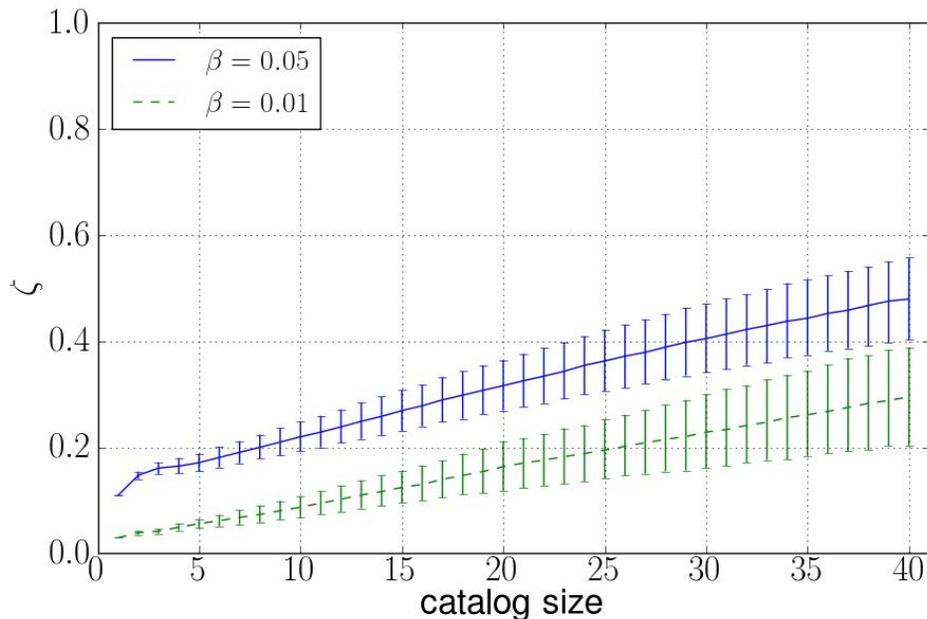
However, the number of sources per catalog, 15 in the examples shown so far, and the chosen false alarm probability, $\beta = 0.05$, are somewhat arbitrary. In reality the size of the catalog will depend on the number of detected sources, and the false alarm probability is set according to the required confidence. It is therefore of interest to investigate the effects of both of these factors. In Fig. B.10 we show the behavior of the efficiency as a function of the catalog size and the false alarm probability. To account for the arbitrariness in which the sources are combined to form a catalog, we show the median and 68% confidence interval of the efficiencies from 5000 random source orderings as the central curves and the error bars, respectively. Results are shown for $\beta \in \{0.05, 0.01\}$.

As is evident from Fig. B.10, the efficiency rises as a function of the catalog size. This highlights the importance of combining all available sources in the advanced detector era, when one looks for deviations from GR. The maximum catalog sizes shown are

comparable to the ‘realistic’ estimates of the number of detections of binary neutron star inspirals in a the span of a year [4].

We see that $\delta\chi_3 = 0.025$ is a borderline case in terms of discernability of a GR violation. Later on, when we show posterior PDFs, it will become evident that indeed, $\delta\chi_3$ can typically be measured with an an accuracy of this order.

Figure B.10 The efficiency of detecting a GR violation for sources with $\delta\chi_3 = 0.025$, as a function of catalog size for false alarm probabilities $\beta \in \{0.05, 0.01\}$. The median and the 68% confidence interval from 5000 random catalog orderings are shown as the central curve and the error bars, respectively. The efficiency increases as a function of catalog size, once again underscoring the benefit of combining all available data.



B.4.2 Effect of number of hypotheses used

It is of interest to see what would have happened if we had used a smaller number of testing coefficients; say, $\{\psi_1, \psi_2\}$, so that the hypotheses to be tested are H_1 , H_2 , and H_{12} . In the example with $\delta\chi_3 = 0.025$ as in the previous subsection, the PN order where the deviation occurs, namely 1.5PN, would then be higher than the PN orders

associated with our testing coefficients, which are 0.5PN and 1PN.

In Fig. B.11, the following two things are shown:

- In the case where only $\{\psi_1, \psi_2\}$ are testing coefficients, we compute the thresholds $\ln^{(2)}\mathcal{O}_\beta$ corresponding to false alarm probabilities (FAPs) $\beta \in \{0.05, 0.01\}$. Next, we re-calculate the false alarm probabilities for the same thresholds, but now for the case where there are three testing parameters, $\{\psi_1, \psi_2, \psi_3\}$, and show the difference in false alarm probabilities;
- On the other hand, one can compare the efficiencies $^{(2)}\zeta$ and $^{(3)}\zeta$ for the two and three parameter cases, for fixed false alarm probabilities $\beta \in \{0.05, 0.01\}$. This is shown in the bottom panel.

As expected, in the first case (fixed thresholds for the odds ratios), the false alarm probabilities increase in going from two to three testing parameters, but only moderately so. On the other hand, for fixed false alarm probabilities, there is no appreciable change in efficiency. Indeed, the spread in the GR ‘background’ will increase with an increase in hypotheses to be tested against GR; yet, having more hypotheses does not really hurt us in terms of our ability to detect a deviation from GR.

Fig. B.11 indicates the typical behavior for catalogs with a specific deviation from GR, in this case $\delta\chi_3 = 0.025$. It is worth repeating, however, that especially when there is only marginal evidence for a GR violation, it is important to use as many hypotheses as is computationally feasible; see Fig. B.9 (and also Fig. B.16 and B.17below). Also, we will obviously not know beforehand what the nature of the GR violation is.

One may nevertheless wonder how our 3-parameter test would compare with a ‘targeted search’ that only looks for a deviation in ψ_3 , which in this example happens to be

where the deviation actually is. With our choice of $\alpha = 1$, this corresponds to setting $\mathcal{O}_{\text{GR}}^{\text{modGR}} = {}^{(\text{cat})}B_{\text{GR}}^3$. Fig. B.12 shows the change in false alarm probabilities in going from testing only H_3 to the full test for fixed log odds ratio thresholds, as well as the change in efficiencies for fixed false alarm probabilities. The results are as follows:

- The change in false alarm probabilities for fixed log odds ratio thresholds is minor;
- However, especially for a large number of sources per catalog, the efficiencies show a clear rise. This can be accounted for by the fact that, for a small violation of GR, it will not always be the case that the Bayes factor against GR for H_3 is the largest, but our method is able to compensate for that.

We conclude that for this particular example, our method with $N_T = 3$ testing parameters will tend to outperform a ‘targeted search’ that happens to look for the violation actually present. However, we do not expect this to be true for more complicated deviations from GR.

B.4.3 Measurability of deviations with non-PN frequency dependences

The aim of the previous subsections was to get a rough idea of the sensitivity of our method to deviations in post-Newtonian coefficients, and in order to gauge this we assumed a constant relative offset in the physically interesting parameter ψ_3 . However, we stress once again that we do not expect a violation of GR to manifest itself as a simple constant relative shift in one of the post-Newtonian coefficients. Even if modifications are confined to the PN coefficients, the $\delta\chi_i$ in the signals can be dependent on (\mathcal{M}, η) , in addition to whatever charges and coupling constants may be present. Moreover, a

Figure B.11 Top: The change in false alarm probabilities (FAPs) in going from two to three testing parameters, but keeping the odds ratio thresholds fixed. Bottom: The change in efficiencies when keeping the false alarm probabilities fixed. The plots shown are for the case where the signals have $\delta\chi_3 = 0.025$. We see that increasing the number of testing parameters has only a moderate effect on the FAPs, while for fixed FAPs, the efficiencies do not change appreciably. Note, however, that when the evidence for a deviation from GR is marginal, the use of as many hypotheses as possible can be pivotal in finding the violation (see Fig. B.9, and also Fig. B.16 and B.17).

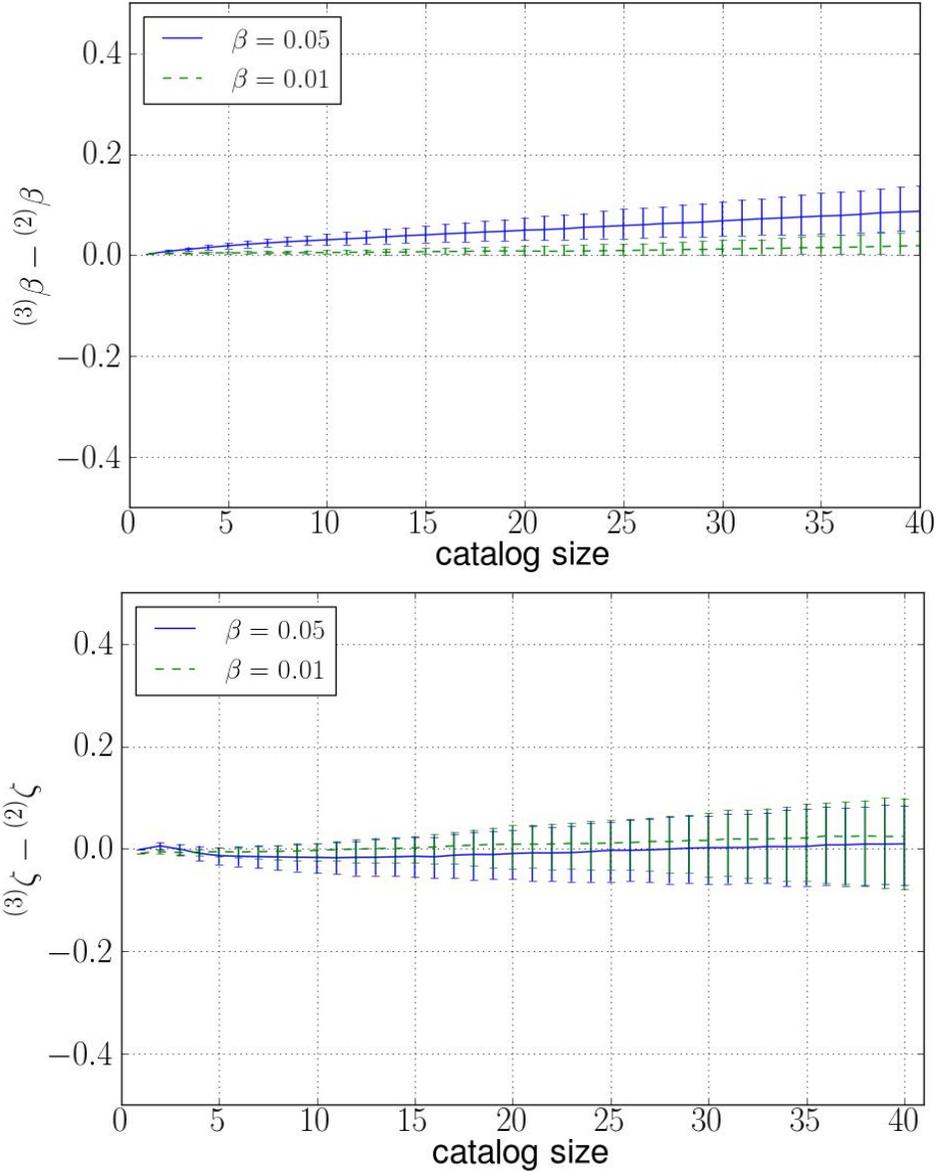
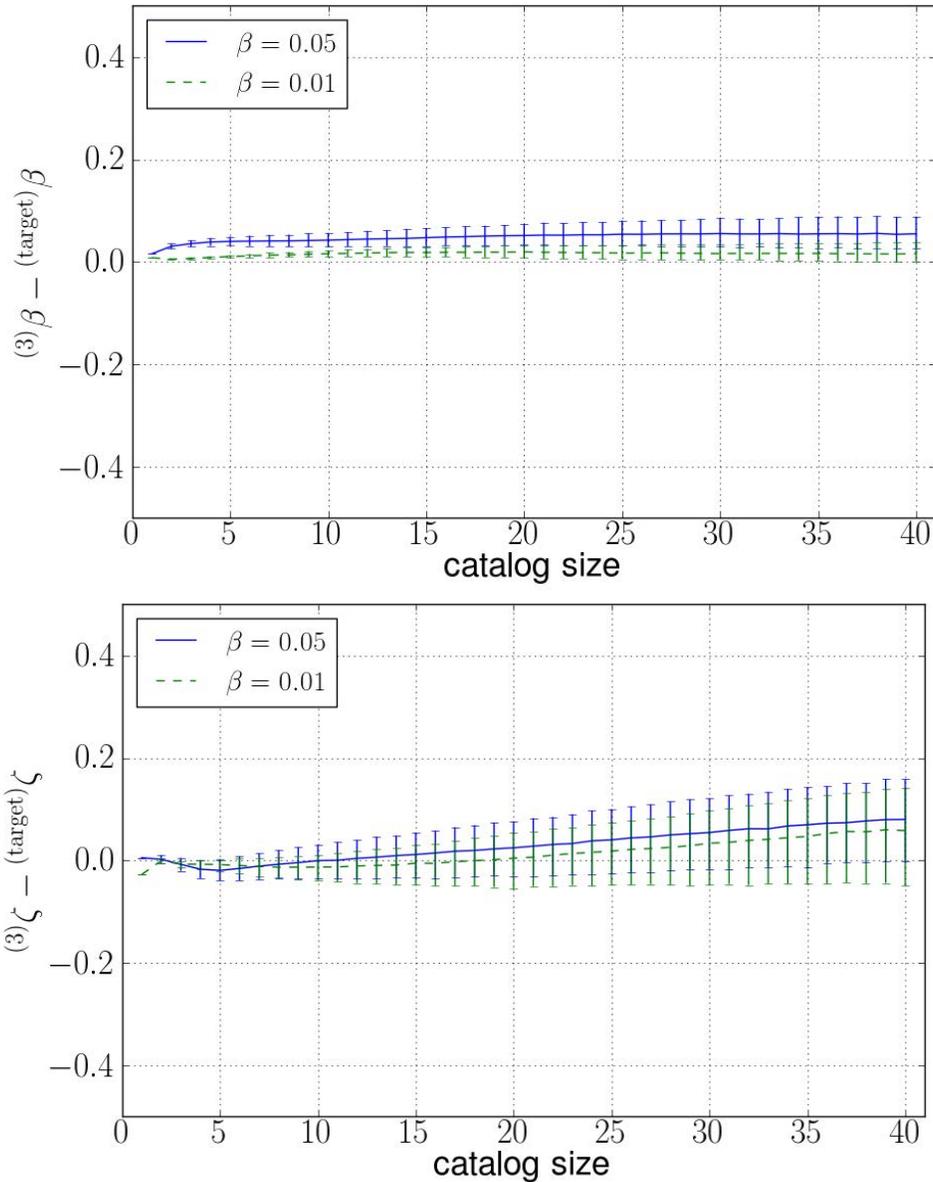


Figure B.12 Top: The difference of false alarm probabilities, ${}^{(3)}\beta - {}^{(\text{target})}\beta$, for fixed log odds ratio thresholds and signals having $\delta\chi_3 = 0.025$, between our 3-parameter test and a ‘targeted search’ which only looks for a deviation in ψ_3 , i.e., only tests the hypothesis H_3 against GR. We see that the difference is minor. Bottom: More important is the difference in efficiencies, ${}^{(3)}\zeta - {}^{(\text{target})}\zeta$, for fixed false alarm probabilities. Especially for a large number of sources per catalog, our 3-parameter test is actually more efficient than the ‘targeted search’, at least for this particular example. This is because the Bayes factor against GR for H_3 will not be the largest in every catalog, but our method naturally compensates for that. Of course, we do not expect our method to outperform a targeted search in the case of a more complicated deviation from GR.



deviation from GR could introduce terms in the phase with frequency dependences that do not correspond to any of the PN contributions. We now show that the method can also be sensitive to violations of that kind, even though the model waveforms we use in our analyses only have deformations of PN terms. Let us give an heuristic example where the phase of the simulated signals contains a term with an anomalous frequency dependence in between that of the 1PN and 1.5PN contributions. Specifically,

$$\frac{3}{128\eta}(\pi\mathcal{M}\eta^{-3/5})^{-5/6}\delta\chi_A f^{-5/6}, \quad (\text{B.57})$$

and we note that the 1PN term goes like f^{-1} and the 1.5PN term like $f^{-2/3}$; thus, the deviation introduced here could be dubbed ‘1.25PN’. However, for the recovery, we will continue to use the same model waveforms as before, which can only have shifts in the phase coefficients at 0.5PN, 1PN, and 1.5PN. Our aim in this subsection is to show that they will nevertheless allow us to find a deviation in the signal of the form (B.57).

We now need to make a choice for $\delta\chi_A$. We aim to show that even if there is a deviation in the phase that is not represented in any of our model waveforms, it can be recovered, if near the ‘bucket’ of the noise curve (at $f \sim 150$ Hz) the amount by which it affects the phase is on a par with a shift in the PN coefficients of more than a few percent.

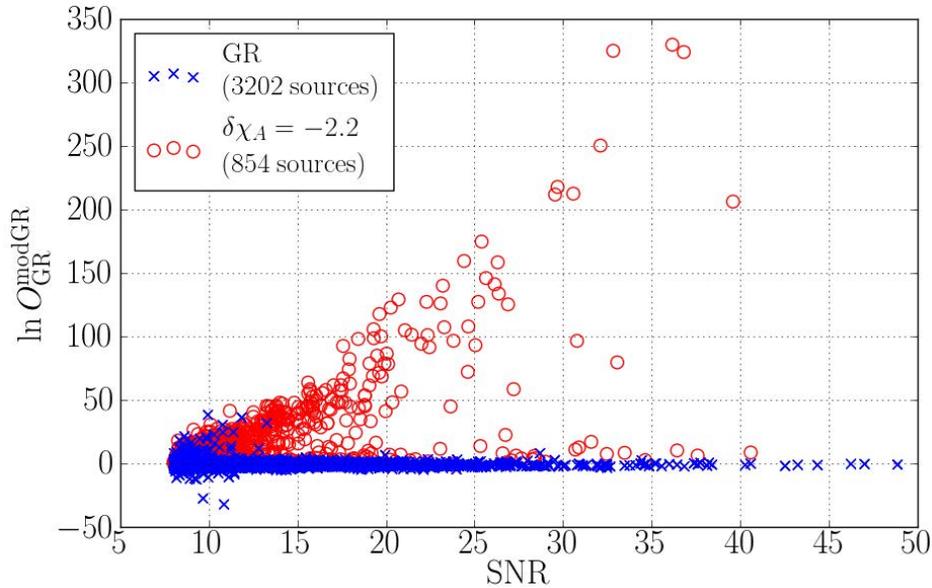
For definiteness, let us take $\delta\chi_A$ to be constant, and such that at $f = 150$ Hz and for a system of $(1.5, 1.5) M_\odot$, the contribution (B.57) to the phase is equal to the change caused by a shift in the 1.5PN contribution with $\delta\chi_3 = 0.1$:

$$\begin{aligned} & (\pi 3 M_\odot)^{-5/6} \delta\chi_A (150 \text{ Hz})^{-5/6} \\ &= g_3(0.25) (\pi 3 M_\odot)^{-2/3} \times 0.1 \times (150 \text{ Hz})^{-2/3}, \end{aligned} \quad (\text{B.58})$$

leading to $\delta\chi_A = -2.2$.

As before, we first give results for the odds ratios of individual sources with increasing SNR; see Fig. B.13. We see that even at small SNR there is already a good separation between the GR injections and the injections with a modification in the structure of the phase.

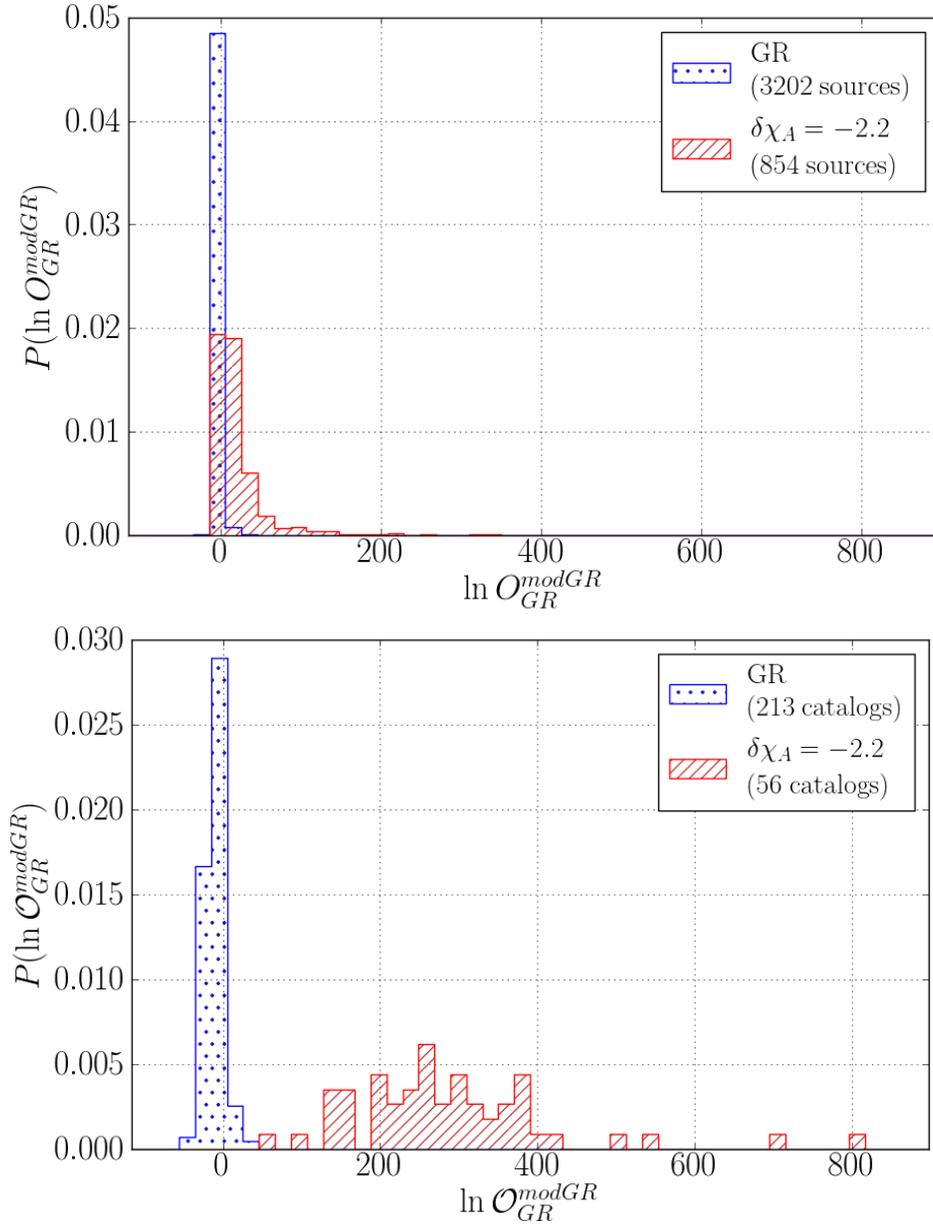
Figure B.13 The log odds ratios for individual sources for injections with an anomalous frequency dependence in the phase. The blue crosses represent GR injections, while the red circles are for signals that have a contribution to the phase with a frequency dependence in between that of the 1PN and 1.5PN terms (‘1.25PN’). As with the $\delta\chi_3 = 0.1$ example, a separation between the two is visible for SNR greater than ~ 10 and becomes more pronounced as the SNR increases.



Next we show normalized distributions of the log odds ratios, both for individual sources and for catalogs of 15 sources each: Fig. B.14. As expected, for the catalogs there is an excellent separation between the GR injections and injections with a modification in the phase.

Now let us look at the cumulative number of times that the Bayes factor against noise for a particular hypothesis is the largest, for individual sources with $\delta\chi_A = -2.2$, arranged with increasing SNR (Fig. B.15). From SNR greater than ~ 15 , the Bayes factor B_{noise}^2

Figure B.14 Top: The normalized distribution $P(\ln O_{GR}^{modGR})$ of log odds ratios for individual sources, where the injections are either GR or have the anomalous frequency dependence. Bottom: The normalized distribution $P(\ln O_{GR}^{modGR})$ of logs of the combined odds ratios for GR injections and injections for catalogs of 15 sources each.



starts to dominate, followed by B_{noise}^{23} and B_{noise}^3 , with the latter two crossing over between $\text{SNR} \sim 20$ and $\text{SNR} \sim 25$. Already at $\text{SNR} \sim 9$, all of the $B_{\text{noise}}^{i_1 i_2 \dots i_k}$ dominate the Bayes factor $B_{\text{noise}}^{\text{GR}}$ for the GR hypothesis. However, near the SNR threshold, no single hypothesis dominates clearly, which again shows that as many hypotheses as possible should be included in the analysis.

Especially in this case, it is interesting to look at the growth of cumulative Bayes factors against GR for individual hypotheses, as well as of the odds ratio, as sources with increasing SNR are being added within catalogs of 15 sources. This is shown for a few example catalogs in Fig. B.16 and B.17. The salient features are:

- Even if all 15 sources only have modest SNR, by their cumulative contributions they can cause a relatively large odds ratio for the catalog as a whole;
- In catalogs containing a source with a particularly high SNR, it is by no means a given that the contribution of this source will dominate the odds ratio compared to the cumulative contributions of the other sources;
- Which hypothesis comes out on top will vary from one catalog to another; in the examples of Fig. B.16 and B.17 we see H_{12} , H_2 , or H_{23} giving the largest contribution, respectively, but there are examples where any of the other four sub-hypotheses contributes the most. In this respect we note that the odds ratio for a catalog is proportional to the average of the cumulative Bayes factors themselves, not of their logarithms. If one were to a priori favor particular (subsets of) hypotheses, the log odds ratio could be lowered by as much as 100. This could have a large effect on the false alarm probability; see Fig. B.14. These are again arguments for using as many sub-hypotheses as possible, and give them equal relative prior odds.

Figure B.15 Top; curves and left vertical axis: The cumulative number of times that the Bayes factor against noise for a particular component hypothesis is the largest, with increasing SNR, for individual sources with $\delta\chi_A = -2.2$. All 854 simulated sources were used. Histogram and right vertical axis: The number of sources per SNR bin. Bottom: The same as above, but now for sources with SNR < 12. Close to threshold, no single hypothesis is the dominant one.

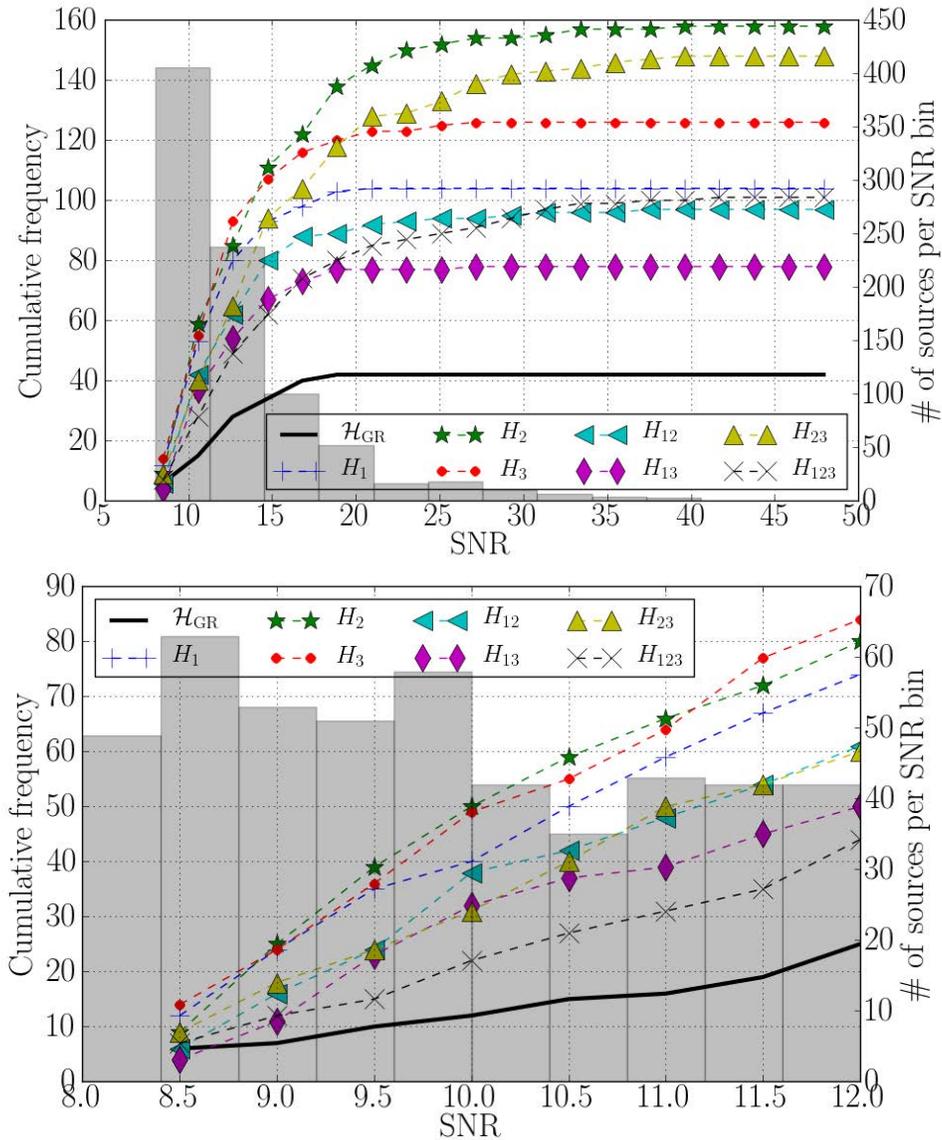


Figure B.16 A few examples of how cumulative Bayes factors against GR for individual hypotheses, and the odds ratio, grow as sources with increasing SNR are being added within three different catalogs of 15 sources in total. These catalogs have sources with only modest SNRs (< 20). Note the large differences in contributions from different hypotheses, and in the ordering of Bayes factors, between these two catalogs.

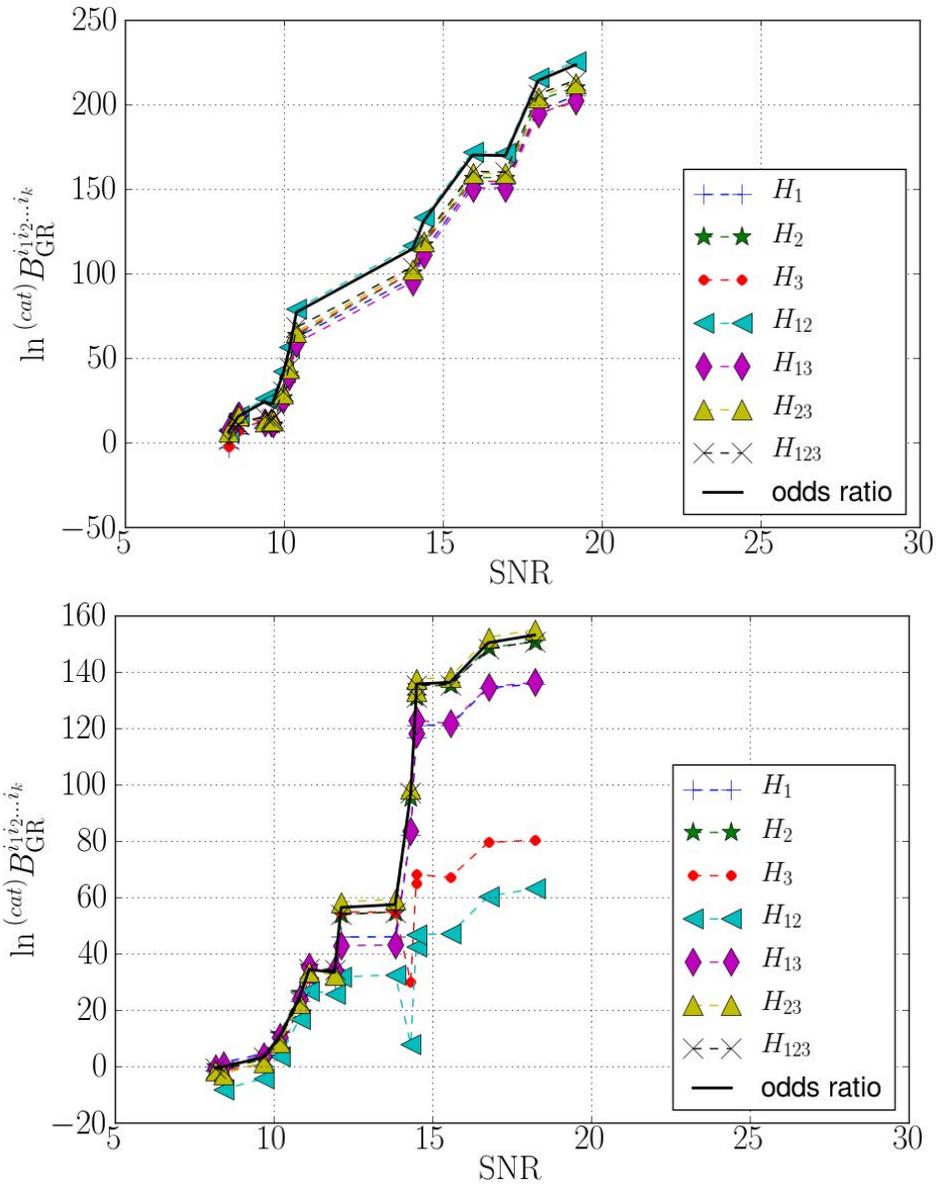
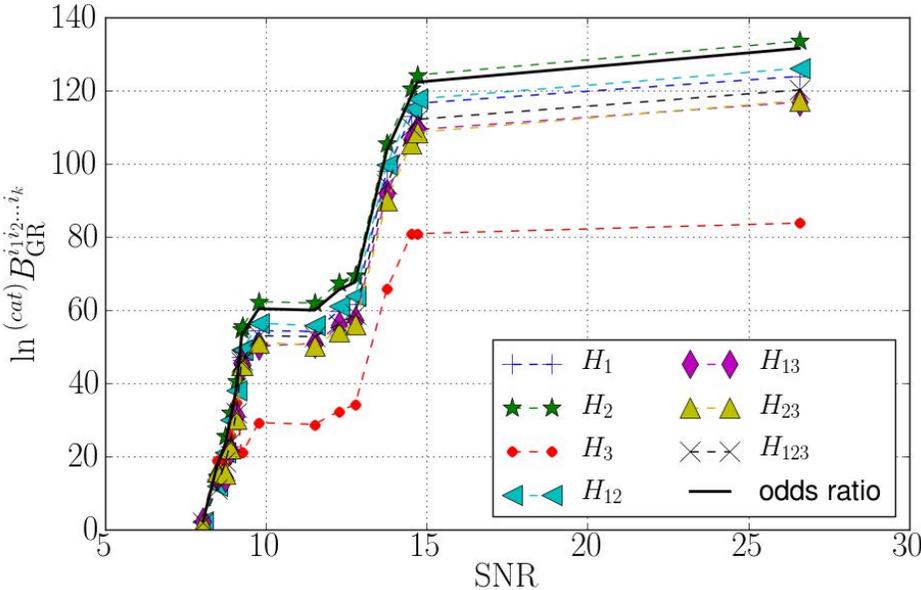


Figure B.17 A third examples of how cumulative Bayes factors against GR for individual hypotheses, and the odds ratio, grow as sources with increasing SNR are being added within three different catalogs of 15 sources in total. This catalog has a high SNR source; note however that the source with the highest SNR does not cause a particularly large ‘boost’, and the cumulative log odds ratio ends up being considerably lower than the examples in Figure B.16.



Note that in principle we could have extended our model waveforms with more free parameters so as to be more sensitive to deviations of this kind, e.g. by including a term in the phase of the form $\mathcal{A}f^a$, similar to what one has in the so-called parameterized post-Einsteinian (ppE) waveforms [216]. However, the point of our method is to search for generic deviations from GR. In the future, we will want to search with more sophisticated (time domain) waveforms whose use is computationally more demanding in a Bayesian setting, and we will not be able to allow for an arbitrarily large number of deformations in the model waveforms. However, here we have an example where the injections have a contribution to the phase that is not present in any of the model waveforms, yet is clearly observable. Recall that the overall magnitude of the anomalous phase contribution was chosen so that, at $f \sim 150$ Hz, it changes the phase by a similar amount as a shift in ψ_3 of 10%. Our results make it plausible that generically, when there is a deviation in the phase which, at frequencies where the detectors are the most sensitive, causes a phase change on a par with a change in one of the (low order) PN coefficients of more than a few percent, it will be detectable.

B.4.4 Parameter estimation

Finally, let us look at some posterior PDFs for the $\delta\chi_i$. We stress that unlike Bayes factors, the PDFs can not be combined across sources since we should not expect the $\delta\chi_i$ to be independent of the component masses; they can differ from source to source. Even looking at the PDFs for a single source may then be misleading: even if the deviation is exactly in one or more of the PN coefficients, a given source will have values for the $\delta\chi_i$ that are representative just for the (\mathcal{M}, η) of that source, and possibly also the values of additional charges that may appear in an alternative theory of gravity. In a given catalog, there may be only one source with sufficient SNR to allow for accurate

parameter estimation, in which case the posteriors will not tell us much even if they are strongly peaked. More generally, the deviation from GR may manifest itself by the appearance of terms in the phase that do not have the frequency dependence of any of the PN contributions. However, in the event that the log odds ratio and Bayes factors strongly favor the GR hypothesis, posterior PDFs will allow us to constrain deviations in the PN coefficients, thereby adding further support that GR is the correct theory. Hence we start with an analysis of pure GR injections.

Example: A GR injection

Let us first look at a GR source with $(\mathcal{M}, \eta, D) = (1.31 M_{\odot}, 0.243, 131 \text{ Mpc})$, and a LIGO-Virgo network SNR of 23.0. The Bayes factors for the various component hypotheses are:

$$\begin{aligned}
 \ln B_{\text{GR}}^1 &= -2, & \ln B_{\text{GR}}^2 &= -2, & \ln B_{\text{GR}}^3 &= -2, \\
 \ln B_{\text{GR}}^{12} &= -3, & \ln B_{\text{GR}}^{13} &= -1, & \ln B_{\text{GR}}^{23} &= -1, \\
 \ln B_{\text{GR}}^{123} &= -2.
 \end{aligned}
 \tag{B.59}$$

We recall that these are the Bayes factors for a particular deviation from GR versus GR. The GR hypothesis is favored in all cases. We can also look at the Bayes factors

for all of the hypotheses against noise:

$$\begin{aligned}
\ln B_{\text{noise}}^{\text{GR}} &= 211, \\
\ln B_{\text{noise}}^1 &= 209, \quad \ln B_{\text{noise}}^2 = 209, \quad \ln B_{\text{noise}}^3 = 209, \\
\ln B_{\text{noise}}^{12} &= 208, \quad \ln B_{\text{noise}}^{13} = 210, \quad \ln B_{\text{noise}}^{23} = 210, \\
\ln B_{\text{noise}}^{123} &= 209.
\end{aligned}
\tag{B.60}$$

Hence the signal is picked up very well by the waveforms of all of the hypotheses, with the GR waveform doing slightly better.

Let us now look at some posterior PDFs. In Fig. B.19 and B.19, we show the PDFs for $\delta\chi_1$, $\delta\chi_2$, and $\delta\chi_3$, respectively for the waveforms that have free parameters $\{\vec{\theta}, \delta\chi_1\}$, $\{\vec{\theta}, \delta\chi_2\}$, and $\{\vec{\theta}, \delta\chi_3\}$, with $\vec{\theta}$ the parameters of the GR waveform. We see that the distributions are all narrowly peaked around the correct value of zero.

Example: A signal with $\delta\chi_3 = 0.1$.

We now consider an example with $(\mathcal{M}, \eta, D) = (1.18 M_\odot, 0.244, 196 \text{ Mpc})$, with a non-zero relative shift in ψ_3 of $\delta\chi_3 = 0.1$, and network SNR 23.2. The Bayes factors are:

$$\begin{aligned}
\ln B_{\text{GR}}^1 &= 117, \quad \ln B_{\text{GR}}^2 = 124, \quad \ln B_{\text{GR}}^3 = 124, \\
\ln B_{\text{GR}}^{12} &= 123, \quad \ln B_{\text{GR}}^{13} = 124, \quad \ln B_{\text{GR}}^{23} = 125, \\
\ln B_{\text{GR}}^{123} &= 114.
\end{aligned}
\tag{B.61}$$

This time the GR hypothesis is very much disfavored. However, we note that the Bayes factor for the hypothesis that only ψ_3 differs from its GR value is not the largest. In

Figure B.18 Posterior PDFs for a single GR injection with network SNR of 23.0. Top: $\delta\chi_1$ measured with a waveform that has $\{\vec{\theta}, \delta\chi_1\}$ as free parameters; bottom: $\delta\chi_2$ measured with a waveform that has $\{\vec{\theta}, \delta\chi_2\}$ free; In each case the distribution is tightly centered on zero, with standard deviations of 0.014, 0.015, and 0.019, respectively.

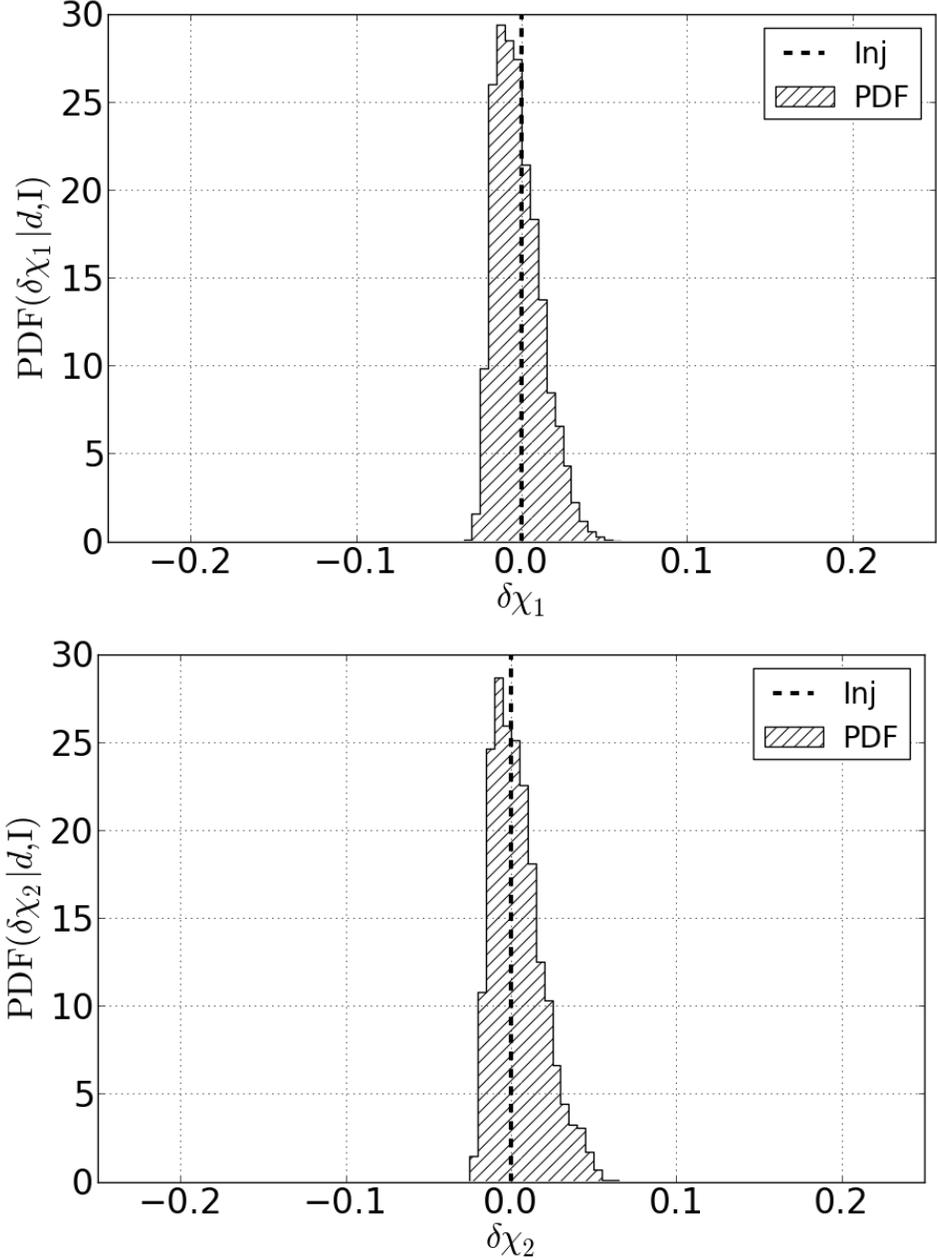
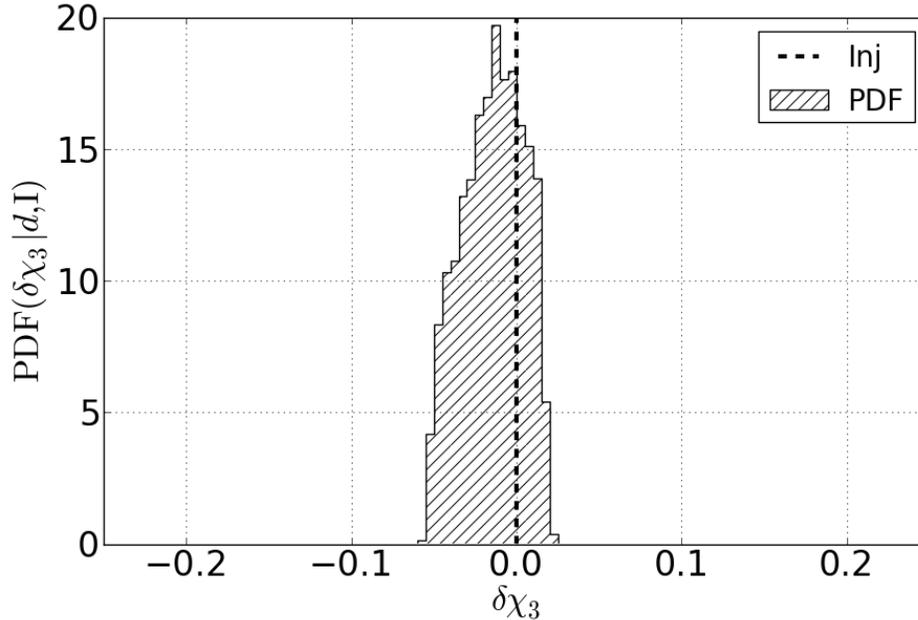


Figure B.19 Posterior PDFs for a single GR injection with network SNR of 23.0. Here $\delta\chi_3$ measured with a waveform that has $\{\vec{\theta}, \delta\chi_3\}$ free. As in Fig. the distribution is tightly centered on zero, with standard deviations of 0.014, 0.015, and 0.019, respectively.



fact, all the Bayes factors except for B_{GR}^1 and B_{GR}^{123} are rather similar in magnitude, and no clear conclusions can be drawn from them regarding the underlying nature of the deviation from GR.

When looking at the Bayes factors against noise, we see that the signal is clearly detected for all hypotheses:

$$\begin{aligned}
 \ln B_{\text{noise}}^{\text{GR}} &= 128, \\
 \ln B_{\text{noise}}^1 &= 245, \quad \ln B_{\text{noise}}^2 = 252, \quad \ln B_{\text{noise}}^3 = 252, \\
 \ln B_{\text{noise}}^{12} &= 251, \quad \ln B_{\text{noise}}^{13} = 252, \quad \ln B_{\text{noise}}^{23} = 253, \\
 \ln B_{\text{noise}}^{123} &= 242.
 \end{aligned}$$

(B.62)

Now let us consider posterior PDFs. We expect the PDF of $\delta\chi_3$ for the hypothesis H_3 , where only $\{\vec{\theta}, \delta\chi_3\}$ are allowed to vary, to be peaked at the injected value of 0.1, and this is the case with very good accuracy, as shown in the bottom panel of Fig. B.20 and B.21.

In the upper and middle panels of the same figure, the PDFs of $\delta\chi_1$ for the hypothesis H_1 , and of $\delta\chi_2$ for the hypothesis H_2 are shown. In these cases the parameter in the signal that has the shift is now not represented; in the first case only $\delta\chi_1$ is allowed to vary on top of the parameters $\vec{\theta}$ of GR, and in the second case only $\delta\chi_2$. In the nested sampling process, the waveform will still try to adapt itself to the deformation in the signal. The result is that $\delta\chi_1$ and $\delta\chi_2$ are strongly peaked, but away from the correct values $\delta\chi_1 = \delta\chi_2 = 0$. Thus, if one were to study the data only using waveforms from a specific alternative theory of gravity (e.g. a ‘massive graviton’ model with a deviation in ψ_2 only), one might find a violation of GR but draw the wrong conclusions about the nature of the deviation.

We can also look at the PDF for the hypothesis H_{123} , where the waveforms have $\delta\chi_1$, $\delta\chi_2$, $\delta\chi_3$ free; see Fig. B.22. Once again the peak is more or less at the correct value of $\delta\chi_3$, but we now have a much bigger spread. This too is as expected; parameter estimation degrades if one tries to measure too many parameters at once.

Finally, we look at the two-dimensional PDF for $\{\delta\chi_2, \delta\chi_3\}$, in the case where the waveform is the one that tests the hypothesis H_{123} ; Fig. B.23. Here too there is little to learn about the underlying nature of the deviation.

Example: A signal with $\delta\chi_3 = 0.025$

Let us consider an example with $(\mathcal{M}, \eta, D) = (1.14 M_\odot, 0.242, 216 \text{ Mpc})$, $\delta\chi_3 = 0.025$, and a network SNR of 20.6. As expected, the Bayes factors for the component hypothe-

Figure B.20 Posterior PDFs for a single injection with $\delta\chi_3 = 0.1$ and network SNR 23.2. Top: $\delta\chi_1$ measured with a waveform that has $\{\vec{\theta}, \delta\chi_1\}$ free; bottom: $\delta\chi_2$ measured with a waveform that has $\{\vec{\theta}, \delta\chi_2\}$ free. The one test parameter that is used differs from the parameter in the signal that has the shift. The parameters in the waveform will rearrange themselves such as to best accommodate the properties of the signal. Both $\delta\chi_1$ and $\delta\chi_2$ end up being sharply peaked, but not at the correct value of zero.

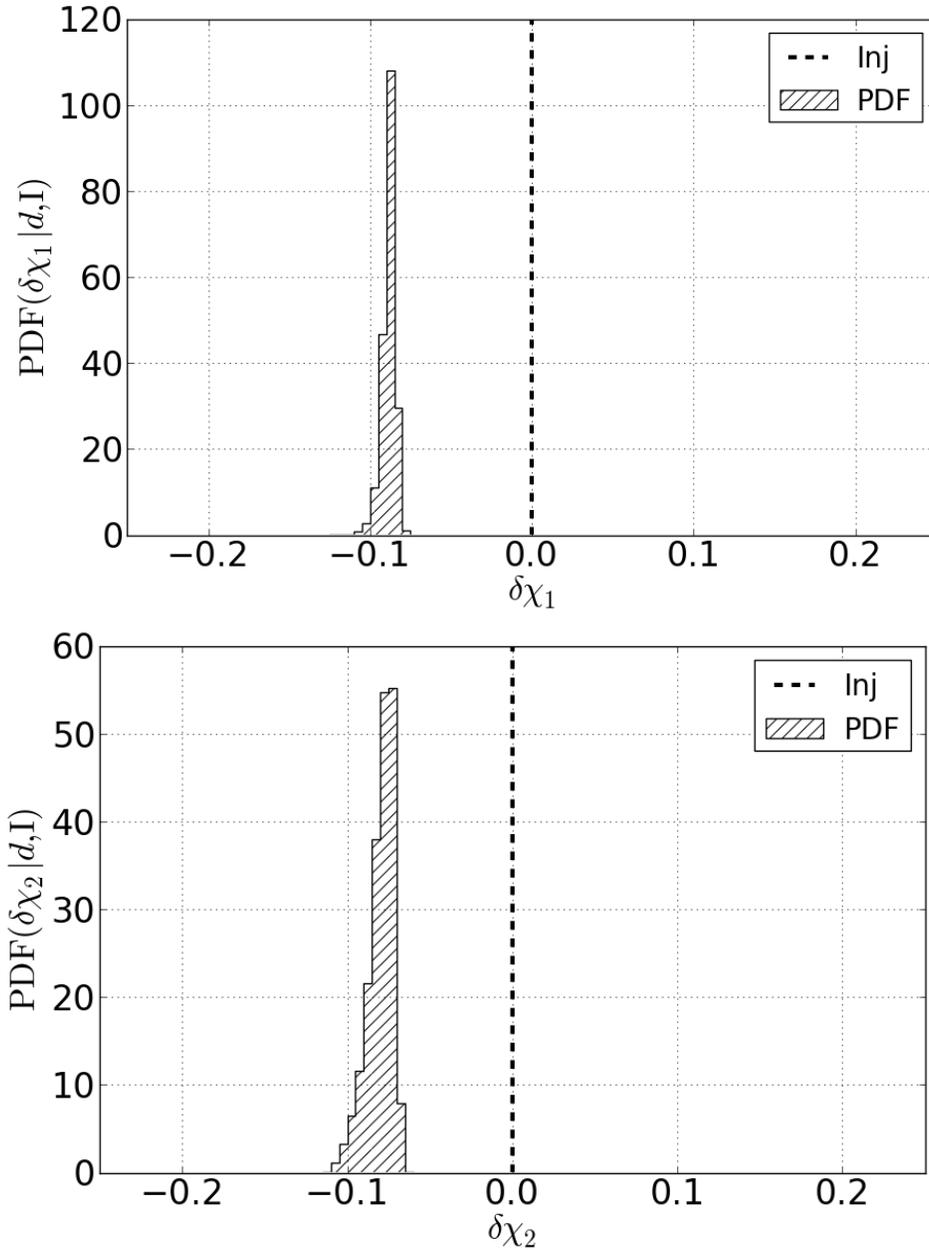


Figure B.21 Posterior PDFs for a single injection with $\delta\chi_3 = 0.1$ and network SNR 23.2. $\delta\chi_3$ measured with a waveform that has $\{\vec{\theta}, \delta\chi_3\}$ free. As expected, the PDF is sharply peaked at the correct value of $\delta\chi_3 = 0.1$, with a standard deviation of 0.012.

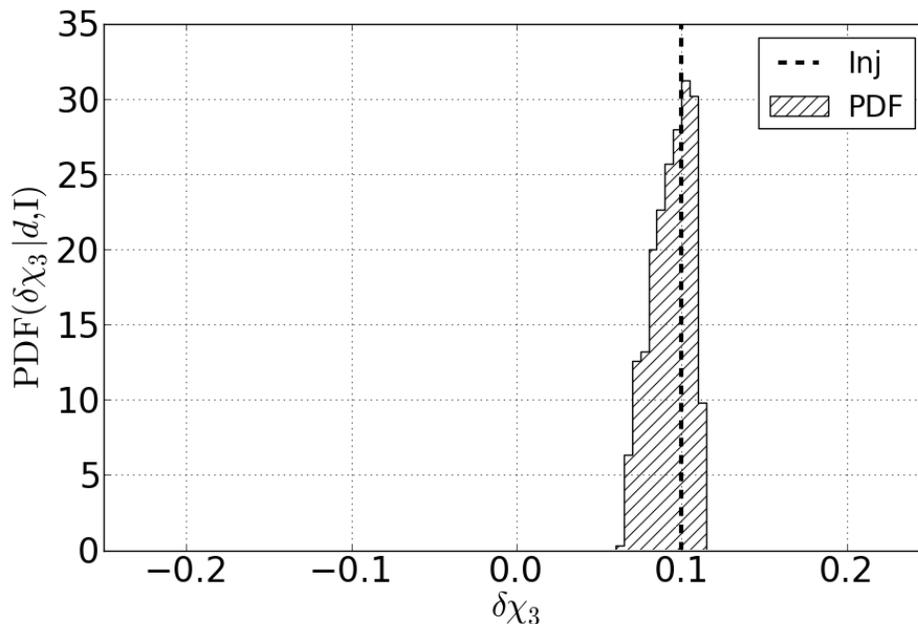
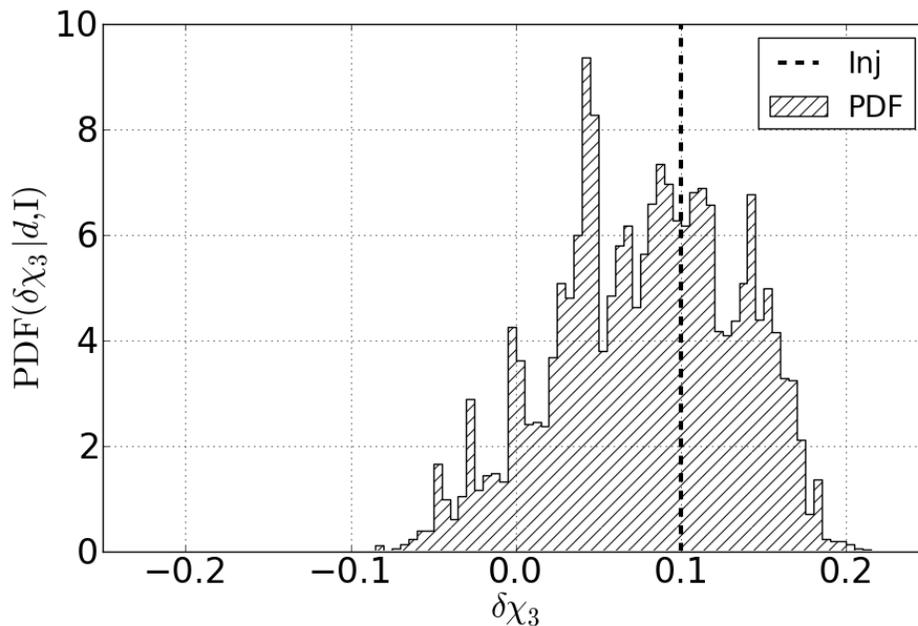


Figure B.22 The posterior PDF for the injection with $\delta\chi_3 = 0.1$ as in the previous figure, but recovered with waveforms where $\{\vec{\theta}, \delta\chi_1, \delta\chi_2, \delta\chi_3\}$ are all free. The peak is near the correct value of $\delta\chi_3$ (with a median of 0.083), but this time the spread is considerably larger (with a standard deviation of 0.055), as we are trying to measure more parameters at the same time.



Also as expected, the signal is easily found by all of the model waveforms:

$$\begin{aligned}
\ln B_{\text{noise}}^{\text{GR}} &= 186, \\
\ln B_{\text{noise}}^1 &= 197, \quad \ln B_{\text{noise}}^2 = 198, \quad \ln B_{\text{noise}}^3 = 198, \\
\ln B_{\text{noise}}^{12} &= 196, \quad \ln B_{\text{noise}}^{13} = 197, \quad \ln B_{\text{noise}}^{23} = 197, \\
\ln B_{\text{noise}}^{123} &= 197.
\end{aligned}
\tag{B.64}$$

As before we look at the posterior PDF of $\delta\chi_3$ for the hypothesis H_3 , where only $\{\vec{\theta}, \delta\chi_3\}$ are allowed to vary: see the bottom plot in Fig. B.24 and B.25. The distribution is peaked near the correct value and stays away from zero; however, one should not expect the same to happen for lower-SNR sources.

Let us also look at the PDF for $\delta\chi_1$ when $\{\vec{\theta}, \delta\chi_1\}$ are free parameters, and of $\delta\chi_2$ when $\{\vec{\theta}, \delta\chi_2\}$ are free; see the top and middle plots of Fig. B.24 and B.25. As before, $\delta\chi_1$ and $\delta\chi_2$ are not peaked at the right values of $\delta\chi_1 = \delta\chi_2 = 0$.

Example: A signal with non-PN frequency dependence in the phasing

We now look at a signal with a non-standard contribution to the phase, with a frequency dependence between 1PN and 1.5PN, as in Eq. (B.57). In the example we use here, $(\mathcal{M}, \eta, D) = (1.29 M_\odot, 0.250, 208 \text{ Mpc})$, with a network SNR of 22.4. The Bayes factors for the component hypotheses against GR are:

$$\begin{aligned}
\ln B_{\text{GR}}^1 &= 91, \quad \ln B_{\text{GR}}^2 = 93, \quad \ln B_{\text{GR}}^3 = 89, \\
\ln B_{\text{GR}}^{12} &= 92, \quad \ln B_{\text{GR}}^{13} = 91, \quad \ln B_{\text{GR}}^{23} = 92, \\
\ln B_{\text{GR}}^{123} &= 91.
\end{aligned}
\tag{B.65}$$

Figure B.24 The posterior PDFs for $\delta\chi_1$ (top) and $\delta\chi_2$ (bottom) for a single injection with $\delta\chi_3 = 0.025$ and network SNR 20.6, recovered with waveforms where, respectively, $\{\vec{\theta}, \delta\chi_1\}$ and $\{\vec{\theta}, \delta\chi_2\}$ are free. Both $\delta\chi_1$ and $\delta\chi_2$ are strongly peaked at incorrect values.

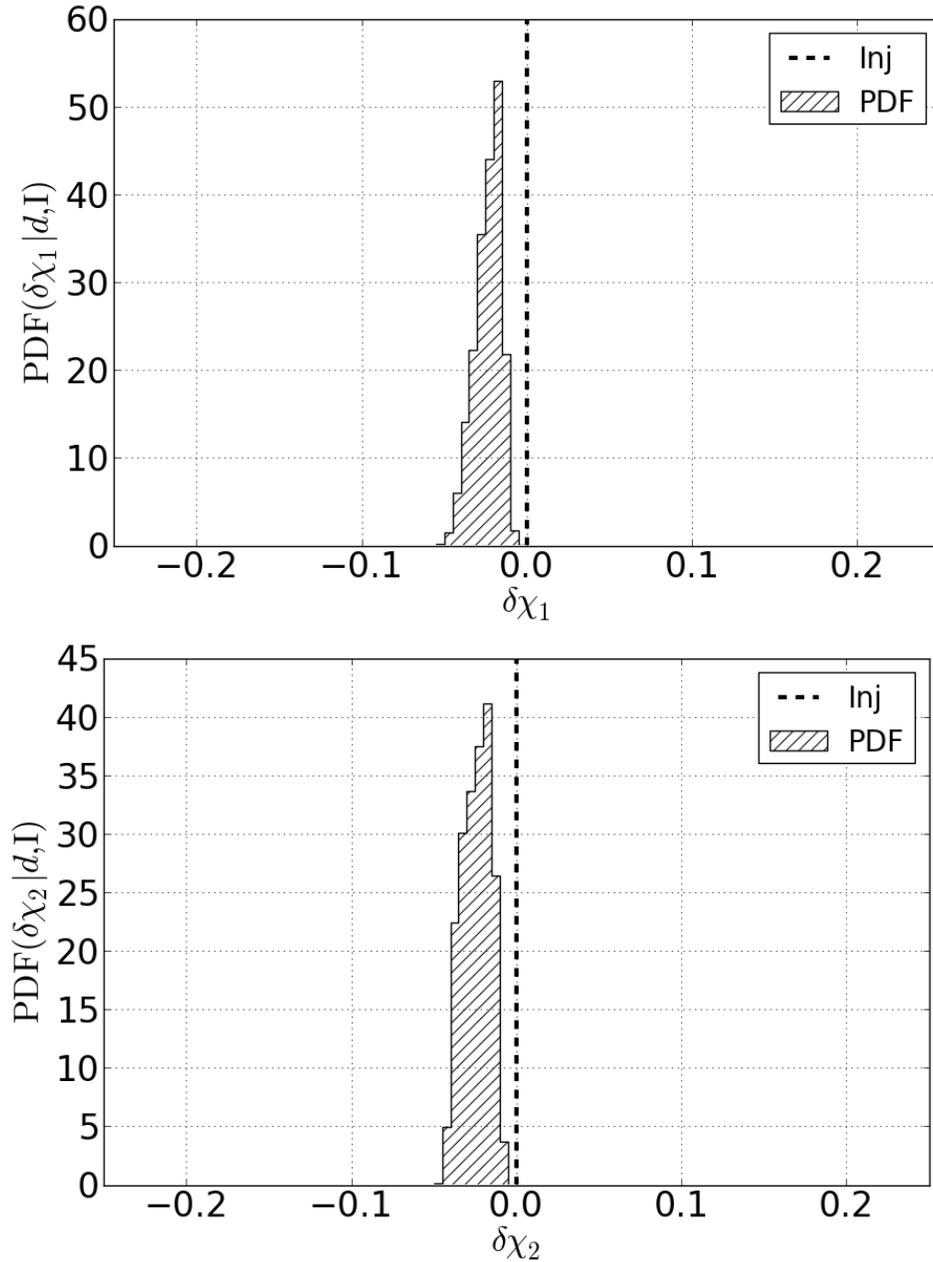
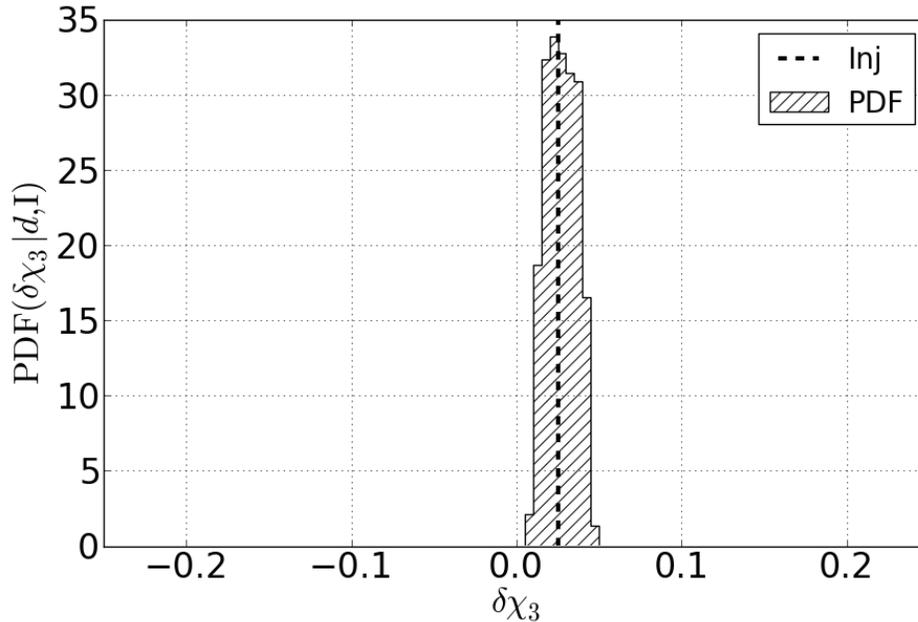


Figure B.25 The posterior PDFs for $\delta\chi_1$ (top), $\delta\chi_2$ (middle), and $\delta\chi_3$ (bottom) for a single injection with $\delta\chi_3 = 0.025$ and network SNR 20.6, recovered with waveforms where $\{\vec{\theta}, \delta\chi_3\}$ is free. Again $\delta\chi_3$ is peaked at close to the correct value, with a median of 0.027 and a standard deviation of 0.0092.



Thus, also in this case the GR hypothesis is very much disfavored, despite the fact that none of our model waveforms contain the anomalous frequency dependence which is present in the phase of the signal. We can also look at the Bayes factors against noise:

$$\begin{aligned}
 \ln B_{\text{noise}}^{\text{GR}} &= 148, \\
 \ln B_{\text{noise}}^1 &= 239, \quad \ln B_{\text{noise}}^2 = 241, \quad \ln B_{\text{noise}}^3 = 237, \\
 \ln B_{\text{noise}}^{12} &= 240, \quad \ln B_{\text{noise}}^{13} = 239, \quad \ln B_{\text{noise}}^{23} = 239, \\
 \ln B_{\text{noise}}^{123} &= 239.
 \end{aligned}$$

(B.66)

It is interesting to look at the posterior PDF of $\delta\chi_3$ for the case where $\{\vec{\theta}, \delta\chi_3\}$ are allowed to vary (Fig. B.27). The distribution looks uncannily like the analogous one

for a signal with $\delta\chi_3 = 0.1$; see the bottom panel of Fig. B.20 and B.21. We can also look at the PDF of $\delta\chi_1$ in the case where $\{\vec{\theta}, \delta\chi_1\}$ are free parameters, and the PDF of $\delta\chi_2$ when $\{\vec{\theta}, \delta\chi_2\}$ are free; see Fig. B.26. Here too there is an interesting resemblance to the analogous panels in Fig. B.20 and B.21, for an injection with $\delta\chi_3 = 0.1$.

In summary,

- Also here, the Bayes factors against GR clearly disfavor the GR hypothesis, despite the fact that none of our model waveforms has the kind of non-PN contribution to the phase that the signal contains;
- As before, the Bayes factors against GR are quite close to each other, and one cannot conclude much from them about the nature of the underlying deviation from GR;
- The Bayes factors against noise indicate that the signal will not be missed;
- The posteriors are quite similar to the ones where the deviation from GR is purely in the 1.5PN coefficient, with $\delta\chi_3 = 0.1$.

Finally, let us look at the two-dimensional PDF for $\{\delta\chi_2, \delta\chi_3\}$ in the case where the waveform is the one that tests H_{123} ; Fig. B.28. Unlike the one-dimensional PDFs, here there is not much resemblance with the two-dimensional PDF for $\delta\chi_3 = 0.1$ (Fig. B.23). Still, nothing much can be learned about the actual nature of the violation.

This shows once again that we will be able to also discern violations of GR of a kind that has no analog in the model waveforms. In the posteriors, these can even ‘masquerade’ as deviations in one of the post-Newtonian coefficients, if, for example, one would only be looking for a ‘massive graviton’ with a waveform model that has $\{\vec{\theta}, \delta\chi_2\}$ as free parameters.

Figure B.26 The posterior PDFs for $\delta\chi_1$ (top) and $\delta\chi_2$ (bottom) or a single injection with $\delta\chi_A = -2.2$ and network SNR 22.4, recovered with waveforms where, respectively, $\{\vec{\theta}, \delta\chi_1\}$ and $\{\vec{\theta}, \delta\chi_2\}$ are free. Note the remarkable resemblance with Fig. B.20, where the signal had $\delta\chi_3 = 0.1$. Also for $\delta\chi_1$ and $\delta\chi_2$, the distributions are very similar to the ones for a signal with $\delta\chi_3 = 0.1$.

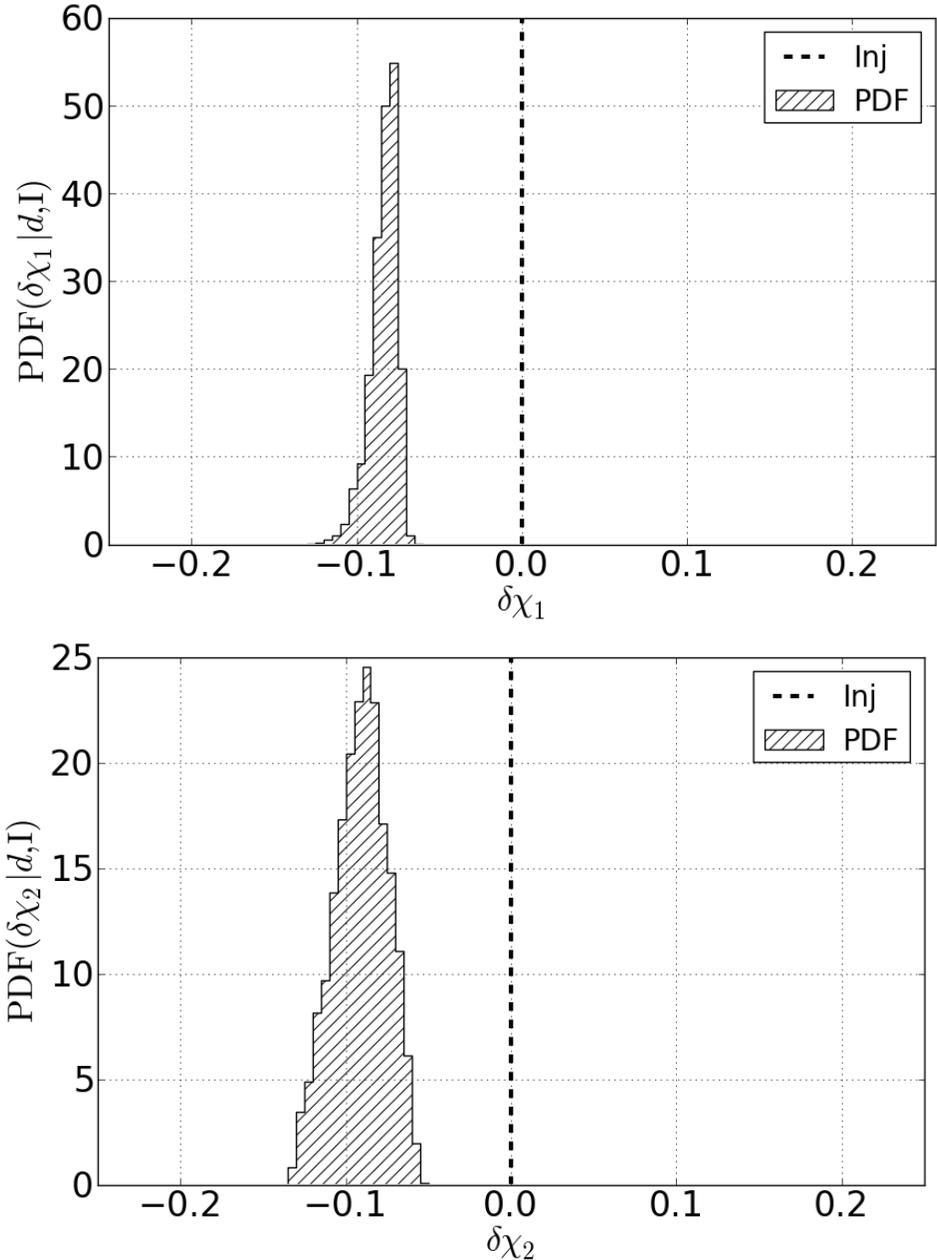


Figure B.27 The posterior PDFs for $\delta\chi_3$ (bottom) for a single injection with $\delta\chi_A = -2.2$ and network SNR 22.4, recovered with waveforms where $\{\vec{\theta}, \delta\chi_3\}$ are free. The distribution of $\delta\chi_3$ has its median at 0.11 and a standard deviation of 0.017. Note the remarkable resemblance with Fig. B.21, where the signal had $\delta\chi_3 = 0.1$.

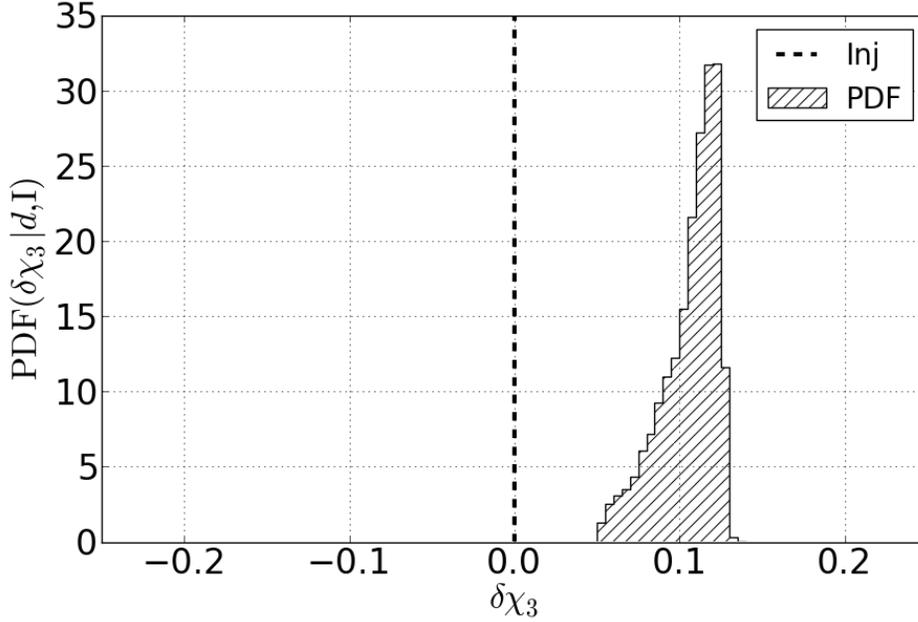
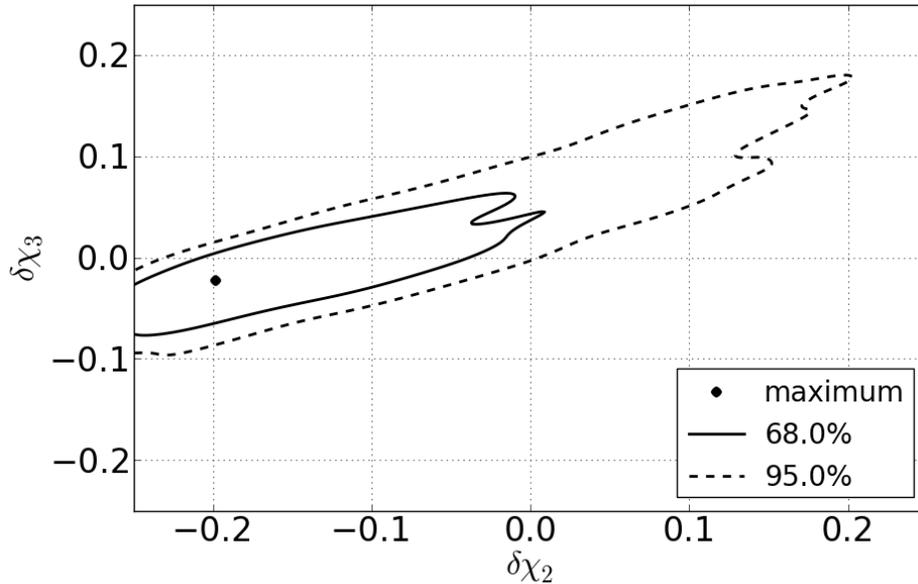


Figure B.28 The 68% and 95% confidence contours of the two-dimensional PDF for $\{\delta\chi_2, \delta\chi_3\}$, for an injection with $\delta\chi_A = -2.2$, and with a waveform that has $\{\vec{\theta}, \delta\chi_1, \delta\chi_2, \delta\chi_3\}$ as free parameters. The maximum of the PDF is given by the black dot. Here the distribution is quite different from the case with $\delta\chi_3 = 0.1$ (Fig. B.23).



B.4.5 A note on parameter estimation and multiple sources

We end this section with some cautionary remarks on the use of parameter estimation in testing GR. As we have seen in subsection B.3.4, model selection allows for combining the information from multiple sources to compute a single odds ratio, but for parameter estimation the situation is quite different.

In the examples presented in this paper, the deviations $\delta\chi_3$ and $\delta\chi_A$ were taken to be the same for all sources. Hence, in principle we could have combined PDFs from multiple sources, simply by multiplying them, to arrive at more accurate measurements. However, we reiterate that in reality, one cannot expect the deviations to be constant in this fashion; rather, they might vary from source to source, with dependence on the masses as well as whatever additional charges may be present in the correct theory of gravity. If it so happens that deviations in individual sources can go both ways, making positive or negative contributions to the overall phase depending on the parameters of the source, then a combined PDF may not show any significant deviation at all. Moreover, since most sources will have SNRs near threshold, it could be that only a few sources will allow for accurate parameter estimation; here we only showed PDFs for single, relatively ‘loud’ sources with SNR greater than ~ 20 . However, if there is significant dependence of the GR deviation on source parameters (unlike in the heuristic examples shown here), the particular deviation exhibited by the loudest source may not be representative. Hence, in contrast to model selection, parameter estimation alone does not provide a solid foundation to look for deviations from GR.

B.5 Conclusions and future directions

As we showed at the beginning, it is possible for a signal to contain very significant deviations from General Relativity while still being detectable with a template family of GR waveforms; this is the ‘fundamental bias’ discussed in [63, 216]. A violation of GR can in principle take any form, and the question which then presents itself is how to search for generic deviations.

We have developed a general method to search for deviations from General Relativity using signals from compact binary coalescence events. To this end we constructed an odds ratio $O_{\text{GR}}^{\text{modGR}}$ for modifications to GR against GR, which is the posterior probability that there is a deviation from GR, versus GR being correct. This odds ratio can be written as a linear combination of Bayes factors $B_{\text{GR}}^{i_1 i_2 \dots i_k}$ for hypotheses $H_{i_1 i_2 \dots i_k}$, in each of which one or more of the phase parameters ψ_i is assumed to deviate from the GR value, without actually assuming any specific dependence on the frequency and/or physical parameters pertinent to a given theory. Since this includes hypotheses where only a single one of the ψ_i is non-GR, our method will be particularly well-suited in low-SNR scenarios, which we expect to be in with the upcoming advanced detector network. Finally, information from multiple sources can easily be combined to arrive at an odds ratio $\mathcal{O}_{\text{GR}}^{\text{modGR}}$ for the ‘catalog’ of all observed events.

The method we developed, applied to phase coefficients in inspiral waveforms, essentially addresses the question “Do one or more of the coefficients differ from the values predicted by GR?” This is in contrast with previous Bayesian analyses such as [55, 63, 83], where effectively, the question being asked was limited to “Do all of the additional free parameters introduced differ from their GR values?” In addition to being

better adapted to a low-SNR environment, the test proposed here is far more general.

In order to gauge how large a deviation might be detectable, we first considered signals with a constant fractional deviation in the 1.5PN coefficient ψ_3 . This coefficient is of particular interest, since it incorporates the so-called ‘tail effects’ [47, 48] (as well as spin-orbit coupling [45], although we did not consider spin here), which are not empirically accessible with binary pulsar observations and can only be studied through direct detection of gravitational waves. When considering catalogs of only 15 binary neutron star sources, we saw that a deviation in ψ_3 at the 10% level would easily be detectable. In fact, even a deviation at the few percent level can be discernable. This is confirmed by posterior PDFs for ψ_3 in the case where this is the only parameter that is assumed to deviate from its GR value.

We also considered a deviation in the phase with a frequency dependence that does not match any of the post-Newtonian terms, and hence is not present in any of the recovery waveforms that we used. More precisely, we looked at signals whose phase has an additional contribution, with a frequency dependence in between that of the 1PN and 1.5PN terms (‘1.25PN’). The magnitude of the deviation was chosen such that near $f \sim 150$ Hz, where the detectors are the most sensitive, the change in phase is roughly the same as the change caused by a 10% shift in the 1.5PN coefficient. The deviation was clearly detectable in the log odds ratios, the Bayes factors, and the posterior PDFs. We expect this to be an instance of a more general fact. Namely, even if there is a deviation in the phase which the model waveforms technically do not allow for, it will typically be observable, on condition that near the ‘bucket’ it causes a change in the phase that is on a par with the effect of a shift in the (low order) PN phase coefficients of more than a few percent.

In order to establish the basic validity and usefulness of our method for testing GR, we

considered constant fractional deviations in ψ_3 , and a non-PN frequency dependence in the signal. However, even if there is a deviation in one or more of the PN phase contributions only, these may depend on (\mathcal{M}, η) as well as whatever additional charges and coupling constant might be present. It would be of great interest to study the effects of more general deviations from GR on the odds ratio $\mathcal{O}_{\text{GR}}^{\text{modGR}}$, the cumulative Bayes factors $\prod_A {}^{(A)}B_{\text{GR}}^{i_1 i_2 \dots i_k}$ for the component hypotheses $H_{i_1 i_2 \dots i_k}$ against \mathcal{H}_{GR} , and the cumulative Bayes factors $\prod_A {}^{(A)}B_{\text{noise}}^{i_1 i_2 \dots i_k}$ against noise. Additionally, in future one should consider deviations in the amplitude as well, and use model waveforms which have such freedom. A priori there is no reason not to use an arbitrary number of free coefficients in both the phase and the amplitude of recovery waveforms, the only limiting factor being computational time.

Should no deviation from GR be convincingly found through model selection, one will still be interested in constraining the theory. Advanced gravitational wave detectors will then take us considerably beyond the binary pulsar observations. We recall that even the 1PN phase coefficient ψ_2 is not fully constrained by the latter, since the dissipative dynamics is only probed to leading PN order. As we have seen from PDFs, with a single compact binary coalescence event in Advanced LIGO/Virgo at SNR ~ 20 , the coefficient ψ_2 can be constrained to better than 2%, and the same is true of the 1.5PN coefficient ψ_3 . Mainly for computational reasons, we did not study constraints on the 2PN and higher-order terms, but it would clearly be of interest to see how well one can pin down the corresponding coefficients. We also draw attention to the results for the 0.5PN contribution, which in General Relativity is identically zero. For simplicity we restricted the mass range of our simulated sources to that of binary neutron stars. Given the very encouraging results, in future work the BHNS and BBH ranges should also be investigated.

The preliminary results presented here motivate the construction of a full data analysis pipeline for testing or constraining General Relativity. To have a real chance of finding a deviation, much more sophisticated GR waveforms will have to be used, with inclusion of merger and ringdown, higher harmonics both in the inspiral and ringdown parts, dynamical spins, and residual eccentricity. The development of such waveforms, with input from numerical simulations, is currently a subject of intense investigations [18, 20, 35, 50–52, 57, 59, 149, 174, 175]. We note that the method we presented is not tied to any particular waveform approximant. Moreover, the deformations need not be in the phase. Indeed, in the future one would presumably want to use time domain waveforms, for which it may be more convenient (and physically more appropriate) to introduce parameterized deformations directly in coefficients appearing in, e.g., a Hamiltonian used to evolve the inspiral part of the waveform. Irrespective of the parameterization, one would still be able to associate with it an exhaustive set of logically disjoint hypotheses $H_{i_1 i_2 \dots i_k}$.

Once sufficiently accurate waveforms are available, a test of GR on Advanced LIGO/Virgo could go as follows:

- Starting from the best available GR waveforms, introduce parameterized deformations, leading to disjoint hypotheses like our $H_{i_1 i_2 \dots i_k}$, which together form $\mathcal{H}_{\text{modGR}}$;
- Use many injections of GR waveforms in real or realistic data, arranged into simulated ‘catalogs’, to investigate the distributions of the cumulative odds ratio $\mathcal{O}_{\text{GR}}^{\text{modGR}}$ as well as of the cumulative Bayes factors when GR is correct. Use these to set thresholds for the measured odds ratio and Bayes factors to overcome;
- Apply our method to the catalog of sources actually found by the detectors. If the

measured cumulative odds ratio $\mathcal{O}_{\text{GR}}^{\text{modGR}}$ is below threshold, then there is no real reason to believe that a deviation from GR is present. The posterior PDFs for the free phase and amplitude coefficients in the model waveforms, taken from the highest-SNR sources, will provide (potentially very strong) constraints on these parameters;

- If the measured odds ratio is above threshold, then a violation of GR is likely. As we have seen, Bayes factors and PDFs can be misleading in trying to find out what the precise nature of the deviation may be. However, one may be able to follow up on the violation by again using our method, this time with waveforms with more complicated deformations and a larger number of free parameters, inspired by particular alternative theories of gravity, similar to what is done in ppE [216].

Thus, although much work remains to be done, we have the basics of a very general method for testing General Relativity using compact binary coalescence events to be detected by the upcoming Advanced LIGO and Advanced Virgo observatories.

Appendix C

TESTING GENERAL RELATIVITY WITH COMPACT COALESCING BINARIES: COMPARING EXACT AND PREDICTIVE METHODS TO COMPUTE THE BAYES FACTOR

The work in this Appendix is presented in the form of the paper Published in General Relativity and Quantum Cosmology [62]

Authorship of paper: W. Del Pozzo, K. Grover, I. Mandel, A. Vecchio. K. Grover did not write the text but was involved in all discussions and work.

The second generation of gravitational-wave detectors is scheduled to start operations in 2015. Gravitational-wave signatures of compact binary coalescences could be used to accurately test the strong-field dynamical predictions of general relativity. Computationally expensive data analysis pipelines, including TIGER, have been developed to carry out such tests. As a means to cheaply assess whether a particular deviation from general relativity can be detected, Cornish et al. [55] and Vallisneri [181] recently proposed an approximate scheme to compute the Bayes factor between a general-relativity gravitational-wave model and a model representing a class of alternative theories of gravity parametrised by one additional parameter. This approximate scheme is based on only two easy-to-compute quantities: the signal-to-noise ratio of the signal and the fitting factor between the signal and the manifold of possible waveforms within general relativity.

In this work, we compare the prediction from the approximate formula against an exact numerical calculation of the Bayes factor using the `lalinferenc` library. We find that, using frequency-domain waveforms, the approximate scheme predicts exact results with good accuracy, providing the correct scaling with the signal-to-noise ratio at a fitting factor value of 0.992 and the correct scaling with the fitting factor at a signal-to-noise ratio of 20, down to a fitting factor of ~ 0.9 . We extend the framework for the approximate calculation of the Bayes factor which significantly increases its range of validity, at least to fitting factors of ~ 0.7 or higher.

C.1 Introduction

The upgraded versions of the ground-based gravitational wave detectors LIGO [11, 87] and Virgo [2, 12–14] are expected to detect gravitational-wave signals from the coa-

lescence of compact binary systems. The prospect of frequent detections, with expected rates between one per few years and a few hundred per year [4] promises to yield a variety of scientific discoveries. Among these, the possibility of testing the strong field dynamics of general relativity (GR) has received increasing attention (e.g., [16, 55, 63, 112, 114, 148]). In fact, during the latest phase of the inspiral, typical orbital velocities are an appreciable fraction of the speed of light ($v/c \sim 0.4$); following merger, the *compactness* $GM/(Rc^2)$ of the newly formed black hole that is undergoing quasinormal ringing is close to 1. By comparison, the orbital velocity of the double pulsar J0737-3039 is $O(10^{-3}c)$ and its compactness is $\sim 10^{-6}$ [54]. Consequently, efforts have concentrated on the development of robust frameworks to reliably detect deviations from GR using gravitational-wave signatures of compact-binary mergers.

One of these frameworks is the so-called *Test Infrastructure for General Relativity* (TIGER) [16, 112, 113]. TIGER operates by computing the odds ratio between GR and a test model in which one or more of the post-Newtonian coefficients are allowed to deviate from the value predicted by GR. The interested reader is referred to Refs. [16, 112, 113] for the details of the method and for analysis of its robustness against various potential systematic effects. To account for unmodelled effects, TIGER constructs a “background” distribution of odds ratios between GR and the test hypothesis by analysing $O(10^3)$ simulated GR signals. The background distribution defines the null hypothesis against which any particular observation (or catalog of observations) is tested. For validation purposes, the sensitivity of the algorithm to a specific deviation from GR is currently assessed by comparing it with a “foreground” odds ratio distribution. The foreground distribution is constructed by simulating a variety of signals in which the chosen deviation from GR is introduced. If the *integrated* overlap between the foreground and background distributions is smaller than a given false alarm probability, sensitivity to that particular deviation can be claimed.

The process described in the previous paragraph is extremely computationally expensive. If arbitrary combinations of k post-Newtonian coefficients are allowed to deviate from GR values, the total number of simulations that necessary to construct the background is 2^k for each synthetic source.

As a means to cheaply evaluate the detectability of particular deviations from GR, Cornish et al. [55] proposed an approximate formula to calculate the odds ratio between GR and an alternative model for gravity (AG). Subsequently, Vallisneri [181] proposed a similar approximation derived from the Fisher matrix formalism. Vallisneri’s approximation considers the distribution of the odds ratio in the presence of noise and characterises the efficiency and false alarm of a Bayesian detection scheme for alternative theories of gravity. Whilst neither of these approaches can replace the necessary analysis for real data, the possibility of having a quick and easily understandable formalism to check the performance of complex pipelines such as TIGER and assess whether a specific type of deviation is detectable without having to run thousands of simulations seems quite attractive.

In this work we investigate, in an idealised and controlled scenario, whether the predictions from Refs. [55, 181] are in agreement with the output of a numerical Bayesian odds-ratio calculation. We find in particular that the the analytical prescription of Ref. [181] is in reasonable agreement with the numerical result when the fitting factor (FF) between AG and GR waveforms is ≥ 0.9 , and that for $\text{FF} \leq 0.8$, both analytical prescriptions overestimate the exact odds ratio.

Nevertheless, when the analytical odds ratio is regarded as an upper limit, useful indications of the detectability of a given deviation from GR can be drawn.

We analytically correct the approximate framework for computing the Bayes factor by

introducing terms that are negligible at $FF \sim 1$, reproducing the proposed analytical expressions given in [55, 181], but become significant at lower values of the fitting factor. We show that these corrections extend the range of validity of the approximate expressions at least down to fitting factor values of ~ 0.7 .

The rest of the paper is organised as follows: in section C.2 we briefly review the Bayesian definition of the odds ratio; in section C.2.1 we introduce the formula from Ref. [181]. In section C.3 we present our findings and finally we discuss them in section C.4.

C.2 Bayesian Inference for gravitational wave signals

In a Bayesian context, the relative probability of two or more alternative hypotheses given observed data d is described by the odds ratio (see, e.g., [63]). If GR is the general relativity hypothesis and AG is the hypothesis corresponding to some alternative theory of gravity, the odds ratio is given by:

$$O_{\text{AG,GR}} = \frac{p(\text{AG}|d)}{p(\text{GR}|d)} = \frac{p(\text{AG}) p(d|\text{AG})}{p(\text{GR}) p(d|\text{GR})} \equiv \frac{p(\text{AG})}{p(\text{GR})} B_{\text{AG,GR}} \quad (\text{C.1})$$

where we introduced the Bayes factor $B_{\text{AG,GR}}$, which is the ratio of the marginalised likelihoods (or evidences). The marginal likelihood is the expectation value of the likelihood of observing the data given the specific model H under consideration over of the prior probability distribution for all the model parameters θ :

$$p(d|H) \equiv Z = \int d\theta p(d|\theta, H) p(\theta|H). \quad (\text{C.2})$$

With the exception of a few idealised cases, the integral (C.2) is, in general, not tractable analytically. In gravitational-wave data analysis, the parameter space is at least 9-dimensional (for binaries with components that are assumed to have zero spin), and up to 15-dimensional for binaries with arbitrary component spins, and the integrand is a complex function of the data and the waveform model. For stationary Gaussian noise

$$p(d|\theta, H) \propto \exp[-(d - h(\theta)|d - h(\theta))/2], \quad (\text{C.3})$$

where $h(\theta) \equiv h(\theta|H)$ is the model waveform given parameters θ and we introduced the scalar product

$$(a|b) \equiv 2 \int_0^\infty df \frac{a(f)b(f)^* + a(f)^*b(f)}{S(f)} \quad (\text{C.4})$$

with the one-sided noise power spectral density $S(f)$. We analysed data from a single detector with a noise spectral density corresponding to the zero-detuning, high-power Advanced LIGO design configuration [163].

C.2.1 Analytical Approximation

Vallisneri [181] proposed an analytical approximation to the integral (C.2). He considered the following assumptions:

- linear signal approximation leading to a quadratic approximation of the log likelihood;
- only one additional dimension is necessary to describe the AG model;
- uniform prior distributions for all parameters describing both GR and AG models;
- the distance between the AG waveform and the manifold of GR waveforms is

small so that the fitting factor (FF) between the two, defined as

$$FF = \left[\frac{(h_{AG}|h_{GR}(\theta))}{\sqrt{(h_{AG}|h_{AG})(h_{GR}(\theta)|h_{GR}(\theta))}} \right]_{\max \text{ over } \theta}, \quad (\text{C.5})$$

is close to unity.

With the above assumptions, the integral Eq. (C.2) can be approximately computed analytically and the Bayes factor (C.1) is then given by:

$$B_{AG,GR} \approx \sqrt{2\pi} \frac{\Delta\theta_{est}^a}{\Delta\theta_{prior}^a} e^{\rho^2(1-FF)}, \quad (\text{C.6})$$

where ρ denotes the optimal signal-to-noise ratio (SNR)¹:

$$\rho \equiv 2\sqrt{\int_0^\infty df \frac{|h(\theta_{true})|^2}{S(f)}}. \quad (\text{C.7})$$

The terms $\Delta\theta_{prior}^a$ and $\Delta\theta_{est}^a$ are the width of the prior distribution and of the Fisher matrix 1- σ uncertainty estimate for the additional AG parameter, respectively. Eq. (C.6) given here is valid for the case in which a zero realisation of the noise is present in the data. In Ref. [181] noise is considered and the appropriate formulae for the distribution of the Bayes factor over noise realisations can be found there. We opted for a zero-noise case for ease of comparison.

¹Note that the definition by Vallisneri of the signal-to-noise ratio is different from ours. In Ref. [181] the signal-to-noise ratio quantity that appears in Eq. (C.6) is the norm of an hypothetical GR signal whose parameters are exactly the same as the “true” AG waveform, but with the extra AG parameter set to zero. In our case, the signal-to-noise ratio corresponds to the power in the AG signal (in the ideal case when it is filtered with AG templates). However, when the AG parameter is present only in the phase of the gravitational wave these two signal-to-noise ratios coincide.

C.3 Comparison between the exact calculation and the analytical approximation

We compare the prediction from Eq. (C.6) with the evidence calculated by the Nested Sampling algorithm [167] as implemented in `lalInference` [190] in a simple experiment. Using the test waveform model presented in [112], we generate inspiral signals which would span a range of FFs. The testing waveform is a frequency-domain stationary phase approximation waveform, based on the TaylorF2 approximant [53], that has been modified in such way that the post-Newtonian coefficients are allowed to vary around the GR values within a given range. The TaylorF2 waveform for a face-on, overhead binary is given by:

$$h(f) = \frac{1}{D} \sqrt{\frac{5}{24}} \pi^{-2/3} \mathcal{M}^{5/6} f^{-7/6} e^{i\Psi(f)}, \quad (\text{C.8})$$

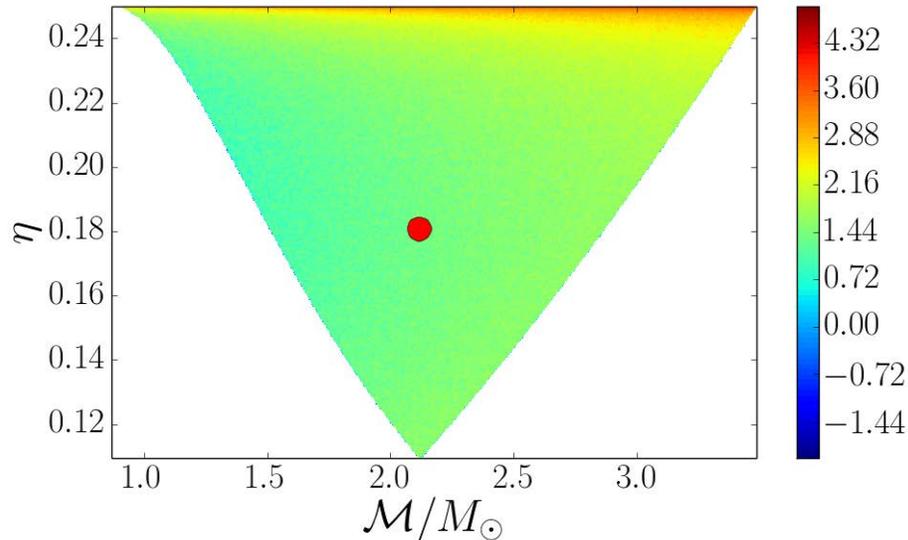
where D is the luminosity distance, \mathcal{M} is the chirp mass and the phase $\Psi(f)$ is:

$$\begin{aligned} \Psi(f) &= 2\pi f t_c - \phi_c - \pi/4 \\ &+ \sum_{i=0}^7 \left[\psi_i + \psi_i^{(l)} \ln f \right] f^{(i-5)/3}. \end{aligned} \quad (\text{C.9})$$

The explicit forms of the coefficients ψ_i and $\psi_i^{(l)}$ in (\mathcal{M}, η) , where η is the symmetric mass ratio, can be found in [131]. In all our experiments we kept the parameters of the simulated sources fixed with the exception of the 1.5 post-Newtonian coefficient ψ_3 which we varied between $[0.5, 1.5]$ times its GR value by adding an arbitrary shift $d\chi_3$ between $[-0.5, 0.5]$:

$$\psi_3 \rightarrow \psi_3(1 + d\chi_3). \quad (\text{C.10})$$

Figure C.1 Two-dimensional logarithmic prior probability distribution in the \mathcal{M} and η space. The shape of the region is set by the allowed ranges of individual and total masses. The red marker indicates the location of the $1.4M_{\odot} + 4.5M_{\odot}$ system simulated for the analysis.



The Nested Sampling algorithm was set up to sample from the following prior:

- the component masses were allowed to vary uniformly $\in [1, 7]M_{\odot}$ with the total mass constrained to the range $\in [2, 8]M_{\odot}$. This choice results in an allowed region of triangular shape in the \mathcal{M}, η plane, see Fig. C.1;
- uniform on the 2-sphere for sky position and orientation parameters;
- uniform in Euclidean volume for the luminosity distance;
- for recovery with AG templates, we used only one free testing parameter ($d\chi_3$) which was allowed to vary uniformly between $[-0.5, 0.5]$ times its GR value.

The FFs were computed from the maximum likelihood values obtained from the `lal inference` simulations, see Appendix A. The parameter uncertainty for Eq. (C.6) was computed using a 5-dimensional Fisher matrix calculation in which we varied the two mass pa-

rameters, the time of coalescence, the phase at coalescence and the deviation parameter $d\chi_3$.

Our experiments were performed analysing simulated signals from a system whose component masses were chosen to be $1.4M_\odot + 4.5M_\odot$. We chose this system because it lies in the centre of our prior probability distribution over the masses, far away from prior boundaries. This minimises the impact of the prior on the fitting factor and Bayes factor computations, which ensures that we can make a fair comparison with Eq. (C.6) derived under the assumption of a uniform prior.²

The approximate formula in Eq. (C.6) depends essentially on two quantities: the signal-to-noise ratio ρ and the FF. Below, we describe our investigations of the dependence of the Bayes factor on these two quantities.

For Nested Sampling calculations, there is a Poisson uncertainty on the number of iterations taken to find the high posterior region [193],

$$n_{tot} = nH \pm \sqrt{nH} . \quad (\text{C.11})$$

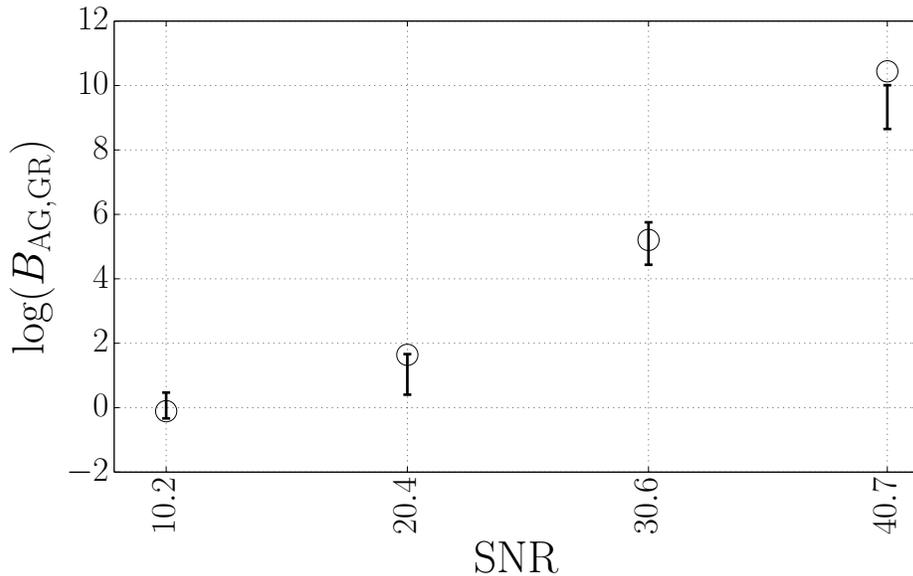
where H is the Kullback–Leibler divergence (or relative entropy) between the posterior distribution and the prior distribution and n is the number of live points used for Nested Sampling. The uncertainty on the calculated value of the evidence Z is evaluated as [167]:

$$\Delta \log Z \simeq \sqrt{\frac{H}{n}} \quad (\text{C.12})$$

H is computed by the Nested Sampling along side the evidence Z . Typical values for

²For example, an equal mass system would lie exactly on the prior boundary at $\eta = 0.25$. For this reason, the GR model has very little room in the η direction to accommodate the additional phase shift due to a non-zero $d\chi_3$. The net result is a very rapid drop in FF towards negative $d\chi_3$.

Figure C.2 Logarithmic Bayes factor from `lalInference` (errorbars) and from Eq. (C.6) (empty circles) as a function of the SNR, for FF fixed to 0.992. The `lalInference` errors are computed from Eq. (C.12). The circles are the values of the logarithmic Bayes factor obtained from Eq. (C.6) using the FF extracted from the maximum likelihood values as computed by `lalInference`.



$\Delta \log Z$ are $O(10^{-1})$.

Finally, it was recently pointed out that when a signal terminates abruptly in the detector band, measurement uncertainty may be significantly smaller than predicted by the Fisher matrix calculation [124]. To avoid these complications, we limited our analysis to frequencies between 30 and 512 Hz.

C.3.1 Scaling with the signal-to-noise ratio

We investigated the dependence of the Bayes factor on SNR by comparing the output of `lalInference` and Eq. (C.6) for a $1.4M_{\odot} + 4.5M_{\odot}$ system at SNRs of 10, 20, 30 and 40 at a fixed value of the FF, 0.992. Fig. C.2 shows the Bayes factors from the two calculations.

The quadratic dependence of the Bayes factor on the SNR was verified by means of a simple chi-squared fit to an expression of the form

$$\ln B_{\text{AG,GR}} = \alpha \text{SNR}^\beta + \gamma. \quad (\text{C.13})$$

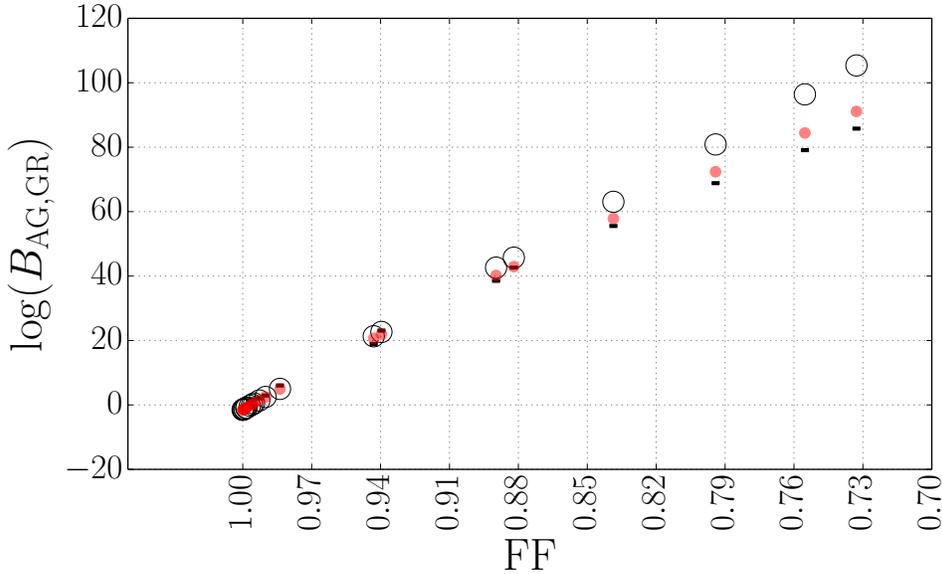
The scaling of $\log(B_{\text{AG,GR}})$ with the SNR of the appears to be consistent with the expected quadratic dependence: we find $\beta = 1.95 \pm 0.4$.

C.3.2 Scaling with the fitting factor

We evaluate the dependence of the Bayes factor on FF by again injecting a signal from a $1.4M_\odot + 4.5M_\odot$ binary, now at a fixed SNR of 20 but with varying FF. As in the previous section, we vary the FF by adding arbitrary deviations from the GR value to the 1.5 post-Newtonian phase coefficient. In particular, we varied $d\chi_3$ between -0.5 and 0.5 , leading to $\text{FF} \in [0.7, 1.0]$. We verified that our injection was sufficiently far from prior boundaries by confirming that the Bayes factor is the same for positive and negative values of $d\chi_3$ that yield the same FF.

Fig. C.3 shows the logarithmic Bayes factor computed by `lal inference` and from Eq. (C.6). The two methods agree for $\text{FF} \sim 1$. At $\text{FF} \leq 0.9$, the analytical approximation overestimates the value of the Bayes factor compared to `lal inference`. Moreover, the disagreement gets worse with decreasing FF, suggesting a nonlinear dependence on the FF. In the next section, we investigate the approximate analytical expression in greater detail and derive additional corrections that extend its validity to lower fitting factors.

Figure C.3 Logarithmic Bayes factor from `lalinference` (error bars) and from Eq. (C.6) (empty circles) as a function of the FF. The SNR was fixed to 20. The Bayes factors computed with the two approaches agree for $FF \sim 1$ but tend to diverge for decreasing FF. The red dots indicate the value of the logarithmic Bayes factor obtained by using a quadratic dependence on the FF, Eq. (C.17), rather than the linear dependence of Eq. (C.6).



C.3.3 Correcting the analytical expression for lower fitting factors

Under the assumption that the region of likelihood support on the parameter space is small, and that over this region the prior does not vary significantly, the evidence for any of the models H_i , depending on parameters θ , under consideration can be approximated as (e.g., [189], correcting for a typo in the exponent of (2π)):

$$Z(H_i) \propto [L_{H_i}]_{\max \text{ over } \theta} (2\pi)^{N/2} \prod_i^N \frac{\Delta\theta_{est}^i}{\Delta\theta_{prior}^i}, \quad (\text{C.14})$$

where N is the number of parameters. Strictly speaking, the equation above is only valid when parameters are uncorrelated. In the general case of correlated parameters, $\prod_i^N \Delta\theta_{est}^i$ should be replaced with the uncertainty volume in which the likelihood has support, while $\prod_i^N \Delta\theta_{prior}^i$ is shorthand for the total prior volume. However, such correlations do not affect the scaling of the uncertainty with the SNR, and do not impact our conclusions.

Therefore, the Bayes factor between the AG and GR model can be approximated as the ratio of the maximum likelihoods times the product of the ratios of posterior widths to prior supports:

$$B_{AG,GR} \approx \frac{[L_{AG}]_{\max \text{ over } \theta'}}{[L_{GR}]_{\max \text{ over } \theta}} \sqrt{2\pi} \frac{\prod_{i=0}^N \frac{\Delta\theta'_{est}}{\Delta\theta'_{prior}}}{\prod_{i=0}^{N-1} \frac{\Delta\theta_{est}}{\Delta\theta_{prior}}} \quad (\text{C.15})$$

where θ' and θ are parameter vectors within the AG and GR models, respectively, and N is the dimensionality of the AG parameter space.

We begin by considering just the first term in Eq. (C.15), which scales exponentially with the SNR in contrast to the components of the second term, which scale inversely

with the SNR. Neglecting the second term, we find:

$$\log(B_{AG,GR}) \propto \log([L_{AG}]_{\max \text{ over } \theta'}) - \log([L_{GR}]_{\max \text{ over } \theta}). \quad (\text{C.16})$$

Using Eq. (C.31), we find:

$$\log(B_{AG,GR}) \propto \frac{\rho^2}{2}(1 - FF^2). \quad (\text{C.17})$$

which is the expression originally proposed in [55]. At FF close to unity, $(1 - FF^2) = (1+FF)(1-FF) \approx 2(1-FF)$, the approximation implicitly made in [181], and we recover Eq. (C.6). However, we expect (C.17) to lead to a better fit at low fitting factors. The filled (red) dots in Fig. C.3 show the Bayes factor computed via Eq. (C.17), with the proportionality constant fixed to be the same as in Eq. (C.6). Indeed, Eq. (C.17) predicts Bayes factors that are in closer agreement with the exact ones than Eq. (C.6). In this case, disagreements with the exact result can be seen for $FF \sim 0.75$, when the differences in the local shapes of the GR and AG manifolds can become significant.

Vallisneri [181] further assumed that the priors and measurement uncertainties on all parameters except the one describing the deviation from GR, θ^a , are the same for the AG and GR models (which, in turn, is a statement about the similarity in the shape of the two waveform manifolds near the maximum likelihood locations). In this case, the Bayes factor between the two models is [cf. (C.6)]:

$$B_{AG,GR} \propto \frac{[L_{AG}]_{\max \text{ over } \theta'}}{[L_{GR}]_{\max \text{ over } \theta}} \sqrt{2\pi} \frac{\Delta\theta_{est}^a}{\Delta\theta_{prior}^a}, \quad (\text{C.18})$$

where a again refers to the one additional AG parameter which describes the deviation from GR.

However, we should not expect that the posterior widths will be identical in the AG and GR models for all parameters except the additional AG parameter are the same in the AG and GR models. At high SNRs where the log likelihood can be approximated by a quadratic, posterior widths should scale inverse with the signal-to-noise ratio ρ . While ρ is the optimal SNR recovered when AG templates are used within the AG model, the maximal SNR recoverable when using GR templates within the GR model is lower. By definition, this GR SNR is

$$\rho_{\text{GR}} \equiv \left[\frac{(h_{\text{AG}}|h_{\text{GR}}(\theta))}{\sqrt{(h_{\text{GR}}(\theta)|h_{\text{GR}}(\theta))}} \right]_{\text{max over } \theta} = \text{FF} \rho. \quad (\text{C.19})$$

Assuming the inverse SNR scaling of the posteriors, and using identical priors on common parameters in the AG and GR models, Eq. (C.15) reduces to

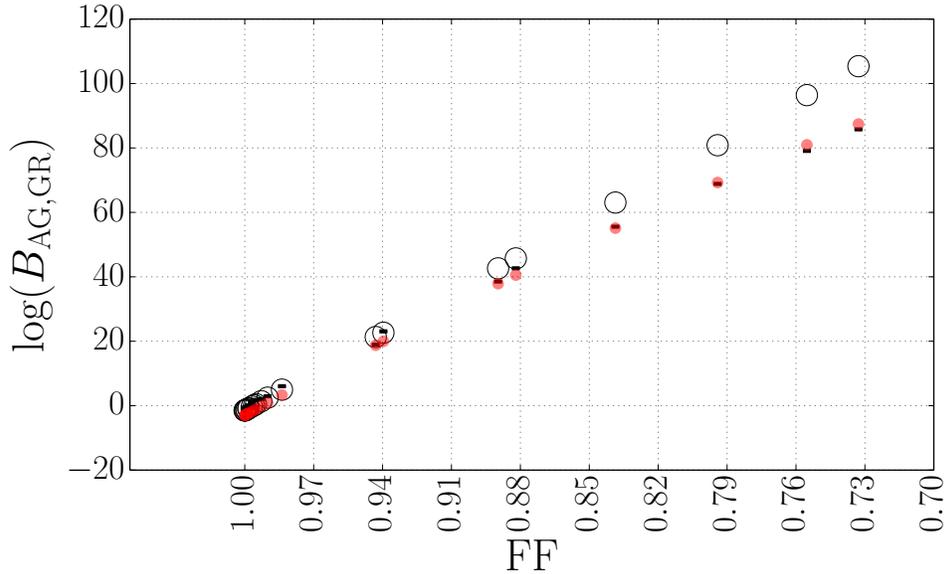
$$B_{\text{AG,GR}} \approx \frac{[L_{\text{AG}}]_{\text{max over } \theta'}}{[L_{\text{GR}}]_{\text{max over } \theta}} \text{FF}^{N-1} \sqrt{2\pi} \frac{\Delta\theta_{\text{est}}^a}{\Delta\theta_{\text{prior}}^a}. \quad (\text{C.20})$$

Taking a logarithm of this equation and again using $\frac{\rho^2}{2}(1 - \text{FF}^2)$ for the difference between maximum likelihoods (C.31), we find

$$\begin{aligned} \log(B_{\text{AG,GR}}) &\approx \frac{\rho^2}{2}(1 - \text{FF}^2) + (N - 1) \log(\text{FF}) \\ &+ \log\left(\sqrt{2\pi} \frac{\Delta\theta_{\text{est}}^a}{\Delta\theta_{\text{prior}}^a}\right). \end{aligned} \quad (\text{C.21})$$

Eq. (C.21) reduces to Eq. (C.6) for $\text{FF} \sim 1$. However, it is accurate for a much wider range of fitting factors. Fig. C.4 shows the comparison between the log-Bayes factors from `lalinference` (error bars), the ones from Eq. (C.6) (circles) and finally the ones from Eq. (C.21) (red dots). Indeed, the log-Bayes factors from Eq. (C.21) show a very close agreement with the numerical values. Thus, Eq. (C.21) provides a good

Figure C.4 Logarithmic Bayes factor from `lalinference` (error bars), from Eq. (C.6) (empty circles), and from Eq. (C.21) (red dots) as a function of the FF. The SNR is fixed to 20. The `lalinference` errors are computed from Eq. (C.12). The extended analytical expression of Eq. (C.21) which includes corrections for lower fitting factors is in good agreement with the exact calculation from `lalinference`.



approximation to the exact values of the log-Bayes factors.

Another merit of Eq. (C.21) is that it can be generalised to an arbitrary number of extra non-GR parameters. If we have k non-GR parameters, Eq. (C.21) becomes:

$$\begin{aligned} \log(B_{\text{AG,GR}}) &\approx \frac{\rho^2}{2}(1 - \text{FF}^2) + (N - k) \log(\text{FF}) \\ &+ \log \left((2\pi)^{k/2} \prod_i^k \frac{\Delta\theta_{est}^{a_i}}{\Delta\theta_{prior}^{a_i}} \right). \end{aligned} \quad (\text{C.22})$$

Throughout this work, we have restricted our attention to the zero-noise realisation. Vallisneri analysed the distribution of the Bayes factor under different noise realisations and showed (see Eq. (15) of [181]) that fluctuations in the logarithm of the Bayes factor have a standard deviation of $\sim \sqrt{2\rho\sqrt{1 - \text{FF}}}$. While our additional corrections to the

Bayes factor also lead to corrections in this quantity, we neglect these second-order effects.

We can compare the two systematic corrections discussed above to the level of these statistical fluctuations due to noise. The difference in the log-Bayes factor between Eq. (C.6) and Eq. (C.17), i.e., the difference between the approximations of Refs. [181] and [55], is $(1/2)\rho^2(1 - \text{FF}^2) - \rho^2(1 - \text{FF}) = -(1/2)\rho^2(1 - \text{FF})^2$. This difference is approximately equal to the statistical fluctuation in the log-Bayes factor for $\rho = 20$ and $\text{FF} \sim 0.73$, corresponding to the rightmost points in Fig. C.4. Meanwhile, the new correction to the log-Bayes factor which we introduced in Eq. (C.21) has a magnitude of $(N - 1)\log \text{FF}$; for $N = 9$ and other parameters as above, it is several times smaller than the noise-induced fluctuations.

Therefore, these corrections are unlikely to impact the detectability of a deviation from GR; in any case, in practice the detectability of the deviation would be determined by an analysis of the data and a numerical computation of the Bayes factor, not approximate predictive techniques. However, these corrections are useful in explaining the apparent difference between numerical and analytical calculations, and therefore help validate both approaches by enabling a successful cross-check.

C.4 Discussion

We computed the Bayes factor between a GR model and an alternative gravity model for a gravitational-wave signature of an inspiraling compact binary. We compared two calculations of the Bayes factor: an exact numerical computation with `lalinference` and an approximate analytical prediction due to Vallisneri [181]. We verified that the

analytical approximation yields the correct scaling of the logarithm of the Bayes factor with the square of the signal-to-noise ratio at high fitting factor values. However, the predicted scaling of the Bayes factor with the fitting factor is inaccurate for $\text{FF} \leq 0.9$.

We extended the regime of validity of the analytical approximation of [181] to lower fitting factors by including additional FF-dependent terms and by extending to multiple non-GR parameters. We confirmed that the more complete analytical prediction that we derived in this work, Eq. (C.21), remains valid down to fitting factors of ≤ 0.7 .

It is worth noting that Eq. (C.6) loses accuracy precisely in the regime where it becomes possible to differentiate GR and alternative gravity models. The FF is very close to unity in the regime in which the GR waveform can still match a signal which violates GR through different choices of the values of the binary’s parameters within the GR framework. The Bayes factor in this case is not significantly different from 1, thus no decision on the nature of the signal can be made at an acceptably low false alarm probability³. Therefore, our extension of the analytical expression for the Bayes factor to lower fitting factors provides a useful, easy-to-compute approximate technique precisely in the regime of interest in the case of a zero noise realisation.

The analytical expressions presented in [55, 181] and in this work are predicated on the assumption that the $(N - 1)$ -dimensional GR manifold and the N -dimensional AG manifold are sufficiently similar near the maximum-likelihood values that the parameter uncertainties can be assumed to be equal (up to scaling with the inverse SNR) on the two manifolds. Differences in the local curvature of the two manifolds could become significant when the distance between them is large, or the systematic bias between true and best-fit parameters is significant relative to statistical measurement uncertainty.

³This regime is a case of the so-called “fundamental bias” [216]. It is treated using the analytical approximation presented in [181] by [182]. A numerical study with the `lalInference` code can be found in [196].

Therefore, this assumption could (although need not) break down either at small fitting factors, or, somewhat paradoxically, at large SNR for a fixed fitting factor. Specifically, when $\rho^2(1 - FF^2) \gg N$, the uncertainty region within a manifold is much smaller than the distance between manifolds or between true and best-fit parameters within a manifold, and the AG and GR manifolds may no longer yield similar parameter uncertainties.

Another possible cause of the breakdown of the analytical approximation is the impact of priors. If the prior distribution is very non-uniform within the region of likelihood support, particularly if a sharp prior boundary is present within this region, the analytical approach described above is no longer valid. A further limitation is the restriction to high SNRs. The widths of the posterior distributions are inversely proportional to the SNR only when the linearized-signal approximation is valid (i.e., when the covariance matrix is well approximated by the inverse of the Fisher matrix).

In summary, the analytical approximation presented by Cornish et al. [55] and Vallisneri [181], and its extensions as given in Eqs. (C.21) and (C.22), provide a computationally cheap way of predicting the detectability of a deviation from GR for a given AG theory without the need to run expensive numerical simulations, subject to the limitations outlined above. Hence, these analytical approximations can be a very useful tool to get quick indications of whether a particular class and magnitude of one-parameter deviations from GR are detectable. However, these methods are merely predictive, and inference on actual data must rely on parameter estimation and model comparison with complete data-analysis pipelines.

Acknowledgements

We thank Michele Vallisneri, Neil Cornish, John Veitch, Will Farr, Christopher Berry, Carl-Johan Haster and Zachary Hafen for useful comments and discussions. The work was funded in part by a Leverhulme Trust research project grant. The numerical simulations were performed on the Tsunami cluster of the University of Birmingham.

C.5 Appendix: Computing Fitting Factors from log-Likelihoods

The fitting factor Eq. (C.5) can be extracted directly from the Nested Sampling runs without the need to search over a parameter grid. Begin by writing the logarithmic likelihood in a zero noise realisation:

$$\log(L) = \text{const} + (h_{true}|h(\theta)) - \frac{(h_{true}|h_{true})}{2} - \frac{(h(\theta)|h(\theta))}{2} \quad (\text{C.23})$$

where h_{true} is the gravitational wave signal in the data stream, $h(\theta)$ is the search template, and const is a constant. Consider the difference $\Delta\lambda$ between the maximum log likelihoods given for the AG and GR models given an AG signal:

$$\Delta\lambda = \log(L_{GR})_{\max \text{ over } \theta} - \log(L_{AG})_{\max \text{ over } \theta'} , \quad (\text{C.24})$$

where

$$\log(L_{GR}) = \text{const} + (h_{true}|h_{GR}(\theta)) - \frac{(h_{true}|h_{true})}{2} - \frac{(h_{GR}(\theta)|h_{GR}(\theta))}{2} \quad (\text{C.25})$$

and $\log(L_{AG})_{\max \text{ over } \theta'} = \text{const}$ since this likelihood is maximized for $h_{AG}(\theta') = h_{true}$.

We can maximise the GR log-likelihood analytically over the amplitude of $h_{GR}(\theta)$ by defining

$$h_{GR}(\theta) = A\hat{h}_{GR}(\xi) \quad (\text{C.26})$$

with $\xi \equiv \theta \setminus A$ being the set of parameters other than the amplitude. One can solve for the value of the amplitude that satisfies

$$\frac{\partial \log(L_{AG})}{\partial A} = 0 \quad (\text{C.27})$$

and obtain:

$$A = \frac{(h_{true}|\hat{h}_{GR}(\xi))}{(\hat{h}_{GR}(\xi)|\hat{h}_{GR}(\xi))}. \quad (\text{C.28})$$

Substituting this into Eq. (C.25) and setting $(h_{true}|h_{true}) \equiv \rho^2$ yields:

$$\Delta\lambda = \frac{1}{2} \left[\frac{(h_{true}|\hat{h}_{GR}(\xi))^2}{(\hat{h}_{GR}(\xi)|\hat{h}_{GR}(\xi))} \right]_{\max \text{ over } \xi} - \frac{\rho^2}{2}. \quad (\text{C.29})$$

Meanwhile, the fitting factor FF can be similarly written as:

$$FF = \left[\frac{(h_{true}|\hat{h}_{GR}(\xi))}{\rho\sqrt{(\hat{h}_{GR}(\xi)|\hat{h}_{GR}(\xi))}} \right]_{\max \text{ over } \xi}. \quad (\text{C.30})$$

Thus,

$$\Delta\lambda = \frac{1}{2}\rho^2 FF^2 - \frac{\rho^2}{2} = -\frac{\rho^2}{2}(1 - FF^2) \quad (\text{C.31})$$

and

$$FF = \sqrt{\frac{2\Delta\lambda}{\rho^2} + 1}. \quad (\text{C.32})$$

The numerically computed values of $\Delta\lambda$ have an intrinsic variability due to the stochastic nature of the sampler. The standard deviation for $\Delta\lambda$ derived from our simulations is $\sigma_{\Delta\lambda} = 0.016$. The corresponding uncertainty in our FF estimate is given by

$$\sigma_{FF} = \frac{1}{\rho^2 FF} \sigma_{\Delta\lambda}. \quad (\text{C.33})$$

For an SNR of 20 and a FF of 1, $\sigma_{FF} = 4 \times 10^{-5}$. For all practical purposes we can consider our estimated FFs to be exact.

REFERENCES

- [1] www.advancedligo.mit.edu.
- [2] wwwcascina.virgo.infn.it/advirgo.
- [3] www.skatelescope.org.
- [4] Abadie, J. et al. (the LIGO Scientific Collaboration and the Virgo Collaboration). Predictions for the rates of compact binary coalescences observable by ground-based gravitational-wave detectors. Class. Quantum Gravity, 27(17):173001, 2010.
- [5] Aasi, J. et al. (LIGO-Virgo Scientific Collaboration). Parameter estimation for compact binary coalescence signals with the first generation gravitational-wave detector network. Phys. Rev. D, 88(6):062001, September 2013.
- [6] J. Abadie, B. P. Abbott, R. Abbott, T. D. Abbott, M. Abernathy, T. Accadia, F. Acernese, C. Adams, R. Adhikari, C. Affeldt, and et al. All-sky search for gravitational-wave bursts in the second joint LIGO-Virgo run. Phys. Rev. D, 85(12):122007, June 2012.
- [7] J. Abadie, B. P. Abbott, R. Abbott, T. D. Abbott, M. Abernathy, T. Accadia, F. Acernese, C. Adams, R. Adhikari, C. Affeldt, and et al. Search for gravitational waves from low mass compact binary coalescence in LIGO's sixth science run and Virgo's science runs 2 and 3. Phys. Rev. D, 85(8):082002, April 2012.
- [8] J. Abadie, B. P. Abbott, R. Abbott, T. D. Abbott, M. Abernathy, T. Accadia, F. Acernese, C. Adams, R. Adhikari, C. Affeldt, and et al. Upper limits on a stochastic gravitational-wave background using LIGO and Virgo interferometers at 600-1000 Hz. Phys. Rev. D, 85(12):122001, June 2012.
- [9] J. et al. Abadie. First low-latency LIGO+Virgo search for binary inspirals and their electromagnetic counterparts. Astron. Astrophys., 541:A155, May 2012.

- [10] B. Abbott, R. Abbott, R. Adhikari, A. Ageev, B. Allen, R. Amin, S. B. Anderson, W. G. Anderson, M. Araya, H. Armandula, and et al. Detector description and performance for the first coincidence observations between LIGO and GEO. *Nuclear Instruments and Methods in Physics Research A*, 517:154–179, January 2004.
- [11] B. P. Abbott, R. Abbott, R. Adhikari, P. Ajith, B. Allen, G. Allen, R. S. Amin, S. B. Anderson, W. G. Anderson, M. A. Arain, and et al. LIGO: the Laser Interferometer Gravitational-Wave Observatory. *Reports on Progress in Physics*, 72(7):076901, July 2009.
- [12] T Accadia, F Acernese, M Alshourbagy, P Amico, F Antonucci, S Aoudia, N Arnaud, C Arnault, K G Arun, P Astone, S Avino, D Babusci, G Ballardini, F Barone, G Barrand, L Barsotti, M Barsuglia, A Basti, Th S Bauer, F Beauville, M Bebronne, M Bejger, M G Beker, F Bellachia, A Belletoile, J L Beney, M Bernardini, S Bigotta, R Bilhaut, S Birindelli, M Bitossi, M A Bizouard, M Blom, C Boccara, D Boget, F Bondu, L Bonelli, R Bonnand, V Boschi, L Bosi, T Bouedo, B Bouhou, A Bozzi, L Bracci, S Braccini, C Bradaschia, M Branchesi, T Briant, A Brilliet, V Brisson, L Brocco, T Bulik, H J Bulten, D Buskulic, C Buy, G Cagnoli, G Calamai, E Calloni, E Campagna, B Canuel, F Carbognani, L Carbone, F Cavalier, R Cavalieri, R Cecchi, G Cella, E Cesarini, E Chassande-Mottin, S Chatterji, R Chiche, A Chincarini, A Chiummo, N Christensen, A C Clapson, F Cleva, E Coccia, P F Cohadon, C N Colacino, J Colas, A Colla, M Colombini, G Conforto, A Corsi, S Cortese, F Cottone, J P Coulon, E Cuoco, S D’Antonio, G Daguin, A Dari, V Dattilo, P Y David, M Davier, R Day, G Debreczeni, G De Carolis, M Dehamme, R Del Fabbro, W Del Pozzo, M del Prete, L Derome, R De Rosa, R DeSalvo, M Dialinas, L Di Fiore, A Di Lieto, M Di Paolo Emilio, A Di Virgilio, A Dietz, M Doets, P Dominici, A Dominjon, M Drago, C Drezen, B Dujardin, B Dulach, C Eder, A Eleuteri, D Enard, M Evans, L Fabbroni, V Fafone, H Fang, I Ferrante, F Fidecaro, I Fiori, R Flaminio, D Forest, L A Forte, J D Fournier, L Fournier, J Franc, O Francois, S Frasca, F Frasconi, A Freise, A Gaddi, M Galimberti, L Gammaitoni, P Ganau, C Garnier, F Garufi, M E Gáspár, G Gemme, E Genin, A Gennai, G Gennaro, L Giacobone, A Giazotto, G Giordano, L Giordano, C Girard, R Gouaty, A Grado, M Granata, V Granata, X Grave, C Greverie, H Groenstege, G M Guidi, S Hamdani, J F Hayau, S Hebri, A Heidmann, H Heitmann, P Hello, G Hemming, E Hennes, R Hermel, P Heusse, L Holloway, D Huet, M Iannarelli, P Jaranowski, D Jehanno, L Journet, S Karkar, T Ketel, H Voet, J Kovalik, I Kowalska, S Kreckelbergh, A Krolak, J C Lacotte, B Lagrange, P La Penna, M Laval, J C Le Marec, N Leroy, N Letendre, T G F Li, B Lieunard, N Liguori, O Lodygensky, B Lopez, M Lorenzini, V Lorette, G Losurdo, M Loupias, J M Mackowski, T Maiani, E Majorana, C Magazzù, I Maksimovic, V Malvezzi, N Man, S Mancini, B Mansoux, M Mantovani, F Marchesoni, F Marion, P Marin, J Marque, F Martelli, A Masserot, L Massonnet, G Matone,

L Matone, M Mazzoni, F Menzinger, C Michel, L Milano, Y Minenkov, S Mitra, M Mohan, J L Montorio, R Morand, F Moreau, J Moreau, N Morgado, A Morgia, S Mosca, V Moscatelli, B Mours, P Mugnier, F A Mul, L Naticchioni, I Neri, F Nocera, E Pacaud, G Pagliaroli, A Pai, L Palladino, C Palomba, F Paoletti, R Paoletti, A Paoli, S Pardi, G Parguez, M Parisi, A Pasqualetti, R Passaquieti, D Passuello, M Perciballi, B Perniola, G Persichetti, S Petit, M Pichot, F Piergiovanni, M Pietka, R Pignard, L Pinard, R Poggiani, P Popolizio, T Pradier, M Prato, G A Prodi, M Punturo, P Puppo, K Qipiani, O Rabaste, D S Rabeling, I Rácz, F Raffaelli, P Rapagnani, S Rapisarda, V Re, A Reboux, T Regimbau, V Reita, A Remilleux, F Ricci, I Ricciardi, F Richard, M Ripepe, F Robinet, A Rocchi, L Rolland, R Romano, D Rosińska, P Roudier, P Ruggi, G Russo, L Salconi, V Sannibale, B Sassolas, D Sentenac, S Solimeno, R Sottile, L Sperandio, R Stanga, R Sturani, B Swinkels, M Tacca, R Taddei, L Taffarello, M Tarallo, S Tissot, A Toncelli, M Tonelli, O Torre, E Tournefier, F Travasso, C Tremola, E Turri, G Vajente, J F J van den Brand, C Van Den Broeck, S van der Putten, M Vasuth, M Vavoulidis, G Vedovato, D Verkindt, F Vetrano, O Véziant, A Viceré, J Y Vinet, S Vilalte, S Vitale, H Vocca, R L Ward, M Was, K Yamamoto, M Yvert, J P Zendri, and Z Zhang. Virgo: a laser interferometer to detect gravitational waves. *Journal of Instrumentation*, 7(03):P03012, 2012.

- [13] T. Accadia, F. Acernese, F. Antonucci, P. Astone, G. Ballardin, F. Barone, M. Barsuglia, A. Basti, T. S. Bauer, M. Bebronne, M. G. Beker, A. Belletoile, S. Birindelli, M. Bitossi, M. A. Bizouard, M. Blom, F. Bondu, L. Bonelli, R. Bonnard, V. Boschi, L. Bosi, B. Bouhou, S. Braccini, C. Bradaschia, M. Branchesi, T. Briant, A. Brillet, V. Brisson, R. Budzyński, T. Bulik, H. J. Bulten, D. Buskulic, C. Buy, G. Cagnoli, E. Calloni, B. Canuel, F. Carbognani, F. Cavalier, R. Cavalieri, G. Cella, E. Cesarini, O. Chaïbi, E. Chassande Mottin, A. Chincarini, F. Cleva, E. Coccia, P.-F. Cohadon, C. N. Colacino, J. Colas, A. Colla, M. Colombini, A. Corsi, J.-P. Coulon, E. Cuoco, S. D’Antonio, V. Dattilo, M. Davier, R. Day, R. De Rosa, G. Debreczeni, W. Del Pozzo, M. del Prete, L. Di Fiore, A. Di Lieto, M. D. P. Emilio, A. Di Virgilio, A. Dietz, M. Drago, V. Fafone, I. Ferrante, F. Fidecaro, I. Fiori, R. Flaminio, L. A. Forte, J.-D. Fournier, J. Franc, S. Frasca, F. Frasconi, M. Galimberti, L. Gammaitoni, F. Garufi, M. E. Gáspár, G. Gemme, E. Genin, A. Gennai, A. Giazotto, R. Gouaty, M. Granata, C. Greverie, G. M. Guidi, J.-F. Hayau, A. Heidmann, H. Heitmann, P. Hello, D. Huet, P. Jaranowski, I. Kowalska, A. Królak, N. Leroy, N. Letendre, T. G. F. Li, N. Liguori, M. Lorenzini, V. Loriette, G. Losurdo, E. Majorana, I. Maksimovic, N. Man, M. Mantovani, F. Marchesoni, F. Marion, J. Marque, F. Martelli, A. Masserot, C. Michel, L. Milano, Y. Minenkov, M. Mohan, N. Morgado, A. Morgia, S. Mosca, V. Moscatelli, B. Mours, F. Nocera, G. Pagliaroli, L. Palladino, C. Palomba, F. Paoletti, M. Parisi, A. Pasqualetti, R. Passaquieti, D. Passuello, G. Persichetti, M. Pichot, F. Piergiovanni, M. Pietka, L. Pinard, R. Poggiani, M. Prato, G. A. Prodi, M. Punturo, P. Puppo, D. S. Rabeling, I. Rácz, P. Rapag-

nani, V. Re, T. Regimbau, F. Ricci, F. Robinet, A. Rocchi, L. Rolland, R. Romano, D. Rosińska, P. Ruggi, B. Sassolas, D. Sentenac, L. Sperandio, R. Sturani, B. Swinkels, M. Tacca, L. Taffarello, A. Toncelli, M. Tonelli, O. Torre, E. Tournefier, F. Travasso, G. Vajente, J. F. J. van den Brand, C. Van Den Broeck, S. van der Putten, M. Vasuth, M. Vavoulidis, G. Vedovato, D. Verkindt, F. Vetrano, A. Viceré, J.-Y. Vinet, S. Vitale, H. Vocca, R. L. Ward, M. Was, M. Yvert, and J.-P. Zendri. Status of the Virgo project. *Classical and Quantum Gravity*, 28(11):114002, June 2011.

- [14] F. Acernese, M. Alshourbagy, P. Amico, F. Antonucci, S. Aoudia, K. G. Arun, P. Astone, S. Avino, L. Baggio, G. Ballardín, F. Barone, L. Barsotti, M. Barsuglia, T. S. Bauer, S. Bigotta, S. Birindelli, M. A. Bizouard, C. Boccara, F. Bondu, L. Bosi, S. Braccini, C. Bradaschia, A. Brillet, V. Brisson, D. Buskulic, G. Cagnoli, E. Calloni, E. Campagna, F. Carbognani, F. Cavalier, R. Cavalieri, G. Cella, E. Cesarini, E. Chassande-Mottin, S. Chatterji, F. Cleva, E. Coccia, C. Corda, A. Corsi, F. Cottone, J.-P. Coulon, E. Cuoco, S. D’Antonio, A. Dari, V. Dattilo, M. Davier, R. De Rosa, M. DelPrete, L. Di Fiore, A. Di Lieto, M. D. P. Emilio, A. Di Virgilio, M. Evans, V. Fafone, I. Ferrante, F. Fidecaro, I. Fiori, R. Flaminio, J.-D. Fournier, S. Frasca, F. Frasconi, L. Gammaitoni, F. Garufi, E. Genin, A. Gennai, A. Giazotto, V. Granata, C. Greverie, D. Grosjean, G. Guidi, S. Hamedani, S. Hebri, H. Heitmann, P. Hello, D. Huet, P. La Penna, M. Laval, N. Leroy, N. Letendre, B. Lopez, M. Lorenzini, V. Loriette, G. Losurdo, J.-M. Mackowski, E. Majorana, N. Man, M. Mantovani, F. Marchesoni, F. Marion, J. Marque, F. Martelli, A. Masserot, F. Menzinger, L. Milano, Y. Minenkov, M. Mohan, J. Moreau, N. Morgado, S. Mosca, B. Mours, I. Neri, F. Nocera, G. Pagliaroli, C. Palomba, F. Paoletti, S. Pardi, A. Pasqualetti, R. Passaquieti, D. Passuello, F. Piergiovanni, L. Pinard, R. Poggiani, M. Punturo, P. Puppò, O. Rabaste, P. Rapagnani, T. Regimbau, A. Remillieux, F. Ricci, I. Ricciardi, A. Rocchi, L. Rolland, R. Romano, P. Ruggi, D. Sentenac, S. Solimeno, B. L. Swinkels, R. Terenzi, A. Toncelli, M. Tonelli, E. Tournefier, F. Travasso, G. Vajente, J. F. J. van den Brand, S. van der Putten, D. Verkindt, F. Vetrano, A. Viceré, J.-Y. Vinet, H. Vocca, and M. Yvert. Virgo status. *Classical and Quantum Gravity*, 25(18):184001, September 2008.

- [15] F. Acernese, P. Amico, N. Arnaud, D. Babusci, R. Barillé, F. Barone, L. Barsotti, M. Barsuglia, F. Beauville, M. A. Bizouard, C. Boccara, F. Bondu, L. Bosi, C. Bradaschia, L. Bracci, S. Braccini, A. Brillet, V. Brisson, L. Brocco, D. Buskulic, G. Calamai, E. Calloni, E. Campagna, F. Cavalier, G. Cella, E. Chassande-Mottin, F. Cleva, T. Cokelaer, G. Conforto, C. Corda, J.-P. Coulon, E. Cuoco, V. Dattilo, M. Davier, R. De Rosa, L. Di Fiore, A. Di Virgilio, B. Dujardin, A. Eleuteri, D. Enard, I. Ferrante, F. Fidecaro, I. Fiori, R. Flaminio, J.-D. Fournier, S. Frasca, F. Frasconi, L. Gammaitoni, A. Gennai, A. Giazotto, G. Giordano, G. Guidi, H. Heitmann, P. Hello, P. Heusse, L. Holloway, S. Kreckelberg,

- P. La Penna, V. Lorette, M. Loupias, G. Losurdo, J.-M. Mackowski, E. Majorana, C. N. Man, F. Marion, F. Martelli, A. Masserot, L. Massonnet, M. Mazzoni, L. Milano, J. Moreau, F. Moreau, N. Morgado, F. Mornet, B. Mours, J. Pacheco, A. Pai, C. Palomba, F. Paoletti, R. Passaquieti, D. Passuello, B. Perniola, L. Pinard, R. Poggiani, M. Punturo, P. Puppo, K. Qipiani, J. Ramonet, P. Rapagnani, V. Reita, A. Remillieux, F. Ricci, I. Ricciardi, G. Russo, S. Solimeno, R. Stanga, A. Toncelli, M. Tonelli, E. Tournefier, F. Travasso, H. Trinquet, M. Varvella, D. Verkindt, F. Vetrano, O. Veziat, A. Viceré, J.-Y. Vinet, H. Vocca, and M. Yvert. Status of VIRGO. Classical and Quantum Gravity, 21:385, March 2004.
- [16] M. Agathos, W. Del Pozzo, T. G. F. Li, C. Van Den Broeck, J. Veitch, and S. Vitale. Testing general relativity using gravitational waves from binary neutron stars: Effect of spins. ArXiv e-prints, May 2013.
- [17] M. Agathos, W. Del Pozzo, T. G. F. Li, C. Van Den Broeck, J. Veitch, and S. Vitale. TIGER: A data analysis pipeline for testing the strong-field dynamics of general relativity with gravitational wave signals from coalescing compact binaries. Phys. Rev. D, 89(8):082001, April 2014.
- [18] P. Ajith, S. Babak, Y. Chen, M. Hewitson, B. Krishnan, J. T. Whelan, B. Brügmann, P. Diener, J. Gonzalez, M. Hannam, S. Husa, M. Koppitz, D. Pollney, L. Rezzolla, L. Santamaría, A. M. Sintes, U. Sperhake, and J. Thornburg. A phenomenological template family for black-hole coalescence waveforms. Classical and Quantum Gravity, 24:689, October 2007.
- [19] P. Ajith, M. Boyle, D. A. Brown, B. Brügmann, L. T. Buchman, L. Cadonati, M. Campanelli, T. Chu, Z. B. Etienne, S. Fairhurst, M. Hannam, J. Healy, I. Hinder, S. Husa, L. E. Kidder, B. Krishnan, P. Laguna, Y. T. Liu, L. London, C. O. Lousto, G. Lovelace, I. MacDonald, P. Marronetti, S. Mohapatra, P. Mösta, D. Müller, B. C. Mundim, H. Nakano, F. Ohme, V. Paschalidis, L. Pekowsky, D. Pollney, H. P. Pfeiffer, M. Ponce, M. Pürrer, G. Reifenberger, C. Reisswig, L. Santamaría, M. A. Scheel, S. L. Shapiro, D. Shoemaker, C. F. Sopuerta, U. Sperhake, B. Szilágyi, N. W. Taylor, W. Tichy, P. Tsatsin, and Y. Zlochower. The NINJA-2 catalog of hybrid post-Newtonian/numerical-relativity waveforms for non-precessing black-hole binaries. Classical and Quantum Gravity, 29(12):124001, June 2012.
- [20] P. Ajith, M. Hannam, S. Husa, Y. Chen, B. Brügmann, N. Dorband, D. Müller, F. Ohme, D. Pollney, C. Reisswig, L. Santamaría, and J. Seiler. Inspiral-Merger-Ringdown Waveforms for Black-Hole Binaries with Nonprecessing Spins. Phys. Rev. Lett., 106(24):241101, June 2011.

- [21] S. Alexander, L. S. Finn, and N. Yunes. Gravitational-wave probe of effective quantum gravity. Phys. Rev. D, 78(6):066005, September 2008.
- [22] B. Allen, W. G. Anderson, P. R. Brady, D. A. Brown, and J. D. E. Creighton. FINDCHIRP: An algorithm for detection of gravitational waves from inspiraling compact binaries. Phys. Rev. D, 85(12):122006, June 2012.
- [23] T. A. Apostolatos, C. Cutler, G. J. Sussman, and K. S. Thorne. Spin-induced orbital precession and its modulation of the gravitational waveforms from merging binaries. Phys. Rev. D, 49:6274–6297, June 1994.
- [24] Theocharis A. Apostolatos. Search templates for gravitational waves from precessing, inspiraling binaries. Phys. Rev. D, 52(2):605–620, Jul 1995.
- [25] K. G. Arun, B. R. Iyer, M. S. S. Qusailah, and B. S. Sathyaprakash. Probing the nonlinear structure of general relativity with black hole binaries. Phys. Rev. D, 74:024006, Jul 2006.
- [26] K. G. Arun, Bala R Iyer, B. S. Sathyaprakash, and Pranesh A Sundararajan. Parameter estimation of inspiralling compact binaries using 3.5 post-newtonian gravitational wave phasing: The nonspinning case. Phys. Rev. D, 71:084008, Apr 2005.
- [27] K. G. Arun and C. M. Will. Bounding the mass of the graviton with gravitational waves: effect of higher harmonics in gravitational waveform templates. Classical and Quantum Gravity, 26(15):155002, August 2009.
- [28] K.G. Arun, B.R. Iyer, M.S.S. Qusailah, and B.S. Sathyaprakash. Testing post-newtonian theory with gravitational wave observations. Class. Quantum Grav., 23:L37–L43, 2006.
- [29] Y. Aso, Y. Michimura, K. Somiya, M. Ando, O. Miyakawa, T. Sekiguchi, D. Tatsumi, and H. Yamamoto. Interferometer design of the KAGRA gravitational wave detector. Phys. Rev. D, 88(4):043007, August 2013.
- [30] B. Aylott, J. G. Baker, W. D. Boggs, M. Boyle, P. R. Brady, D. A. Brown, B. Brügmann, L. T. Buchman, A. Buonanno, L. Cadonati, J. Camp, M. Campanelli, J. Centrella, S. Chatterji, N. Christensen, T. Chu, P. Diener, N. Dorband, Z. B. Etienne, J. Faber, S. Fairhurst, B. Farr, S. Fischetti, G. Guidi, L. M. Goggin, M. Hannam, F. Herrmann, I. Hinder, S. Husa, V. Kalogera, D. Keppel, L. E. Kidder, B. J. Kelly, B. Krishnan, P. Laguna, C. O. Lousto, I. Mandel, P. Marronetti, R. Matzner, S. T. McWilliams, K. D. Matthews, R. A. Mercer, S. R. P. Mohapatra, A. H. Mroué, H. Nakano, E. Ochsner, Y. Pan, L. Pekowsky, H. a. P. Pfeiffer, D. Pollney, F. Pretorius, V. Raymond, C. Reisswig, L. Rezzolla, O. Rinne,

- C. Robinson, C. Röver, L. Santamaría, B. Sathyaprakash, M. A. Scheel, E. Schnetter, J. Seiler, S. L. Shapiro, D. Shoemaker, U. Sperhake, A. Stroeer, R. Sturani, W. Tichy, Y. T. Liu, M. van der Sluys, J. R. van Meter, R. Vaulin, A. Vecchio, J. Veitch, A. Viceré, J. T. Whelan, and Y. Zlochower. Testing gravitational-wave searches with numerical relativity waveforms: results from the first Numerical INjection Analysis (NINJA) project. Classical and Quantum Gravity, 26(16):165008, August 2009.
- [31] B. Aylott, J. Veitch, and A. Vecchio. Bayesian inference on the Numerical INjection Analysis (NINJA) data set using a nested sampling algorithm. Classical and Quantum Gravity, 26(11):114011, June 2009.
- [32] S. Babak and A. Sesana. Resolving multiple supermassive black hole binaries with pulsar timing arrays. Phys. Rev. D, 85(4):044034, February 2012.
- [33] L. Barack and C. Cutler. Using LISA extreme-mass-ratio inspiral sources to test off-Kerr deviations in the geometry of massive black holes. Phys. Rev. D, 75(4):042003, February 2007.
- [34] Leor Barack and Curt Cutler. Lisa capture sources: Approximate waveforms, signal-to-noise ratios, and parameter estimation accuracy. Phys. Rev. D, 69:082005, Apr 2004.
- [35] E. Barausse and A. Buonanno. Extending the effective-one-body Hamiltonian of black-hole binaries to include next-to-next-to-leading spin-orbit couplings. Phys. Rev. D, 84(10):104027, November 2011.
- [36] K. Belczynski, V. Kalogera, F. A. Rasio, R. E. Taam, A. Zezas, T. Bulik, T. J. Maccarone, and N. Ivanova. Compact Object Modeling with the StarTrack Population Synthesis Code. The Astrophysical Journal Supplement Series, 174:223–260, January 2008.
- [37] E. Berti, A. Buonanno, and C. M. Will. Estimating spinning binary parameters and testing alternative theories of gravity with LISA. Phys. Rev. D, 71(8):084025, April 2005.
- [38] E. Berti, A. Buonanno, and C. M. Will. Testing general relativity and probing the merger history of massive black holes with LISA. Classical and Quantum Gravity, 22:943, September 2005.
- [39] E. Berti, V. Cardoso, and C. M. Will. Gravitational-wave spectroscopy of massive black holes with the space interferometer LISA. Phys. Rev. D, 73(6):064030, March 2006.

- [40] E. Berti, J. Gair, and A. Sesana. Graviton mass bounds from space-based gravitational-wave observations of massive black hole populations. Phys. Rev. D, 84(10):101501, November 2011.
- [41] E. Berti and M. Volonteri. Cosmological Black Hole Spin Evolution by Mergers and Accretion. Astrophys. J., 684:822–828, September 2008.
- [42] B. Bertotti, L. Iess, and P. Tortora. A test of general relativity using radio links with the Cassini spacecraft. Nature, 425:374–376, September 2003.
- [43] BICEP2/Keck and Planck Collaborations, P. A. R. Ade, N. Aghanim, Z. Ahmed, R. W. Aikin, K. D. Alexander, M. Arnaud, J. Aumont, C. Baccigalupi, A. J. Banday, and et al. Joint Analysis of BICEP2/Keck Array and Planck Data. Physical Review Letters, 114(10):101301, March 2015.
- [44] L. Blanchet, A. Buonanno, and G. Faye. Higher-order spin effects in the dynamics of compact binaries. II. Radiation field. Phys. Rev. D, 74(10):104034, November 2006.
- [45] L. Blanchet, T. Damour, B. R. Iyer, C. M. Will, and A. G. Wiseman. Gravitational-Radiation Damping of Compact Binary Systems to Second Post-Newtonian Order. Phys. Rev. Lett., 74:3515–3518, May 1995.
- [46] L. Blanchet, G. Faye, B. R. Iyer, and S. Sinha. The third post-Newtonian gravitational wave polarizations and associated spherical harmonic modes for inspiralling compact binaries in quasi-circular orbits. Classical and Quantum Gravity, 25(16):165003, August 2008.
- [47] L. Blanchet and B. S. Sathyaprakash. Signal analysis of gravitational wave tails. Classical and Quantum Gravity, 11:2807–2831, November 1994.
- [48] L. Blanchet and B. S. Sathyaprakash. Detecting a Tail Effect in Gravitational-Wave Experiments. Phys. Rev. Lett., 74:1067–1070, February 1995.
- [49] Luc Blanchet. Gravitational radiation from post-newtonian sources and inspiralling compact binaries. Living Rev. Relativity, 9(4), 2002.
- [50] A. Buonanno, Y. Chen, and T. Damour. Transition from inspiral to plunge in precessing binaries of spinning black holes. Phys. Rev. D, 74(10):104005, November 2006.
- [51] A. Buonanno and T. Damour. Effective one-body approach to general relativistic two-body dynamics. Phys. Rev. D, 59(8):084006, April 1999.

- [52] A. Buonanno and T. Damour. Transition from inspiral to plunge in binary black hole coalescences. Phys. Rev. D, 62(6):064015, September 2000.
- [53] A. Buonanno, B. R. Iyer, E. Ochsner, Y. Pan, and B. S. Sathyaprakash. Comparison of post-Newtonian templates for compact binary inspiral signals in gravitational-wave detectors. Phys. Rev. D, 80(8):084043, October 2009.
- [54] M. Burgay, N. D’Amico, A. Possenti, R. N. Manchester, A. G. Lyne, B. C. Joshi, M. A. McLaughlin, M. Kramer, J. M. Sarkissian, F. Camilo, V. Kalogera, C. Kim, and D. R. Lorimer. An increased estimate of the merger rate of double neutron stars from observations of a highly relativistic system. Nature, 426:531–533, December 2003.
- [55] N. Cornish, L. Sampson, N. Yunes, and F. Pretorius. Gravitational wave tests of general relativity with the parameterized post-Einsteinian framework. Phys. Rev. D, 84(6):062003, September 2011.
- [56] Curt Cutler and Éanna E. Flanagan. Gravitational waves from merging compact binaries: How accurately can one extract the binary’s parameters from the inspiral waveform? Phys. Rev. D, 49:2658–2697, Mar 1994.
- [57] T. Damour. Coalescence of two spinning black holes: An effective one-body approach. Phys. Rev. D, 64(12):124013, December 2001.
- [58] T. Damour and G. Esposito-Farèse. Gravitational-wave versus binary-pulsar tests of strong-field gravity. Phys. Rev. D, 58(4):042001, August 1998.
- [59] T. Damour, P. Jaranowski, and G. Schäfer. Determination of the last stable orbit for circular general relativistic binaries at the third post-Newtonian approximation. Phys. Rev. D, 62(8):084011, October 2000.
- [60] K. et al. Danzmann. Lisa: Laser interferometer space antenna for gravitational wave measurements. Class. Quantum Grav., 13:A247–A250, 1996.
- [61] C. de Rham, G. Gabadadze, and A. J. Tolley. Resummation of Massive Gravity. Phys. Rev. Lett., 106(23):231101, June 2011.
- [62] W. Del Pozzo, K. Grover, I. Mandel, and A. Vecchio. Testing general relativity with compact coalescing binaries: comparing exact and predictive methods to compute the Bayes factor. Class. Quantum Grav., 31(20):205006, October 2014.
- [63] 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. Phys. Rev. D, 83(8):082002, Apr 2011.

- [64] A. T. Deller, J. P. W. Verbiest, S. J. Tingay, and M. Bailes. Extremely High Precision VLBI Astrometry of PSR J0437-4715 and Implications for Theories of Gravity. *Astrophys. J. Lett.*, 685:L67–L70, September 2008.
- [65] P. B. Demorest, R. D. Ferdman, M. E. Gonzalez, D. Nice, S. Ransom, I. H. Stairs, Z. Arzoumanian, A. Brazier, S. Burke-Spolaor, S. J. Chamberlin, J. M. Cordes, J. Ellis, L. S. Finn, P. Freire, S. Giampanis, F. Jenet, V. M. Kaspi, J. Lazio, A. N. Lommen, M. McLaughlin, N. Palliyaguru, D. Perrodin, R. M. Shannon, X. Siemens, D. Stinebring, J. Swiggum, and W. W. Zhu. Limits on the Stochastic Gravitational Wave Background from the North American Nanohertz Observatory for Gravitational Waves. *Astrophys. J.*, 762:94, January 2013.
- [66] P. B. Demorest, R. D. Ferdman, M. E. Gonzalez, D. Nice, S. Ransom, I. H. Stairs, Z. Arzoumanian, A. Brazier, S. Burke-Spolaor, S. J. Chamberlin, J. M. Cordes, J. Ellis, L. S. Finn, P. Freire, S. Giampanis, F. Jenet, V. M. Kaspi, J. Lazio, A. N. Lommen, M. McLaughlin, N. Palliyaguru, D. Perrodin, R. M. Shannon, X. Siemens, D. Stinebring, J. Swiggum, and W. W. Zhu. Limits on the Stochastic Gravitational Wave Background from the North American Nanohertz Observatory for Gravitational Waves. *Astrophys. J.*, 762:94, January 2013.
- [67] Riccardo DeSalvo. Path-finding towards a cryogenic interferometer for ligo. *Classical and Quantum Gravity*, 19(7):2021, 2002.
- [68] S. Detweiler. Pulsar timing measurements and the search for gravitational waves. *Astrophys. J.*, 234:1100–1104, December 1979.
- [69] M. Dotti, M. Volonteri, A. Perego, M. Colpi, M. Ruzkowski, and F. Haardt. Dual black holes in merger remnants - II. Spin evolution and gravitational recoil. *Mon. Not. Royal Astron. Soc.*, 402:682–690, February 2010.
- [70] J. C. Driggers, J. Harms, and R. X. Adhikari. Subtraction of Newtonian noise using optimized sensor arrays. *Phys. Rev. D*, 86(10):102001, November 2012.
- [71] J. A. Ellis, F. A. Jenet, and M. A. McLaughlin. Practical Methods for Continuous Gravitational Wave Detection Using Pulsar Timing Data. *Astrophys. J.*, 753:96, July 2012.
- [72] J. A. Ellis, X. Siemens, and J. D. E. Creighton. Optimal Strategies for Continuous Gravitational Wave Detection in Pulsar Timing Arrays. *Astrophys. J.*, 756:175, September 2012.
- [73] F. B. Estabrook and H. D. Wahlquist. Response of Doppler spacecraft tracking to gravitational radiation. *General Relativity and Gravitation*, 6:439–447, October 1975.

- [74] S. Fairhurst. Triangulation of gravitational wave sources with a network of detectors. New Journal of Physics, 11(12):123006, December 2009.
- [75] S. Fairhurst. Source localization with an advanced gravitational wave detector network. Class. Quantum Gravity, 28(10):105021, May 2011.
- [76] S. Fairhurst. In Preparation , 2013.
- [77] R. D. Ferdman, R. van Haasteren, C. G. Bassa, M. Burgay, I. Cognard, A. Corongiu, N. D’Amico, G. Desvignes, J. W. T. Hessels, G. H. Janssen, A. Jessner, C. Jordan, R. Karuppusamy, E. F. Keane, M. Kramer, K. Lazaridis, Y. Levin, A. G. Lyne, M. Pilia, A. Possenti, M. Purver, B. Stappers, S. Sanidas, R. Smits, and G. Theureau. The European Pulsar Timing Array: current efforts and a LEAP toward the future. Classical and Quantum Gravity, 27(8):084014, April 2010.
- [78] Éanna É Flanagan and Scott A Hughes. The basics of gravitational wave theory. New Journal of Physics, 7(1):204, 2005.
- [79] Andreas Freise and Kenneth Strain. Interferometer techniques for gravitational-wave detection. Living Rev. Relativity, 13(2), 09 2010.
- [80] J. R. Gair, C. Li, and I. Mandel. Observable properties of orbits in exact bumpy spacetimes. Phys. Rev. D, 77(2):024035, January 2008.
- [81] C. F. Gammie, S. L. Shapiro, and J. C. McKinney. Black Hole Spin Evolution. Astrophys. J., 602:312–319, February 2004.
- [82] W.W.R. Gilks, S. Richardson, and D.J. Spiegelhalter. Markov Chain Monte Carlo in Practice. Interdisciplinary Statistics Series. Chapman and Hall, 1996.
- [83] S. Gossan, J. Veitch, and B. S. Sathyaprakash. Bayesian model selection for testing the no-hair theorem with black hole ringdowns. Phys. Rev. D, 85(12):124056, June 2012.
- [84] H. Grote and LIGO Scientific Collaboration. The status of GEO 600. Classical and Quantum Gravity, 25(11):114043, June 2008.
- [85] H. Grote and LIGO Scientific Collaboration. The GEO 600 status. Classical and Quantum Gravity, 27(8):084003, April 2010.
- [86] K. Grover, S. Fairhurst, B. F. Farr, I. Mandel, C. Rodriguez, T. Sidery, and A. Vecchio. Comparison of gravitational wave detector network sky localization approximations. Phys. Rev. D, 89(4):042004, February 2014.

- [87] G. M. Harry and the LIGO Scientific Collaboration. Advanced LIGO: the next generation of gravitational wave detectors. Class. Quantum Gravity, 27(8):084006, April 2010.
- [88] R. W. Hellings and G. S. Downs. Upper limits on the isotropic gravitational radiation background from pulsar timing analysis. Astrophys. J. Lett., 265:L39–L42, February 1983.
- [89] C.W. Helstrom. Statistical Theory of Signal Detection, volume 9 of International Series of Monographs in Electronics and Instrumentation. Pergamon Press, Oxford; New York, 2nd edition, 1968.
- [90] G. Hobbs. The Parkes Pulsar Timing Array. Classical and Quantum Gravity, 30(22):224007, November 2013.
- [91] G. Hobbs, A. Archibald, Z. Arzoumanian, D. Backer, M. Bailes, N. D. R. Bhat, M. Burgay, S. Burke-Spolaor, D. Champion, I. Cognard, W. Coles, J. Cordes, P. Demorest, G. Desvignes, R. D. Ferdman, L. Finn, P. Freire, M. Gonzalez, J. Hessels, A. Hotan, G. Janssen, F. Jenet, A. Jessner, C. Jordan, V. Kaspi, M. Kramer, V. Kondratiev, J. Lazio, K. Lazaridis, K. J. Lee, Y. Levin, A. Lommen, D. Lorimer, R. Lynch, A. Lyne, R. Manchester, M. McLaughlin, D. Nice, S. Osłowski, M. Pilia, A. Possenti, M. Purver, S. Ransom, J. Reynolds, S. Sanidas, J. Sarkissian, A. Sesana, R. Shannon, X. Siemens, I. Stairs, B. Stappers, D. Stinebring, G. Theureau, R. van Haasteren, W. van Straten, J. P. W. Verbiest, D. R. B. Yardley, and X. P. You. The International Pulsar Timing Array project: using pulsars as a gravitational wave detector. Classical and Quantum Gravity, 27(8):084013, April 2010.
- [92] D. W. Hogg. Distance measures in cosmology. ArXiv Astrophysics e-prints, May 1999.
- [93] S. A. Hughes. (Sort of) Testing relativity with extreme mass ratio inspirals. In S. M. Merkowitz and J. C. Livas, editors, Laser Interferometer Space Antenna: 6th International LISA Symposium, volume 873 of American Institute of Physics Conference Series, pages 233–240, November 2006.
- [94] S. A. Hughes and R. D. Blandford. Black Hole Mass and Spin Coevolution by Mergers. Astrophys. J. Lett., 585:L101–L104, March 2003.
- [95] Scott A. Hughes and Kip S. Thorne. Seismic gravity-gradient noise in interferometric gravitational-wave detectors. Phys. Rev. D, 58:122002, Nov 1998.
- [96] R. A. Hulse and J. H. Taylor. Discovery of a pulsar in a binary system. Astrophys. J., 195(2):L51–L53, Jan 1975.

- [97] K.S. Thorne in. Three Hundred Years of Gravitation. Cambridge University Press, 1987.
- [98] A. H. Jaffe and D. C. Backer. Gravitational Waves Probe the Coalescence Rate of Massive Black Hole Binaries. Astrophys. J., 583:616–631, February 2003.
- [99] F. Jenet, L. S. Finn, J. Lazio, A. Lommen, M. McLaughlin, I. Stairs, D. Stinebring, J. Verbiest, A. Archibald, Z. Arzoumanian, D. Backer, J. Cordes, P. Demorest, R. Ferdman, P. Freire, M. Gonzalez, V. Kaspi, V. Kondratiev, D. Lorimer, R. Lynch, D. Nice, S. Ransom, R. Shannon, and X. Siemens. The North American Nanohertz Observatory for Gravitational Waves. ArXiv e-prints, September 2009.
- [100] F. A. Jenet, G. B. Hobbs, W. van Straten, R. N. Manchester, M. Bailes, J. P. W. Verbiest, R. T. Edwards, A. W. Hotan, J. M. Sarkissian, and S. M. Ord. Upper Bounds on the Low-Frequency Stochastic Gravitational Wave Background from Pulsar Timing Observations: Current Limits and Future Prospects. Astrophys. J., 653:1571–1576, December 2006.
- [101] F. A. Jenet, A. Lommen, S. L. Larson, and L. Wen. Constraining the Properties of Supermassive Black Hole Systems Using Pulsar Timing: Application to 3C 66B. Astrophys. J., 606:799–803, May 2004.
- [102] V. Kalogera, K. Belczynski, C. Kim, R. O’Shaughnessy, and B. Willems. Formation of double compact objects. Phys. Rep., 442:75–108, April 2007.
- [103] I. Kamaretsos, M. Hannam, S. Husa, and B. S. Sathyaprakash. Black-hole hair loss: Learning about binary progenitors from ringdown signals. Phys. Rev. D, 85(2):024018, January 2012.
- [104] M. M. Kasliwal and S. Nissanke. On Discovering Electromagnetic Emission from Neutron Star Mergers: The Early Years of Two Gravitational Wave Detectors. Astrophys. J. Lett., 789:L5, 2014.
- [105] D. Keppel and P. Ajith. Constraining the mass of the graviton using coalescing black-hole binaries. Phys. Rev. D, 82(12):122001, December 2010.
- [106] L. E. Kidder. Coalescing binary systems of compact objects to (post)^{5/2}-Newtonian order. V. Spin effects. Phys. Rev. D, 52:821–847, July 1995.
- [107] L. E. Kidder, C. M. Will, and A. G. Wiseman. Spin effects in the inspiral of coalescing compact binaries. Phys. Rev. D, 47:4183, May 1993.

- [108] S. M. Koushiappas and A. R. Zentner. Testing Models of Supermassive Black Hole Seed Formation through Gravity Waves. *Astrophys. J.*, 639:7–22, March 2006.
- [109] M. Kramer, I. H. Stairs, R. N. Manchester, M. A. McLaughlin, A. G. Lyne, R. D. Ferdman, M. Burgay, D. R. Lorimer, A. Possenti, N. D’Amico, J. M. Sarkissian, G. B. Hobbs, J. E. Reynolds, P. C. C. Freire, and F. Camilo. Tests of General Relativity from Timing the Double Pulsar. *Science*, 314:97–102, October 2006.
- [110] K. J. Lee, N. Wex, M. Kramer, B. W. Stappers, C. G. Bassa, G. H. Janssen, R. Karuppusamy, and R. Smits. Gravitational wave astronomy of single sources with a pulsar timing array. *Mon. Not. Royal Astron. Soc.*, 414:3251–3264, July 2011.
- [111] L. Lentati, S. R. Taylor, C. M. F. Mingarelli, A. Sesana, S. A. Sanidas, A. Vecchio, R. N. Caballero, K. J. Lee, R. van Haasteren, S. Babak, C. G. Bassa, P. Brem, M. Burgay, D. J. Champion, I. Cognard, G. Desvignes, J. R. Gair, L. Guillemot, J. W. T. Hessels, G. H. Janssen, R. Karuppusamy, M. Kramer, A. Lassus, P. Lazarus, K. Liu, S. Osłowski, D. Perrodin, A. Petiteau, A. Possenti, M. B. Purver, P. A. Rosado, R. Smits, B. Stappers, G. Theureau, C. Tiburzi, and J. P. W. Verbiest. European Pulsar Timing Array limits on an isotropic stochastic gravitational-wave background. *Mon. Not. Royal Astron. Soc.*, 453:2576–2598, November 2015.
- [112] T. G. F. Li, W. Del Pozzo, S. Vitale, C. Van Den Broeck, M. Agathos, J. Veitch, K. Grover, T. Sidery, R. Sturani, and A. Vecchio. Towards a generic test of the strong field dynamics of general relativity using compact binary coalescence. *Phys. Rev. D*, 85(8):082003, April 2012.
- [113] T. G. F. Li, W. Del Pozzo, S. Vitale, C. Van Den Broeck, M. Agathos, J. Veitch, K. Grover, T. Sidery, R. Sturani, and A. Vecchio. Towards a generic test of the strong field dynamics of general relativity using compact binary coalescence: Further investigations. *Journal of Physics Conference Series*, 363(1):012028, June 2012.
- [114] T.G.F. Li, W.Del Pozzo, S. Vitale, C.Van Den Broeck, J.Veitch, K.Grover, T.Sidery, R.Sturani, and A. Vecchio. A generic test of the strong field dynamics of general relativity using compact binary coalescence. *In Preparation*, 2011.
- [115] LIGO Scientific Collaboration, Virgo Collaboration, J. Aasi, J. Abadie, B. P. Abbott, R. Abbott, T. D. Abbott, M. Abernathy, T. Accadia, F. Acernese, and et al. Prospects for Localization of Gravitational Wave Transients by the Advanced LIGO and Advanced Virgo Observatories. *ArXiv e-prints*, April 2013.

- [116] K. Liu, J. P. W. Verbiest, M. Kramer, B. W. Stappers, W. van Straten, and J. M. Cordes. Prospects for high-precision pulsar timing. Mon. Not. Royal Astron. Soc., 417:2916–2926, November 2011.
- [117] A. N. Lommen and D. C. Backer. Using Pulsars to Detect Massive Black Hole Binaries via Gravitational Radiation: Sagittarius A* and Nearby Galaxies. Astrophys. J., 562:297–302, November 2001.
- [118] Duncan R. Lorimer. Binary and millisecond pulsars. Living Rev. Relativity, 11(8), 2008.
- [119] LSC Algorithm Library software. <http://www.lsc-group.phys.uwm.edu/1a1>.
- [120] A. G. Lyne, M. Burgay, M. Kramer, A. Possenti, R. N. Manchester, F. Camilo, M. A. McLaughlin, D. R. Lorimer, N. D’Amico, B. C. Joshi, J. Reynolds, and P. C. C. Freire. A Double-Pulsar System: A Rare Laboratory for Relativistic Gravity and Plasma Physics. Science, 303:1153–1157, February 2004.
- [121] M. Maggiore. Gravitational wave experiments and early universe cosmology. Phys. Rep., 331:283–367, July 2000.
- [122] M. Maggiore. Gravitational Waves: Theory and experiments. Gravitational Waves. Oxford University Press, 2007.
- [123] R. K. Malbon, C. M. Baugh, C. S. Frenk, and C. G. Lacey. Black hole growth in hierarchical galaxy formation. Mon. Not. Royal Astron. Soc., 382:1394–1414, December 2007.
- [124] I. Mandel, C. P. L. Berry, F. Ohme, S. Fairhurst, and W. M. Farr. Parameter estimation on compact binary coalescences with abruptly terminating gravitational waveforms. Classical and Quantum Gravity, 31(15):155005, August 2014.
- [125] I. Mandel, W. Farr, S. Fairhurst, and C. Berry. The fisher information matrix for a gravitational waveform that abruptly terminates in the domain of integration. In Preparation, 2014.
- [126] I. Mandel and R. O’Shaughnessy. Compact binary coalescences in the band of ground-based gravitational-wave detectors. Class. Quantum Gravity, 27(11):114007–+, June 2010.
- [127] I. Mandel and R. O’Shaughnessy. Compact binary coalescences in the band of ground-based gravitational-wave detectors. Classical and Quantum Gravity, 27(11):114007, June 2010.

- [128] D. Merritt and R. D. Ekers. Tracing Black Hole Mergers Through Radio Lobe Morphology. Science, 297:1310–1313, August 2002.
- [129] B. D. Metzger and E. Berger. What is the Most Promising Electromagnetic Counterpart of a Neutron Star Binary Merger? The Astrophysical Journal, 746:48, February 2012.
- [130] C. M. F. Mingarelli, K. Grover, T. Sidery, R. J. E. Smith, and A. Vecchio. Observing the Dynamics of Supermassive Black Hole Binaries with Pulsar Timing Arrays. Physical Review Letters, 109(8):081104, August 2012.
- [131] Chandra Kant Mishra, K. G. Arun, Bala R. Iyer, and B. S. Sathyaprakash. Parametrized tests of post-newtonian theory using advanced ligo and einstein telescope. Phys. Rev. D, 82:064010, Sep 2010.
- [132] C. W. Misner, K. S. Thorne, and J. A. Wheeler. Gravitation. W.H. Freeman and Co., 1973.
- [133] H. Mukhopadhyay, N. Sago, H. Tagoshi, S.V. Dhurandhar, H. Takahashi, and N. Kanda. Detecting gravitational waves from inspiraling binaries with a network of detectors: Coherent versus coincident strategies. Phys. Rev. D, 74, 2006.
- [134] S. Nissanke, M. Kasliwal, and A. Georgieva. Identifying Elusive Electromagnetic Counterparts to Gravitational Wave Mergers: An End-to-end Simulation. Astrophys. J., 767:124, April 2013.
- [135] C. D Ott. TOPICAL REVIEW: The gravitational-wave signature of core-collapse supernovae. Classical and Quantum Gravity, 26(6):063001, March 2009.
- [136] Benjamin J. Owen. Maximum elastic deformations of compact stars with exotic equations of state. Phys. Rev. Lett., 95:211101, Nov 2005.
- [137] A. Perego, M. Dotti, M. Colpi, and M. Volonteri. Mass and spin co-evolution during the alignment of a black hole in a warped accretion disc. Mon. Not. Royal Astron. Soc., 399:2249–2263, November 2009.
- [138] P. C. Peters. Gravitational Radiation and the Motion of Two Point Masses. Physical Review, 136:1224–1232, November 1964.
- [139] Matthew Pitkin, Stuart Reid, Sheila Rowan, and Jim Hough. Gravitational wave detection by interferometry (ground and space). Living Rev. Relativity, 14(5), 2011.
- [140] D. Psaltis. Probes and Tests of Strong-Field Gravity with Observations in the Electromagnetic Spectrum. Living Rev. Relativity, 11:9, November 2008.

- [141] M. Punturo and et al. The einstein telescope: a third-generation gravitational wave observatory. Classical and Quantum Gravity, 27(19):194002, 2010.
- [142] M. Rajagopal and R. W. Romani. Ultra-Low-Frequency Gravitational Radiation from Massive Black Hole Binaries. Astrophys. J., 446:543, June 1995.
- [143] T. Regimbau and J. A. de Freitas Pacheco. Cosmic background of gravitational waves from rotating neutron stars. Astron. Astrophys., 376:381–385, September 2001.
- [144] K. Riles. Searches for continuous gravitational waves with the LIGO and Virgo detector. In van Leeuwen, J., editor, IAU Symposium, volume 291 of IAU Symposium, pages 477–479, 2013.
- [145] C. L. Rodriguez, B. Farr, W. M. Farr, and I. Mandel. Inadequacies of the Fisher Information Matrix in gravitational-wave parameter estimation. Phys. Rev. D, 88(8):084013, 2013.
- [146] C. L. Rodriguez, B. Farr, V. Raymond, W. M. Farr, T. Littenberg, D. Fazi, and V. Kalogera. Basic Parameter Estimation of Binary Neutron Star Systems by the Advanced LIGO/Virgo Network. Astrophys. J., 784:119, 2014.
- [147] Christian Röver, Renate Meyer, and Nelson Christensen. Coherent bayesian inference on compact binary inspirals using a network of interferometric gravitational wave detectors. Phys. Rev. D, 75:062004, Mar 2007.
- [148] L. Sampson, N. Cornish, and N. Yunes. Mismodeling in gravitational-wave astronomy: The trouble with templates. Phys. Rev. D, 89(6):064037, March 2014.
- [149] L. Santamaría, F. Ohme, P. Ajith, B. Brügmann, N. Dorband, M. Hannam, S. Husa, P. Mösta, D. Pollney, C. Reisswig, E. L. Robinson, J. Seiler, and B. Krishnan. Matching post-Newtonian and numerical relativity waveforms: Systematic errors and a new phenomenological model for nonprecessing black hole binaries. Phys. Rev. D, 82(6):064016, September 2010.
- [150] B. S. Sathyaprakash. Scientific benefits of ligo-india. LSC internal report, G1100991, 2011.
- [151] B. S. Sathyaprakash and S. V. Dhurandhar. Choice of filters for the detection of gravitational waves from coalescing binaries. Phys. Rev. D, 44:3819–3834, December 1991.
- [152] B. S. Sathyaprakash and B. F. Schutz. Physics, Astrophysics and Cosmology with Gravitational Waves. Living Rev. Relativity, 12:2–+, March 2009.

- [153] M. V. Sazhin. Opportunities for detecting ultralong gravitational waves. Soviet Astronomy, 22:36–38, February 1978.
- [154] P. D. Scharre and C. M. Will. Testing scalar-tensor gravity using space gravitational-wave interferometers. Phys. Rev. D, 65(4):042002, February 2002.
- [155] B. F. Schutz and F. Ricci. Gravitational waves, sources and detectors. ArXiv e-prints, May 2010.
- [156] A. Sesana. A Practical Guide to the Massive Black Hole Cosmic History. Advances in Astronomy, 2012, 2012.
- [157] A. Sesana, F. Haardt, P. Madau, and M. Volonteri. Low-Frequency Gravitational Radiation from Coalescing Massive Black Hole Binaries in Hierarchical Cosmologies. Astrophys. J., 611:623–632, August 2004.
- [158] A. Sesana, C. Roedig, M. T. Reynolds, and M. Dotti. Multimessenger astronomy with pulsar timing and X-ray observations of massive black hole binaries. Mon. Not. Royal Astron. Soc., 420:860–877, February 2012.
- [159] A. Sesana, A. Vecchio, and C. N. Colacino. The stochastic gravitational-wave background from massive black hole binary systems: implications for observations with Pulsar Timing Arrays. Mon. Not. Royal Astron. Soc., 390:192–209, October 2008.
- [160] A. Sesana, A. Vecchio, and M. Volonteri. Gravitational waves from resolvable massive black hole binary systems and observations with pulsar timing arrays. Mon. Not. Royal Astron. Soc., 394(4):2255–2265, 2009.
- [161] Alberto Sesana and Alberto Vecchio. Measuring the parameters of massive black hole binary systems with pulsar timing array observations of gravitational waves. Phys. Rev. D, 81(10):104008, May 2010.
- [162] R. M. Shannon, V. Ravi, W. A. Coles, G. Hobbs, M. J. Keith, R. N. Manchester, J. S. B. Wyithe, M. Bailes, N. D. R. Bhat, S. Burke-Spolaor, J. Khoo, Y. Levin, S. Osłowski, J. M. Sarkissian, W. van Straten, J. P. W. Verbiest, and J.-B. Will. Gravitational-wave limits from pulsar timing constrain supermassive black hole evolution. Science, 342:334–337, October 2013.
- [163] D. Shoemaker. Advanced ligo anticipated sensitivity curves. LIGO Document T0900288-v3, The LIGO Scientific Collaboration, 2009.
- [164] T. Sidery, B. Aylott, N. Christensen, B. Farr, W. Farr, F. Feroz, J. Gair, K. Grover, P. Graff, C. Hanna, V. Kalogera, I. Mandel, R. O’Shaughnessy, M. Pitkin, L. Price, V. Raymond, C. Roever, L. Singer, M. Van der Sluys, R. J. E.

Smith, A. Vecchio, J. Veitch, and S. Vitale. Reconstructing the sky location of gravitational-wave detected compact binary systems: methodology for testing and comparison. Phys. Rev. D, 89(8):084060, 2014.

- [165] T. Sidery, J. Gair, I. Mandel, and W. Farr. In Preparation, 2013.
- [166] L. P. Singer, L. R. Price, B. Farr, A. L. Urban, C. Pankow, S. Vitale, J. Veitch, W. M. Farr, C. Hanna, K. Cannon, T. Downes, P. Graff, C.-J. Haster, I. Mandel, T. Sidery, and A. Vecchio. The First Two Years of Electromagnetic Follow-up with Advanced LIGO and Virgo. Astrophys. J., 795:105, November 2014.
- [167] J. Skilling. Nested Sampling. In R. Fischer, R. Preuss, and U. V. Toussaint, editors, American Institute of Physics Conference Series, volume 735 of American Institute of Physics Conference Series, pages 395–405, November 2004.
- [168] J. Skilling. Nested Sampling for Bayesian Computations. In World Meeting on Bayesian Statistics, 2006.
- [169] R. Smits, S. J. Tingay, N. Wex, M. Kramer, and B. Stappers. Prospects for accurate distance measurements of pulsars with the Square Kilometre Array: Enabling fundamental physics. Astron. Astrophys., 528:A108, April 2011.
- [170] C. F. Sopuerta and N. Yunes. Extreme- and intermediate-mass ratio inspirals in dynamical Chern-Simons modified gravity. Phys. Rev. D, 80(6):064006, September 2009.
- [171] V. Springel and L. Hernquist. The history of star formation in a Λ cold dark matter universe. Mon. Not. Royal Astron. Soc., 339:312–334, February 2003.
- [172] I. H. Stairs. Testing General Relativity with Pulsar Timing. Living Rev. Relativity, 6:5, September 2003.
- [173] A. Stavridis and C. M. Will. Bounding the mass of the graviton with gravitational waves: Effect of spin precessions in massive black hole binaries. Phys. Rev. D, 80(4):044002, August 2009.
- [174] R. Sturani, S. Fischetti, L. Cadonati, G. M. Guidi, J. Healy, D. Shoemaker, and A. Viceré. Complete phenomenological gravitational waveforms from spinning coalescing binaries. Journal of Physics Conference Series, 243(1):012007, August 2010.
- [175] R. Sturani, S. Fischetti, L. Cadonati, G. M. Guidi, J. Healy, D. Shoemaker, and A. Viceré. Phenomenological gravitational waveforms from spinning coalescing binaries. ArXiv e-prints, December 2010.

- [176] T. Tanaka, K. Menou, and Z. Haiman. Electromagnetic counterparts of supermassive black hole binaries resolved by pulsar timing arrays. Mon. Not. Royal Astron. Soc., 420:705–719, February 2012.
- [177] The LIGO Scientific Collaboration, J. Aasi, B. P. Abbott, R. Abbott, T. Abbott, M. R. Abernathy, K. Ackley, C. Adams, T. Adams, P. Addesso, and et al. Advanced LIGO. Classical and Quantum Gravity, 32(7):074001, April 2015.
- [178] The LIGO Scientific Collaboration and The Virgo Collaboration. Sensitivity Achieved by the LIGO and Virgo Gravitational Wave Detectors during LIGO’s Sixth and Virgo’s Second and Third Science Runs. ArXiv e-prints, March 2012.
- [179] The Virgo Collaboration. Advanced virgo baseline design. Technical Report VIR-0027A-09, Virgo Technical Report, 2009.
- [180] M. Vallisneri. Use and abuse of the Fisher information matrix in the assessment of gravitational-wave parameter-estimation prospects. Phys. Rev. D, 77(4):042001, February 2008.
- [181] M. Vallisneri. Testing general relativity with gravitational waves: A reality check. Phys. Rev. D, 86(8):082001, October 2012.
- [182] M. Vallisneri and N. Yunes. Stealth bias in gravitational-wave parameter estimation. Phys. Rev. D, 87(10):102002, May 2013.
- [183] H. van Dam and M. Veltman. Massive and mass-less Yang-Mills and gravitational fields. Nuclear Physics B, 22:397–411, September 1970.
- [184] C. Van Den Broeck and A. S. Sengupta. Binary black hole spectroscopy. Classical and Quantum Gravity, 24:1089–1113, 2007.
- [185] C. Van Den Broeck and A. S. Sengupta. Phenomenology of amplitude-corrected post-Newtonian gravitational waveforms for compact binary inspiral: I. Signal-to-noise ratios. Classical and Quantum Gravity, 24:155–176, January 2007.
- [186] M. van der Sluys, V. Raymond, I. Mandel, C. Röver, N. Christensen, V. Kalogera, R. Meyer, and A. Vecchio. Parameter estimation of spinning binary inspirals using markov chain monte carlo. Class. Quantum Gravity, 25(18):184011, September 2008.
- [187] R. van Haasteren, Y. Levin, G. H. Janssen, K. Lazaridis, M. Kramer, B. W. Stappers, G. Desvignes, M. B. Purver, A. G. Lyne, R. D. Ferdman, A. Jessner, I. Cognard, G. Theureau, N. D’Amico, A. Possenti, M. Burgay, A. Corongiu, J. W. T. Hessels, R. Smits, and J. P. W. Verbiest. Placing limits on the stochastic

gravitational-wave background using European Pulsar Timing Array data. Mon. Not. Royal Astron. Soc., 414:3117–3128, July 2011.

- [188] R. van Haasteren, Y. Levin, G. H. Janssen, K. Lazaridis, M. Kramer, B. W. Stappers, G. Desvignes, M. B. Purver, A. G. Lyne, R. D. Ferdman, A. Jessner, I. Cognard, G. Theureau, N. D’Amico, A. Possenti, M. Burgay, A. Corongiu, J. W. T. Hessels, R. Smits, and J. P. W. Verbiest. Placing limits on the stochastic gravitational-wave background using European Pulsar Timing Array data. Mon. Not. Royal Astron. Soc., 414:3117–3128, July 2011.
- [189] J. Veitch, I. Mandel, B. Aylott, B. Farr, V. Raymond, C. Rodriguez, M. van der Sluys, V. Kalogera, and A. Vecchio. Estimating parameters of coalescing compact binaries with proposed advanced detector networks. Phys. Rev. D, 85:104045, May 2012.
- [190] J. Veitch, V. Raymond, B. Farr, W. Farr, P. Graff, S. Vitale, B. Aylott, K. Blackburn, N. Christensen, M. Coughlin, W. Del Pozzo, F. Feroz, J. Gair, C.-J. Haster, V. Kalogera, T. Littenberg, I. Mandel, R. O’Shaughnessy, M. Pitkin, C. Rodriguez, C. Röver, T. Sidery, R. Smith, M. Van Der Sluys, A. Vecchio, W. Vousden, and L. Wade. Parameter estimation for compact binaries with ground-based gravitational-wave observations using the LALInference software library. Phys. Rev. D, 91(4):042003, February 2015.
- [191] J. Veitch and A. Vecchio. Assigning confidence to inspiral gravitational wave candidates with Bayesian model selection. Classical and Quantum Gravity, 25(18):184010, September 2008.
- [192] J. Veitch and A. Vecchio. Bayesian approach to the follow-up of candidate gravitational wave signals. Phys. Rev. D, 78(2):022001, July 2008.
- [193] J. Veitch and A. Vecchio. Bayesian coherent analysis of in-spiral gravitational wave signals with a detector network. Phys. Rev. D, 81:062003, Mar 2010.
- [194] J. P. W. Verbiest, M. Bailes, N. D. R. Bhat, S. Burke-Spolaor, D. J. Champion, W. Coles, G. B. Hobbs, A. W. Hotan, F. Jenet, J. Khoo, K. J. Lee, A. Lommen, R. N. Manchester, S. Osłowski, J. Reynolds, J. Sarkissian, W. van Straten, D. R. B. Yardley, and X. P. You. Status update of the Parkes pulsar timing array. Classical and Quantum Gravity, 27(8):084015, April 2010.
- [195] J. P. W. Verbiest, M. Bailes, W. A. Coles, G. B. Hobbs, W. van Straten, D. J. Champion, F. A. Jenet, R. N. Manchester, N. D. R. Bhat, J. M. Sarkissian, D. Yardley, S. Burke-Spolaor, A. W. Hotan, and X. P. You. Timing stability of millisecond pulsars and prospects for gravitational-wave detection. Mon. Not. Royal Astron. Soc., 400:951–968, December 2009.

- [196] S. Vitale and W. Del Pozzo. How serious can the stealth bias be in gravitational wave parameter estimation? Phys. Rev. D, 89(2):022002, January 2014.
- [197] S. Vitale and M. Zanolin. Parameter estimation from gravitational waves generated by nonspinning binary black holes with laser interferometers: Beyond the Fisher information. Phys. Rev. D, 82(12):124065, December 2010.
- [198] M. Volonteri. The Formation and Evolution of Massive Black Holes. Science, 337:544–, August 2012.
- [199] M. Volonteri, F. Haardt, and P. Madau. The Assembly and Merging History of Supermassive Black Holes in Hierarchical Models of Galaxy Formation. Astrophys. J., 582:559–573, January 2003.
- [200] M. Volonteri, P. Madau, E. Quataert, and M. J. Rees. The Distribution and Cosmic Evolution of Massive Black Hole Spins. Astrophys. J., 620:69–77, February 2005.
- [201] J. M. Weisberg and J. H. Taylor. The Relativistic Binary Pulsar B1913+16. In M. Bailes, D. J. Nice, & S. E. Thorsett, editor, Radio Pulsars, volume 302 of Astronomical Society of the Pacific Conference Series, pages 93–+, 2003.
- [202] Linqing Wen and Yanbei Chen. Geometrical expression for the angular resolution of a network of gravitational-wave detectors. Phys. Rev. D, 81:082001, Apr 2010.
- [203] Z. L. Wen, F. A. Jenet, D. Yardley, G. B. Hobbs, and R. N. Manchester. Constraining the Coalescence Rate of Supermassive Black-hole Binaries Using Pulsar Timing. Astrophys. J., 730:29, March 2011.
- [204] C. M. Will. Bounding the mass of the graviton using gravitational-wave observations of inspiralling compact binaries. Phys. Rev. D, 57:2061–2068, February 1998.
- [205] C. M. Will and N. Yunes. Testing alternative theories of gravity using LISA. Classical and Quantum Gravity, 21:4367–4381, September 2004.
- [206] Clifford M. Will. Testing scalar-tensor gravity with gravitational-wave observations of inspiralling compact binaries. Phys. Rev. D, 50:6058–6067, Nov 1994.
- [207] C.M. Will. The confrontation between general relativity and experiment. Living Rev. Relativity, 9(3), 2006.
- [208] W. Winkler, K. Danzmann, H. Grote, M. Hewitson, S. Hild, J. Hough, H. Lück, M. Malec, A. Freise, K. Mossavi, S. Rowan, A. Rüdiger, R. Schilling, J. R.

Smith, K. A. Strain, H. Ward, and B. Willke. The geo 600 core optics. Optics Communications, 280(2):492, December 2007.

- [209] J. S. B. Wyithe and A. Loeb. Low-Frequency Gravitational Waves from Massive Black Hole Binaries: Predictions for LISA and Pulsar Timing Arrays. Astrophys. J., 590:691–706, June 2003.
- [210] D. R. B. Yardley, G. B. Hobbs, F. A. Jenet, J. P. W. Verbiest, Z. L. Wen, R. N. Manchester, W. A. Coles, W. van Straten, M. Bailes, N. D. R. Bhat, S. Burke-Spolaor, D. J. Champion, A. W. Hotan, and J. M. Sarkissian. The sensitivity of the Parkes Pulsar Timing Array to individual sources of gravitational waves. Mon. Not. Royal Astron. Soc., 407:669–680, September 2010.
- [211] J. Yoo, J. Miralda-Escudé, D. H. Weinberg, Z. Zheng, and C. W. Morgan. The Most Massive Black Holes in the Universe: Effects of Mergers in Massive Galaxy Clusters. Astrophys. J., 667:813–825, October 2007.
- [212] N. Yunes and L. S. Finn. Constraining effective quantum gravity with LISA. Journal of Physics Conference Series, 154(1):012041, March 2009.
- [213] N. Yunes and S. A. Hughes. Binary pulsar constraints on the parametrized post-Einsteinian framework. Phys. Rev. D, 82(8):082002, October 2010.
- [214] N. Yunes, R. O’Shaughnessy, B. J. Owen, and S. Alexander. Testing gravitational parity violation with coincident gravitational waves and short gamma-ray bursts. Phys. Rev. D, 82(6):064017, September 2010.
- [215] N. Yunes, F. Pretorius, and D. Spergel. Constraining the evolutionary history of Newton’s constant with gravitational wave observations. Phys. Rev. D, 81(6):064018, March 2010.
- [216] Nicolás Yunes and Frans Pretorius. Fundamental theoretical bias in gravitational wave astrophysics and the parametrized post-einsteinian framework. Phys. Rev. D, 80:122003, Dec 2009.