Substantiating the void

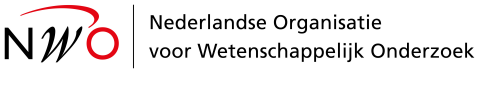
Strong-field tests of general relativity
with the first direct detection of gravitational waves
and
Fast likelihood evaluations for future tests

Jeroen Meidam

Promotiecommissie:

Promotor:	Prof. dr. J.F.J. van den Brand
Copromotor:	Dr. C.F.F. Van Den Broeck
Overige Leden:	Prof. dr. P.J. Mulders
	Prof. dr. B.S. Sathyaprakash
	Prof. dr. F.L. Linde
	Dr. A.P. Colijn
	Dr. J. Veitch

© 2018 Jeroen Meidam
ISBN: 978-94-6332-292-8
Printed in The Netherlands by: GVO printers & designers
Cover design and other cartoons by: Jeroen Meidam | www.behance.net/jmeidam



This work is part of the research program of the Stichting voor Fundamenteel onderzoek der Materie (FOM), which is partially supported by the Nederlandse organisatie voor Wetenschappelijk Onderzoek (NWO). It was carried out at the Nationaal Instituut voor Subatomaire Fysica (Nikhef) in Amsterdam, the Netherlands.

VRIJE UNIVERSITEIT

Substantiating the void

Strong-field tests of general relativity
with the first direct detection of gravitational waves
and
Fast likelihood evaluations for future tests

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. V. Subramaniam,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Bètawetenschappen
op woensdag 7 februari 2018 om 11.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Jeroen Meidam

geboren te Oss

promotor:	prof.dr. J.F.J. van den Brand
copromotor:	dr. C.F.F. Van Den Broeck

Wat is nou de wetenschap?
Volgens de wetenschap stijgt de zeespiegel met 10 cm per eeuw!!
Ik heb het voor de aardigheid eens nagemeten,
ik kwam op een stijging van anderhalve meter in 6 uur.
- Herman Finkers

Contents

Preface

Astrophysics is a fascinating topic of scientific research as it not only combines various physical principles, astrophysical sources also provide us with laboratories that could not conceivably be replicated on Earth. The brightest, heaviest, largest, most powerful systems are observed out there in the Universe. When we observe our Galaxy and any other beyond, the main source of information reaches our pale blue dot by ways of electromagnetic radiation and even subatomic particles. What we observe is the universe as it evolves from diffuse clouds of gas to galaxies containing billions of stars. We can see galaxies merging, planets forming and stars collapsing. Most stars we observe, even by the naked eye are in fact part of binary systems; a star system where two stars revolve around one another under their mutual gravitational influence. As these stars patrol their confined corner of space, they evolve, much the same as any other star; some stars will eventually collapse into a neutron star, balancing the forces of gravity and pressure by yet unknown states of matter, while the more heavy stars will lose this battle and submit under the force of gravity to collapse into a black hole. The binary systems that have evolved to the state where they consist of two black holes can meet a future much like the now famous GW150914; the binary system that ended its evolution as the two black hole components crashed into each other producing the most powerful event ever observed by mankind. In a mere fraction of a second the final orbital revolutions and consecutive collapse of the binary system produced a wave through the fabric of space-time containing the energy equivalent to three times the mass of the sun, leaving a single black hole remnant ringing away the last traces of its fate and the evidence of what transpired. Approximately one and a half billion years later, we observed the gravitational waves produced by this event with the two Advanced LIGO detectors here on Earth. The fabric of space-time itself can twist bend and stretch, and for the first time and certainly not the last, we were capable of observing this.

How the flexibility of the fabric of space-time works and interacts with matter and itself is described by the theory of general relativity, which characterizes gravitational interactions as a purely geometrical effect. In essence, this means that whether it is a planet or photon that follows the curvature of space-time, they both follow it in the same fashion. This core principle of the theory of general relativity is referred to as the equivalence principle. Ever since Einstein published his now celebrated theory of relativity, physicists around the world have set out to test its predictions. This has been done in and around our own solar system ranging from Eddington traveling half across the globe to observe a solar eclipse to satellites sent into orbit to measure how the earth drags the fabric of space along with its rotation. Also observations of stars in the center of our galaxy revolving around the super-massive black hole we expect to find there and pulsars (rotating neutron stars emitting strong radiation in a narrow cone) in binary systems have been used to further support the validity of the theory. The latter provided the first indirect evidence supporting the existence of gravitational waves as the separation between the two neutron star components decreased over time precisely according to the dynamics predicted by the theory of general relativity. Even though general relativity passed all tests to date with unprecedented precision, the complexity and dynamics it has to offer have never been

reached observationally. Systems boasting such complexity are generally hidden from us by matter surrounding them or because they simply do not emit electromagnetic radiation and are dark to us. Concerning the extent to which the theory can make predictions on the behavior of space-time, we can indeed say that merely the shroud of the dark side has fallen. Until September 14th 2015, we were blind concerning the genuinely strong-field dynamics of general relativity. On this day, only just after the first Advanced LIGO observing run began, GW150914 – the signal from an enormously powerful merger of two black holes – raced through the detectors. We could not have wished for a more ideal laboratory to put the thus far unexplored regime of general relativity to the test; black holes are in essence pure space-time phenomenon where only geometry superabounds the dynamics of the system. Indeed, for the first time ever, we were able to substantiate the void itself.

From the beginning of my doctoral studies, I knew that testing general relativity would be at the core. Although the development of TIGER (test infrastructure for general relativity) was in a mature state when I joined the group at Nikhef, I was able to contribute to a lot of extensions and improvements and I am proud to have been part of this. From very early on in my research, the goal was to extend the pipeline to not only work with signals originating from binary neutron stars but also black holes. This direction of research turned out to be right on the money as not long after we started implementing the testing framework to binary black hole sources, the first detection of gravitational waves was a fact and the source was a perfect binary black hole merger. Due to our efforts toward extending TIGER to perform analyses on binary black holes, we were able to quickly set up the tests of general relativity that were later published as one of the companion papers of the discovery. I could not have hoped for a better finale to my time at Nikhef.

During my visits to conferences and collaboration meetings I learned a lot from my colleagues in the LIGO-Virgo Collaboration. The final part of my thesis was sparked mostly thanks to the discussions I had there. The prospect of making a significantly faster TIGER pipeline was exciting. I hope that the work I present on this topic has made a worthy contribution to the development of a quick and successful testing infrastructure that will be able to cope with the many detections we expect to see in the future.

This dissertation is structured as follows: The first chapter provides the reader with the essential background to follow the core of the research presented in the remaining chapters. The second chapter is based on the first publication that came out of this thesis; testing the no-hair theorem using binary black hole systems. Chapter 3 lays down the groundwork for extending our TIGER pipeline to include light binary black hole systems. Chapter 4 introduces the parameterized test of general relativity and presents our finding on the first two detections of gravitational waves. Chapter 5 presents the proof of principle by which we can significantly increase the calculational efficiency of all the tests of general relativity presented in the preceding chapters. The final chapter will summarize our findings and provide a glimpse into the future of testing general relativity using gravitational waves.

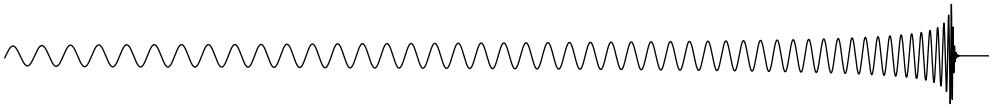
I would like to end this preface with a fun fact: Note the binary black hole merger that is animated in the bottom right corner of this book; this is a simplified

simulation of the final few orbits and merger of GW150914. The masses I used in the simulation are consisted with this event, the radii are calculated from the masses and the decreasing orbital separation and increasing rotation frequency are calculated with a genuine waveform generator. As you read though my dissertation you might also notice the waveform shown at the start of each chapter; this is the gravitational wave shape that would result from the merger shown in the animation if we were to observe it starting from 15 Hz.

Jeroen Meidam, Amsterdam, March 2017

Chapter 1

Foundation



1.1 General relativity

The common thread throughout this thesis is testing the theory of general relativity (GR). This section will give a brief introduction to GR to give the reader an understanding of its basic building blocks. Throughout this work a geometrized unit system will be used where the speed of light in vacuum c and the gravitational constant G are equal to one, *i.e.* $G = c = 1$. We will also be using the Einstein summation convention where repeated indices are summed over *e.g.* if $i = \{1, 2, 3\}$, then $x_i y_i = x_1 y_1 + x_2 y_2 + x_3 y_3$.

The theory of general relativity describes gravitational interactions as a result of curvature in space-time, which in turn is caused by the presence of energy. Space-time is represented by a differentiable manifold where space-time points, also referred to as events, can be labeled using a coordinate system

$$x^\mu = \left(x^0, x^1, x^2, x^3\right), \tag{1.1}$$

which can be given a sense of scale by defining a metric $g_{\mu\nu}$ on the manifold. Once a metric is defined, distances may be calculated between nearby space-time events. Consider for example the Minkowski metric that determines distances on a flat space-time $g_{\mu\nu} = \eta_{\mu\nu} = \mathrm{diag}(-1, 1, 1, 1)$. The distance, or distance element is then defined as

$$\mathrm{d}s^2 = \eta_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu = -\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2. \tag{1.2}$$

In general it not possible to bring the metric $g_{\mu\nu}$ into Minkowski form, but it is possible to do this at a given point, which is then referred to as a local Lorentz frame (LLF). Considering a local Lorentz frame can be used in many applications, but when studying the genuinely strong-field properties of the theory such as encountered around a black hole, a specialized metric describing a curved space-time needs to be derived. The metric $g_{\mu\nu}$, may be transformed into any other coordinate system $g'_{\mu\nu}$ in the following way

$$g_{\mu\nu}(x) \to g'_{\mu\nu} = \frac{\partial x^\rho}{\partial x'^\mu}\frac{\partial x^\sigma}{\partial x'^\nu} g_{\rho\sigma}(x). \tag{1.3}$$

Curvature in general relativity can be quantified by the Riemann tensor

$$R^\alpha{}_{\beta\mu\nu} = \partial_\mu \Gamma^\alpha_{\beta\nu} - \partial_\nu \Gamma^\alpha_{\mu\beta} + \Gamma^\alpha_{\mu\gamma}\Gamma^\gamma_{\beta\nu} - \Gamma^\alpha_{\gamma\nu}\Gamma^\gamma_{\mu\beta}, \tag{1.4}$$

where

$$\Gamma^{\rho}_{\mu\nu} = \frac{1}{2} g^{\rho\sigma} \left(\partial_\mu g_{\sigma\nu} + \partial_\nu g_{\mu\sigma} - \partial_\sigma g_{\mu\nu} \right) \tag{1.5}$$

is called the Christoffel symbol. The Riemann tensor is a measure of how derivatives on a manifold fail to commute. Consider parallel transport of a vector along some path in curved space-time. During parallel transport of vector v^β along a path parameterized by λ,

$$v^\beta u^\alpha_{;\beta} = 0, \tag{1.6}$$

where $u^\mu = \mathrm{d}x^\mu / \mathrm{d}\lambda$ is the tangent along a path parameterized by λ and

$$u^\alpha_{;\beta} \equiv \frac{\partial u^\alpha}{\partial x^\beta} + u^\mu \Gamma^\alpha_{\mu\beta}, \tag{1.7}$$

is the *covariant derivative*, the angle between v^β and u^α may change. This can only happen on a curved manifold. A classic example of how this works on a curved manifold is by considering parallel transport of a vector along a loop on a sphere. After completing the loop, the parallel transported vector $v^{\beta'}$ will not have the same orientation as the original v^β. The degree by how much they differ can be quantified by the Riemann tensor, which is thus a measure of curvature. By contracting indices in the Riemann tensor one can construct the Ricci tensor

$$R_{\alpha\beta} \equiv R^\mu{}_{\alpha\mu\beta}, \tag{1.8}$$

and the Ricci scalar

$$R \equiv R^\alpha_\alpha. \tag{1.9}$$

Finally, an important property of the Riemann tensor is captured by the Bianchi identity

$$R_{\alpha\beta\gamma\delta;\mu} + R_{\alpha\beta\delta\mu;\gamma} + R_{\alpha\beta\mu\gamma;\delta} = 0. \tag{1.10}$$

In Newtonian physics, the separation between two nearby particles in free fall is quantified by the tidal tensor

$$\mathcal{E}^i_j = \frac{\partial^2 \Phi}{\partial x_i \partial x^j}, \tag{1.11}$$

where Φ is the Newtonian potential and the Newtonian field equation can be written in terms of its trace

$$\mathcal{E}^i_i = \partial_i \partial^i \Phi = \nabla^2 \Phi = 4\pi G \rho. \tag{1.12}$$

In 1912 Einstein proposed a relativistic version of the above equation, making use of the Riemann tensor capturing the curving nature of space-time

$$R_{\alpha\beta} = \frac{4\pi G}{c^4} T_{\alpha\beta}, \tag{1.13}$$

where $T_{\alpha\beta}$ is a tensor containing all the energy and momentum within the considered part of space-time. This equation however still has 10 degrees of freedom while there are only 6 independent metric components. By using the Bianchi identities in Eq. 1.10 the Einstein tensor

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R, \tag{1.14}$$

can be constructed and we can define the Einstein field equations (EFE)

$$G_{\mu\nu} \equiv \frac{8\pi G}{c^4} T_{\mu\nu}. \tag{1.15}$$

1.1.1 Linearized general relativity

The Einstein field equations can only be analytically solved for certain choices of metrics. One such an example is the flat space Minkowski metric or the Schwarzschild metric that describes space-time around non-spinning black holes. It is also possible to find solutions for weak gravitational fields, i.e.

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \tag{1.16}$$

where $|h_{\mu\nu}| \ll 1$ is a small perturbation on an otherwise flat Minkowski space-time, so that higher orders of $h_{\mu\nu}$ may be neglected when substituting into the EFE. The EFE are invariant under general coordinate transformations, but this invariance is broken as a result of the choice of background $\eta_{\mu\nu}$. As we have decided on numerical values for the components of the metric, namely $\mathrm{diag}(-1, 1, 1, 1)$ plus some small perturbation, we can only consider a sufficiently large *specific* reference frame where Eq. (1.16) holds. This means that we are now restricted in how much we can change the coordinates; in particular, we are restricted to a limited set of coordinate transformations called "gauge transformations"

$$x^\mu \to x'^\mu = x^\mu + \xi(x^\mu). \tag{1.17}$$

We limit the extent of the transformations so that all derivatives $|\partial_\rho \xi_\mu|$ are at most of the same order as $|h_{\mu\nu}|$ so that terms containing higher orders of this quantity may be neglected. When we substitute the transformation in Eq. (1.17) into the transformation rule of the metric Eq. (1.3) keeping only lowest-order terms, we find

$$h_{\mu\nu}(x) \to h'_{\mu\nu}(x') = h_{\mu\nu}(x) - (\partial_\mu \xi_\nu + \partial_\nu \xi_\mu). \tag{1.18}$$

Using the above gauge freedom and limiting coordinate transformations in such a way that the requirement $|h_{\mu\nu}| \ll 1$ remains satisfied, we can linearize the Einstein equations. First, we find that to leading order in $h_{\mu\nu}$ the Riemann tensor reads

$$R_{\mu\nu\rho\sigma} = \frac{1}{2}(\partial_\nu \partial_\rho h_{\mu\sigma} + \partial_\mu \partial_\sigma h_{\nu\rho} - \partial_\mu \partial_\rho h_{\nu\sigma} - \partial_\nu \partial_\sigma h_{\mu\rho}), \tag{1.19}$$

which is invariant under the gauge transformation defined in Eq. (1.18). Before we substitute the Riemann tensor into the EFE we define

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2} \eta_{\mu\nu} h, \tag{1.20}$$

where $h = \eta^{\mu\nu} h_{\mu\nu}$. Using this definition and its inverse

$$h_{\mu\nu} = \bar{h}_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}\bar{h}, \tag{1.21}$$

it can be shown that the EFE now take the form

$$\Box\bar{h}_{\mu\nu} + \eta_{\mu\nu}\partial^\rho\partial^\sigma\bar{h}_{\rho\sigma} - \partial^\rho\partial_\nu\bar{h}_{\mu\rho} - \partial^\rho\partial_\mu\bar{h}_{\nu\rho} = -\frac{16\pi G}{c^4}T_{\mu\nu}, \tag{1.22}$$

where $\Box \equiv \partial_\mu\partial^\mu$ is the d'Alembertian. The gauge freedom in Eq. (1.17) can now be used to further simplify the EFE by imposing the so-called harmonic gauge

$$\partial^\nu\bar{h}_{\mu\nu} = 0, \tag{1.23}$$

after which the last three terms in the left hand side of Eq. (1.22) vanish and we end up with the linearized Einstein equations

$$\Box\bar{h}_{\mu\nu} = -\frac{16\pi G}{c^4}T_{\mu\nu}. \tag{1.24}$$

In the linearized theory, the general expression of energy-momentum conservation $T^{;\nu}_{\mu\nu} = 0$, simplifies to $\partial^\nu T_{\mu\nu} = 0$.

1.1.2 Gravitational waves

A conceptually important feature of Eq. (1.24) is what happens outside the source, where $T_{\mu\nu} = 0$: In this case, the EFE reduce to $\Box\bar{h}_{\mu\nu} = 0$, or

$$\left(-\frac{1}{c^2}\partial t^2 + \Delta\right)\bar{h}_{\mu\nu} = 0, \tag{1.25}$$

which is a wave equation for waves propagating with the speed of light c. Solutions to this wave equation can be written as superpositions of plane waves traveling with wave vectors $\mathbf{k}$ and frequency $\omega = c|\mathbf{k}|$,

$$h(t) = A_{\mu\nu}\cos\left(\omega t - \mathbf{k}\cdot\mathbf{x}\right). \tag{1.26}$$

This last result illustrates that the gravitational field, or space-time has dynamics of its own, independent of matter; even though some matter distribution must have generated the current solution, it can still exists far away from the source where $T_{\mu\nu} = 0$.

We have just learned that small perturbations on a background gravitational field propagate as waves with the speed of light c. Next, we consider what would cause gravitational waves to be generated. As we wish to learn what can physically generate gravitational waves, we must now include the energy-momentum tensor of matter $T_{\mu\nu}$.

To gain a conceptual understanding of the generation of gravitational waves we start with the linearized EFE from Eq. 1.24. The linear equations can be solved using the retarded Green's function assuming no incoming radiation from infinity. For a given $T_{\mu\nu}$ the solution is

$$\bar{h}_{\mu\nu}(t,\mathbf{x}) = \frac{4\pi G}{c^4}\int \mathrm{d}^3x'\,\frac{1}{|\mathbf{x}-\mathbf{x}'|}T_{\mu\nu}\left(t - \frac{|\mathbf{x}-\mathbf{x}'|}{c}, \mathbf{x}'\right), \tag{1.27}$$

where $\mathbf{x}$ and $\mathbf{x}'$ are any two space-time points.

Recall that in order to arrive at the linearized EFE in Eq. 1.24, we imposed the harmonic gauge. We can utilize an additional gauge freedom by imposing the radiation gauge: $h = 0$, $h_{0i} = 0$. Imposing both the harmonic and radiation gauge allows us to write Eq. (1.27) in the *transverse traceless* (TT) gauge

$$h_{ij}^{\mathrm{TT}}(t, \mathbf{x}) = \frac{4\pi G}{c^4} \Lambda_{ij,kl}(\hat{\mathbf{n}}) \int \mathrm{d}^3 x' \, \frac{1}{|\mathbf{x} - \mathbf{x}'|} T_{kl}\left(t - \frac{|\mathbf{x} - \mathbf{x}'|}{c}, \mathbf{x}'\right), \tag{1.28}$$

where $\mathbf{n}$ is the direction of propagation of the gravitational wave and the projector $\Lambda_{ij,kl}(\hat{\mathbf{n}})$ is given by

$$\Lambda_{ij,kl}(\hat{\mathbf{n}}) = P_{ik} P_{jl} - \frac{1}{2} P_{ij} P_{kl}, \tag{1.29}$$

where

$$P_{ij} \equiv \delta_{ij} - n_i n_j. \tag{1.30}$$

Eq. (1.29) can be used to bring $h_{\mu\nu}$ outside the source into the TT gauge. The perturbation $h_{ij}^{\mathrm{TT}}(t, \mathbf{x})$ is evaluated outside the source at $\mathbf{x}$, while $\mathbf{x}'$ is a point inside the source, *i.e.* $T_{kl}(t - |\mathbf{x} - \mathbf{x}'|/c, \mathbf{x}') \neq 0$.

For the moment we are particularly interested in the behavior of $h_{ij}^{\mathrm{TT}}(t, \mathbf{x})$ far away from the source at a distance r that is much larger than the size of the source d. In that case it is possible to expand $\Delta\mathbf{x} = r - \mathbf{x}' \cdot \hat{n} + \mathcal{O}(d^2/r)$ and Eq. (1.28) can be written as

$$h_{ij}^{\mathrm{TT}}(t, \mathbf{x}) = \frac{4\pi G}{c^4} \Lambda_{ij,kl}(\hat{\mathbf{n}}) \int \mathrm{d}^3 x' \, \frac{1}{|\mathbf{x} - \mathbf{x}'|} T_{kl}\left(t - \frac{r}{c} + \frac{\mathbf{x}' \cdot \hat{n}}{c}, \mathbf{x}'\right). \tag{1.31}$$

Additionally, if the source is non-relativistic *i.e.* $v/c \ll 1$, the source T_{kl} can be expanded as

$$T_{kl}\left(t - \frac{r}{c} + \frac{\mathbf{x}' \cdot \hat{n}}{c}, \mathbf{x}'\right) = T_{kl}(t - \frac{r}{c}, \mathbf{x}') + \frac{x'^i n^i}{c} \partial_0 T_{kl} + \frac{1}{2c^2} x'^i x'^j n^i n^j \partial_0^2 T_{kl} + \ldots, \tag{1.32}$$

where derivatives are evaluated at $(t - r/c, \mathbf{x}')$. Substituting this expansion into Eq. (1.31) we can write

$$\begin{aligned} h_{ij}^{\mathrm{TT}}(t, \mathbf{x}) = & \frac{1}{r} \frac{4\pi G}{c^4} \Lambda_{ij,kl}(\hat{\mathbf{n}}) \Big[S^{kl} + \frac{1}{c} n_m \dot{S}^{kl,m} \\ & + \frac{1}{2c^2} n_m n_p \ddot{S}^{kl,mp} + \ldots \Big]_{\mathrm{ret}}, \end{aligned} \tag{1.33}$$

where $[\ldots]_{\mathrm{ret}}$ means the expression in brackets is evaluated at the retarded time

$t_{\rm ret} = 1 - r/c$, and

$$\begin{aligned} S^{ij} &= \int \mathrm{d}^3x\, T^{ij}(t,\mathbf{x}), \\ S^{ij,k} &= \int \mathrm{d}^3x\, T^{ij}(t,\mathbf{x})x^k, \\ S^{ij,kl} &= \int \mathrm{d}^3x\, T^{ij}(t,\mathbf{x})x^k x^l, \end{aligned} \tag{1.34}$$

are the *multipole moments* of the stress tensor T_{ij}. At this point it is interesting to note that the expansion in multipole moments in Eq. (1.33) is in fact an expansion in v/c. For example, compared to S^{ij}, the multipole moment $S^{ij,k}$ depends on an additional term $x^k \sim \mathcal{O}(d)$ and each time derivative ∂_t introduces a factor proportional to the characteristic frequency of the source $\mathcal{O}(\omega_s)$. When one defines $v \equiv \omega_s d$, we find that the term $\frac{1}{c} n_m \dot{S}^{kl,m}$ can indeed be written as an additional power of v/c compared to S^{kl}. In the same way it can be shown that the term $(1/2c^2) n_m n_p \ddot{S}^{kl,mp}$ is a correction of the order v^2/c^2 and so on for higher order corrections.

For the purpose of this introduction it is convenient to define the momenta of the energy density T^{00}/c^2 known as the mass moments,

$$\begin{aligned} M &= \frac{1}{c^2}\int \mathrm{d}^3x\, T^{00}(t,\mathbf{x}), \\ M^i &= \frac{1}{c^2}\int \mathrm{d}^3x\, T^{00}(t,\mathbf{x})x^i, \\ M^{ij} &= \frac{1}{c^2}\int \mathrm{d}^3x\, T^{00}(t,\mathbf{x})x^i x^j, \\ &\ldots, \end{aligned} \tag{1.35}$$

and the momenta of the momentum density T^{0i}/c

$$\begin{aligned} P^i &= \frac{1}{c}\int \mathrm{d}^3x\, T^{0i}(t,\mathbf{x}), \\ P^{i,j} &= \frac{1}{c}\int \mathrm{d}^3x\, T^{0i}(t,\mathbf{x})x^j, \\ P^{i,jk} &= \frac{1}{c}\int \mathrm{d}^3x\, T^{0i}(t,\mathbf{x})x^j x^k, \\ &\ldots \end{aligned} \tag{1.36}$$

If we now consider energy-momentum conservation within a volume encompassing the source (so that $T^{\mu\nu}$ vanishes at its boundary) and remember we are working in the linearized regime in which $\partial_\nu T^{\mu\nu} = 0$, it is possible to derive a relation between the multipole moments in Eq. (1.34) and the mass moments in Eq. (1.35) reading

$$S^{ij} = \frac{1}{2}\ddot{M}^{ij}. \tag{1.37}$$

Using this relation we can eliminate the multipole moments in Eq. (1.33) in favor of the mass moments: To leading order in v/c, Eq. (1.33) becomes

$$\left[h_{ij}^{\mathrm{TT}}(t,\mathbf{x})\right]_{\mathrm{quad}} = \frac{1}{r}\frac{2\pi G}{c^4}\Lambda_{ij,kl}(\hat{\mathbf{n}})\ddot{M}^{kl}(t-r/c). \tag{1.38}$$

By letting $\hat{\mathbf{n}}$ point in the z direction, *i.e.* $\hat{\mathbf{n}} = \hat{\mathbf{z}}$, the projector in Eq. 1.30 becomes $P = \mathrm{diag}(1,1,0)$ and we can write the mass quadrupole radiation in Eq. (1.38) as

$$\left[h_{ij}^{\mathrm{TT}}(t,\mathbf{x})\right]_{\mathrm{quad}} = \begin{pmatrix} h_+ & h_\times & 0 \\ h_\times & -h_+ & 0 \\ 0 & 0 & 0 \end{pmatrix}_{ij}, \tag{1.39}$$

where h_+ and $h_\times$ are the two gravitational wave polarizations

$$\begin{aligned} h_+ &= \frac{1}{r}\frac{2\pi G}{c^4}\left(\ddot{M}_{11} - \ddot{M}_{22}\right), \\ h_\times &= \frac{2}{r}\frac{2\pi G}{c^4}\ddot{M}_{12}, \end{aligned} \tag{1.40}$$

where again the right hand side is evaluated at the retarded time $t - r/c$.

We have now arrived at expressions with which the gravitational wave polarizations may be calculated to leading order in v/c when the mass quadrupole moments are known, provided we are far away from the source where the background spacetime curvature may be treated as flat and the perturbation caused by the gravitational wave is small, *i.e.* when we can cast the EFE into their linearized form.

For the type of source that is at the heart of this thesis – *compact binaries* – we shall now derive the leading order contribution in v/c to the polarizations using Eq. (1.40).

Consider two compact objects with masses m_1 and m_2 revolving around a common center of mass, where the separation between the two objects R is such that they may be regarded as point particles, *i.e.* R is much larger than the Schwarzschild radius of either object. As the objects revolve around the center of mass with an orbital frequency ω_{orb} they retain a circular orbit. Without loss of generality, the coordinate system can be chosen such that the direction toward the observer is along the z axis. The orbit itself may be tilted with respect to the line of sight at an angle ι. For such a coordinate system it is easy to show that the paths of the objects are described by

$$\begin{aligned} \mathbf{x}_1(t) &= \frac{m_2 R}{M}\left(\cos\omega_{\mathrm{orb}}t, \cos\iota\sin\omega_{\mathrm{orb}}t, -\sin\iota\sin\omega_{\mathrm{orb}}t\right), \\ \mathbf{x}_2(t) &= -\frac{m_1 R}{M}\left(\cos\omega_{\mathrm{orb}}t, \cos\iota\sin\omega_{\mathrm{orb}}t, -\sin\iota\sin\omega_{\mathrm{orb}}t\right), \end{aligned} \tag{1.41}$$

where $M = m_1 + m_2$ is the system's total mass. Using the expressions for $\mathbf{x}_i$ the mass moments M^{ij} and their derivatives can be calculated: T^{00} is the energy density of the system consisting of two point particles, hence

$$T^{00} = m_1\delta^3(\mathbf{x}-\mathbf{x}_1) + m_2\delta^3(\mathbf{x}-\mathbf{x}_2), \tag{1.42}$$

and thus

$$\begin{aligned} M^{11} &= M\eta R^2 \cos^2 \omega_{\rm orb} t, \\ M^{22} &= M\eta R^2 \cos^2 \iota \sin^2 \omega_{\rm orb} t, \\ M^{21} &= M\eta R^2 \cos \iota \cos \omega_{\rm orb} t \sin \omega_{\rm orb} t, \end{aligned} \tag{1.43}$$

and their second time derivatives

$$\begin{aligned} \ddot{M}^{11} &= -2M\eta R^2 \omega_{\rm orb}^2 \cos 2\omega_{\rm orb} t, \\ \ddot{M}^{22} &= 2M\eta R^2 \omega_{\rm orb}^2 \cos^2 \iota \cos^2 2\omega_{\rm orb} t, \\ \ddot{M}^{21} &= -2M\eta R^2 \omega_{\rm orb}^2 \cos \iota \sin 2\omega_{\rm orb} t, \end{aligned} \tag{1.44}$$

where

$$\eta = \frac{m_1 m_2}{M^2}, \tag{1.45}$$

is the *symmetric mass ratio*. We then substitute the expressions in Eq. (1.44) into Eq. (1.40) to find the plus and cross polarizations for this system

$$\begin{aligned} h_+(t) &= -\frac{2}{r} M\eta R^2 \omega_{\rm orb}^2 \left(1 + \cos^2 \iota\right) \cos 2\omega_{\rm orb} t_{\rm ret}, \\ h_\times(t) &= -\frac{4}{r} M\eta R^2 \omega_{\rm orb}^2 \cos \iota \sin 2\omega_{\rm orb} t_{\rm ret}. \end{aligned} \tag{1.46}$$

If the objects are sufficiently far apart, we may write the separation R in terms of the orbital frequency $\omega_{\rm orb}$ using Kepler's law

$$\omega_{\rm orb}^2 = \frac{M}{R^3}, \tag{1.47}$$

and we find

$$\begin{aligned} h_+(t) &= -\frac{2}{r} \mathcal{M}_c^{5/3} \eta \omega_{\rm orb}^{2/3} \left(1 + \cos^2 \iota\right) \cos 2\omega_{\rm orb} t_{\rm ret}, \\ h_\times(t) &= -\frac{4}{r} \mathcal{M}_c^{5/3} \eta \omega_{\rm orb}^{2/3} \cos \iota \sin 2\omega_{\rm orb} t_{\rm ret}, \end{aligned} \tag{1.48}$$

where

$$\mathcal{M}_c = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}}, \tag{1.49}$$

is the *chirp mass*.

Lastly, it is of interest to consider the gravitational wave power emitted by the point-particle system introduced above. The power emitted per unit solid angle Ω is given by

$$\frac{{\rm d}P_{\rm gw}}{{\rm d}\Omega} = \frac{r^2 c^3}{16\pi G} \left\langle \dot{h}_+^2 + \dot{h}_\times^2 \right\rangle. \tag{1.50}$$

Substituting Eq. (1.48) and integrating over the sphere yields the total radiated power

$$P_{\rm gw} = \frac{32}{5}\frac{c^5}{G}\left(\frac{G\mathcal{M}_c\pi f_{\rm gw}}{c^3}\right)^{10/3}. \tag{1.51}$$

In this section we have linearized the EFE which demonstrated that space-time by itself has dynamics regardless of sources being present. We then considered a source $T_{\mu\nu}$ and derived the resulting small perturbation $h_{\mu\nu}$ on a Minkowski background $\eta_{\mu\nu}$ far away from the source to leading order in characteristic velocity v/c. From the resulting expression we were able to derive the two gravitational wave polarizations h_+ and $h_\times$ also to leading order in v/c as a function of the mass moments in Eq. (1.35). Finally we have provided the reader with a specific example source of gravitational waves; a system of two compact objects revolving around each other. We will further elaborate on this class of systems in Sec. 1.4 by considering what happens to the system as it loses energy by radiating gravitational waves and we will discuss the construction of much more advanced waveform models. Now let us consider how gravitational waves influence matter far away from the source.

1.1.3 Gravitational waves acting on matter

The most illustrative way to discuss the effect gravitational waves have on matter is by considering two neighboring free-falling particles at $x^\mu(\tau)$ and $x^\mu(\tau)+\zeta^\mu(\tau)$, where τ is proper time. Free-falling test particles obey the geodesic equation

$$\frac{{\rm d}^2x^\mu}{{\rm d}\tau^2} + \Gamma^\mu_{\nu\rho}(x)\frac{{\rm d}x^\nu}{{\rm d}\tau}\frac{{\rm d}x^\rho}{{\rm d}\tau} = 0. \tag{1.52}$$

We can write down the geodesic equation for both particles and consider their difference expanded to leading order in ζ^μ

$$\frac{{\rm d}^2\zeta^\mu}{{\rm d}\tau^2} + 2\Gamma^\mu_{\nu\rho}(x)\frac{{\rm d}x^\nu}{{\rm d}\tau}\frac{{\rm d}\zeta^\rho}{{\rm d}\tau} + \zeta^\sigma\partial_\sigma\Gamma^\mu_{\nu\rho}(x)\frac{{\rm d}x^\nu}{{\rm d}\tau}\frac{{\rm d}x^\rho}{{\rm d}\tau} = 0. \tag{1.53}$$

Using the covariant derivative of a vector the vector field ζ^μ along a curve $x^\mu(\tau)$

$$\frac{{\rm D}\zeta^\mu}{{\rm D}\tau} = \frac{{\rm d}\zeta^\mu}{{\rm d}\tau} + \Gamma^\mu_{\nu\rho}\zeta^\nu\frac{{\rm d}x^\rho}{{\rm d}\tau}, \tag{1.54}$$

and the definition of the Riemann tensor in Eq. (1.4), we can write the above as

$$\frac{{\rm D}^2\zeta^\mu}{{\rm D}\tau^2} = -R^\mu{}_{\nu\rho\sigma}\zeta^\rho\frac{{\rm d}x^\nu}{{\rm d}\tau}\frac{{\rm d}x^\sigma}{{\rm d}\tau}, \tag{1.55}$$

which is called the equation of geodesic deviation and it describes the relative motion between two nearby particles on some background curvature. By performing a coordinate transformation into a local Lorentz frame and assuming the particles move non-relativistically so that their motion ${\rm d}x^i/{\rm d}\tau \ll {\rm d}x^0/{\rm d}\tau$ we can simplify Eq. (1.55) into

$$\frac{{\rm D}^2\zeta^i}{{\rm D}\tau^2} = -R^i{}_{0j0}\zeta^j\left(\frac{{\rm d}x^0}{{\rm d}\tau}\right)^2. \tag{1.56}$$

Since we are now considering non-relativistic velocities of the two particles $\mathrm{d}x^0/\mathrm{d}\tau \simeq c$ and the proper time becomes the time associated with the flat background space-time, *i.e.* $\tau = t$. This allows us to write

$$\ddot{\zeta}^i = -c^2 R^i{}_{0j0}\zeta^j. \tag{1.57}$$

The components of the Riemann tensor may be calculated in any frame due to its invariance in linearized theory, and we compute them in the TT frame. This allows us to relate the geodesic deviation to the perturbation h_{ij} caused by a passing gravitational wave;

$$\ddot{\zeta}^i = \frac{1}{2}\ddot{h}^{\mathrm{TT}}_{ij}\zeta^j. \tag{1.58}$$

It is now possible to calculate how the separation ζ^i between two nearby particles in the (x, y) plane changes when a gravitational wave traveling in the z direction passes by. To study the effect each individual polarization has on matter, first consider the $+$ polarization, which is the only component of h^{TT}_{ij} if the plane of orbital motion is edge-on, *i.e.* $\iota = \pi/2$. In other words, we consider a passing gravitational wave of the form

$$h^{\mathrm{TT}}_{ij}(t) = -\frac{2}{r}\mathcal{M}_c^{5/3}\eta\omega^{2/3}\cos 2\omega t \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}_{ij}, \tag{1.59}$$

where for brevity I dropped the $()_{\mathrm{orb}}$ subscript. Now take $\zeta^i = (x_0 + \delta x(t), y_0 + \delta y(t), 0)$, where (x_0, y_0) is the original unperturbed position of the particle and $(\delta x(t), \delta y(t))$ its displacement caused by the gravitational wave. If the particles are assumed to stay on nearby geodesics and the displacement $(\delta x(t), \delta y(t))$ is sufficiently small compared to the original position, one can derive the following from Eq. (1.57);

$$\begin{aligned} \delta x(t) &= -\mathcal{A}_+\omega^{2/3}\cos(2\omega t)x_0, \\ \delta y(t) &= \mathcal{A}_+\omega^{2/3}\cos(2\omega t)y_0, \end{aligned} \tag{1.60}$$

where $\mathcal{A}_+ = -\frac{2}{r}\mathcal{M}_c^{5/3}\eta$ for the example compact binary source. In a similar fashion the effect of the $\times$ polarization can be calculated;

$$\begin{aligned} \delta x(t) &= \mathcal{A}_\times\omega^{2/3}\cos(2\omega t)x_0, \\ \delta y(t) &= \mathcal{A}_\times\omega^{2/3}\cos(2\omega t)y_0, \end{aligned} \tag{1.61}$$

where $\mathcal{A}_\times = -\frac{2}{r}\mathcal{M}_c^{5/3}\eta$ for the same source. The associated effect on a ring of particles is illustrated in Fig. 1.1. Note that this example is only to illustrate the effect the plus and cross polarizations have by themselves; in reality, unlike the cross polarization, the plus polarization can never be fully suppressed no matter how the binary is oriented due to the $(1 + \cos^2\iota)$ contribution.

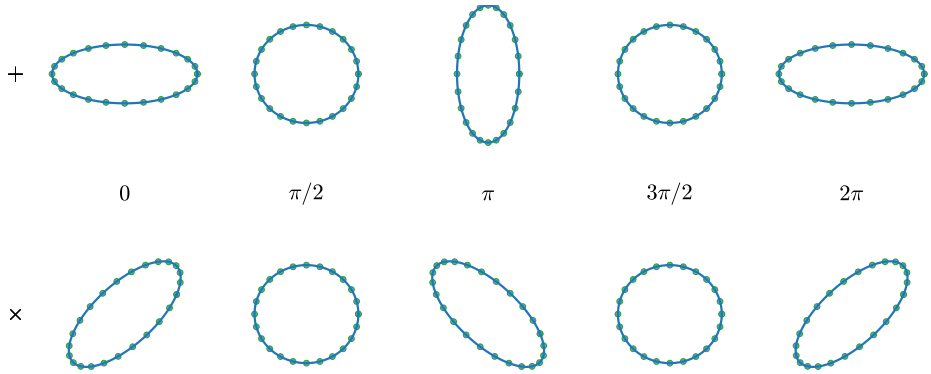


Figure 1.1: An (exaggerated) illustration of the effect a passing gravitational wave traveling into the page has on a ring of particles. The top figure shows the effect of the + polarization as a function of gravitational wave phase $\phi = [0, 2\pi]$ and the bottom figure shows the effect of the × polarization.

1.2 Tests of general relativity

Not long after the publication of Einstein's theory of general relativity, the first tests of the theory were performed. Einstein himself demonstrated how the theory accounts for the anomalous perihelion precession of the planet Mercury. Shortly after that, Eddington set up an expedition which confirmed the predictions made about the deflection of light around a massive object by observing the deflection of starlight around the sun during a solar eclipse. In the following decades the theory of general relativity has passed many more tests with exceptional precision. These tests involved the confirmation of predicted frame dragging in Earth's gravitational field caused by the rotation of the planet; the Shapiro time delay predicting a redshift in electromagnetic waves passing through a gravitational field; and the orbital decay in the Hulse-Taylor binary predicted by the emission of gravitational waves, to name a few.

Even though the theory of general relativity stands strong as ever after passing the observational tests above, only a small region of the full extent of the theory has been covered. All tests above were performed for systems that are here referred to as weak-field and/or stationary systems as opposed to the highly relativistic strong-field dynamical systems that the theory can also describe. To get an idea of what is meant by this, it is useful to follow the notation from e.g. [1] and consider a system's characteristic curvature $\mathcal{R} = M/\mathcal{L}^3$ and characteristic time scale $|\mathcal{T}| = |E_b/\dot{E}_b|$, where M and $\mathcal{L}$ are the characteristic mass and size of the system respectively and E_b is the characteristic gravitational binding energy. In Fig. 1.2 we placed a few sources in the $(1/|\mathcal{T}|, \mathcal{R})$ plane and note that as a system moves more toward the right it means that its characteristic time scale is shorter while if a system is placed more toward the top it boasts a stronger gravitational field. The points depicted in this figure are that

of four systems with which general relativity has been tested so far: The LAGEOS satellite; the perihelion precession of Mercury; the distance between Earth and the Moon and orbital decay of binary pulsars.

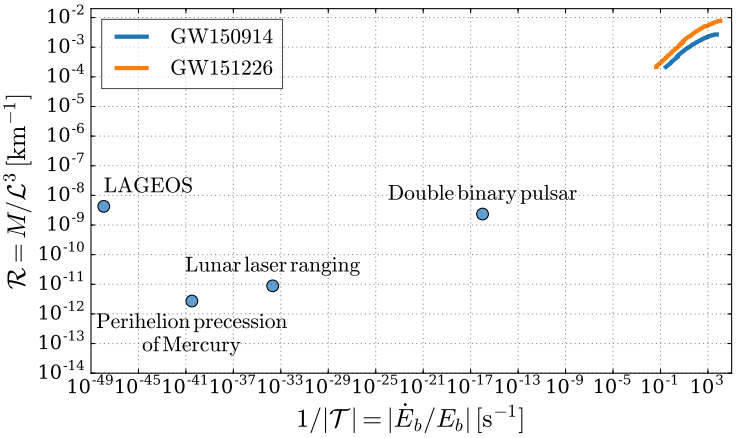


Figure 1.2: Indication of the gravitational field strength and dynamics of a few systems that have been used to test the theory of general relativity [1]. The y-axis shows the characteristic curvature $\mathcal{R} = M/\mathcal{L}^3$ and the x-axis the characteristic time scale $1/|\mathcal{T}| = |\dot{E}_b/E_b|$. The figure demonstrates that the two gravitational wave events GW150914 and GW151226 boast highly strong-field dynamical systems even compared to binary pulsars.

Even though there is no clear distinction between what one calls a strong-field dynamical system or a weak-field slow moving system, we can consider the most dynamical strong-field system in this plot that has been used to test GR prior to the detection of gravitational waves – the binary pulsar with $M = 2.59\,\mathrm{M}_\odot$ and $\mathcal{L} = 8.7 \times 10^8\,\mathrm{m}$ [2] – and see how it compares to the recent detections made by the Advanced LIGO detectors. First of all, in Fig. 1.2 the events GW150914 and GW151226 which will be introduced in Sec. 1.7, are many orders of magnitude separated from the double pulsar system; they reach velocities near the speed of light with $v/c > 0.5$ as opposed to $v/c \sim 0.001$ for the double pulsar and their gravitational potential is $M/R > 0.2$ vs. $M/R \sim 10^{-6}$. Second, the two gravitational wave events sweep from weaker to strong fields in a matter of milliseconds justifying the term "dynamical". The strong-field dynamical aspects also become apparent in the description of the orbital dynamics of the systems: The double pulsar can be accurately approximated by leading order expansion in v/c, whereas for the gravitational wave events, since $v/c \sim \mathcal{O}(1)$, this is no longer accurate and higher order terms are required. In fact, as will be discussed in Sec. 1.4, at some point the black holes in the gravitational wave events merged and the simplified signal we considered so far changes completely. This is where the theory of general relativity enters the *genuinely strong-field dynamics* and is a regime that has so far been completely unexplored by observations as the only means of accessing

this regime is through the detection of gravitational waves.

1.3 Detecting gravitational waves

In the LIGO-Virgo Collaboration we make use of large interferometric detectors for the search of gravitational waves. These detectors are typically on the order of a few kilometers in size and are sensitive between tens of Hz and a few kHz. Fig. 1.1 demonstrates graphically what happens to particles arranged in a ring as a gravitational wave passes by. The gravitational wave (GW) interferometers essentially consist of three free-falling particles, creating an L-shaped arrangement. By looking at the plus-polarization in Fig. 1.1 and keeping the L-shaped detector in mind, it can be seen that as the gravitational wave passes by, one arm gets shorter while the other gets longer, depending on the point in time one considers. Now imagine these free-falling points consist of a beam-splitter in the middle and mirrors at the end points making up a simple Michelson interferometer as depicted in Fig. 1.3. By analyzing the interference pattern of the laser beam as it made its way across the two arms simultaneously and recombines in the photodetector, it is possible to detect differences in the lengths of the detector arms.

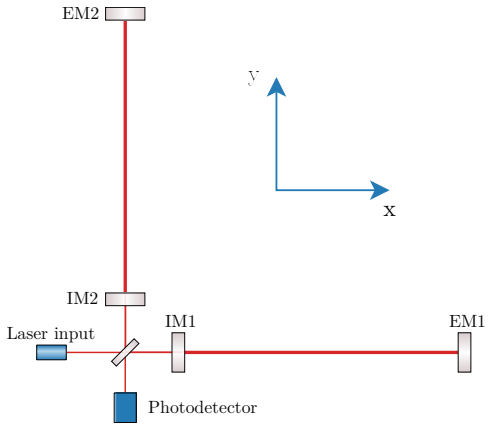


Figure 1.3: Schematic representation of a basic L-shaped gravitational wave interferometer. The tilted center mirror is the beam splitter which sends the laser into the two detector arms which each consist of an input mirror (IM) and end mirror (EM).

This is of course a simplified picture of how an actual gravitational wave interferometer works. In reality there is at least one crucial difference in the general set-up of the interferometer: Additional to the end mirrors (EM) there is a mirror at the start of each arm called the input mirror (IM) to let the laser bounce back and forth inside the cavity of each arm to build up power (also shown in Fig. 1.3). Aside from the many technological challenges in filtering out any undesired sources of vibration that could distort a true GW signal, this simplified picture of a Michelson interferometer with two resonant cavities to increase power is our tool to observe the gravitational

universe.

One can derive the response of an interferometric GW detector, by considering the limit in which the wavelength is much larger than the size L of the detector. In terms of the GW frequency this means $\omega_{\rm GW} L \ll 1$. For most astrophysical sources and certainly the ones considered in this thesis this assumption is more than justified. In making this assumption we can say that the mirrors and beam splitter are on nearby geodesics where the length of the arm as a function of time evolves according to Eq. (1.58)

$$\ddot{\zeta}^i_A = \frac{1}{2}\ddot{h}^{\rm TT}_{ij}\zeta^j_A, \tag{1.62}$$

where we now added an additional index $A = 1, 2$ to label the mirrors.

Let us first assume that the detector is in the (x, y) plane with the beam splitter at the origin and the end-mirrors at $(L, 0, 0)$ and $(0, L, 0)$ when no gravitational wave is present. In this case we are only interested in the motion in the x and y direction of mirror $A = 1$ and $A = 2$ respectively. For the separation we have $\zeta^i_A = L + \delta\zeta^i_A$, where $i = x, y$. Then when using for simplicity a monochromatic wave $h_{ij} = A_{ij}\cos(\omega t)$ and considering small displacements $\delta\zeta^i_A \ll L$ we find

$$\begin{aligned}\delta\zeta^x_1 &= \frac{1}{2}A_{xx}\cos(\omega t) = \frac{1}{2}h_{xx},\\ \delta\zeta^y_2 &= \frac{1}{2}A_{yy}\cos(\omega t) = \frac{1}{2}h_{yy}.\end{aligned} \tag{1.63}$$

Effectively, the detector output, also called *strain*, is the difference between the arm lengths in the x and y directions;

$$h(t) = (L + \delta\zeta^x_1) - (L + \delta\zeta^y_2) = \frac{1}{2}\left(h_{xx} - h_{yy}\right). \tag{1.64}$$

Consider now the general case where the detector is placed at an arbitrary location on the ground with an arbitrary orientation. First we drop the assumption that the detector arms are perpendicular to each other, which is achieved by letting the arms point along unit vectors $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$. In that case the strain in Eq. (1.64) can be written as

$$h(t) = \frac{1}{2}\left(h_{uu} - h_{vv}\right) \equiv D^{ij}h_{ij}, \tag{1.65}$$

where

$$D^{ij} = \frac{1}{2}(u^i u^j - v^i v^j), \tag{1.66}$$

is called the detector tensor.

To fully generalize our expression for the strain, we also need to take into account that the signal can come from any direction. The detector frame is described in the (x, y, z) right-handed coordinate system. The coordinate system of the source (x', y', z') is such that h_{ij} is in the TT frame:

$$h^{\hat{\mathbf{n}}=\hat{\mathbf{z}}'}_{ij} = \begin{pmatrix} h_+ & h_\times & 0\\ h_\times & -h_+ & 0\\ 0 & 0 & 0\end{pmatrix}_{ij}, \tag{1.67}$$

where $\hat{\mathbf{n}}$ is the direction of propagation. To write the strain in the general case one therefore needs to apply linear transformations to rotate the source frame (x', y', z') into the detector frame;

$$h_{ij}^{\text{det}} = \mathcal{R}_{ik}\mathcal{R}_{jl}h_{ij}^{\hat{\mathbf{n}}=\hat{\mathbf{z}}'}, \tag{1.68}$$

where $\mathcal{R}_{ij}$ are the appropriate rotation matrices. We have seen that the projection of h_{ij} onto the detector arms with Eq. (1.65) is also a linear operation and one can therefore write the general expression for the strain $h(t)$ as

$$h(t) = D^{ij}h_{ij}^{\text{det}} = D^{ij}\mathcal{R}_{ik}\mathcal{R}_{jl}h_{ij}^{\hat{\mathbf{n}}=\hat{\mathbf{z}}'} = F_+h_+ + F_\times h_\times, \tag{1.69}$$

where F_+ and $F_\times$ are called the antenna pattern functions, which for an L-shaped detector can be written as

$$\begin{aligned} F_+(\theta,\phi,\psi) &= \frac{1}{2}(1+\cos^2\theta)\cos 2\phi\cos 2\psi - \cos\theta\sin 2\phi\sin 2\psi, \\ F_\times(\theta,\phi,\psi) &= \frac{1}{2}(1+\cos^2\theta)\cos 2\phi\sin 2\psi + \cos\theta\sin 2\phi\cos 2\psi. \end{aligned} \tag{1.70}$$

In the antenna pattern functions θ and ϕ determine the sky position of the source with respect to a coordinate system adapted to the detector and ψ is the rotation of the y' axis with respect to a vertical line on the sky (line from zenith to nadir).

1.3.1 Extracting signals from noise

In reality the detector's output is not simply $h(t)$, but a combination of the signal and noise $n(t)$. The main contributions to the noise in ground-based interferometric detectors are *thermal noise* from thermal vibrations of atoms in the optics of the detector, *shot noise* due to the fact that light has particle as well as wave properties and *seismic noise* ranging from trucks driving past to ocean waves hitting the coast. The final raw signal measured with a detector is called the strain $s(t)$ and consists of noise $n(t)$ and possibly a signal $h(t)$; $s(t) = n(t) + h(t)$. The sources considered in this work are all short – in the order of seconds – and we know what they look like, should GR be correct. In this case it is possible to find a signal $h(t)$ hidden in noise by integrating it against the strain and dividing by observation time T

$$\frac{1}{T}\int_0^T \mathrm{d}t\, s(t)h(t) = \frac{1}{T}\int_0^T \mathrm{d}t\, n(t)h(t) + \frac{1}{T}\int_0^T \mathrm{d}t\, h(t)^2. \tag{1.71}$$

To get a feel for how well one might dig a certain signal out of the noise, it is useful to consider the following approximation to Eq. (1.71);

$$\frac{1}{T}\int_0^T \mathrm{d}t\, s(t)h(t) \sim \sqrt{\frac{\tau_0}{T}}n_0h_0 + h_0^2, \tag{1.72}$$

where h_0 and n_0 are the characteristic amplitudes of the signal and noise respectively and τ_0 is the characteristic period of the signal. In the limit of $T \to \infty$ the first term in the right hand side of Eq. (1.72) goes to zero and one is left with only the

contribution of the amplitude of the signal. In practice T is limited by the length of the signals unless a continuous wave source is analyzed. For the sources we are currently interested in, T is at most in the order of minutes (the time a coalescing binary can be seen within the frequency bandwidth of the detector). By comparing the two terms in the right hand side of Eq. (1.72) we see that one does not need $h_0 > n_0$ in order to find the signal; instead one requires $h_0 > n_0\sqrt{\tau_0/T}$. Coalescing compact binaries, which will be introduced in Sec. 1.4 have a characteristic frequency of $\sim 100\,\mathrm{Hz}$, or $\tau_0 \sim 0.01\,\mathrm{s}$. When we assume an observation time of 100 s, this means that the characteristic amplitude $h_0 \gtrsim 10^{-2} n_0$. This basic approximation already indicates that one can potentially dig quite deep in the data.

In practice we do not simply integrate the detector output against the signal, instead the signal is weighted by the noise *power spectral density* (PSD) $S_n(f)$ – which describes in which way the noise is correlated with itself – and the integration takes place in the frequency domain. The motivation for this approach is that it allows us to maximize the signal-to-noise ratio (SNR). Let us briefly discuss how this comes about. First, instead of integrating the strain against a particular signal, consider a yet to be determined *filter* $K(t)$ and define

$$\hat{s} \equiv \int_{-\infty}^{\infty} \mathrm{d}t\, s(t) K(t). \tag{1.73}$$

The SNR is defined as S/N, where $S = \langle \hat{s} \rangle$ if the signal is present and N is its root-mean-squared value if the signal is absent, *i.e.* $N = \left[\langle \hat{s}^2 \rangle - \langle \hat{s} \rangle^2 \right]_{h=0}^{1/2}$. In the frequency domain the SNR can be written as

$$\frac{S}{N} = \frac{\int_{-\infty}^{\infty} \mathrm{d}f\, \tilde{h}(f) \tilde{K}^*(f)}{\left[\int_{-\infty}^{\infty} \mathrm{d}f\, \frac{1}{2} S_n(f) |\tilde{K}(f)|^2 \right]^{1/2}}, \tag{1.74}$$

where a tilde denotes the Fourier transform, an asterisk a complex conjugation. To arrive at Eq. (1.74) from the definition of $\hat{s}$ we used the stationarity $\langle n(t) \rangle = 0$ of the noise and by assuming Gaussianity, $S_n(f)$ is defined implicitly through

$$\langle \tilde{n}^*(f) \tilde{n}(f') \rangle = \frac{1}{2} \delta(f - f') S_n(f), \tag{1.75}$$

and has units of Hz^{-1}. It is now possible to find a $K(t)$ that maximizes the SNR. In order to do so, we first introduce the *noise-weighted inner product*

$$\begin{aligned} (a|b) &\equiv \Re \int_{-\infty}^{\infty} \mathrm{d}f\, \frac{\tilde{a}^*(f) \tilde{b}(f)}{\frac{1}{2} S_n(f)}, \\ &= 4\Re \int_{0}^{\infty} \mathrm{d}f\, \frac{\tilde{a}^*(f) \tilde{b}(f)}{S_n(f)}. \end{aligned} \tag{1.76}$$

Using this definition the SNR can be written as

$$\frac{S}{N} = \frac{(\mathcal{K}|h)}{(\mathcal{K}|\mathcal{K})^{1/2}}, \tag{1.77}$$

where

$$\mathcal{K} = \frac{1}{2} S_n(f) \tilde{K}(f). \tag{1.78}$$

In this notation, the SNR can be interpreted as the inner product of the signal waveform with the unit vector $\mathcal{K}$, suggesting that in order to optimize the SNR, $\mathcal{K}$ needs to point in the direction of $\tilde{h}(f)$. This is achieved by choosing the filter's Fourier transform to be the so-called *Wiener filter*

$$\tilde{K}(f) \propto \frac{\tilde{h}(f)}{S_n(f)}. \tag{1.79}$$

Using this filter, the optimal SNR becomes

$$\frac{S}{N} = (h|h)^{1/2} = 4 \int_0^\infty \mathrm{d}f \, \frac{|h(f)|^2}{S_n(f)}. \tag{1.80}$$

It is now clear that the optimal filter does not only depend on the signal itself, but also on the properties of the noise; at frequencies where the noise is stronger, the filter is suppressed and at frequencies with weaker noise, the filter receives more weight.

In practice the optimal SNR is never reached. For the type of signal considered in this thesis, we know which family of waveforms to use, but the signal may be any waveform from a continuous set covering the entire parameter space. Consider for example a system of two black holes. These may have a wide variety in component masses and spins, which all lead to different waveforms. It is therefore necessary to consider a wide range in filters $\{K(f;\boldsymbol{\theta})\}$ that cover the desired parameter space, each of which has different values for $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_N\}$. Out of necessity, the amount of filters considered has to be finite and in order to cover the continuous parameter space, filters are distributed across it. The resulting collection of filters is called a *template bank* and the waveforms associated to each point are called *template waveforms* or simply templates.

1.3.2 Detector noise

How well a gravitational wave detector performs is characterized by the power spectral density (PSD) of its noise background. The noise PSD of a detector can be constructed by considering the dimensionless strain $s(t) = n(t)$, *i.e.* the strain in the absence of a signal. To discuss the typical shape of a PSD from a large ground-based interferometric gravitational wave detector, we will continue considering only stationary Gaussian noise for now. The one-sided noise PSD $S_n(f)$ in that case is the Fourier transform of the noise autocorrelation function $\kappa(\tau) = \langle n(t)n(t+\tau)\rangle$, where $\tau \equiv |t_1 - t_2|$. The PSD is called one-sided since it is only defined for positive frequencies

$$\begin{aligned} S_n(f) &\equiv \frac{1}{2} \int_{-\infty}^{\infty} \mathrm{d}\tau \, \kappa(\tau) e^{2\pi \mathrm{i} f \tau} \quad \text{for} \quad f \geq 0, \\ S_n(f) &\equiv 0 \quad \text{for} \quad f < 0. \end{aligned} \tag{1.81}$$

It can be shown that by using the Fourier transform of $n(t)$ in the autocorrelation function κ and substituting the result into Eq. (1.81), that one obtains Eq. (1.75). The latter justifies the name PSD as it describes how the power in the noise is distributed over frequency. Even though $S_n(f)$ has dimension of time, conventionally Hz^{-1} is used since the PSD is defined in the frequency domain. Often, the *noise amplitude* $\sqrt{S_n(f)}$ is presented in figures rather than the PSD itself, as will be done in this chapter.

Fig. 1.4 shows the noise amplitude $\sqrt{S_n(f)}$ for a few interferometric detectors that will be introduced shortly in Sec. 1.3.3 and 1.3.4. Note that the shape of all PSDs are similar. Each of the curves shown in Fig. 1.4 has a steeply decreasing slope as frequency increases, and increases again after a few hundred Hz. The region in the middle where the detectors are at their most sensitive is often referred to as the *bucket*. All the curves shown here are analytical predictions based on various modeled contributions of noise. The most important sources of noise are the thermal vibrations of atoms in the optics of the interferometer (thermal noise), seismic noise and shot noise due to the fact that light has a particle as well as a wave character. The three sources constitute the bulk of the shape of the PSDs shown in Fig. 1.4.

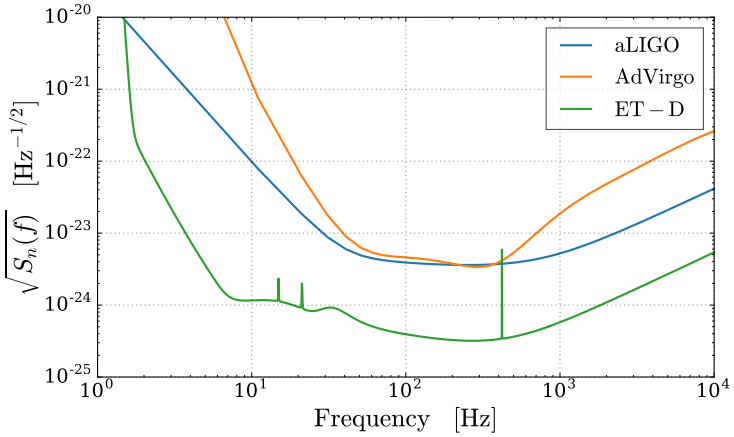


Figure 1.4: The anticipated noise amplitude $\sqrt{S_n(f)}$ as a function of frequency for the two advanced ground-based interferometers Advanced LIGO and Advanced Virgo and the third generation ground-based observatory ET. Shown for Advanced LIGO is the amplitude of the so-called "zero detuned, high power spectral density" [3] and is an analytic prediction of the PSD at final design sensitivity. Also for Advanced Virgo, the PSD at final design sensitivity is shown. Currently "ET-D" corresponds to the most detailed assessment of the possible noise budget of Einstein Telescope [4].

So far we limited the discussion to stationary noise, *i.e.* the expectation value of the noise is independent of time. In reality this is not the case and the data stream coming from a detector may contain *glitches*; short bursts of power, not related to GW events. It is not always a trivial task to label a signal as a glitch, but detector characterization

pipelines analyzing hundreds of different diagnostic channels are effective in properly recognizing glitches and allow us to exclude (veto) them from our analyses. There are 5 veto categories ranked from high severity (CAT1) to the lowest (CAT5) and these are described in great detail in [5]. For completeness the veto categories are briefly described as

CAT1 The detectors were not taking data in their design configuration.

CAT2 Clearly and reliably determined transients with a clear coupling to the GW channel.

CAT3 Contains vetoes significantly correlated with transients, but there is less understanding about the coupling with the GW channel.

CAT4 Vetoes with low statistical significance, but often with long duration. An example is a signal caused by an aircraft passing within a few kilometers of the detectors.

CAT5 Advisory flag for checking on the status of a detector, but without any direct influence on the GW channel.

For the analyses performed in this thesis, we exclude any stretches of data that have been vetoed with CAT4 or lower.

For proof-of-principle investigations or certain basic analysis tests, it may be sufficient to use the analytic or numerically estimated PSDs as depicted in Fig. 1.4. In a realistic scenario however, the noise is not stationary and the PSD varies over time, hence a PSD needs to be calculated that reflects the status of the detector output for the stretch of data that is analyzed. To obtain such an estimation of the PSD a sufficiently long section of unvetoed detector output where no GW signal is present is required. For the analyses performed in the following chapters we typically demand the segment of data from which we estimate the PSD to be 512 seconds long and no more than ~ 5 minutes before or after the GW event to be analyzed. This segment is then split into $N = 512/32$ sub-segments each 32 s long and for which N PSDs are calculated using Eq. (1.75). The final PSD $S_n(f)$ is then calculated by taking the point-wise median of the N individual PSDs.

For all investigations performed on the effects of realistic data on our analyses, we have used S6-VSR2/3 data "recolored" to early advanced detector sensitivities. The recolored noise combines a noise PSD that corresponds to that of the early-advanced detectors and real data from the 6th science run S6-VSR2/3. This allows us to take into account non-Gaussianity and non-stationarity of realistic noise, complete with glitches and veto definer files. The left panel in Fig. 1.5 shows two example PSDs that were generated using the recolored S6 data. The right panel shows the PSDs of the two detectors at the time of the first detection (O1 data) for comparison. Note that the recolored S6 data closely approaches the general trend of actual O1 data especially around 150 Hz where the detectors are most sensitive. At lower frequencies the O1 PSDs show slightly better sensitivity, meaning that recolored S6 data can be considered a conservative estimate.

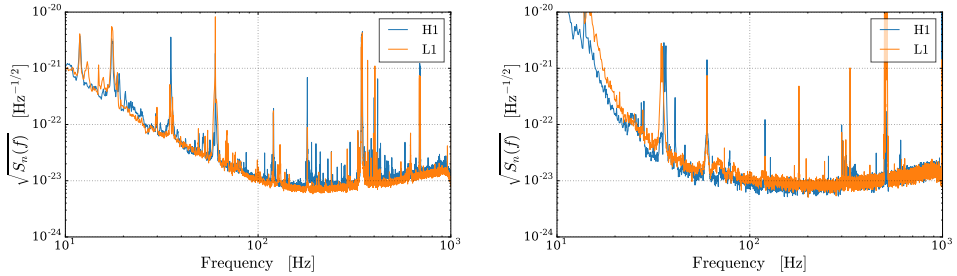


Figure 1.5: The noise amplitude $\sqrt{S_n(f)}$ as a function of frequency for the two Advanced LIGO interferometers at the Hanford (H1) and Livingston (L1) sites. The left panel shows the PSDs estimated from 512 s of recolored S6-VSR2/3 data. The right panel shows the PSDs of the first observing run (O1) at the time of the first detection. Note that the recolored PSDs do a decent job in simulating O1 data, even though they are more conservative at low frequencies.

1.3.3 Advanced detector network

Currently there are three primary ground-based L-shaped gravitational wave interferometers and a fourth for research and development purposes. In the United States there are two LIGO detectors each with an arm length of 4 km; one in Hanford, WA (H1) and another in Livingston, LA (L1). In Italy a third large interferometer with an arm length of 3 km called Virgo is located in Cascina near Pisa. Finally the much smaller GEO-HF [6, 7] with arm length 600 m is located near Sarstedt in the South of Hannover, Germany. The three main observatories, LIGO-Hanford, LIGO-Livingston and Virgo started operating in 2002, 2002 and 2007 respectively. The developmental status of the observatories at that time is referred to as the *initial* configuration.

In 2011 the VSR3 science run of initial Virgo and the S6 run of initial LIGO were finished and the detectors were scheduled to go off-line for major upgrades. These upgrades are the first in a series of several, each time aimed at an increase in performance until design sensitivity is reached. The detectors at design sensitivity are dubbed Advanced LIGO[8, 9] and Advanced Virgo[10]. After the first major upgrade the two LIGO detectors began their first observation run labeled O1 on September 12, 2015 at a sensitivity of about three times that of initial LIGO. After a successful run as described in Sec. 1.7, O1 ended on January 19, 2016 [11] to allow for the next step in upgrading LIGO to design sensitivity. The second observing run lasted approximately six months and started in October 2016. A third observing run lasting nine months is currently scheduled to start somewhere around 2018 [12].

The advanced detectors should reach sensitivities of about a factor of ten more than the initial configuration, allowing us to observe a volume in space about one thousand times larger than before. Such an increase in volume means that many more galaxies are included within our observational reach. By assuming a particular type of source and some threshold SNR, the distance at which these signals can be seen, the *horizon distance*, can be calculated. This is done by considering an optimal

orientation of the source and averaging the result over the sky position angles. For binary neutron star sources with a total mass of $2.8\,M_\odot$, initial LIGO could see out to an angle averaged distance of $\sim 20\,\mathrm{Mpc}$ and Advanced LIGO is anticipated to see out to $\sim 200\,\mathrm{Mpc}$. For binary black holes the increase is even larger where initial detectors would be sensitive to binary black hole sources with total mass $10\,M_\odot$ and equal mass ratio, out to $50\,\mathrm{Mpc}$ and the advanced detectors are anticipated to see these sources out to about $1\,\mathrm{Gpc}$.

In the near future the current detector network consisting of the two LIGO detectors and Virgo, will be expanded with the addition of a cryogenic LIGO-like underground interferometer called KAGRA [13, 14], near the Kamiokande experiment in Japan. In addition to KAGRA there will be a fifth detector, LIGO-India [15], which will be built in India and will be similar to the Advanced LIGO detectors in the US.

1.3.4 Einstein Telescope

Within the next two decades the construction of a third-generation gravitational wave detector is expected somewhere in Europe. This proposed detector is called Einstein Telescope (ET) [16]. ET differs in a lot of ways from the second-generation Advanced LIGO and Advanced Virgo. One clear distinction is its triangular configuration as opposed to the L-shape of the current detectors. This has for example the advantage of being able to disentangle the two gravitational wave polarizations with a single detector (the triangle essentially consists of 3 co-located interferometers with a 60° angle between the arms). The detector will also be built entirely underground at a depth of a few hundreds of meters, suppressing seismic noise originating at the surface. Aside from these obvious differences there are many more advanced technologies that will be implemented in ET to truly make a third-generation detector. Its final estimated sensitivity based on the current design [4] is shown in Fig. 1.4 and comes down to a further increase over advanced detectors of about a factor of 10 in the most sensitive part of the frequency band and Einstein Telescope (ET) will be sensitive at lower frequencies down to approximately $2\,\mathrm{Hz}$.

1.4 Sources and signals

In Sec. 1.1.2 we explained how gravitational waves can be produced by a system of two compact objects orbiting around each other. In this section we will expand on this by considering what happens to the system as it loses energy through the emission of gravitational waves. We will find that such systems will revolve around each other at increasingly smaller orbits before finally merging into a single object. This entire process is referred to as compact binary coalescence (CBC).

1.4.1 The basics of compact binary coalescence

In reality, the plus and cross polarizations of gravitational waves emitted from two compact objects revolving around each other derived in Sec. 1.1.2 are much more complicated and so far we only considered a snap-shot the complete evolution of such

a system. As gravitational radiation is emitted, the system loses energy and angular momentum which is countered by a decrease in orbital separation R and by Kepler's law an increase in orbital frequency $\omega_{\rm orb}$. From Eq. (1.51) we see that this increase in frequency results in an increase in the total power emitted; the system is losing more energy which is again balanced by a further decrease in orbital separation, which results in yet more power being radiated away from the system and a runaway effect follows where the separation between binary components decreases to a point where they inevitably merge.

Let us consider the basic dynamics a bit more concretely by specifying the orbital energy for a non-relativistic system:

$$E_{\rm orb} = E_{\rm kin} + E_{\rm pot} = -\frac{Gm_1m_2}{2R}. \tag{1.82}$$

Gravitational waves carry away energy from the system and $E_{\rm orb}$ becomes more and more negative, which can only be achieved by decreasing R. From Eq. (1.47) we can derive

$$\begin{aligned}\dot{R} &= -\frac{2}{3}\frac{GM}{R^2}\frac{\dot{\omega}_{\rm orb}}{\omega_{\rm orb}^3},\\ &= -\frac{2}{3}R\frac{\dot{\omega}_{\rm orb}}{\omega_{\rm orb}} = -\frac{2}{3}\left(R\omega_{\rm orb}\right)\frac{\dot{\omega}_{\rm orb}}{\omega_{\rm orb}^2},\end{aligned} \tag{1.83}$$

where $\dot{R}$ is the radial velocity and $R\omega_{\rm orb}$ the tangential velocity. If the radial velocity is much smaller than the tangential velocity, or in terms of orbital frequency, if

$$\dot{\omega}_{\rm orb} \ll \omega_{\rm orb}^2, \tag{1.84}$$

the motion of the system is said to be *quasi-circular* (also referred to as *adiabatic*) and the binary is in the *inspiral* regime. At this point we are still treating the orbital motion in the Newtonian regime even though we are considering gravitational radiation. One is allowed to do this as long as Eq. (1.84) is satisfied. In a full relativistic treatment there exists an *innermost stable circular orbit* (ISCO), also referred to as last stable orbit (LSO), which for a test mass m_1 orbiting in a Schwarzschild geometry with mass $m_2 = M$ occurs at

$$R_{\rm isco} \simeq \frac{6GM}{c^2}, \tag{1.85}$$

or in terms of GW frequency using Eq. 1.47

$$f_{\rm gw,isco} = \frac{c^3}{6^{3/2}\pi GM}. \tag{1.86}$$

Also in the case of comparable component masses there will be an ISCO frequency beyond which no stable orbits exist and the simplified form in Eq. (1.48) no longer holds. It turns out that even in this case Eq. (1.86) is a reasonable approximation and can be used as a rule of thumb to determine up to which frequency the quasi-circular regime is still a valid approximation. Beyond the ISCO, the compact binary components will fall toward each other (plunge), collide and merge into a single object.

We know now that as the binary evolves, the orbital separation decreases, and taking the time dependence of $\omega_{\rm orb}(t)$ and $R(t)$ into account, the expressions for the gravitational waveform polarizations can now be written as

$$h_+(t) = \frac{4}{r}\left(\frac{G\mathcal{M}_c}{c^2}\right)^{5/3}\left(\frac{\pi f_{\rm gw}(\tau)}{c}\right)^{2/3}\frac{1+\cos^2\iota}{2}\cos\Phi(\tau),$$
$$h_\times(t) = \frac{4}{r}\left(\frac{G\mathcal{M}_c}{c^2}\right)^{5/3}\left(\frac{\pi f_{\rm gw}(\tau)}{c}\right)^{2/3}\cos\iota\cos\Phi(\tau), \tag{1.87}$$

where we introduced the time to coalescence $\tau = t_c - t$ with t_c the time at coalescence and we defined

$$\Phi(t) = 2\int_{t_0}^{t} {\rm d}t'\,\omega_{\rm orb}(t') = \int_{t_0}^{t} {\rm d}t'\,\omega_{\rm gw}(t'), \tag{1.88}$$

with t_0 an initial time.

Using the quasi-circular regime of compact binary evolution, interesting observational quantities can be derived. How long for example would a typical binary system be visible within a detector's frequency bandwidth? In order to answer this question one can first obtain the gravitational wave frequency as function of time and invert the result. To this end we first consider the energy balance $-{\rm d}E_{\rm orb}/{\rm d}t = P_{\rm gw}$ and find an expression for $\dot{f}$ by substituting Eq. (1.82) and Eq. (1.51), eliminating R in favor of $f_{\rm gw}$ using Eq. (1.47):

$$\dot{f}_{\rm gw} = \frac{96}{5}\pi^{8/3}\left(\frac{G\mathcal{M}_c}{c^3}\right)^{5/3} f_{\rm gw}^{11/3}. \tag{1.89}$$

We can now integrate this expression to find the time T the signal spends between some lower frequency $f_{\rm min}$ and an upper frequency $f_{\rm max}$,

$$T = 2.19\,{\rm s}\left(\frac{1.21\,{\rm M}_\odot}{\mathcal{M}_c}\right)^{5/3}\left[\left(\frac{100\,{\rm Hz}}{f_{\rm min}}\right)^{8/3} - \left(\frac{100\,{\rm Hz}}{f_{\rm max}}\right)^{8/3}\right], \tag{1.90}$$

where $1.21\,{\rm M}_\odot$ is the chirp mass of a binary system consisting of two neutron stars with $m_1 = m_2 = 1.4\,{\rm M}_\odot$. For $f_{\rm min} = 20\,{\rm Hz}$ and $f_{\rm max} = f_{\rm gw,isco} = 1566\,{\rm Hz}$ this gives $T \simeq 160\,{\rm s}$, while for a binary black hole system this quickly decreases as for a $m_1 = m_2 = 5\,{\rm M}_\odot$ system $T \simeq 19\,{\rm s}$.

Another quantity of interest is the number of gravitational wave cycles that are within the detector's bandwidth. The number of cycles is given by

$$\mathcal{N}_{\rm cyc} = \int_{t_{\rm min}}^{t_{\rm max}} {\rm d}t\, f_{\rm gw} = \int_{f_{\rm min}}^{f_{\rm max}} {\rm d}f_{\rm gw}\,\frac{f_{\rm gw}}{\dot{f}_{\rm gw}}, \tag{1.91}$$

which, using Eq. (1.89) can be rewritten in terms of $f_{\rm gw}$ alone and integrated yielding

$$\mathcal{N}_{\rm cyc} = 5\times10^3\left(\frac{1.21\,{\rm M}_\odot}{\mathcal{M}_c}\right)^{5/3}\left(\frac{20\,{\rm Hz}}{f_{\rm min}} - \frac{20\,{\rm Hz}}{f_{\rm max}}\right)^{5/3}. \tag{1.92}$$

From this expression we can see that for a binary neutron star source, a few thousand cycles can be observed between 20 Hz and 1566 Hz. This presses the need for a good understanding of the waveform and especially its phasing so that parameters describing the source can be extracted with great accuracy.

Ground-based observatories such as LIGO and Virgo are typically sensitive down to about 20 Hz. For high frequencies they are sensitive up to a few kHz. This sensitive frequency band determines the type of sources these detectors can see. We can for example determine the most massive system a ground-based observatory can see by considering a system with an ISCO corresponding to 20 Hz: By Eq. (1.86), the observatories can therefore detect sources of masses up to $\sim 200\,\mathrm{M}_\odot$. The lowest mass systems we can observe is determined mainly astrophysically: The lowest mass binary neutron star (BNS) systems can have a mass of approximately $1.35\,\mathrm{M}_\odot$ in order for the components to form neutron stars in the first place. BNS systems have component masses ranging between approximately $1\,\mathrm{M}_\odot$ and $2\,\mathrm{M}_\odot$ which means that many cycles will be in band for such systems (Eq. (1.92)). Components with masses $\lesssim 1.35\,\mathrm{M}_\odot$ form white dwarfs and are not detectable by ground-based observatories as they will merge before 20 Hz, due to their physical size[1]. Sources with component masses exceeding $5\,\mathrm{M}_\odot$ are binary black holes (BBH). Note there is a mass gap between $2\,\mathrm{M}_\odot$ and $5\,\mathrm{M}_\odot$, where it is difficult to be certain whether one is dealing with black holes, or neutron stars. A third class of CBC signals that are detectable with ground-based observatories are systems with a black hole component and a neutron star component, so called neutron star and black hole (NSBH) systems, as long as their total mass lies below the $200\,\mathrm{M}_\odot$ limit.

From Eq.(1.92), we learn that the more massive the system, the fewer cycles will be in band. Of course this is only an approximation and the waveform signal does not abruptly terminate at the last stable orbit. Up to around the moment of LSO the waveform is in the *inspiral* regime. After LSO, the binary enters the *merger* regime where the components are no longer on quasi-circular orbits and they plunge toward one another and merge into a single remnant object. Depending on the combined mass of the initial components, this may be a neutron star or black hole; however, when considering systems with component masses $m_i > 1\,\mathrm{M}_\odot$, the final object will probably collapse into a black hole as the resulting mass will exceed $2\,\mathrm{M}_\odot$, which is currently the observational upper limit for neutron star masses[2]. After the merger regime, when we consider for simplicity a binary black hole (BBH) system, an asymmetric rotating remnant black hole will settle down into a Kerr black hole through the emission of GW in the form of a superposition of damped sinusoids. This process, which is referred to as the *ringdown* regime, can be modeled by considering linear perturbations of the Kerr metric, or quasi-normal modes, which are characterized by frequencies ω_{lm} and damping times τ_{lm} [18, 19, 20, 21] (see also Chapter 2). There is no strict definition as to where the inspiral ends and the merger begins, or the merger ends and the ringdown begins: In most cases this definition is model-dependent and has no intrinsic physical

[1] White dwarfs are much larger than neutron stars as they are not forced to be compressed to such densities as neutron stars are. Because they are larger, they touch much sooner and merge at frequencies below 20 Hz

[2] After merger, a super-massive neutron star may be stable for a while until it spins down and the balance between outward pressure induced by spin and inward pressure by gravitational pull no longer holds. For a more extensive coverage of the topic see e.g. [17].

significance; however, it is useful to be able to refer to the different regimes as rules of thumb.

A typical complete inspiral-merger-ringdown (IMR) waveform is shown in Fig. 1.6, where inspiral region is determined by the time of last stable orbit and the start of the ringdown is just a rough indication. Also indicated in this figure is the coalescence time t_c, which occurs generally after the LSO, except for large objects which can be the case for neutron stars with a specific equation of state. The ringdown always occurs after the coalescence time as the ringdown is generally taken to be the region from where the signal is well described with linear perturbations of the Kerr metric, which is not the case at or close after the merger.

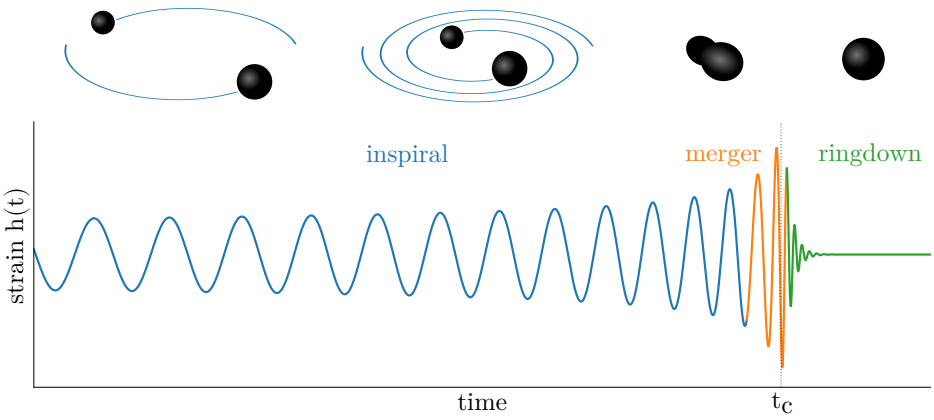


Figure 1.6: A gravitational wave inspiral-merger-ringdown waveform resulting from a compact binary coalescence. Highlighted in blue, orange and green are the three regions that are often referred to. As explained in the text, there is no strict definition of where the regions start and end; in this case, the inspiral ends at the last stable orbit and the ringdown start is just a rough indication. Also indicated is the coalescence time t_c, which occurs before the ringdown and after the LSO (can happen before if objects are large, *e.g.* for some neutron stars).

1.4.2 Waveform models

So far we have only considered approximations in the Newtonian limit and arrived at non-relativistic representations of the full relativistic expressions that go into describing a gravitational waveform accurately. One obvious feature the approximations lack, is a representation for the waveform beyond the LSO; furthermore, as orbital velocities increase, relativistic effects become more important and our simple representations will get less accurate the closer the system gets to the LSO. Clearly more advanced models are required to perform accurate gravitational wave analysis.

In CBC analysis there are essentially 4 ways to generate a waveform with varying degree of accuracy depending on the model's particularities;

- **A.** Solving the Einstein equations directly by numerical integration: *numerical waveforms*.
- **B.** Post-Newtonian (PN) expansion alone resulting in the *TaylorX* family of waveforms.
- **C.** Effective one body formalism which combines PN theory, re-summation techniques and numerical fitting: *EOB waveforms*.
- **D.** *Phenomenological* waveforms, providing fast to evaluate analytical fits modeled on numerical waveforms and EOB inspirals.

When it comes to waveform models in the context of parameter estimation, speed and accuracy are the main aspects to consider when deciding on which model to use; speed is of the essence since millions of waveforms need to be generated in order to cover the large parameter space and accuracy is indeed required to produce meaningful results. The degree of accuracy required in a model is entirely dependent on the type of source that is analyzed, or which method is used to what end: We will see for example in Sec. 1.6 that for certain tests of GR we only require a waveform to be accurate up to 400 Hz in the case of BNS sources. The code for generating the various waveform models (waveform generators), with the exception of numerical waveforms, is part of the LIGO Algorithm Library (LALSuite) [22].

Numerical integration is the most accurate and reliable method to produce waveforms, but is also computationally the most expensive. For parameter estimation, numerical waveforms are therefore not feasible to use. They do however serve as a benchmark for producing faster analytical representations of the waveforms. In the following sections each of the above mentioned waveform generation techniques are briefly discussed. We will start by giving a qualitative description of some of the numerical nomenclature used throughout this work in Sec. 1.4.2-A. The next section covers how expressions for gravitational wave luminosity and binding energy resulting from PN theory are used to construct one of the TaylorX waveform models. Sec. 1.4.2-C will provide the reader with a brief overview of the effective one body (EOB) framework and more sophisticated models building on that principle. In Sec. 1.4.2-D an overview is given of various phenomenological waveform models that are currently available and highlights one particular model that is at this time the preferred one for tests of GR.

1.4.2-A Numerical relativity

The most expensive, yet most accurate method to produce gravitational waves corresponding to a particular source is by numerically evolving the Einstein field equations starting from a set of initial conditions. There exist many methods and codes for accomplishing this among which are the BAM code [23, 24] and the Spectral Einstein Code (SpEC). The latter is used for the publicly available numerical waveforms from the Simulating Extreme Spacetimes (SXS) Collaboration [25].

Numerical simulations are used to model the space-time surrounding the two binary components and the gravitational wave radiation we are interested in is calculated from the Weyl scalar Ψ_4 [26, 27]; a complex function that encodes the two

polarizations of the outgoing transverse radiation. At large distances the Weyl scalar is related to the + and × polarizations through two time derivatives

$$\Psi_4 = \frac{\mathrm{d}^2}{\mathrm{d}t^2}\left[h_+(t) - \mathrm{i}h_\times(t)\right]. \tag{1.93}$$

In the time domain, obtaining the two polarizations means performing two integrations of Ψ_4 and requires fixing two integration constants. Calculating the strain $h = h_+ - \mathrm{i}h_\times$ in the frequency domain is more straightforwardly done by division

$$\tilde{h}^{\mathrm{num}}(f) = -\frac{\tilde{\Psi}_4^{\mathrm{num}}(f)}{4\pi^2 f^2} = A^{\mathrm{num}}(f)e^{\mathrm{i}\Phi^{\mathrm{num}}(f)}, \tag{1.94}$$

where $\tilde{a}(f)$ denotes the Fourier transform of $a(t)$. For numerical waveforms it is common to use dimensionless times and frequencies t/M and Mf where M is the system's total mass; in numerical simulations M is set to unity as afterwards the waveforms can simply be scaled to any value of M.

Since generating numerical waveforms requires a lot of computational resources, they typically only produce waveforms starting between 5 to a few tens of orbits before merger, where the latter case is for relatively simple configurations (*i.e.* no strongly precessing spins or large mass ratios). Therefore, in order to use these waveforms for improvement and verification of other models such as Effective One Body or phenomenological ones, *hybrids* need to be created. Hybrid waveforms are numerical waveforms extended down to lower frequencies by smoothly connecting a reliable inspiral model. One example of a reliable inspiral model (restricted to systems where component spins are aligned) is the EOB approximant by Taracchini *et al.* [28] (labeled SEOBNRv2 in LALSuite), which shows the most consistent agreement with results from numerical simulations within the part of parameter space it was developed for [29, 30]. Depending on the extent of the numerical part of the hybrid, a relatively basic inspiral model such as PN approximants can be of sufficient accuracy and have been used in e.g. [31]. Using an analytical model calibrated to numerical data such as SEOBNRv2 to create a hybrid waveform which itself is built with numerical data is typically not desired. Instead, as is done in [32], the calibration terms in SEOBNRv2 are set to zero so that the resulting hybrid is independent of previous numerically informed EOB tuning.

Recently a numerical waveform injection framework [33, 34] has become available for LALSuite, which allows an end-user to implement any numerical waveform from the public SXS database in a similar fashion as other waveform approximants that are available in the algorithm library. This means for example that a numerical waveform can be produced at arbitrary sample rates by way of spline interpolation, which in turn allows it to be used as a simulated signal in evaluating parameter estimation pipelines.

1.4.2-B Post-Newtonian approximants

The most important tool for non-phenomenological analytical representations of gravitational waves resulting from CBC sources is the post-Newtonian (PN) formalism. The PN formalism provides a method to calculate the inspiral phasing of a binary

system that is sufficiently accurate up to a certain frequency approaching LSO. The idea of the PN formalism is to expand the expression of the phasing $\Phi(t)$ in terms of a small parameter typically taken as $v = (\pi M f_{\rm gw})^{1/3}$, which is called the characteristic velocity of the binary. Here M is the total mass of the binary in seconds (setting $c = G = 1$) and $f_{\rm gw}$ is the gravitational wave frequency. In a special case, where the GW phase is twice the orbital phase and Eq. (1.84) is satisfied (adiabatic, restricted regime), the phasing can be written as the pair of differential equations

$$\begin{aligned} \dot{\phi} &= \frac{v^3}{M}, \\ \dot{v} &= -\frac{\mathcal{F}(v)}{M E'(v)}, \end{aligned} \tag{1.95}$$

where $E'(v)$ is the derivative of the binding energy per unit total mass with respect to v and $\mathcal{F}(v)$ is the gravitational-wave luminosity. The assumption in Eq. (1.84) is valid during most of the evolution, but begins to fail as the system approaches the LSO [35].

The fraction $\mathcal{F}(v)/E'(v)$ can be written in different ways leading to different PN waveform families. Because we are dealing with a perturbative series, we can "re-sum" or "re-expand" the series any way we like as long as we keep terms to the correct order. For example, in the family called TaylorT1, both $\mathcal{F}$ and E are left as they are, while in the TaylorT4 family the fraction as a whole is expanded in v to consistent order (*i.e.* leaving out any cross-terms that exceed the highest power of v in either $\mathcal{F}(v)$ or $E'(v)$). The PN order always indicates a power in v up to which a model or expression is expanded; in general nPN corresponds to v^{2n} so that *e.g.* an expression expanded to v^3 is indicated as 1.5PN. In [35], six different PN families are described and compared. All these families solve Eq. (1.95) in a different way, but all start with the same expressions for the conserved 3PN energy ((per unit total mass) [36, 37, 38, 39, 40, 41]

$$\begin{aligned} E_3(v) = &-\frac{1}{2}\eta v^2\Bigg[1 - (\frac{3}{4} + \frac{1}{12}\eta)v^2 - \left(\frac{27}{8} - \frac{19}{8}\eta + \frac{1}{24}\eta^2\right)v^4 \\ &- \left\{\frac{675}{64} - \left(\frac{34445}{576} - \frac{205}{96}\pi^2\right)\eta + \frac{155}{96}\eta^2 + \frac{35}{5184}\eta^3\right\}v^6\Bigg], \end{aligned} \tag{1.96}$$

and 3.5PN energy flux [42, 43, 44, 45, 46] which represents the system's energy loss

$$\begin{aligned}\mathcal{F}_{3.5}(v) =& \frac{32}{5}\eta^2 v^{10}\Bigg[1 - \left(\frac{1247}{336} + \frac{35}{12}\eta\right)v^2 + 4\pi v^3 \\ &- \left(\frac{44711}{9072} - \frac{9271}{504}\eta - \frac{65}{18}\eta^2\right)v^4 \\ &- \left(\frac{8191}{672} + \frac{583}{24}\eta\right)\pi v^5 + \Bigg\{\frac{6643739519}{69854400} \\ &+ \frac{16}{3}\pi^2 - \frac{1712}{105}\gamma + \left(\frac{41}{48}\pi^2 - \frac{134543}{7776}\right)\eta \\ &- \frac{94403}{3024}\eta^2 - \frac{775}{324}\eta^3 - \frac{856}{105}\ln(16v^2)\Bigg\}v^6 \\ &- \left(\frac{16285}{504} - \frac{214745}{1728}\eta - \frac{193385}{3024}\eta^2\right)\pi v^7\Bigg],\end{aligned} \tag{1.97}$$

where $\gamma = 0.577216\ldots$ is the Euler constant. The PN approximation is only accurate enough up to mildly relativistic velocities. Re-summation improves this, but still does not lead to a unique model of phase evolution. In the adiabatic approximation, which is equivalent to satisfying Eq. (1.84), one expects the luminosity in gravitational waves to come from the change in orbital energy averaged over one period. If we then assume circular orbits[3], one can use the energy balance equation

$$\mathcal{F}(t) = -M\frac{\mathrm{d}E(t)}{\mathrm{d}t}, \tag{1.98}$$

and Kepler's law Eq. (1.47) to arrive at the evolution of the orbital phase introduced in Eq. (1.95).

In the following we will highlight one widely used PN waveform model; the closed-form analytical frequency domain TaylorF2. In the discussion that follows we will restrict ourselves to the non-spinning model, even though at this time TaylorF2 does allow for aligned spins in the LALSuite implementation.

The TaylorF2 approximant is obtained by analytically Fourier transforming $h(t)$ using the stationary phase approximation (SPA). The waveform polarizations in the time-domain may be written as

$$\begin{aligned}h_+(t) &= \frac{1+\cos^2\iota}{2}a_{\mathrm{gw}}(t)\cos\phi_{\mathrm{gw}}(t), \\ h_\times(t) &= \cos\iota a_{\mathrm{gw}}(t)\sin\phi_{\mathrm{gw}}(t)\end{aligned} \tag{1.99}$$

where a_{gw} is a short-hand notation for the gravitational wave amplitude from Eq. (1.87). The signal observed by a detector is a combination of the two polarizations, mixed in

[3] The binaries considered in this thesis start their evolution at separations much larger than their gravitational radius. For such systems any initial eccentricity will have been lost by the time they enter LIGO and Virgo's sensitivity band due to gravitational radiation reaction which tends to circularize a binary [47, 48].

a specific way determined by the detector's antenna pattern functions F_+ and $F_\times$

$$\begin{aligned} h(t) &= F_+ h_+ + F_\times h_\times \\ &= \frac{1}{2} X_a(F_+, F_\times, \iota) a_{\rm gw}(t) \cos\left[\phi_{\rm gw}(t) + X_\phi(F_+, F_\times, \iota)\right], \end{aligned} \tag{1.100}$$

where $X_a(F_+, F_\times, \iota)$ and $X_\phi(F_+, F_\times, \iota)$ are constants only dependent on the antenna pattern functions and inclination angle ι. This notation shows that $h(t)$ may also be written in the simple form

$$h(t) \equiv a(t) \cos\phi(t). \tag{1.101}$$

Using the SPA it is possible to derive a closed-form analytical approximation to the Fourier transform of $h(t)$

$$\begin{aligned} \tilde{h}(f) &= \int_{-\infty}^{\infty} {\rm d}t\, a(t) \cos\phi(t) e^{2\pi {\rm i} f t}, \\ &= \frac{1}{2} \int_{-\infty}^{\infty} {\rm d}t\, a(t) \left[e^{{\rm i}(2\pi f t + \phi(t))} + e^{{\rm i}(2\pi f t - \phi(t))} \right]. \end{aligned} \tag{1.102}$$

We note that the first term in Eq. (1.102) oscillates rapidly for all t and averages to zero as opposed to the second term which does have a maximum contribution around t_f, which can be found by solving

$$\left. \frac{{\rm d}\phi(t)}{{\rm d}t} \right|_{t_f} = 2\pi f \equiv f_{\rm gw}. \tag{1.103}$$

Then in order to approximate Eq. (1.102), one can expand the phase up to $(t - t_f)^2$ and make the additional assumption that the amplitude also varies slowly around t_f. The latter allows us to take the amplitude out of the integral, resulting in

$$\tilde{h}(f) = \frac{1}{2} a(t_f) e^{{\rm i}(2\pi f t_f - \phi(t_f))} \sqrt{\frac{2}{\ddot{\phi}(t_f)}} \int_{-\infty}^{\infty} {\rm d}x\, e^{{\rm i}x^2}, \tag{1.104}$$

where $x = \sqrt{\phi(t_f)/2}$. With the remaining integral solved as

$$\int {\rm d}x\, e^{{\rm i}x^2} = \sqrt{\pi} e^{-{\rm i}\pi/4}, \tag{1.105}$$

Eq. (1.102) becomes

$$\tilde{h}(f) = \frac{a(t_f)}{2\sqrt{\ddot{\phi}(t_f)/2\pi}} e^{-{\rm i}[2\pi f t_f - \pi/4 - \phi(t_f)]}, \tag{1.106}$$

One can now substitute the PN time-domain expressions for $\phi(t)$ and $a(t)$ obtained from Eq. (1.95) to arrive at a closed-form frequency domain waveform. In particular, explicitly truncating the ratio between the energy and flux functions $E(v)$ and $\mathcal{F}(v)$ to consistent order constitutes the TaylorF2 waveform family. The highest known PN order for the TaylorF2 phase at this time is 3.5PN.

We may write the phase as function of the PN parameter $v = (\pi M f)^{1/3}$ as

$$\psi(f) = 2\pi f t_c - \phi_c + \sum_{n=0}^{7} \left(\varphi_n + \varphi_n^l\right) v^{n-5}, \tag{1.107}$$

with t_c and ϕ_c arbitrary time and phase shifts and

$$\begin{aligned}
\varphi_0 &= \psi_{\rm N} \equiv \frac{3}{128\eta},\\
\varphi_1 &= 0,\\
\varphi_2 &= \frac{20}{9}\left(\frac{743}{336} + \frac{11}{4}\eta\right)\psi_{\rm N},\\
\varphi_3 &= -16\pi\psi_{\rm N},\\
\varphi_4 &= 10\left(\frac{3058673}{1016064} + \frac{5429}{1008}\eta + \frac{617}{144}\eta^2\right)\psi_{\rm N},\\
\varphi_5 &= \left(\frac{38645}{756} - \frac{65}{9}\eta\right)\pi\psi_{\rm N},\\
\varphi_5^l &= \left(\frac{38645}{252} - \frac{65}{3}\eta\right)\ln\frac{v}{v_{\rm lso}}\pi\psi_{\rm N},\\
\varphi_6 &= \left\{\left(\frac{11583231236531}{4694215680} - \frac{640}{3}\pi^2 - \frac{6848}{21}\gamma\right)\right.\\
&\quad + \left(-\frac{15737765635}{3048192} + \frac{2255}{12}\pi^2\right)\eta + \frac{76055}{1728}\eta^2 - \frac{127825}{1296}\eta^3\\
&\quad \left. - \frac{6848}{21}\ln(4)\right\}\psi_{\rm N},\\
\varphi_6^l &= -\frac{6848}{21}\psi_{\rm N}\ln(v),\\
\varphi_7 &= \left(\frac{77096675}{254016} + \frac{378515}{1512}\eta - \frac{74045}{756}\eta^2\right)\pi\psi_{\rm N},
\end{aligned} \tag{1.108}$$

are the PN coefficients.

TaylorF2 is quick to evaluate since its lacks the requirement to solve differential equations. Furthermore, since it is already in the frequency domain, where likelihood calculations are performed, this waveform does not require a Fourier transform to be calculated. Regardless of its seemingly simple form, the waveform performs well in accuracy during inspiral compared to other approximants [35]. At this time the expressions in Eq. (1.108) have been extended to include aligned spins and leading order tidal effects for neutron star binaries. Lastly, TaylorF2 due to its fast evaluation and closed-form analytical nature, lends itself extremely well for tests of general relativity with binary neutron star sources (where the inspiral constitutes the bulk of the signal). In fact, this is the waveform model used in our test of the GR pipeline introduced in Sec. 1.6.

1.4.2-C Effective one body formalism

The PN formalism can only reproduce the inspiral part of the waveform and becomes less accurate the closer the LSO is approached. Since the GW luminosity is brightest at the end of the inspiral and merger-ringdown phases of the evolution, it is important to develop models that maximize the detection potential by including these final phases. Especially for BBH systems this becomes important as less of the inspiral falls into the detector's sensitive band.

More than a decade ago the use of re-summation techniques lead to the construction of the EOB model. This model analytically provided the complete waveform of the binary evolution. Over the past years the model has been calibrated to numerical relativity (NR) simulations and many important features have been improved since. This section aims to provide the reader with a brief introduction on the underlying theory of EOB models. The basic framework of EOB which I present in the following has since been significantly improved and extended by including spins, tidal effects and calibration using numerical waveforms.

The EOB formalism reduces a two-body problem in GR into an effective one-body problem, meaning that a metric is found that effectively represents the system with a single massive body around which a test particle is allowed to inspiral toward it. The EOB metric takes the form [49]

$$\mathrm{d}s^2_{\mathrm{eff}} = -A(r)\mathrm{d}t^2 + \frac{D(r)}{A(r)}\mathrm{d}r^2 + r^2(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2). \tag{1.109}$$

In the manifold represented by Eq. (1.109), the dynamics of a test particle are represented by the EOB Hamiltonian

$$H^{\mathrm{real}}(r, p_r, p_\phi) \equiv \mu\hat{H}^{\mathrm{real}} = M\sqrt{1 + 2\eta(\hat{H}^{\mathrm{eff}} - 1)}, \tag{1.110}$$

where H^{eff} is the effective Hamiltonian

$$H^{\mathrm{eff}}(r, p_r, p_\phi) \equiv \mu\hat{H}^{\mathrm{eff}} = \mu\sqrt{A(r)\left[1 + \frac{A(r)}{D(r)}p_r^2 + \frac{p_\phi^2}{r^2} + 2(4-3\eta)\eta\frac{p_r^4}{r^2}\right]}, \tag{1.111}$$

and μ and η are the reduced mass m_1m_2/M and the symmetric mass ratio μ/M respectively. Note that the hatted symbols represent dimensionless quantities *i.e.* $\hat{\omega} = \omega M$.

The coefficients $A(r)$ and $D(r)$ are written as

$$A_k(r) = \sum_{i=0}^{k+1}\frac{a_i(\eta)}{r^i}, \tag{1.112a}$$

$$D_k(r) = \sum_{i=0}^{k}\frac{d_i(\eta)}{r^i}, \tag{1.112b}$$

in which the η-dependent coefficients are included up to 3PN order (*i.e.* up to $k = 4$).

The EOB dynamics can be adjusted to achieve better agreement with numerical waveforms by including a so-called pseudo-4PN[4] (p4PN) coefficient $a_5(\eta) = 60\eta$ [50]. To ensure that the effective metric given in Eq. (1.109) contains a horizon (like the Schwarzschild radius in the Schwarzschild metric), a zero is to be factored out from $A(r)$. As one has the freedom to expand and re-sum the expressions as long as it is done to the appropriate order, a zero can be factored out by applying so-called Padé re-summation [51].

Re-summation reduces the ratio of Eqns. (1.112a) and (1.112b) into well-defined closed-form expressions as function of r and η and the dynamics of the system can now be obtained by solving the EOB Hamilton equations. Note that all the equations are written in terms of the reduced quantities in Eq. (1.110), $\hat{t} = t/M$ and $\hat{\omega} = \omega M$, etc. To avoid cluttering the expressions, the hat notation is dropped from this point onward. The EOB Hamilton equations now read

$$\frac{\mathrm{d}r}{\mathrm{d}t} = \frac{\partial H^{\mathrm{real}}}{\partial p_r}(r, p_r, p_\phi), \tag{1.113a}$$

$$\frac{\mathrm{d}\phi}{\mathrm{d}t} = \frac{\partial H^{\mathrm{real}}}{\partial p_\phi}(r, p_r, p_\phi), \tag{1.113b}$$

$$\frac{\mathrm{d}p_r}{\mathrm{d}t} = -\frac{\partial H^{\mathrm{real}}}{\partial r}(r, p_r, p_\phi), \tag{1.113c}$$

$$\frac{\mathrm{d}p_\phi}{\mathrm{d}t} = \mathcal{F}_\phi(r, p_r, p_\phi), \tag{1.113d}$$

where as before we use $\omega \equiv \dot{\phi}(t)$. The last equation describes the decay of the angular momentum by using another critical input to the EOB model; the radiation reaction force $\mathcal{F}$ which can be related to the flux in Eq. (1.97). Different choices to represent $\mathcal{F}$ include using the basic PN approximation as given in Eq. (1.97) and Padé re-summations, but there are other methods.

Constructing the EOB inspiral part is a matter of solving the ODEs in Eq. (1.113). There are some points to take into account when solving Eq. (1.113). For example, the initial values of p_r^0 and p_ϕ^0, which are non-trivial to determine, have a large influence on the way the binary components will inspiral. If the estimation of the initial conditions is off by too much at the point one starts the evolution, the system will show eccentric orbits; when p_ϕ^0 is too large, the binary will never merge as the components will merely approach one another and move off to infinity.

If binaries start out at radii much larger than their gravitational radius (*i.e.* $r \gg GM/c^2$), by the time the system enters the sensitivity band of ground-based detectors, any initial eccentricity will have been lost due to the gravitational radiation reaction, which tends to circularize a binary [35]. Hence in order to make sure the system is provided with correct initial conditions, one needs to start at low frequencies, so that any eccentricity caused by errors will have time to dissipate. Note that this feature adds additional computing time to the waveform even if one would like to only produce the waveform for the final few orbits.

[4]It is called pseudo-4PN because even though it is used as though it is 4PN order, it is not claimed that once the actual 4PN order is derived, it will have the same form [50].

The basics of the EOB framework described above will only generate the inspiral part of the waveform. The EOB models that are used for detection and parameter estimation are complete inspiral-merger-ringdown models. There are various techniques to bridge the gap between the end of the inspiral and the final ringdown; the most common method is described in e.g. [50] and is basically a smooth continuation of the inspiral into an analytical closed form description of the ringdown.

Over the past years, spinning EOB models have been developed, starting with aligned spins in the model called SEOBNRv1 [50], where "NR" indicates it has been further calibrated against numerical waveforms. Then the model was extended further to allow for fully precessing spins in SEOBNRv3 [52]. The development of EOB is still ongoing to keep improving the accuracy of the models. The advantage of accurate EOB models is that they can be generated for arbitrary points in parameter space with good agreement with numerical waveforms, which are only available for a sparse set of parameter choices. In the context of parameter estimation and some of the work presented in this thesis, SEOBNRv3 is particularly well suited as a benchmark waveform, allowing the comparison of other models to something that resembles numerical accuracy.

1.4.2-D Phenomenological models

The phenomenological class of waveform models is one that combines PN theory which accurately and analytically describes the inspiral phase of binary coalescence, and phenomenological representations of the merger and ringdown parts, fitted to a set of numerical waveforms. In the following, two aligned spin approximants, IMRPhenomC and IMRPhenomD, will be introduced which are the foundation of the precessing models IMRPhenomPv1 and IMRPhenomPv2 respectively. The precessing models work on the same principle, the difference being that IMRPhenomPv2 is based on the more accurate IMRPhenomD.

The aligned spin approximant IMRPhenomC introduced by Santamaria *et al.* [31] utilizes the SPA frequency domain phase used in TaylorF2 to describe the inspiral part of the waveform and extends the model phenomenologically to create a complete inspiral-merger-ringdown (IMR) waveform.

The construction and verification of IMRPhenomC is achieved using a set of nine numerical waveforms with varying mass ratios $1 \leq q \leq 4$ and (anti-)aligned spin components $0 \leq |\chi_a| \leq 0.85$. In addition to the numerical waveforms, the model uses the 3.5PN SPA phase $\psi_{\rm SPA}(f)$ introduced in Sec. 1.4.2-B for the inspiral, with the addition of next-to-leading order spin-orbit effects [53, 54, 55] as well as spin-spin effects that appear at relative 2PN order [54, 56, 57]. The complete waveform that depends on physical parameters $\boldsymbol{\lambda} = (M, \eta, \chi_1, \chi_2)$ in the frequency domain is written as

$$\tilde{h}(f;\boldsymbol{\lambda}) = A(f;\boldsymbol{\lambda})e^{-\mathrm{i}\psi(f;\boldsymbol{\lambda})}, \tag{1.114}$$

where in this overview our primary focus will be on the construction of the phase $\psi(f;\boldsymbol{\lambda})$; the method for constructing the amplitude is similar and not directly used in the tests of general relativity that are the main topic of this thesis.

To describe the parts of the waveform following the inspiral, the authors of [31] used two Ansatz functions; one for what they refer to as the post-merger (PM)

$$\psi_{\mathrm{PM}}^{\mathrm{PhenomC}}(f) = \frac{1}{\eta}(\xi_1 f^{-5/3} + \xi_2 f^{-1} + \xi_3^{-1/3} + \xi_4 + \xi_5 f^{2/3} + \xi_6 f), \tag{1.115}$$

and another that models the ringdown (RD)

$$\psi_{\mathrm{RD}}^{\mathrm{PhenomC}}(f) = \zeta_1 + \zeta_2 f. \tag{1.116}$$

The phenomenological coefficients ξ_i of the Ansatz function in Eq. (1.115) are determined by fitting to the Fourier transform of the numerical waveforms. The ringdown coefficients ζ_i are determined by demanding continuity, *i.e.* by matching to the value and slope of the PM phase Ansatz at a particular frequency. The complete phase is then constructed as follows

$$\Phi_{\mathrm{PhenomC}}(f) = \psi_{\mathrm{SPA}}(f) w_{f_1}^{-}(f) + \psi_{\mathrm{PM}}(f) w_{f_1}^{+}(f) w_{f_2}^{-}(f) + \psi_{\mathrm{RD}}(f) w_{f_2}^{+}(f), \tag{1.117}$$

where the transition between the different regimes is smoothed by means of tanh window functions

$$w_{f_0}^{\pm}(f) = \frac{1}{2}\left[1 \pm \tanh\left(\frac{4(f - f_0)}{0.005}\right)\right], \tag{1.118}$$

and the frequencies f_1 and f_2 depend on a predetermined calibration regime. This is, in essence, how IMRPhenomC is constructed, but let us now turn to the next generation IMRPhenomD and go into a bit more detail.

IMRPhenomD [32] is not only an extension of the parameter space in which the waveform is valid with respect to its predecessor IMRPhenomC, it is also structured in a slightly different fashion. First of all, the set of numerical waveforms that are used in calibration and verification of IMRPhenomD has been almost doubled to 19 and the parameter ranges are now extended to $1 \leq q \leq 18$ and $0 \leq |\chi_a| \leq 0.95$. The other notable difference is that next to the SPA phasing for the inspiral (here referred to as the *early*-inspiral), three additional regimes were considered to accommodate more features in the waveform:

- late-inspiral regime
- intermediate regime
- merger-ringdown regime

Together, the early-inspiral and late-inspiral constitute the regime that is simply referred to as the inspiral. The regimes are graphically depicted in Fig. 1.7 together with $-\partial\phi(f)/\partial f = -\phi'(f)$ (top panel) and the amplitude $|\tilde{h}(f)|$ (bottom panel), which are used for fitting the various phase and amplitude coefficients respectively. The extent of each regime is based on the lowest common GW frequency f_{NRmin} for which numerical data are available and the ringdown frequency f_{RD} defined as the central frequency of the dominant quasi-normal mode (Eq. (5.5) in[31]). Each regime overlaps slightly with the connecting regimes to ensure continuity. The first regime

(inspiral) ends at $Mf_{\rm NRmin} = 0.018$ where the model transitions into the intermediate region. The intermediate region extends between $[0.017/M, 0.5f_{\rm RD}]$, and the ringdown between $[0.45, 1.15]f_{\rm RD}$.

Note that there is some ambiguity in naming the different parts of the waveform. There are three regions defined: Inspiral, intermediate and merger-ringdown. However the boundaries of these regions have varying definitions between building the model by performing fits, and constructing the full inspiral-merger-ringdown (IMR) waveform; furthermore, for each of those it depends on whether one is considering the amplitude or the phase. It would seem that the most relevant bounds for the end-user of the waveform model are the ones used in constructing the full IMR waveform, *i.e.* the ones used in the waveform generator. For the phase the transition frequencies are $Mf_1^\phi = 0.018$ and $f_2^\phi = 0.5f_{\rm RD}$ and for the amplitude they are $Mf_1^A = 0.014$ and $f_2^A = f_{\rm peak}$. In this definition the peak frequency $f_{\rm peak}$ is defined in Eq. (20) in [32] where $f_{\rm damp}$ is the imaginary part of the ringdown frequency, which can be read off from Fig. 9 in [58]. There is also a function available in LALSUITE called `XLALSimIMRPhenomDGetPeakFreq` that calculates the peak frequency given component masses and dimensionless spins in the direction of orbital angular momentum χ_1 and χ_2. (The function gives $Mf_{\rm peak} = 0.078$ for $\chi_1 = \chi_2 = -0.95$.)

The representation of the late-inspiral regime is inspired by the PN expansion and introduces 5 additional phenomenological "PN" coefficients σ_i making the Ansatz for the complete inspiral regime

$$\psi_{\rm ins}(f) = \psi_{\rm SPA}(f) + \frac{1}{\eta}\left(\sigma_0 + \sigma_1 f + \frac{3}{4}\sigma_2 f^{4/3} + \frac{1}{2}\sigma_4 f^2\right). \tag{1.119}$$

The intermediate regime is modeled as

$$\psi_{\rm int}(f) = \frac{1}{\eta}\left(\beta_0 + \beta_1 f + \beta_2 \ln(f) - \frac{\beta_3}{3} f^{-3}\right), \tag{1.120}$$

and the merger-ringdown (MR) as

$$\psi_{\rm MR}(f) = \frac{1}{\eta}\left\{\alpha_0 + \alpha_1 f - \alpha_2 f^{-1} + \frac{3}{4}\alpha_3 f^{3/4} + \alpha_4 \tan^{-1}\left(\frac{f - \alpha_5 f_{\rm RD}}{f_{\rm damp}}\right)\right\}. \tag{1.121}$$

The functions of frequency with which the various coefficients are multiplied are summarized in Table 1.1 for future reference.

The final phase of the complete IMR waveform in Eq. (1.114) is constructed by joining the piecewise regions with step functions defined as

$$\theta^{\pm}_{f_0}[1 \pm \theta(f - f_0)], \tag{1.122}$$

where

$$\theta(f - f_0) = \begin{cases} -1 & \text{for } f < f_0, \\ 1 & \text{for } f \geq f_0. \end{cases} \tag{1.123}$$

The coefficients $\alpha_0, \alpha_1, \beta_0, \beta_1$ are fixed by the requirement of $C(1)$ continuity. Furthermore, σ_0 and φ_5 can be absorbed into an arbitrary phase shift and σ_1 into an

Table 1.1: Overview of the various PN and phenomenological phasing coefficients that appear in IMRPhenomPv2. Each parameter p_i multiplies a particular function of frequency in the waveform model which is listed in the f-dependence column. The a and b in α_4 are functions of the component masses and spins, see [32]. The parameters that are degenerate with t_c or ϕ_c, *i.e.* multiplying f or not being frequency dependent, are left out. Finally, α_5 is also left out since it enters in the same term as α_4 (it is part of b).

waveform regime	parameter	$f-$dependence
early-inspiral	φ_0	$f^{-5/3}$
	φ_1	$f^{-4/3}$
	φ_2	f^{-1}
	φ_3	$f^{-2/3}$
	φ_4	$f^{-1/3}$
	φ_{5l}	$\log f$
	φ_6	$f^{1/3}$
	φ_{6l}	$f^{1/3}\log f$
	φ_7	$f^{2/3}$
late-inspiral	σ_2	$f^{4/3}$
	σ_3	$f^{5/3}$
	σ_4	f^2
intermediate	β_2	$\log f$
	β_3	f^{-3}
merger-ringdown	α_2	f^{-1}
	α_3	$f^{3/4}$
	α_4	$\tan^{-1}(af+b)$

arbitrary time shift. The full IMR phase is now given by

$$\Phi_{\rm IMR} = \psi_{\rm ins}(f)\theta^-_{f_1} + \theta^+_{f_1}\psi_{\rm int}(f)\theta^-_{f_2} + \theta^+_{f_2}\psi_{\rm RD}(f), \tag{1.124}$$

with $Mf_1 = 0.018$ and $f_2 = 0.5 f_{\rm RD}$. Note that the known PN inspiral part of the phase $\psi_{\rm ins}$ is calculated using the full two-spin dependence.

So far we have introduced two phenomenological IMR aligned-spin frequency domain models, but in order to perform reliable tests of general relativity using binary black hole sources, the inclusion of non-aligned spins becomes more important as such systems can exhibit orbital precession. In aligned-spin systems the orbital angular momentum $\mathbf{L}$ is fixed and makes an angle ι with the observer. For the more general case $\mathbf{L}$ can precess around the direction to the observer, forming a kind of precession cone. The spread, or opening-angle of this cone can be time-dependent and the entire cone may in itself be time-dependent. This precession adds complicated dynamics to the system and has a major impact on the GW evolution of the system.

In the final part of the discussion on waveform models we will introduce the phenomenological frequency domain model that allows for generic spins and is accurate for a wide range in parameter values, dubbed IMRPhenomP. There are in fact two types: IMRPhenomPv1 and IMRPhenomPv2. Both models essentially "twist up" an existing IMR waveform to accommodate effects of precession. The difference between versions is only that v1 is built on IMRPhenomC, while v2 is built on IMRPhenomD. As the principle for the two variants is similar, we will only mention IMRPhenomPv2 henceforth.

The authors of [59, 60, 61] introduce a framework to construct precessing waveforms from existing IMR spin-aligned waveforms, based on their earlier work in [62], where they introduced the quadrupole aligned (QA) frame. The QA frame approximately follows the instantaneous orbital plane of the binary. In this co-precessing frame, the binary is continuously viewed face-on and the amplitude and phase behavior can be described by (anti-)aligned spin models such as IMRPhenomD. The main novel feature introduced in [59] is a simple mapping between a fully precessing frame and non-precessing waveforms. This mapping may be written as

$$h^{\rm prec}_{lm} = \left(\mathbf{R} h^{\rm aligned}\right)_{lm} \tag{1.125}$$

where $\mathbf{R}$ is a rotation matrix built with Euler rotations.

To compute the GW polarizations it is convenient to adopt a time-independent Cartesian coordinate system attached to the binary source. For non-precessing systems a widely used choice is $\hat{\mathbf{L}} \equiv \hat{z}$. However, for precessing systems this is no longer convenient as $\mathbf{L}$ is now time-dependent. Instead one uses the total angular momentum $\mathbf{J} = \mathbf{L} + \mathbf{S}_1 + \mathbf{S}_2$, which remains approximately fixed throughout the binary evolution and one defines the z direction to be along $\hat{\mathbf{J}}$ at a certain reference frequency $f_{\rm ref}$, *i.e.* $\hat{\mathbf{J}}(f_{\rm ref}) \equiv \hat{z}$, which is depicted in Fig. 1.8.

Next to the masses m_1, m_2 of the binary components, there are 5 additional parameters that together describe the dynamics of precessing systems in the IMRPhenomPv2 model. The first is θ_J, a generalization of the inclination angle ι and is the angle between the total angular momentum $\mathbf{J}$ and the direction to the observer $\hat{N}$. The second is the Euler angle $\alpha(t)$ at some reference frequency $f_{\rm ref}$: α_0. The last three,

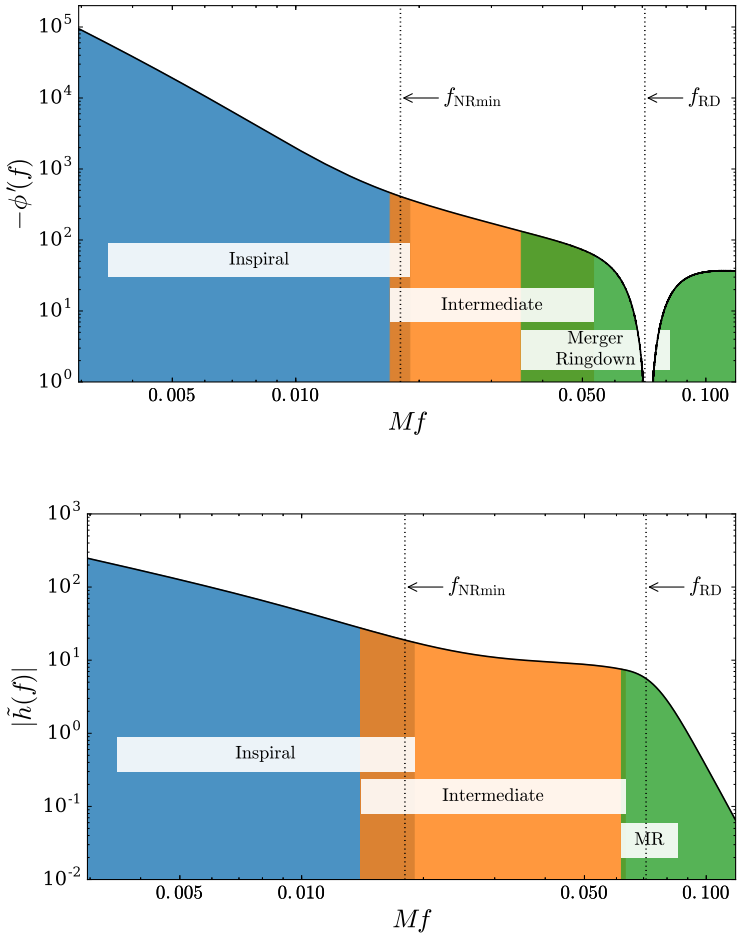


Figure 1.7: Frequency regions of the parameterized waveform model as defined in [32]. The top panel shows the derivative of the phase $-\phi'(f) \equiv -\partial\phi(f)/\partial f$ and the bottom panel the amplitude for an example waveform where $q = 1$, $\chi_1 = \chi_2 = -0.95$. The frequency ranges that were used in the fits for each region are represented by the span of the white boxes. For reference I have included the lowest frequency that was available in the NR waveform calibration set$Mf_{\rm NRmin} = 0.018$ and the ringdown frequency of the 22 mode $Mf_{\rm RD} = 0.071$.

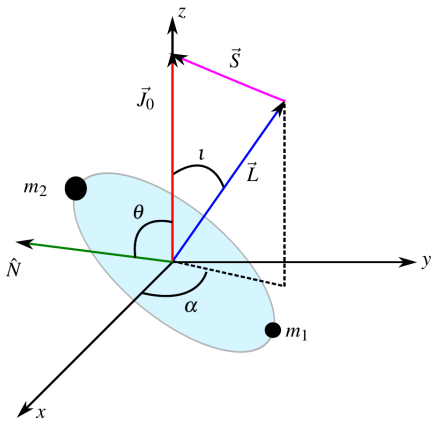


Figure 1.8: The source frame of IMRPhenomPv2 aligned with the total angular momentum $\hat{\mathbf{J}}_0$ at some reference frequency $f_{\rm ref}$. The direction to the observer is indicated with $\hat{N}$. Figure is adapted with permission from [61].

χ_{1L}, χ_{2L} and χ_p, together form the representation of the component spins. χ_{iL} are the projections of the component spins along the $\hat{\mathbf{L}}$ direction and χ_p is the "effective" precessing spin parameter

$$\chi_p = \frac{\max(A_1 m_1^2 S_{1\perp}, A_2 m_2^2 S_{2\perp})}{A_1 m_1^2}, \tag{1.126}$$

where $S_{i\perp}$ are the spin projections onto the orbital plane and

$$\begin{aligned} A_1 &= 2 + \frac{3}{2}\frac{m_2}{m_1}, \\ A_2 &= 2 + \frac{3}{2}\frac{m_1}{m_2}. \end{aligned} \tag{1.127}$$

The effective precessing spin parameter χ_p, when applied to the heavier component m_1, captures the average precession of the system with all in-plane spin components [61].

There are some simplifications to this model that need to be taken into consideration when utilizing it in a data analysis scenario: (i) $\hat{\mathbf{J}}$ is assumed to be constant and (ii) the angle between $\hat{\mathbf{J}}$ and $\hat{\mathbf{L}}$ must be small. This means that IMRPhenomPv2 is not expected to be accurate for systems where $\mathbf{J} \sim 0$ as this results in so-called transitional precession in which $\mathbf{J}$ flips its orientation [53]. This effect is observed for highly anti-aligned spins, moderate mass ratios and small χ_p. The model might also be inaccurate for high mass ratios and large χ_p in which case the angle between $\hat{\mathbf{J}}$ and $\hat{\mathbf{L}}$ can be large.

1.5 Data analysis and parameter estimation

We have so far described how gravitational waves from compact binary sources are detected and what such signals look like given certain parameters. Let us now discuss the inverse challenge of inferring parameter values given a particular signal.

When considering a gravitational wave source described by its extrinsic parameters θ, ϕ, ψ and distance r and a set of intrinsic parameters $\boldsymbol{\theta}$ such as component masses and spins, the gravitational wave strain in the time domain can be written as

$$h(t;\boldsymbol{\lambda}) = F_+(\theta,\phi,\psi)h_+(t;\phi_c,t_c,r,\iota,\boldsymbol{\theta}) + F_\times(\theta,\phi,\psi)h_\times(t;\phi_c,t_c,r,\iota,\boldsymbol{\theta}), \tag{1.128}$$

where ϕ_c and t_c are coalescence phase and time respectively and $\boldsymbol{\lambda}$ represents the set of all intrinsic and extrinsic parameters. The particular parameter estimation pipeline of our choice operates in the frequency domain. The strain in the frequency domain can be obtained by performing a Fourier transform

$$\begin{aligned}\tilde{h}(f;\boldsymbol{\lambda}) &= \int_{-\infty}^{\infty} \mathrm{d}t\, h(t;\boldsymbol{\lambda})e^{2\pi i f t} \\ &= F_+(\theta,\phi,\psi)\tilde{h}_+(f;\phi_c,t_c,r,\iota,\boldsymbol{\theta}) + F_\times(\theta,\phi,\psi)\tilde{h}_\times(t;\phi_c,t_c,r,\iota,\boldsymbol{\theta}) \\ &= e^{-2\pi i f t_c}\left[F_+(\theta,\phi,\psi)\tilde{h}_+(f;\phi_c,0,r,\iota,\boldsymbol{\theta}) + F_\times(\theta,\phi,\psi)\tilde{h}_\times(f;\phi_c,0,r,\iota,\boldsymbol{\theta})\right],\end{aligned} \tag{1.129}$$

where the last equality follows from $h(t;t_c) = h(t-t_c;0)$; a non-zero t_c offsets the signal's time of arrival.

1.5.1 Bayesian inference

Our preferred method of inference is based on Bayes' theorem and follows directly from the basic algebra of probability theory

$$P(A|I) + P(\overline{A}|I) = 1, \tag{1.130a}$$
$$P(A,B|I) = P(A|B,I)P(B|I), \tag{1.130b}$$

with $P(\text{false}) = 0$ and $P(\text{true}) = 1$. In this notation $P(A|B)$ is read as "the probability that A is true, given B". I encodes the relevant background information at hand and is always implied regardless of whether it appears in the equation or not as there is no such thing as an absolute probability. From Eqs. (1.130), one can construct Bayes' theorem

$$P(A|B,I) = \frac{P(A|I)P(B|A,I)}{P(B|I)}. \tag{1.131}$$

Observe that Bayes' theorem allows one to relate the probabilities $P(A|B,I)$ and $P(B|A,I)$. To understand the significance of this let us write Eq. (1.131) in terms of a certain hypothesis $\mathcal{H}$ and (observational) data d

$$P(\mathcal{H}|d,I) = \frac{P(\mathcal{H}|I)P(d|\mathcal{H},I)}{P(d|I)}. \tag{1.132}$$

It is now clear that we may find the quantity of interest $P(\mathcal{H}|d,I)$ also referred to as *posterior probability* – that is the probability that the hypothesis $\mathcal{H}$ is correct given the data – from a probability we are better able to assign $P(d|\mathcal{H},I)$; the probability that one obtains the data d given hypothesis $\mathcal{H}$, also called the *likelihood function*. The likelihood function, assuming Gaussian noise, can be written directly as a function of the data d, the signal h that belongs to hypothesis $\mathcal{H}$ and the parameters $\boldsymbol{\lambda}$ it is described with

$$P(d|\mathcal{H},\boldsymbol{\lambda},I) \propto e^{-\frac{1}{2}(d-h(\boldsymbol{\lambda})|d-h(\boldsymbol{\lambda}))}, \tag{1.133}$$

where $(a|b)$ is the noise-weighted inner product defined in Eq. (1.76).

The *prior* probability $P(\mathcal{H}|I)$ encodes our prior knowledge (or lack thereof) in favor of $\mathcal{H}$ being correct and the *marginal probability* $P(d|I)$ acts as an overall weighting factor. Note that $P(d|I) = \sum_j P(d|\mathcal{H}_j,I)$ can only be calculated if there exists a complete set of independent hypotheses $\mathcal{H}_j$, such that $\sum_j P(\mathcal{H}_j|d,I) = 1$. In practice, finding a complete set of independent hypotheses to calculate $P(d|I)$ can be difficult or even impossible, it is however still possible to consider relative probabilities, or odds ratios, between two competing hypotheses

$$\mathcal{O}^{\mathcal{H}_i}_{\mathcal{H}_j} = \frac{P(\mathcal{H}_i|I)}{P(\mathcal{H}_j|I)}\frac{P(d|\mathcal{H}_i,I)}{P(d|\mathcal{H}_j,I)} = \frac{P(\mathcal{H}_i|I)}{P(\mathcal{H}_j|I)}B^{\mathcal{H}_i}_{\mathcal{H}_j}, \tag{1.134}$$

where

$$B^{\mathcal{H}_i}_{\mathcal{H}_j} = \frac{P(d|\mathcal{H}_i,I)}{P(d|\mathcal{H}_j,I)} \tag{1.135}$$

is called the Bayes factor and we note that $P(d|I)$ has dropped out. In parameter estimation this quantity is unimportant as it acts merely as a normalization constant.

Another important concept is *marginalization*; a powerful tool to *e.g.* integrate out quantities that necessarily enter the analysis, but are of no intrinsic interest (nuisance parameters). An example of where marginalization comes in is the calculation of Bayes factors (Eq. (1.135)); this can be directly found for hypotheses that do not depend on any free parameters, but generally this is not the case. For gravitational waves the hypothesis typically corresponds to a certain waveform model, which as explained in Sec. 1.4 depends on several free parameters $\boldsymbol{\lambda}$. In this case the likelihood of $\mathcal{H}$ ($p(d|\mathcal{H},\boldsymbol{\lambda},I)$) must be marginalized over all parameters in $\boldsymbol{\lambda} \in \Lambda$, where Λ is the parameter space, weighted by their prior probability distribution $p(\boldsymbol{\lambda}|\mathcal{H},I)$, which gives the *evidence* (also know as marginal likelihood)

$$Z \equiv P(d|\mathcal{H},I) = \int_\Lambda \mathrm{d}\boldsymbol{\lambda}\, p(\boldsymbol{\lambda}|\mathcal{H},I)p(d|\mathcal{H},\boldsymbol{\lambda},I). \tag{1.136}$$

The integral in Eq. (1.136) can typically not be evaluated analytically, except for the most trivial cases. The next section will introduce the algorithm that is used to calculate integrals of this sort, in our data analysis efforts.

1.5.2 Parameter estimation with nested sampling

We just introduced the basic principles that are at the core of the parameter estimation pipeline employed throughout this thesis called LALInference [63, 64]. The LALInference pipeline consists of several modules of which we are particularly interested in the nested sampling engine. The very useful feature of nested sampling is that as it calculates the evidence in Eq. (1.136), one obtains the *posterior density function* (PDF)

$$p(\boldsymbol{\lambda}|d,\mathcal{H},I) = \frac{p(\boldsymbol{\lambda}|\mathcal{H},I)p(d|\boldsymbol{\lambda},\mathcal{H},I)}{p(d|\mathcal{H},I)}, \tag{1.137}$$

with negligible additional computational costs. Here, $p(\boldsymbol{\lambda}|d,\mathcal{H},I)$ is the joint PDF on the multi-dimensional space and describes the collective knowledge pertaining to all parameters as well as their relationships. The PDF for a single parameter can be found by marginalizing over the rest: Let $\boldsymbol{\lambda} = \{\lambda_1, \lambda_2, \ldots, \lambda_N\}$, then e.g.

$$p(\lambda_1|d,\mathcal{H},I) = \int \mathrm{d}\lambda_2 \mathrm{d}\lambda_3 \ldots \mathrm{d}\lambda_N \, p(\boldsymbol{\lambda}|d,\mathcal{H},I). \tag{1.138}$$

To calculate the evidence Z and in turn obtain posterior samples, the multi-dimensional integral in Eq. (1.136) is turned into a one dimensional integral using the *nested sampling* method introduced in [65]. The evidence integral can be written as a one-dimensional integral over the prior volume X

$$Z = \int_\Lambda \mathrm{d}X \, p(d|\mathcal{H},\boldsymbol{\lambda},I), \tag{1.139}$$

where $\mathrm{d}X \equiv \mathrm{d}\boldsymbol{\lambda} p(\boldsymbol{\lambda}|\mathcal{H},I)$. In this notation,

$$X(\ell) = \int_{p(d|\mathcal{H},\boldsymbol{\lambda},I)>\ell} \mathrm{d}X, \tag{1.140}$$

where $X(\ell)$ is the total probability volume contained within a likelihood contour $\ell = p(d|\mathcal{H},\boldsymbol{\lambda},I)$. If we now define $L(X)$ as the inverse of Eq. (1.140), we can write Eq. (1.136) as

$$Z = \int_0^1 \mathrm{d}X \, L(X), \tag{1.141}$$

where $L(X)$ is a monotonically increasing function. This equation becomes more intuitive as we evaluate discrete likelihoods $L_i = L(X_i)$ where X_i is now a monotonically decreasing sequence of prior volumes X_i,

$$1 = X_0 > X_1 > X_2 > \cdots > X_M > 0. \tag{1.142}$$

Eq. (1.141) can now be approximated by the trapezium rule

$$Z \simeq \sum_{i=1}^{M} w_i L_i, \tag{1.143}$$

where

$$w_i = \frac{1}{2}\left(X_{i-1} - X_{i+1}\right), \tag{1.144}$$

is the fraction of the prior distribution represented by the i-th sample.

We can now introduce the nested sampling procedure that evaluates Eq. (1.143) with the help of a schematic representation of likelihood contours L_i and corresponding prior volumes X_i in Fig. 1.9. One begins by taking an initial set of so-called "live points" that are drawn from the entire prior distribution. The point with the lowest likelihood is removed and replaced by a new point drawn from the prior that has a higher likelihood. Each of these points of lowest likelihood can be thought of as lying on a contour of constant likelihood as indicated by the solid curves in Fig. 1.9. In turn, the volume enclosed by L_i corresponds to a fraction of prior volume X_i still to be explored. The removing and replacing of points continues until some stopping condition is met, which in LALInference is $L_{\max}X_i/Z_i < e^{0.1}$ by default, where $L_{\max}$ is the maximum likelihood so far discovered by the sampling process and Z_i and X_i are the current estimate of the evidence already accumulated and fraction of prior volume inside the current contour respectively.

Finally, the posterior samples can be obtained by re-sampling the chain of removed points according to their weights

$$p(\boldsymbol{\lambda}_i|d,\mathcal{H},I) = \frac{w_i L_i}{Z}, \tag{1.145}$$

where the collection of all $p(\boldsymbol{\lambda}_i|d,\mathcal{H},I)$ constitutes the PDF for the entire set of parameters $p(\boldsymbol{\lambda}|d,\mathcal{H},I)$. Again, an individual posterior can easily be obtained by marginalizing over the remaining parameters using Eq. (1.138).

There are different methods to determine new live points. The nested sampling algorithm in LALInference includes for example an MCMC sampling of the prior distribution constrained by the likelihood contours. As the algorithm progresses, typically more MCMC points are required to find a new point in parameter space that has a likelihood larger than the one that is replaced and for each such MCMC point, the likelihood needs to be evaluated. In a full LALInference run, depending on the type of source and its strength with respect to the noise, this amounts to about $10^7 - 10^8$ likelihood evaluations. Needless to say, this can become a computational bottleneck in parameter estimation efforts and in Chapter 5 a method to significantly increase the speed of likelihood evaluations is introduced and extended to a specific parameter estimation effort, introduced in the next section, to perform tests of general relativity.

1.6 TIGER

The test infrastructure for general relativity (TIGER) [66, 67, 68, 69] combines the strengths of two worlds: Nested sampling and Bayesian model selection. The goal is to build an infrastructure or pipeline that can perform *model independent* tests of general relativity. GR has passed many tests already with incredible accuracy, but

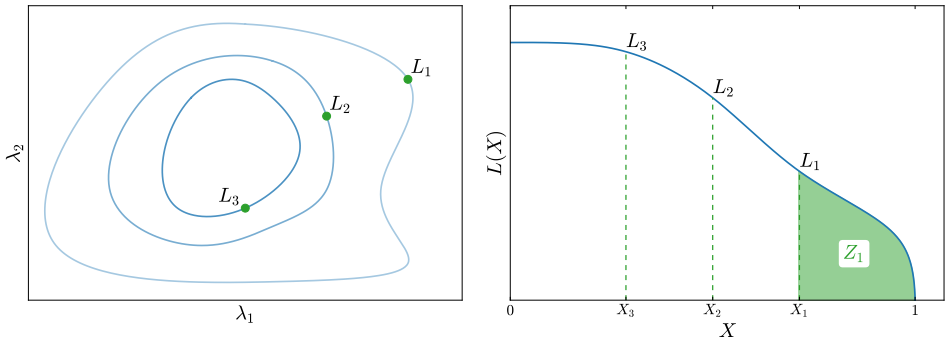


Figure 1.9: A two dimensional visualization of the nested sampling process. The left panel illustrates what the likelihood contours in this two dimensional parameter space might look like. The likelihood enclosed within each contour is such that $L > L_i$, or in practice, all remaining live points have a likelihood higher than L_i. The right panel shows how the likelihood as function of the prior volume $L(X)$ progresses as one moves up the likelihood surface; the remaining prior volume X_i decreases as the likelihood L_i increases. Also shown in the right plot is the evidence Z_i for $i = 1$, which is the integral under the curve and increases as the process proceeds.

as discussed in Sec. 1.2, all of these were performed in relatively weak gravitational fields and slowly evolving systems. This means that thus far the genuinely strong-field non-linear dynamics of GR was unattainable. With the start of O1 (Sec. 1.7) gravitational wave signals originating from BBH mergers now allow us access to this regime for the first time. TIGER has the ability to use these GW signals to test GR in the genuinely strong-field regime without assuming any particular alternative theory, thus arriving at a model independent test.

Waveforms that model GW signals originating from the sources we are sensitive to in aLIGO and adVirgo can be anything from numerically evaluated to post-Newtonian approximations. For TIGER we require waveform models that can be extended with so-called non-GR parameters and that are quick to evaluate. In this section I will present the TIGER pipeline as it was designed for BNS sources, but note that the framework is general and any waveform model that allows additional non-GR parameterization can be plugged in. In chapters 3 and 4 I will focus on TIGER applied to BBH sources and discuss different choices of waveform models.

For TIGER-BNS the waveform model of choice (at least for recovery) is TaylorF2. This closed-form analytical frequency domain waveform is quick to evaluate and easily lends itself to non-GR parameterization. To allow for generic deviations from GR we introduce the non-GR PN shifts $\delta\hat{\varphi}_i$, which act upon each of the GR PN coefficients $\varphi_i^{\rm GR}$ (Eq. (1.108)) as follows:

$$\varphi_i^{\rm GR}(m_1, m_2, \mathbf{S}_1, \mathbf{S}_2) \to \begin{cases} (1 + \delta\hat{\varphi}_i)\varphi_i^{\rm GR}(m_1, m_2, \mathbf{S}_1, \mathbf{S}_2) & \text{if} \quad i \neq 1 \\ \delta\hat{\varphi}_i & \text{if} \quad i = 1 \end{cases}, \tag{1.146}$$

where $i \in \{0, 1, 2, 3, 4, 5l, 6, 6l, 7\}$ and $\delta\hat{\varphi}_1$ is treated special as in GR $\varphi_1 = 0$. In this parameterization GR is represented by all $\delta\hat{\varphi}_i$ being identically zero. Any alternative theory of gravity will have some influence on how the various $\delta\hat{\varphi}_i$ depend on the source parameters such as masses and spins. Using the non-GR parameters these variations can only be captured by adjusting the $\delta\hat{\varphi}_i$ away from zero. Provided these deviations are strong enough, measurements on the $\delta\hat{\varphi}_i$ would indicate a violation of GR as the posteriors will not be centered on zero. This principle is what we will later in chapter 4 refer to as the parameterized test of GR. Of course one also needs to take the effects of noise into account. Any detector noise can mimic a violation of GR and one needs to know when one is dealing with a true violation of GR or with a mere effect of noise. Combining parameter posteriors of multiple detections is an option, but only if the parameter in question is known to be constant across all sources, *i.e.* it does not depend on source parameters such as masses or spins. For the non-GR parameters there is no such guarantee. This is where the strength of the TIGER framework comes in. Instead of measuring posteriors on each of the $\delta\hat{\varphi}_i$, which may vary from source to source and depend strongly on detector noise and SNR, we instead perform model selection. With TIGER we set out to quantify our belief that GR is correct vs. the case where some violation is present. This quantity is called the *odds ratio* and is defined as the evidence for the hypothesis stating that GR is false $\mathcal{H}_{\rm modGR}$ divided by the evidence for the GR hypothesis $\mathcal{H}_{\rm GR}$:

$$\mathcal{O}^{\rm modGR}_{\rm GR} \equiv \frac{P(\mathcal{H}_{\rm modGR}|d, I)}{P(\mathcal{H}_{\rm GR}|d, I)}. \tag{1.147}$$

If we were dealing with a specific non-GR theory, it would be trivial to define $\mathcal{H}_{\rm modGR}$, it would simply be the evidence for one particular waveform model which is calculable. We do however not have any prior assumption on the nature of a possible violation of GR and $\mathcal{H}_{\rm modGR}$ can not be described by one particular waveform model. Instead the modGR hypothesis encodes the general statement "One or more of the non-GR parameters $\delta\hat{\varphi}_i$ deviate from zero". Note that this does not specify *which* parameters deviate from zero. We can however calculate evidences for sub-hypotheses $H_{i_1 i_2 \ldots i_k}$:

> $H_{i_1 i_2 \ldots i_k}$ states that the non-GR parameters $\delta\hat{\varphi}_{i_1}, \ldots, \delta\hat{\varphi}_{i_k}$ deviate from zero, but all other non-GR parameters $\delta\hat{\varphi}_j$, where $j \notin \{i_1, \ldots, i_k\}$ are identically zero.

These sub-hypotheses do correspond to a particular waveform model. For example H_{12} states that $\delta\hat{\varphi}_1 \neq 0$ and $\delta\hat{\varphi}_2 \neq 0$, but all the other $\delta\hat{\varphi}_i$, where $i \neq \{1, 2\}$, are zero. This means that we have a specific waveform to test H_{12}, namely one in which $\delta\hat{\varphi}_1$ and $\delta\hat{\varphi}_2$ are the only additional (as in aside from the usual masses, spins, etc.) free parameters. Note that the definition of the sub-hypotheses $H_{i_1 i_2 \ldots i_k}$ ensures that they are *mutually, logically disjoint*, *i.e.* $H_{i_1 i_2 \ldots i_k} \wedge H_{j_1 j_2 \ldots j_l}$ is always false for $\{i_1, \ldots, i_k\} \neq \{j_1, \ldots, j_l\}$. For example $H_1 \wedge H_2$ is always false since H_1 assumes that $\delta\hat{\varphi}_1$ is not zero whereas H_2 assumes that it is. The hypothesis $\mathcal{H}_{\rm modGR}$ states that one or more of the $\delta\hat{\varphi}_i$ deviate from zero, without specifying which. Using the definition of the sub-hypotheses, this is simply the logical or between these

$$\mathcal{H}_{\rm modGR} = \bigvee_{i_1 < i_2 < \cdots < i_k} H_{i_1 i_2 \ldots i_k}. \tag{1.148}$$

The goal is now to split the odds ratio in Eq. (1.147) into parts that we can evaluate. To illustrate how the sub-hypotheses are used to do this, we will first consider only two non-GR parameters: $\delta\hat{\varphi}_1$ and $\delta\hat{\varphi}_2$. In this case we have three sub-hypotheses: H_1, H_2 and H_{12} and the modGR hypothesis reads

$$\mathcal{H}_{\rm modGR} = H_1 \vee H_2 \vee H_{12}. \tag{1.149}$$

The odds ratio can now be written as

$${}^{(2)}\mathcal{O}^{\rm modGR}_{\rm GR} = \frac{P(H_1 \vee H_2 \vee H_{12}|d,I)}{P(\mathcal{H}_{\rm GR}|d,I)}, \tag{1.150}$$

where the superscript $^{(2)}$ indicates that we are only using two parameters for testing. As noted before, the sub-hypotheses are logically disjoint and we may split the odds ratio into three separate parts

$${}^{(2)}\mathcal{O}^{\rm modGR}_{\rm GR} = \frac{P(H_1|d,I)}{P(\mathcal{H}_{\rm GR}|d,I)} + \frac{P(H_2|d,I)}{P(\mathcal{H}_{\rm GR}|d,I)} + \frac{P(H_{12}|d,I)}{P(\mathcal{H}_{\rm GR}|d,I)}. \tag{1.151}$$

Using Bayes' theorem we can write this as

$${}^{(2)}\mathcal{O}^{\rm modGR}_{\rm GR} = \alpha^1_{\rm GR}B^1_{\rm GR} + \alpha^2_{\rm GR}B^2_{\rm GR} + \alpha^{12}_{\rm GR}B^{12}_{\rm GR}, \tag{1.152}$$

where

$$\begin{aligned} B^1_{\rm GR} &= \frac{P(d|H_1,I)}{P(d|\mathcal{H}_{\rm GR},I)}, \\ B^2_{\rm GR} &= \frac{P(d|H_2,I)}{P(d|\mathcal{H}_{\rm GR},I)}, \\ B^{12}_{\rm GR} &= \frac{P(d|H_{12},I)}{P(d|\mathcal{H}_{\rm GR},I)}, \end{aligned} \tag{1.153}$$

are the Bayes factors and

$$\begin{aligned} \alpha^1_{\rm GR} &= \frac{P(H_1|I)}{P(\mathcal{H}_{\rm GR}|I)}, \\ \alpha^2_{\rm GR} &= \frac{P(H_2|I)}{P(\mathcal{H}_{\rm GR}|I)}, \\ \alpha^{12}_{\rm GR} &= \frac{P(H_{12}|I)}{P(\mathcal{H}_{\rm GR}|I)}, \end{aligned} \tag{1.154}$$

are ratios of prior odds. To evaluate the Bayes factors we employ the nested sampling algorithm in LALInference discussed in Sec. 1.5.2. From this algorithm one does not obtain $B^{i_1 i_2 \ldots i_k}_{\rm GR}$ directly, but rather the Bayes factors between the various sub-

hypothesis vs. the noise-only hypothesis $\mathcal{H}_{\rm noise}$ and GR vs. noise $B^{\rm GR}_{\rm noise}$:

$$
\begin{aligned}
B^{1}_{\rm noise} &= \frac{P(d|H_1, I)}{P(d|\mathcal{H}_{\rm noise}, I)}, \\
B^{2}_{\rm noise} &= \frac{P(d|H_2, I)}{P(d|\mathcal{H}_{\rm noise}, I)}, \\
B^{12}_{\rm noise} &= \frac{P(d|H_{12}, I)}{P(d|\mathcal{H}_{\rm noise}, I)}, \\
B^{\rm GR}_{\rm noise} &= \frac{P(d|\mathcal{H}_{\rm GR}, I)}{P(d|\mathcal{H}_{\rm noise}, I)}.
\end{aligned}
\tag{1.155}
$$

These Bayes factors can be trivially combined to obtain the ones in Eq. (1.152):

$$
\begin{aligned}
B^{1}_{\rm GR} &= \frac{B^{1}_{\rm noise}}{B^{\rm GR}_{\rm noise}}, \\
B^{2}_{\rm GR} &= \frac{B^{2}_{\rm noise}}{B^{\rm GR}_{\rm noise}}, \\
B^{12}_{\rm GR} &= \frac{B^{12}_{\rm noise}}{B^{\rm GR}_{\rm noise}},
\end{aligned}
\tag{1.156}
$$

Now we wish to return to Eq. (1.152) to calculate the odds ratio by combining the sub-hypotheses. In order to do this we need to specify the prior odds for each model against GR $\alpha^{i}_{\rm GR}$ as defined in Eq. (1.154). We will not know a priori what form a violation from GR will take and we do not wish to down-weight any sub-hypothesis with respect to any other. This is achieved by letting all the prior ratios be equal, i.e.

$$
\alpha^{1}_{\rm GR} = \alpha^{2}_{\rm GR} = \alpha^{12}_{\rm GR}. \tag{1.157}
$$

We then define

$$
\frac{P(\mathcal{H}_{\rm modGR}|I)}{P(\mathcal{H}_{\rm GR}|I)} = \frac{P(H_1 \vee H_2 \vee H_{12}|I)}{P(\mathcal{H}_{\rm GR}|I)} \equiv \alpha, \tag{1.158}
$$

so that $\alpha^{1}_{\rm GR} = \alpha^{2}_{\rm GR} = \alpha^{12}_{\rm GR} = \alpha/3$ and we find the final expression for the odds ratio to be

$$
{}^{(2)}\mathcal{O}^{\rm modGR}_{\rm GR} = \frac{\alpha}{3}\left(B^{1}_{\rm GR} + B^{2}_{\rm GR} + B^{12}_{\rm GR}\right). \tag{1.159}
$$

We do not need to specify the value of α since this will end up being an overall scaling of the odds ratio. Furthermore, the odds ratio will later be compared to odds ratios calculated for many GR sources (called the GR background) which have the same scaling factor.

The two-parameter example discussed so far carries over trivially to a generalized case where one considers a larger set of parameters $\{\delta\hat{\varphi}_{i_1}, \ldots, \delta\hat{\varphi}_{i_{N_T}}\}$ with $N_T \leq M$, where M is the number of phasing coefficients φ_i our model has. In the general case the modGR hypothesis becomes

$$
\mathcal{H}_{\rm modGR} = \bigvee_{i_1 < i_2 < \cdots < i_k; k \leq N_T} H_{i_1 i_2 \ldots i_k}, \tag{1.160}
$$

and the odds ratio

$$ {}^{(N_T)}\mathcal{O}^{\rm modGR}_{\rm GR} = \frac{P(\mathcal{H}_{\rm modGR}|d,I)}{P(\mathcal{H}_{\rm GR}|d,I)} = \frac{\bigvee_{i_1<i_2<\cdots<i_k;k\leq N_T} P(H_{i_1 i_2 \ldots i_k}|d,I)}{P(\mathcal{H}_{\rm GR}|d,I)}, \tag{1.161} $$

where as discussed previously, $H_{i_1 i_2 \ldots i_k}$ is the sub-hypothesis stating that $\{\varphi_{i_1}, \ldots, \varphi_{i_k}\}$ do not have the functional dependence on the source parameters as predicted by GR, but all of the remaining coefficients do. When we once more use the logical disjointness of the $2^{N_T}-1$ sub-hypotheses and apply Bayes' theorem, we can write

$$ {}^{(N_T)}\mathcal{O}^{\rm modGR}_{\rm GR} = \sum_{k=1}^{N_T} \sum_{i_1<i_2<\cdots<i_k} \alpha^{i_1 i_2 \ldots i_k}_{\rm GR} B^{i_1 i_2 \ldots i_k}_{\rm GR}, \tag{1.162} $$

where

$$ B^{i_1 i_2 \ldots i_k}_{\rm GR} = \frac{P(d|H_{i_1 i_2 \ldots i_k}, I)}{P(d|\mathcal{H}_{\rm GR}, I)} \quad \text{and} \quad \alpha^{i_1 i_2 \ldots i_k}_{\rm GR} = \frac{P(H_{i_1 i_2 \ldots i_k}|I)}{P(\mathcal{H}_{\rm GR}|I)}. \tag{1.163} $$

As we have done before in Eq. (1.157) we set the prior ratios equal to each other

$$ \alpha^{i_1 i_2 \ldots i_k}_{\rm GR} = \alpha^{j_1 j_2 \ldots j_k}_{\rm GR} \equiv \frac{\alpha}{2^{N_T}-1} \quad \text{for any} \quad k,l \leq N_T. \tag{1.164} $$

The final odds ratio for the general case can now be written as

$$ {}^{(N_T)}\mathcal{O}^{\rm modGR}_{\rm GR} = \frac{\alpha}{2^{N_T}-1} \sum_{k=1}^{N_T} \sum_{i_1<i_2<\cdots<i_k} B^{i_1 i_2 \ldots i_k}_{\rm GR}. \tag{1.165} $$

As mentioned before, TIGER makes use of LALInference to obtain the Bayes factors in Eq. (1.155). As these are calculated, we simultaneously obtain the marginalized posterior density (PDF) function given in Eq. (1.138), where in the case of TIGER $\boldsymbol{\theta} \to \{\boldsymbol{\theta}, \delta\hat{\varphi}_{i_1}, \delta\hat{\varphi}_{i_2}, \ldots, \delta\hat{\varphi}_{i_k}\}$ for sub-hypothesis $H_{i_1 i_2 \ldots i_k}$. At no additional cost we also obtain posteriors on not only the source parameters, but also on any non-GR parameter that was part of the sub-hypothesis.

Even though the TIGER pipeline can be applied to a single detection, its infrastructure is designed to make use of multiple detections. The following section extends the foundation discussed above to make use of multiple sources. Section 1.6.2 explains how the GR background distribution of odds ratios is an essential feature of the pipeline and 1.6.3 demonstrates the robustness of the pipeline against various nuisance effects or unmodeled physics.

1.6.1 Combining information from multiple detections

Taking into account multiple detections allows one to arrive at tighter constraints on PDFs as well as the validity of the GR hypothesis as a whole. Consider $\mathcal{N}$ independent GW detections each corresponding to an independent dataset d_A. We refer to such a

collection of individual detections as a *catalog*. The *combined* odds ratio for a catalog can be written as

$$
\begin{aligned}
{}^{(N_T)}\mathcal{O}^{\mathrm{modGR}}_{\mathrm{GR}} &= \frac{P(\mathcal{H}_{\mathrm{modGR}}|d_1,\ldots,d_{\mathcal{N},I})}{P(\mathcal{H}_{\mathrm{GR}}|d_1,\ldots,d_{\mathcal{N}},I)} \\
&= \frac{\sum_{k=1}^{N_T}\sum_{i_1<i_2<\cdots<i_k} P(H_{i_1 i_2\ldots i_k}|d_1,\ldots,d_{\mathcal{N},I})}{P(\mathcal{H}_{\mathrm{GR}}|d_1,\ldots,d_{\mathcal{N}},I)} \\
&= \sum_{k=1}^{N_T}\sum_{i_1<i_2<\cdots<i_k} \alpha^{i_1 i_2\ldots i_k\,(\mathrm{cat})}_{\mathrm{GR}} B^{i_1 i_2\ldots i_k}_{\mathrm{GR}},
\end{aligned}
\tag{1.166}
$$

where

$$
{}^{(\mathrm{cat})}B^{i_1 i_2\ldots i_k}_{\mathrm{GR}} = \frac{P(d_1,\ldots,d_{\mathcal{N}}|H_{i_1 i_2\ldots i_k},I)}{P(d_1,\ldots,d_{\mathcal{N}}|\mathcal{H}\mathrm{GR},I)}. \tag{1.167}
$$

The Bayes factor above depends on events $d_1,\ldots,d_{\mathcal{N}}$, which are all independent, thus

$$
\begin{aligned}
P(d_1,\ldots,d_{\mathcal{N}}|H_{i_1 i_2\ldots i_k},I) &= \prod_{A=1}^{\mathcal{N}} P(d_A|H_{i_1 i_2\ldots i_k},I), \\
P(d_1,\ldots,d_{\mathcal{N}}|\mathcal{H}_{\mathrm{GR}},I) &= \prod_{A=1}^{\mathcal{N}} P(d_A|H_{\mathrm{GR}},I).
\end{aligned}
\tag{1.168}
$$

This allows us to split the Bayes factor in Eq. (1.167) into individual Bayes factors ${}^{(A)}B^{i_1 i_2\ldots i_k}_{\mathrm{GR}}$ corresponding to event A:

$$
{}^{(\mathrm{cat})}B^{i_1 i_2\ldots i_k}_{\mathrm{GR}} = \prod_{A=1}^{\mathcal{N}} {}^{(A)}B^{i_1 i_2\ldots i_k}_{\mathrm{GR}}, \tag{1.169}
$$

where

$$
{}^{(A)}B^{i_1 i_2\ldots i_k}_{\mathrm{GR}} = \frac{P(d_A|H_{i_1 i_2\ldots i_k},I)}{P(d_A|\mathcal{H}_{\mathrm{GR}},I)}. \tag{1.170}
$$

As before in Eqns. (1.157) and (1.164) we do not a priori prefer one hypothesis over any other, *i.e.* we set the individual prior odds ratios to be equal and obtain

$$
\frac{P(H_{i_1 i_2\ldots i_k}|I)}{P(\mathcal{H}_{\mathrm{GR}}|I)} = \frac{\alpha}{2^{N_T}-1}. \tag{1.171}
$$

Combining the expressions in Eqns. (1.166) and (1.169) results in the final expression for the combined odds ratio

$$
{}^{(N_T)}\mathcal{O}^{\mathrm{modGR}}_{\mathrm{GR}} = \frac{\alpha}{2^{N_T}-1}\sum_{k=1}^{N_T}\sum_{i_1<i_2<\cdots<i_k}\prod_{A=1}^{\mathcal{N}} {}^{(A)}B^{i_1 i_2\ldots i_k}_{\mathrm{GR}}. \tag{1.172}
$$

In Eq. (1.172) we essentially calculate the average of the cumulative Bayes factors from Eq. (1.169) up to an overall scaling factor. So far we have considered the case

where we give all prior odds an equal weight. One can however also consider to investigate the behavior of each cumulative Bayes factor ${}^{(\mathrm{cat})}B^{i_1 i_2 \ldots i_k}_{\mathrm{GR}}$ individually. In this case, no assumptions on the prior odds ratios are made at all. Another option is to only consider the cumulative Bayes factors of individual hypotheses $H_{i_1 i_2 \ldots i_k}$ and the GR hypothesis $\mathcal{H}_{\mathrm{GR}}$ against the noise-only hypothesis $\mathcal{H}_{\mathrm{noise}}$

$$\prod_A {}^{(A)}B^{i_1 i_2 \ldots i_k}_{\mathrm{noise}} = \prod_A \frac{P(d_A|H_{i_1 i_2 \ldots i_k}, I)}{P(d_A|H_{\mathrm{noise}}, I)},$$
$$\prod_A {}^{(A)}B^{\mathrm{GR}}_{\mathrm{noise}} = \prod_A \frac{P(d_A|H_{\mathrm{GR}}, I)}{P(d_A|H_{\mathrm{noise}}, I)}. \tag{1.173}$$

Recall that we can obtain the Bayes factors of any hypothesis against GR as in Eq. (1.169), by a division of the individual Bayes factors against noise

$${}^{(\mathrm{cat})}B^{i_1 i_2 \ldots i_k}_{\mathrm{GR}} = \frac{\prod_A {}^{(A)}B^{i_1 i_2 \ldots i_k}_{\mathrm{noise}}}{\prod_A {}^{(A)}B^{\mathrm{GR}}_{\mathrm{noise}}}. \tag{1.174}$$

In Chapter 2 we not only consider the final combined odds ratio, but also visualize how the cumulative Bayes factors for individual sub-hypotheses evolve as a function of added sources in Fig. 2.5.

1.6.2 The GR background of odds ratios

The previous section described how an odds ratio for a single detection or a collection (catalog) of detections can be evaluated. This odds ratio by itself is not enough to either claim or disregard a possible violation from GR. In the absence of noise one can indeed say that when $\ln \mathcal{O}^{\mathrm{modGR}}_{\mathrm{GR}} > 0$ a violation is present in the detected signal; however, in reality one has to deal with noise in the detector data stream. Noise can distort a pure GR signal, causing the mod-GR hypothesis $\mathcal{H}_{\mathrm{modGR}}$ to be favored and $\ln \mathcal{O}^{\mathrm{modGR}}_{\mathrm{GR}}$ to be positive. How much the mod-GR hypothesis is favored even in the case of a GR signal depends on the nature of the noise. To use the odds ratio as an efficient way of spotting violations, one needs to compare it to a large distribution of typical values for $\ln \mathcal{O}^{\mathrm{modGR}}_{\mathrm{GR}}$ in the case GR signals are analyzed. This distribution is called the GR *background*.

Without considering the details of the detector noise itself, let us assume we have a continuous stream of detector output, devoid of signals, from time t_{start} to t_{end}. A GR background for such a stream of data can then be generated by injecting many – typically on the order of a few thousand – GR gravitational wave sources randomly distributed over a range in intrinsic source parameters and sky locations, orientations and distances. Each source is injected into a different, non-overlapping segment of the detector output, to consider as many different noise realizations as possible. For all the injected sources a corresponding odds ratio is calculated which are then histogrammed. We denote the background distribution as $P(\ln \mathcal{O}|\mathcal{H}_{\mathrm{GR}}, \kappa_{\mathrm{GR}}, I)$, where κ_{GR} is the set of waveforms. Such a background distribution is sketched in the left panel of Fig. 1.10. Shown here is a Gaussian representation of what a typical background for BNS sources looks like. Such cases were investigated in great detail in [68].

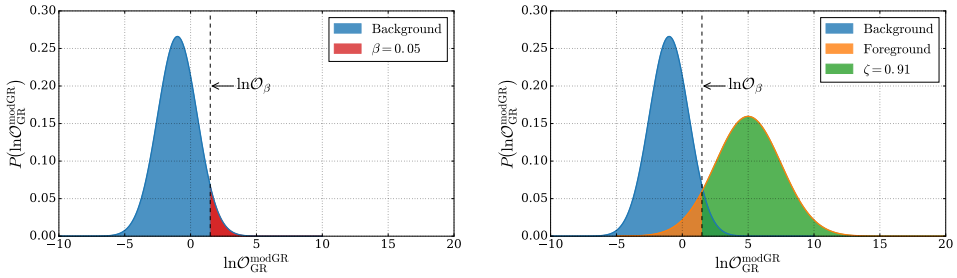


Figure 1.10: A schematic representation of a GR background distribution (blue) and non-GR foreground distribution (orange). The left panel indicates the maximum false alarm probability β and the corresponding threshold $\ln \mathcal{O}_\beta$. When a foreground is generated, an efficiency of recovering a violation may also be calculated. This efficiency ζ is shown in the right panel in green.

We now consider how such a background is used when a detection has been made and thus an odds ratio has been measured. When a background is generated with a sufficiently large amount of GR sources, it is safe to claim a violation of GR when the odds ratio is entirely above the background. We will however pick a *maximum tolerable false alarm probability* β with which we can compute a *threshold* odds ratio $\ln \mathcal{O}_\beta$ that a measured odds ratio needs to overcome. In the limit of infinitely many injections, $\ln \mathcal{O}_\beta$ is defined implicitly by

$$\beta = \int_{\ln \mathcal{O}_\beta}^{\infty} \mathrm{d}(\ln \mathcal{O})\, P(\ln \mathcal{O} | \mathcal{H}_{\rm GR}, \kappa_{\rm GR}, I). \tag{1.175}$$

The left panel of Fig. 1.10 shows the region corresponding to a fraction β of the background in red and the corresponding threshold $\ln \mathcal{O}_\beta$ is indicated by a vertical dashed line. Of course the background needs to be consistent with the type of violations that are considered when measuring the odds ratio of the detected event. For example, if the odds ratio for a detection is calculated using two non-GR parameters ($N_T = 2$), it can only be compared with a background that was generated for the same two parameters. If a measurement is repeated for three parameters, a new background corresponding to this choice of parameters will need to be generated. This also holds for changing the type of sources used, the ranges in parameters considered and the type of noise coming from the detectors. There are however unmodeled effects such as tidal effects to which the shape of the background in the case of TIGER-BNS is insensitive and I will give a brief overview on this in Sec. 1.6.3.

When evaluating how well the TIGER pipeline is capable of uncovering certain types of violations, one also needs to know what values of $\ln \mathcal{O}^{\rm modGR}_{\rm GR}$ are to be expected when the non-GR source is influenced by the same type of noise as the background. This is achieved by generating a *foreground* distribution of log odds $P(\ln \mathcal{O} | \mathcal{H}_{\rm alt}, \kappa_{\rm alt}, I)$, where $\mathcal{H}_{\rm alt}$ is the hypothesis associated with a large set of non-GR

sources κ_{alt}. We can then calculate the *efficiency* of finding such a GR violation as

$$\zeta = \int_{\ln \mathcal{O}_\beta}^{\infty} \mathrm{d}(\ln \mathcal{O})\, P(\ln \mathcal{O}|\mathcal{H}_{\text{alt}}, \kappa_{\text{alt}}, I). \tag{1.176}$$

The efficiency is illustrated in the right panel of Fig. 1.10, where the same background as in the left panel is shown together with a foreground distribution, and the part of the foreground that constitutes the efficiency is highlighted in green. In the previous section we mentioned that the choice of prior odds α in *e.g.* Eq. (1.172) is irrelevant, which can now be seen from the way we calculate thresholds and efficiencies: A change $\alpha \to \alpha'$ causes the same constant shift by $\ln(\alpha/\alpha')$ of both the background and foreground distributions as well as the threshold $\ln \mathcal{O}_\beta$ while leaving the false alarm probability β and efficiency ζ unaffected.

So far we only considered a simple noise model where one could inject sources equidistantly into a long stretch of continuous data. In reality this can not be done so trivially. When considering real detector output to construct a background, only part of the complete observation data (say between t_{start} and t_{end}) can be used due to certain times where the detector was not locked and times that are vetoed as a result of glitches as explained briefly in Sec. 1.3. Because of this the data are split into *segments* of data and only these can be used to inject sources for constructing a background. Furthermore, when considering a background that represents realistic detections, at least two detectors must be included in the analysis as a detection is only claimed when it is *coincident* in two or more detectors, *i.e.* the signal appears in two detectors within a time delay that corresponds to the time it takes a signal to travel from on detector to the other. This complicates matters further as the unvetoed segments in detector A are not necessarily overlapping with the segments in detector B, decreasing the amount of combined segments where both detectors have clean enough data and are long enough to inject a signal into. For example when we built the background for BNS sources in real data, we considered segments of 1024 s long. If we were to use unvetoed coincident times in two detectors we would only have 164 segments left and for three detectors that was further reduced to only 7, far from the order 1000 segments required to construct a background.

In order to overcome this and collect enough segments of coincident unvetoed time, we use *time-slides*. The way this works is illustrated in Fig. 1.11: We start with two sets of detector output, one for detector A and one or detector B. The goal is to gather as many segments of coincident time of length T as possible. We note that there are multiple stretches of data in A and B that have a length T, these are highlighted in green; however, there is only a single stretch of length T that is coincident in both detectors. It is possible to generate many more when the output of detector B is shifted in time as illustrated in cases B'_n, corresponding to the data-stream in B shifted by a time t_n^{B}. After shifting B, the segments that are used are not used again and one moves on to the next segment of length T in A and shifts the time in B to once again find a new coincident time. Each shift in time adds another unique unvetoed coincident segment. Using this method to slide B around, it is possible to find several stretches of unique coincident data instead of just one. In practice we form sets of injection times $\{t_1^a, t_2^a, t_3^a, \dots\}$, where a can be any detector we want to consider and the times t_i^a are randomly drawn from the unvetoed detector

data segments that are long enough to inject the signal, while making sure not to use the same stretch of data twice. In doing so we were able to generate 2500 segments of coincident time of length 1024 s in three detectors using S6 data.

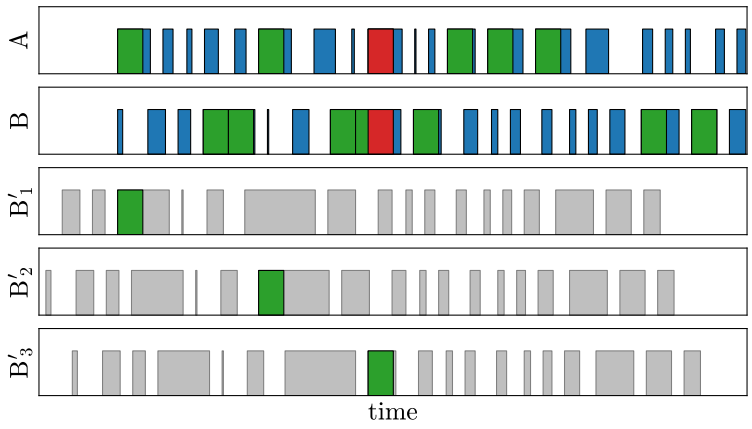


Figure 1.11: An illustration of how time slides can provide many unique stretches of data that are coincident in both detectors (A, B). Each block represents a stretch of data that is unvetoed and thus suitable for injections. Highlighted in red is the only segment of length T that is coincident in both detectors. Highlighted in green are the segments used to generate more coincident data by shifting the output of B with time slides $\{t_1^B, t_2^B, t_3^B\}$ as seen in the bottom three plots B_1', B_2' and B_3' respectively. In practice this is random process.

When constructing a background or foreground distribution, injected sources are distributed randomly in sky-localization, orientation (inclination angle of the orbit) and volume. For the background we only wish to analyze sources that would make it through the detection pipelines and set a lower limit to the optimal network SNR of 8 [70]. This means that due to the random orientation of the sources and distances, many of the injected sources will not pass this SNR bound and we are typically left with about 2/3 of the number of injected sources to form a background. Furthermore, it might happen that a source passes the SNR cut, without being found by the GR waveform model, meaning the Bayes factor for the GR hypothesis versus the noise-only hypothesis $\ln B^{\rm GR}_{\rm noise} \simeq 0$. Such sources are discarded by imposing an additional cut $\ln B^{\rm GR}_{\rm noise} > 32$, motivated by the fact that the main contribution to $\ln B^{\rm GR}_{\rm noise}$ is $(1/2)\,(h_{\rm GR}|h_{\rm GR}) = (1/2)\mathrm{SNR}^2$ with $h_{\rm GR}$ the GR waveform and $(.|.)$ the noise-weighted inner product Eq. (1.76).

1.6.3 Performance and robustness of the TIGER pipeline

The performance of TIGER in the context of BNS sources has been investigated in great detail in Li *et al.* [66, 67]. Here they injected both GR and non-GR waveforms

into Gaussian stationary noise that followed the expected Advanced LIGO and Advanced Virgo PSDs. The resulting backgrounds and foregrounds were analyzed as described in previous sections. It was shown that no matter what choice of maximum false alarm probability was used, a violation of 10% in the 1.5PN phase coefficient ϕ_3 could be seen with essentially 100% efficiency when catalogs of 15 sources were considered. The set of testing parameters used consisted of $\delta\hat{\varphi}_1$, $\delta\hat{\varphi}_2$ and $\delta\hat{\varphi}_3$, but even a violation of 20% in $\delta\hat{\varphi}_4$ could be picked up with high efficiency. Li *et al.* also considered smaller violations in $\delta\hat{\varphi}_3$ that required more sources to be recovered and violations that did not fit the general structure of Eq. (1.95): Two such violations were investigated, one that added an extra "1.25PN" term, *i.e.* $\propto f^{-5/6}$, the other that added a term with a mass dependent power of frequency, *i.e.* $f^{\gamma(M)}$ for an arbitrary function $\gamma(M)$ that effectively ranged from 0.5PN to 1.5PN depending on the source's mass M. To uncover the latter two violations with a 100% efficiency it was found that the phase at $\sim 150\,\mathrm{Hz}$ would differ from the GR phase by more than $\mathcal{O}(10)$ radians. We expect therefore that TIGER is capable of picking up a wide range of GR violations.

Li *et al.* [66, 67] and also Agathos *et al.* [68] demonstrated that potential problems caused by using insufficiently parsimonious non-GR hypotheses are avoided. For example, it could be that the true nature of a violation from GR is described by only a single additional parameter, in which case testing GR with a model containing 3 additional parameters may not be the best approach: Should a model contain too many additional parameters compared to what the signal can be described with, the likelihood function in Eq. (1.136) might still be peaked at the correct values of $\boldsymbol{\theta}$, but the integration against the prior density over a parameter space whose dimensionality is too high, might cause the resulting evidence $P(d|\mathcal{H}_{\mathrm{modGR}})$ to be small compared to $P(d|\mathcal{H}_{\mathrm{GR}})$, event when GR is correct.

TIGER circumvents this problem by not only considering the most inclusive hypothesis *e.g.* H_{12}, but rather the $\mathcal{H}_{\mathrm{modGR}}$ hypothesis which is a combination of sub-hypotheses H_1, H_2 and H_{12}, thus also containing the most parsimonious models. One might expect that a GR violation will affect all the phase coefficients, but this is not necessarily true. One example is the inclusion of a non-zero graviton mass, which primarily affects ϕ_2. It is interesting to see that in *all* cases considered in Li et al., the most general sub-hypothesis $N_{12\ldots N_T}$ was almost always disfavored compared to a number of sub-hypotheses that had a smaller set of free parameters. This was even seen in the rather extreme case where the mass-dependent frequency term $f^{\gamma(M)}$ was introduced. The conclusion is that any number of testing parameters may be used in TIGER without running into problems with insufficient parsimony. Furthermore, Agathos *et al.* have demonstrated that the background itself is largely insensitive to an increase in the amount of non-GR parameters considered.

The TIGER pipeline for BNS sources has been internally reviewed by the LIGO-Virgo Collaboration. During this review, code and post-processing output has been tested for consistency and systematic errors. In conjunction with the review, the pipeline has also undergone tests to verify its robustness against various unmodeled effects [68]. TIGER-BNS uses the aligned-spin inspiral-only waveform model TaylorF2 for *recovery*, *i.e.* the likelihood calculations in the nested sampling algorithm are performed using this model. However, we typically inject sources generated with

a different waveform family that includes more physics to more accurately represent what an expected GR signal would look like. It is possible to see how much the background is influenced by the fact that the recovery waveform is lacking certain physics, by recovering instead with a more advanced model. In [68], many such tests have been performed where in each case two backgrounds were compared: One that represents a "basic" background and one that considers one of the following unmodeled effects:

- Fundamental effects: Difference between waveform approximants and using finite post-Newtonian order.
- Astrophysical effects: Neutron star tidal deformability and spins.
- Instrumental effects: Calibration errors.

For example, currently we do not know precisely the equation of state of neutron stars. The equation of state determines how deformable a neutron star is and in turn what kind of effect this has on the phase evolution of the waveform. In other words, the exact shape of the waveform, especially in the late inspiral stage, is determined by an unknown quantity. We set an upper limit to the frequency range in which the likelihoods are calculated, namely $f_{\max} = 400\,\mathrm{Hz}$, to avoid unmodeled tidal effects to be picked up as a violation of GR. To test if the background still suffers from tidal effects we inject BNS sources assuming some feasible equation of state that allows for the largest effect on the waveform, but consider two cases for recovery: (1) The "basic" TaylorF2 model without tidal effects and (2) the same model that was used to inject the sources, hence containing the perfect description of the tidal effects. The two resulting backgrounds are then compared and found to be nearly identical. They have thus demonstrated that the 400 Hz cut-off renders tidal effects invisible without affecting TIGER's ability to look for GR violations.

For each of the unmodeled effects listed above such background comparisons have been made and in each case the background distributions were found to be identical within statistical uncertainties. Next to presenting the results for individual effects, Agathos *et al.* also showed that TIGER does not suffer from inclusion of all the above effects, which is the most realistic scenario. This proves that the TIGER pipeline for BNS sources is robust against all these unmodeled effects.

All background studies performed in [68] were done with Gaussian noise. The TIGER-BNS pipeline now extends its functionality to non-Gaussian (*e.g.* including glitches), non-stationary noise (the behavior of each detector changes over time) by using recolored S6 data. The impact of non-stationarity in detector noise has been part of the review process of TIGER-BNS. The conclusion is that the shape of the background when using real noise is not much affected aside from a few outliers. These outliers were events that had a disproportionately large $\ln \mathcal{O}_{\mathrm{GR}}^{\mathrm{modGR}}$, but could be removed from the background as the segments these were injected into ostensibly contained large glitches. The topic of handling outliers and recognizing glitches is discussed in more detail in Chapter 3. The latest TIGER-BNS background (without outliers where $\ln \mathcal{O}_{\mathrm{GR}}^{\mathrm{modGR}} > 90$) using spinning TaylorT4 injections is shown in Fig. 1.12. The component spins of the injected waveform have magnitudes $-0.1 \leq |\mathbf{S}_a| \leq 0.1$ and component masses are $1\,\mathrm{M}_\odot \leq m_a \leq 2\,\mathrm{M}_\odot$. Using real data results in the tendency to form a slight tail toward more positive odds ratios, but not to

the extent that it would cause the threshold to be increased by much and hence the ability of TIGER to pick up violations of GR is not affected.

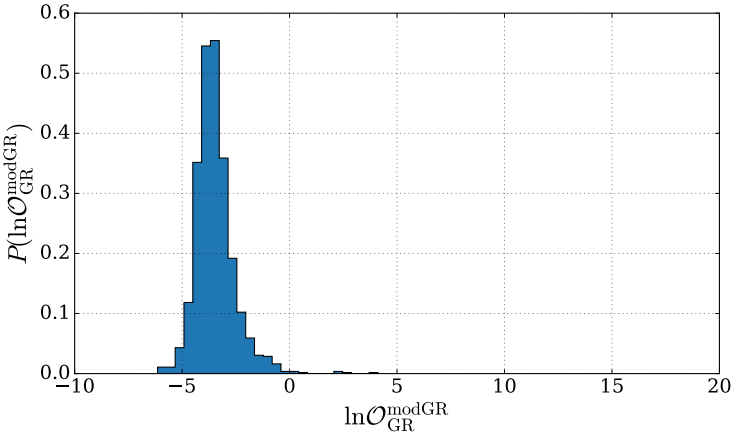


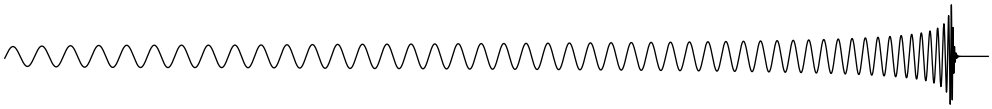
Figure 1.12: Background generated from 1359 GR sources. The sources are spinning BNS with spin magnitudes $-0.1 \leq |\mathbf{S}_a| \leq 0.1$ and component masses $1\,\mathrm{M}_\odot \leq m_a \leq 2\,\mathrm{M}_\odot$. The sources were injected into recolored S6 data.

1.7 O1

The first observing run (O1) of the two Advanced LIGO detectors in the United States which took place from September 12, 2015, to January 19, 2016, delivered the first direct detection of gravitational waves on September 14, 2015 at 09:50:45 UTC [71] labeled GW150914 and a second detection on December 26, 2015 at 03:38:53 [72], labeled GW151226. Both events were identified as binary black hole (BBH) mergers. The first had a total mass of $65.3^{+4.1}_{-3.4}$ $\mathrm{M}_\odot$ with a primary component mass of $36.2^{+5.2}_{-3.8}$ $\mathrm{M}_\odot$ and secondary of $29.1^{+3.7}_{-4.4}$ $\mathrm{M}_\odot$. The second event was much lighter at a total mass of $21.8^{+5.9}_{-1.7}$ $\mathrm{M}_\odot$ and a primary component mass of $14.2^{+8.3}_{-3.7}$ $\mathrm{M}_\odot$ and secondary of $7.5^{+2.3}_{-2.3}$ $\mathrm{M}_\odot$ [11]. After each of the two detection papers [71, 72], the LIGO-Virgo Collaboration has published a whole set of companion papers, two of which contain the first ever tests of the strong-field dynamics of general relativity. One particular test first presented in [73] and used again to combine results of both detections in [11] will also be the topic of Chapter 4.

Chapter 2

Testing the no-hair theorem with black hole ringdowns using TIGER



The main focus of this thesis is performing tests of general relativity (GR) with advanced detectors such as Advanced LIGO and Advanced Virgo. However, we have also investigated the possibilities of using the third generation detector ET (Sec. 1.3.4) to perform tests of GR with different types of sources from what we have seen and expect to see with advanced detectors. ET will have a lower frequency cut-off at around one Hertz as opposed to the advanced ground based gravitational wave observatories which are sensitive down to about 30 Hz. This means that there is a chance ET could see an entirely different class of objects; the hypothesized intermediate-mass black hole systems, for which component masses are around $500 - 1000\,\mathrm{M}_\odot$ [1]. For these signals, only the ringdown part would be within the detectors' sensitivity band. It is therefore interesting to investigate what can be done by using only the quasi-normal modes (QNM) – superpositions of damped sinusoids, each represented by a characteristic mode frequency ω_{lm} and damping time τ_{lm} – from a ringing black hole. By looking at the characteristic frequencies and damping times of these black hole ringdown signals, ET would enable us to perform tests of the no-hair theorem which states that a black hole which has settled down to its final stationary vacuum state is determined only by its mass, spin and electric charge. In previous work it was shown that with a single $500 - 1000\,\mathrm{M}_\odot$ black hole at distance $\lesssim 6\,\mathrm{Gpc}$ (or redshift $z \lesssim 1$), deviations of a few percent in the frequencies and damping times of dominant and sub-dominant modes would be within the range of detectability [74]. Given that such sources may be relatively rare, it is of interest to see how well the no-hair theorem can be tested with events at much larger distances and with smaller signal-to-noise ratios, thus accessing a far bigger volume of space and a larger number of sources. To this end we employ TIGER which is introduced in Sec. 1.6. TIGER is well suited for the regime of low signal-to-noise ratio, and information from a population of sources can be combined so as to arrive at a stronger test. By performing a range of simulations using the expected noise power spectral density of Einstein Telescope, we show that with TIGER, similar deviations from the no-hair theorem as considered in previous work will be detectable with great confidence with $\mathcal{O}(10)$ sources distributed uniformly in co-moving volume out to 50 Gpc ($z \lesssim 5$). In this chapter I will start by briefly describing the definition of the no-hair theorem in Sec. 2.1, followed by a detailed discussion on how we may use the ringdown signal to test the theorem in

[1] The exact mass definition of the term "intermediate-mass" is not well defined, but it typically ranges between $10^2 - 10^5\,\mathrm{M}_\odot$. We only consider masses up to $1000\,\mathrm{M}_\odot$ because sources with a mass any higher than that will merge below 1 Hz and there will be practically no signal left for ET-type detectors.

Sec. 2.2. In Sec. 2.3 we explain our assumptions regarding Einstein Telescope as well as our waveform models for signals and templates, and the set-up of the simulations. In Sec. 2.4 we evaluate TIGER's ability to perform tests of the no-hair theorem. The possibility of precision measurements of the free parameters assuming validity of GR is discussed in Sec. 2.5. Finally, Sec. 2.6 provides a discussion on the possible application to GW150914 followed by a summary and conclusions.

2.1 The no-hair theorem

The no-hair theorem states that a black hole that has settled down to its final stationary vacuum state is determined only by its mass, spin and electric charge [75, 76, 77, 78, 79]. On large scales the Universe appears to be electrically neutral, so astrophysical black holes are also thought to be deprived of net electric charge. This means that only mass and spin need to be considered, leading to the Kerr geometry. When a black hole is formed as a result of the inspiral and merger of two other compact objects, it will undergo 'ringdown' as it evolves towards its quiescent state. This process can be modeled by considering linear perturbations of the Kerr metric, or quasi-normal modes, which are characterized by frequencies ω_{lm} and damping times τ_{lm} [18, 19, 20, 21]. Since the underlying Kerr space-time is characterized only by its mass M and spin J, the frequencies and times are constrained by linearized general relativity to only depend on these quantities through specific functional relationships, *i.e.*

$$\begin{aligned} \omega_{lm} &= \omega_{lm}^{\mathrm{GR}}(M, J), \\ \tau_{lm} &= \tau_{lm}^{\mathrm{GR}}(M, J). \end{aligned} \tag{2.1}$$

Observational tests of these dependences would constitute a test of the no-hair theorem, and hence of general relativity (GR)[2]; this was first hinted at by Detweiler [81], made concrete by Dreyer *et al.* [82], and further explored in [83, 84, 85].

2.2 The road to a no-hair test

Previously Gossan, Veitch, and Sathyaprakash [74] investigated the possibility of performing the test proposed above, using Einstein Telescope (ET), as well as with the space-based LISA (see Sec. 1.3.4 and [86] resp.). These authors evaluated two methods for checking the dependences in Eq. (2.1) predicted by GR: Bayesian parameter estimation and model selection. Specifically, one can write possible deviations from these dependences as

$$\omega_{lm} = \omega_{lm}^{\mathrm{GR}}(M, J)\,(1 + \delta\hat{\omega}_{lm}), \tag{2.2}$$

$$\tau_{lm} = \tau_{lm}^{\mathrm{GR}}(M, J)\,(1 + \delta\hat{\tau}_{lm}), \tag{2.3}$$

[2]For the purposes of this thesis I will treat the no-hair theorem as if it were a prediction of GR, so that a violation of the theorem implies a GR violation. However, it should be noted that existing proofs require analyticity of space-time; see [80] for an overview. For this reason, the no-hair theorem is often referred to as the no-hair conjecture.

and then (a) calculate how well the dimensionless quantities (fractional deviations) $\delta\hat{\omega}_{lm}$, $\delta\hat{\tau}_{lm}$ can be measured, or (b) compare the evidences for two models: one where the $\delta\hat{\omega}_{lm}$, $\delta\hat{\tau}_{lm}$ are free parameters, and another in which they are all identically zero, corresponding to the GR prediction. In practice, the authors of [74] restricted their attention to the set

$$\{\delta\hat{\omega}_{22}, \delta\hat{\omega}_{33}, \delta\hat{\tau}_{22}\}. \tag{2.4}$$

It was found that for black holes with masses in the range $500-1000\,\mathrm{M}_\odot$ at a distance of 6 Gpc, ET would allow for measurements of $\delta\hat{\omega}_{22}$, $\delta\hat{\omega}_{33}$, and $\delta\hat{\tau}_{22}$ with accuracies of a few percent for the first two parameters, and about 10% for the third. (For comparison, boson stars in the same mass range would cause $\delta\hat{\omega}_{22}$ and $\delta\hat{\tau}_{22}$ to be of order 1 [87].) With model selection and assuming a $500\,\mathrm{M}_\odot$ black hole, a deviation of a few percent in $\delta\hat{\omega}_{22}$ could be discriminated from GR with $\ln B^{\mathrm{dev}}_{\mathrm{GR}} > 10$, where $B^{\mathrm{dev}}_{\mathrm{GR}}$ is the *Bayes factor*, or ratio of evidences, for the model that deviates from GR (with the variables in Eq. (2.4) as extra free parameters) versus the GR model.

How frequently might one test GR in this way? Coalescence rates of intermediate-mass binary black holes which would give rise to ringdowns with masses in the above range are highly uncertain [88, 89, 90]; ET may see between a few and a few thousands per year [91, 92]. Gossan *et al.* considered single, relatively loud sources, but one will also want to combine information from multiple, possibly weak signals out to large distances so as to maximally exploit the available set of detections. Since deviations from the no-hair theorem may be such that $\delta\hat{\omega}_{22}$, $\delta\hat{\omega}_{33}$, and/or $\delta\hat{\tau}_{22}$ take on different non-zero values for different sources, when doing parameter estimation it will not be possible to combine posterior probability densities from multiple events unless one already assumes GR to be correct. On the other hand, although Bayesian model selection does lend itself quite easily to the utilization of all available detections, if one lets $\{\delta\hat{\omega}_{22}, \delta\hat{\omega}_{33}, \delta\hat{\tau}_{22}\}$ (and possibly more of the $\delta\hat{\omega}_{lm}$, $\delta\hat{\tau}_{lm}$) vary all at the same time, one may be penalized if the corresponding model is insufficiently parsimonious, *i.e.* if the correct model involves a smaller number of additional parameters.

In [66, 67, 93, 94, 68], a more general algorithm for testing GR was developed, called TIGER. Take a gravitational waveform model as predicted by GR, and introduce deformations parameterized by dimensionless quantities $\delta\xi_i$, $i = 1, 2, \ldots, N_T$ such that all of the $\delta\xi_i$ being zero corresponds to GR being correct. One can then ask the question: "Do one or more of the $\delta\xi_i$ differ from zero?" Let us denote the corresponding hypothesis by $\mathcal{H}_{\mathrm{modGR}}$, and the GR hypothesis by $\mathcal{H}_{\mathrm{GR}}$. Now, there is no waveform model that corresponds to $\mathcal{H}_{\mathrm{modGR}}$. However, as shown in [66], one can define logically disjoint 'sub-hypotheses' $H_{i_1 i_2 \ldots i_k}$, in each of which a *fixed* set of parameters $\{\delta\xi_{i_1}, \delta\xi_{i_2}, \ldots, \delta\xi_{i_k}\}$ are non-zero while $\delta\xi_j = 0$ for $j \notin \{i_1, i_2, \ldots, i_k\}$. There are $2^{N_T} - 1$ such sub-hypotheses, corresponding to the non-empty sub-sets of the full set $\{\delta\xi_1, \delta\xi_2, \ldots, \delta\xi_{N_T}\}$. The $H_{i_1 i_2 \ldots i_k}$ *do* have waveform models associated with them that can be compared with the data, and $\mathcal{H}_{\mathrm{modGR}}$ can be expressed as the logical union of all the sub-hypotheses:

$$\mathcal{H}_{\mathrm{modGR}} = \bigvee_{i_1 < i_2 < \ldots < i_k;\, k \leq N_T} H_{i_1 i_2 \ldots i_k}. \tag{2.5}$$

Given a catalog of detections $d_1, d_2, \ldots, d_{\mathcal{N}}$ and whatever background information I

one may possess, one can then compute the *odds ratio* for $\mathcal{H}_{\rm modGR}$ against $\mathcal{H}_{\rm GR}$:

$$\begin{aligned}\mathcal{O}^{\rm modGR}_{\rm GR} &\equiv \frac{P(\mathcal{H}_{\rm modGR}|d,I)}{P(\mathcal{H}_{\rm GR}|d,I)} \\ &= \frac{\alpha}{2^{N_T}-1} \sum_{i_1<i_2<\ldots<i_k;k\leq N_T} \prod_{A=1}^{\mathcal{N}} {}^{(A)}B^{i_1 i_2 \ldots i_k}_{\rm GR}.\end{aligned} \tag{2.6}$$

Here α is an unimportant scaling factor which below will be set to unity, and the Bayes factors ${}^{(A)}B^{i_1 i_2 \ldots i_k}_{\rm GR}$ for a detection d_A are given by

$${}^{(A)}B^{i_1 i_2 \ldots i_k}_{\rm GR} \equiv \frac{P(d_A|H_{i_1 i_2 \ldots i_k},I)}{P(d_A|\mathcal{H}_{\rm GR},I)}, \tag{2.7}$$

with $P(d_A|H_{i_1 i_2 \ldots i_k},I)$ and $P(d_A|\mathcal{H}_{\rm GR},I)$ the *evidences* for $H_{i_1 i_2 \ldots i_k}$ and $\mathcal{H}_{\rm GR}$, respectively. For basic assumptions and detailed derivations we refer to [66, 67, 93].

The TIGER formalism has been evaluated extensively in the context of binary neutron star inspirals that will be observed by second-generation detectors such as Advanced LIGO, Advanced Virgo, GEO-HF, KAGRA, and LIGO-India (see Sec. 1.3.3). In [66, 67] it was shown that, thanks to the introduction of the sub-hypotheses $H_{i_1 i_2 \ldots i_k}$, the method avoids potential problems due to insufficient parsimony, is well suited to dealing with weak signals, and enables the discovery of a wide range of deviations from GR, including ones that are well outside the particular parameterized waveform family used; moreover, information from multiple sources can trivially be combined.

However, TIGER is not tied to any particular gravitational waveform model and can be applied to testing the no-hair theorem with ringdown signals. Consider the $N_T = 3$ testing parameters of [74],

$$\delta\xi_1 = \delta\hat{\omega}_{22}, \quad \delta\xi_2 = \delta\hat{\omega}_{33}, \quad \delta\xi_3 = \delta\hat{\tau}_{22}. \tag{2.8}$$

$\mathcal{H}_{\rm modGR}$, the hypothesis that one or more of the $\delta\xi_i$ deviate from their GR value, is then the logical union of $2^3 - 1 = 7$ sub-hypotheses H_1, H_2, H_3, H_{12}, H_{13}, H_{23}, and H_{123}. Here H_1 is the hypothesis that $\delta\xi_1 \neq 0$ while $\delta\xi_2 = \delta\xi_3 = 0$, H_{13} the hypothesis that both $\delta\xi_1 \neq 0$ and $\delta\xi_3 \neq 0$ but $\delta\xi_2 = 0$, and similarly for the other sub-hypotheses. In the above language, the model selection set-up of Gossan *et al.* [74] only involved calculating, for a single source, the Bayes factor

$$B^{123}_{\rm GR} = \frac{P(d|H_{123},I)}{P(d|\mathcal{H}_{\rm GR},I)}. \tag{2.9}$$

It would be of great interest to see how our ability to discern violations of the no-hair theorem with ringdown signals would improve if the full formalism of TIGER were brought to bear. This will be the main topic of the present chapter.

When evaluating the odds ratio $\mathcal{O}^{\rm modGR}_{\rm GR}$ of Eq. (2.6) using one or more detected signals, we may find that there is no reason to believe that GR is incorrect. However, in that case it will still be of interest to *measure* $\delta\hat{\omega}_{22}$, $\delta\hat{\omega}_{33}$, and $\delta\hat{\tau}_{22}$ for each source

and combine the resulting posterior density distributions so as to arrive at a joint result for the entire catalog of detections. This we will also do, and as we shall see, potentially tight constraints can be set on these parameters.

2.3 Waveform model and set-up of the simulations

2.3.1 Waveform model

The ringdown signal is given by a superposition of quasi-normal modes characterized by triples of integers (l, m, n), where $l \geq 2$ and $m = -l, -l+1, \ldots, l-1, l$; $n \geq 0$ is an overtone index [95, 96, 20, 21]. Here we will only consider the modes with $n = 0$, as overtones with $n > 0$ are not significantly excited and have much shorter damping times [83]. The 'plus' and 'cross' polarizations read

$$h_+(t) = \frac{M}{D_{\rm L}} \sum_{l,m>0} A_{l|m|} e^{-t/\tau_{lm}} Y_+^{lm}(\iota) \cos(\omega_{lm} t - m\phi),$$
$$h_\times(t) = -\frac{M}{D_{\rm L}} \sum_{l,m>0} A_{l|m|} e^{-t/\tau_{lm}} Y_\times^{lm}(\iota) \sin(\omega_{lm} t - m\phi), \tag{2.10}$$

where the $Y_+^{lm}(\iota)$, $Y_\times^{lm}(\iota)$ can be written in terms of spin-weighted spherical harmonics of weight -2:

$$Y_+^{lm}(\iota) \equiv {}_{-2}Y^{lm}(\iota, 0) + (-1)^l {}_{-2}Y^{l,-m}(\iota, 0),$$
$$Y_\times^{lm}(\iota) \equiv {}_{-2}Y^{lm}(\iota, 0) - (-1)^l {}_{-2}Y^{l,-m}(\iota, 0). \tag{2.11}$$

In the above, M is the observed mass of the black hole, which is related to the intrinsic mass by $M = (1+z)\, M_{\rm intr}$, with z the redshift; $D_{\rm L}$ is the luminosity distance to the source; ι is the angle between the black hole's spin and the line of sight; and ϕ is the azimuth angle of the black hole with respect to the observer. Note that in principle there will be additional phase offsets ϕ_{lm} in Eq. (2.10); since analytic fits for their dependence on progenitor parameters are not yet available, we set them to zero, as was also done in [74, 97]. The $\omega_{lm}(M, j)$ and $\tau_{lm}(M, j)$ are the characteristic frequencies and damping times of the modes, respectively, as functions of the mass and of the dimensionless spin $j = J/M^2$.

As in [74], we only consider the modes $(l, m) = (2, 2)$, $(2, 1)$, $(3, 3)$, $(4, 4)$, which are among the most dominant ones. Analytic expressions for the mode amplitudes $A_{l|m|}$ are not available, but there exist accurate fits to numerical simulations. The authors of [74] took the *progenitor* black holes to be non-spinning, in which case one can use the approximate expressions for the $A_{l|m|}$ in terms of the symmetric mass ratio $\nu = m_1 m_2/(m_1 + m_2)^2$ (with m_1, m_2 the progenitor component masses) from Kamaretsos *et al.* [85].[3] Here we will relax this assumption and include the effect of non-zero progenitor spins in the waveforms, using more recent results. For spinning

[3] More recent fits for the $A_{l|m|}$ in the case of non-spinning progenitors can be found in [98].

progenitors, Kamaretsos, Hannam, and Sathyaprakash [99] found that mainly A_{21} is strongly affected, and a good fit for all the relevant amplitudes is given by

$$A_{22}(\nu) = 0.864\nu, \tag{2.12}$$
$$A_{21}(\nu) = 0.43\,\left[\sqrt{1-4\nu} - \chi_{\rm eff}\right] A_{22}(\nu), \tag{2.13}$$
$$A_{33}(\nu) = 0.44(1-4\nu)^{0.45} A_{22}(\nu). \tag{2.14}$$
$$A_{44}(\nu) = \left[5.4(\nu-0.22)^2 + 0.04\right] A_{22}(\nu), \tag{2.15}$$

where

$$\chi_{\rm eff} = \frac{1}{2}\left(\sqrt{1-4\nu}\,\chi_1 + \chi_-\right), \tag{2.16}$$

with

$$\chi_- = \frac{m_1\chi_1 - m_2\chi_2}{M_{\rm in}}. \tag{2.17}$$

Here (m_1, m_2) and (χ_1, χ_2) are, respectively, the progenitor component masses and dimensionless spin magnitudes, and $M_{\rm in}$ is the initial total mass of the system, which to reasonable approximation we can take to be equal to the mass of the final black hole.

For the frequencies ω_{lm} and damping times τ_{lm} there also exist good fits, which can be expressed through the *quality factors* $Q_{lm} = \omega_{lm}\tau_{lm}/2$:

$$M\omega = f_1 + f_2(1-j)^{f_3}, \tag{2.18}$$
$$Q = q_1 + q_2(1-j)^{q_3}, \tag{2.19}$$

where for the values of the coefficients f_1, f_2, f_3, q_1, q_2, q_3 we refer to Tables VIII, IX and X in [83]. Finally, there exists a fit for the spin j of the final black hole in terms of the progenitor component masses (m_1, m_2) and spin vectors $(\mathbf{a}_1, \mathbf{a}_2)$ [100, 101],

$$\begin{aligned} j = \frac{1}{(1+q)^2}\,&\left[\chi_1^2 + \chi_2^2 q^4 + 2\chi_1\chi_2 q^2\cos\alpha\right. \\ &\left. + 2(\chi_1\cos\beta + \chi_2 q^2\cos\gamma)|\mathbf{l}|q + |\mathbf{l}|^2 q^2\right], \end{aligned} \tag{2.20}$$

where $q = m_2/m_1$, $\chi_i = |\mathbf{a}_i| = \sqrt{a_{ix}^2 + a_{iy}^2 + a_{iz}^2}$ and the angles are defined as

$$\cos\alpha \equiv \hat{\mathbf{a}}_1 \cdot \hat{\mathbf{a}}_2, \quad \cos\beta \equiv \hat{\mathbf{a}}_1 \cdot \hat{\mathbf{l}}, \quad \cos\gamma \equiv \hat{\mathbf{a}}_2 \cdot \hat{\mathbf{l}}. \tag{2.21}$$

In the equations above we used

$$\begin{aligned} |\mathbf{l}| = {}& 2\sqrt{3} + t_3\nu + t_3\nu^2 + \frac{s_4}{(1+q^2)^2}\left(\chi_1^2 + \chi_2^2 q^4 + 2\chi_1\chi_2 q^2\cos\alpha\right) \\ &+ \frac{s_5\nu + t_0 + 2}{1+q^2}\left(\chi_1\cos\beta + \chi_2 q^2\cos\gamma\right), \end{aligned} \tag{2.22}$$

where $t_0 = -2.8904$, $t_3 = 2.5763$, $s_4 = -0.1229$ and $s_5 = 0.4537$.

For the simulated signals, or *injections*, we choose progenitor spins χ_1, χ_2 from a distribution with isotropic directions, and a Gaussian distribution for the magnitudes centered on 0.7, with standard deviation 0.2 and hard cut-offs at 0.5 and 0.99

[102]; note that the value of 0.7 roughly corresponds to what one gets from the coalescence of non-spinning, equal mass binary black holes. The mass M is drawn from a uniform distribution between 500 and $1000\,\mathrm{M}_\odot$, and the mass ratio $q = m_1/m_2$ from a uniform distribution between 0.3 and 1. Amplitudes are computed as in Eqs. (2.12)-(2.17), where we take $\chi_{1,2} = |\chi_{1,2}|$, and the final spin j is calculated from component masses and spins using Eqns. (2.20), (2.21) and (2.22). With these choices for masses and spins, the characteristic frequency $f_{22} = \omega_{22}/(2\pi)$ of the dominant ringdown mode ranges from about 15 to 100 Hz, while the inspiral signal, which ends roughly at $f_{\rm LSO} = (6^{3/2}\pi M)^{-1}$, stays below a few Hz and hence is never in the sensitive frequency band. Redshifts are taken to be between 1.5 and 5, and sources are placed uniformly in co-moving volume assuming a ΛCDM cosmology with $(\Omega_{\rm M}, \Omega_\Lambda, h_0) = (0.27, 0.73, 0.70)$, so that luminosity distances approximately range from 10 to 50 Gpc. Since part of the exercise is to stress-test the TIGER framework, we only analyze sources with signal-to-noise ratio (SNR) < 30, corresponding to a minimum angle-averaged distance of 14.97 Gpc ($z = 1.90$). Sky positions (θ, φ) and orientations (ι, ψ) are drawn from uniform distributions on the sphere. To gauge our sensitivity to deviations in $\omega_{22}(M, j)$, $\omega_{33}(M, j)$, and $\tau_{22}(M, j)$, we introduce constant relative shifts $\delta\hat{\omega}_{22}$, $\delta\hat{\omega}_{33}$, and $\delta\hat{\tau}_{22}$ as explained in the introduction.

For the *templates*, we only take $\chi_{\rm eff}$ and j to be the spin-related free parameters, as the progenitor component spins χ_1 and χ_2 will not be separately measurable from a ringdown signal alone.[4] The free parameters for the waveform model corresponding to the GR hypothesis $\mathcal{H}_{\rm GR}$ are then

$$\theta_{\rm GR} = \{M, \nu, j, \chi_{\rm eff}, D_{\rm L}, \theta, \varphi, \psi, \iota, \phi, t_0\}\,, \tag{2.23}$$

where t_0 is the time of arrival of the signal at the detector. The prior on M is chosen to be uniform between 300 and $1200\,\mathrm{M}_\odot$, and the one for the symmetric mass ratio ν is flat between 0.01 and 0.25; in terms of the mass ratio $q = m_1/m_2$ this range corresponds to $0.01 \lesssim q \leq 1$. The prior on j is uniform between 0.01 and 0.99, and the one on $\chi_{\rm eff}$ is uniform between -1 and 1. Sky positions and orientations are taken to be uniform on the sphere, and the prior on distance is uniform in co-moving volume between 8 and 60 Gpc. t_0 is taken to be in a window of width 100 ms.

2.3.2 TIGER for ringdown

To apply TIGER in the context of ringdown, we introduce the same parameterized deformations of the waveform as in [74], namely the ones of Eqs. (2.2)-(2.4). The parameter spaces corresponding to the various sub-hypotheses $H_{i_1 i_2 \ldots i_k}$ of $\mathcal{H}_{\rm modGR}$

[4] The progenitor spins may become measurable if more modes are included than the ones considered here.

are given by

$$\begin{aligned} H_1 &\longleftrightarrow \{\theta_{\rm GR}, \delta\hat{\omega}_{22}\},\\ H_2 &\longleftrightarrow \{\theta_{\rm GR}, \delta\hat{\omega}_{33}\},\\ H_3 &\longleftrightarrow \{\theta_{\rm GR}, \delta\hat{\tau}_{22}\},\\ H_{12} &\longleftrightarrow \{\theta_{\rm GR}, \delta\hat{\omega}_{22}, \delta\hat{\omega}_{33}\},\\ H_{13} &\longleftrightarrow \{\theta_{\rm GR}, \delta\hat{\omega}_{22}, \delta\hat{\tau}_{22}\},\\ H_{23} &\longleftrightarrow \{\theta_{\rm GR}, \delta\hat{\omega}_{33}, \delta\hat{\tau}_{22}\},\\ H_{123} &\longleftrightarrow \{\theta_{\rm GR}, \delta\hat{\omega}_{22}, \delta\hat{\omega}_{33}, \delta\hat{\tau}_{22}\}. \end{aligned} \tag{2.24}$$

Given a detection d_A, the corresponding Bayes factors ${}^{(A)}B^{1}_{\rm GR}$, ${}^{(A)}B^{2}_{\rm GR}$, ${}^{(A)}B^{3}_{\rm GR}$, ${}^{(A)}B^{12}_{\rm GR}$, ${}^{(A)}B^{13}_{\rm GR}$, ${}^{(A)}B^{23}_{\rm GR}$, and ${}^{(A)}B^{123}_{\rm GR}$ are calculated using

$${}^{(A)}B^{i_1 i_2 \ldots i_k}_{\rm GR} = \frac{{}^{(A)}B^{i_1 i_2 \ldots i_k}_{\rm noise}}{{}^{(A)}B^{\rm GR}_{\rm noise}}, \tag{2.25}$$

where ${}^{(A)}B^{i_1 i_2 \ldots i_k}_{\rm noise}$, ${}^{(A)}B^{\rm GR}_{\rm noise}$ are, respectively, the Bayes factors for $H_{i_1 i_2 \ldots i_k}$ and $\mathcal{H}_{\rm GR}$ against the hypothesis that the data contain only noise. The latter are computed using an appropriate adaptation of the nested sampling algorithm as implemented by Veitch and Vecchio [103, 104, 63].

For completeness, we give the expression for the odds ratio $\mathcal{O}^{\rm modGR}_{\rm GR}$ in the present context; given a catalog of $\mathcal{N}$ ringdown signals, it reads

$$\begin{aligned} \mathcal{O}^{\rm modGR}_{\rm GR} = \frac{1}{7}\Bigg[&\prod_{A=1}^{\mathcal{N}} {}^{(A)}B^{1}_{\rm GR} + \prod_{A=1}^{\mathcal{N}} {}^{(A)}B^{2}_{\rm GR} + \prod_{A=1}^{\mathcal{N}} {}^{(A)}B^{3}_{\rm GR}\\ &+ \prod_{A=1}^{\mathcal{N}} {}^{(A)}B^{12}_{\rm GR} + \prod_{A=1}^{\mathcal{N}} {}^{(A)}B^{13}_{\rm GR} + \prod_{A=1}^{\mathcal{N}} {}^{(A)}B^{23}_{\rm GR}\\ &+ \prod_{A=1}^{\mathcal{N}} {}^{(A)}B^{123}_{\rm GR}\Bigg]. \end{aligned} \tag{2.26}$$

In practice it is often convenient to work with the *logarithm* of the odds ratio, $\ln \mathcal{O}^{\rm modGR}_{\rm GR}$. If GR is correct, then naively one would expect $\mathcal{O}^{\rm modGR}_{\rm GR} < 1$, or $\ln \mathcal{O}^{\rm modGR}_{\rm GR} < 0$. However, features in the noise can have a detrimental effect on the measurement of the log odds ratio, and in practice one can obtain slightly positive values of $\ln \mathcal{O}^{\rm modGR}_{\rm GR}$ even if no deviation from GR is present. For this reason one usually constructs a *background distribution* $P(\ln \mathcal{O}|\mathcal{H}_{\rm GR}, \kappa_{\rm GR}, I)$ [66, 67, 93, 94, 68]. Here $\kappa_{\rm GR}$ denotes a large number of (catalogs of) injections with GR waveforms, for each of which one computes $\ln \mathcal{O}^{\rm modGR}_{\rm GR}$, whose normalized distribution constitutes $P(\ln \mathcal{O}|\mathcal{H}_{\rm GR}, \kappa_{\rm GR}, I)$. Given a maximum false alarm probability β that one is willing to tolerate, one can use this background to set a threshold $\ln \mathcal{O}_\beta$ for the *measured* log odds ratio to overcome; this threshold is defined such that β is the fraction of the

background distribution that is above $\ln \mathcal{O}_\beta$:

$$\beta = \int_{\ln \mathcal{O}_\beta}^{\infty} \mathrm{d}(\ln \mathcal{O})\, P(\ln \mathcal{O}|\mathcal{H}_{\rm GR}, \kappa_{\rm GR}, I). \tag{2.27}$$

In reality there will only be a single value for the measured log odds ratio, computed from the signals one actually detects. However, if one wants to assess how likely it is that a particular type of deviation from GR, denoted by $\mathcal{H}_{\text{non-GR}}$, will lead to a log odds ratio above threshold, then one can construct a *foreground distribution* $P(\ln \mathcal{O}|\mathcal{H}_{\text{non-GR}}, \kappa_{\text{non-GR}}, I)$, where this time $\kappa_{\text{non-GR}}$ is a set of injections whose waveforms are in accordance with the given GR violation. One can then define the *efficiency* ζ as the fraction of the foreground that is above threshold:

$$\zeta = \int_{\ln \mathcal{O}_\beta}^{\infty} \mathrm{d}(\ln \mathcal{O})\, P(\ln \mathcal{O}|\mathcal{H}_{\text{non-GR}}, \kappa_{\text{non-GR}}, I). \tag{2.28}$$

This can be viewed as the probability that the particular kind of deviation from GR considered will be discovered with a false alarm probability of at most β.

In what follows, we will consider both the case where only a single ringdown detection is ever made by ET, so that $\mathcal{N} = 1$, and the case where multiple detections are made. As mentioned before, the event rate for ringdowns with mass in the range $500 - 1000\,\mathrm{M}_\odot$ is highly uncertain, but a few tens of detections out to tens of Gpc is consistent with expectations in the literature [88, 89, 90, 91, 92]. Below we will show results where injections are randomly combined into catalogs of $\mathcal{O}(10)$ sources each.

To evaluate TIGER's ability to find deviations from the no-hair theorem, we will mostly study its behavior in the following cases:

1. There is a 10% deviation in the dominant-mode frequency ω_{22}, but other mode frequencies as well as the damping times are unaffected; *i.e.*, the injections have $(\delta\hat{\omega}_{22}, \delta\hat{\omega}_{33}, \delta\hat{\tau}_{22}) = (0.1, 0, 0)$.

2. There is a 10% deviation in the $(3,3)$ mode frequency ω_{33}, but no deviation in other frequencies or in the damping times; *i.e.*, $(\delta\hat{\omega}_{22}, \delta\hat{\omega}_{33}, \delta\hat{\tau}_{22}) = (0, 0.1, 0)$.

3. There is a 10% deviation in the dominant-mode damping time τ_{22}, but no deviation in other damping times or in the frequencies: $(\delta\hat{\omega}_{22}, \delta\hat{\omega}_{33}, \delta\hat{\tau}_{22}) = (0, 0, 0.1)$.

4. There is a 25% deviation in τ_{22}, but no deviation in other damping times or in the frequencies: $(\delta\hat{\omega}_{22}, \delta\hat{\omega}_{33}, \delta\hat{\tau}_{22}) = (0, 0, 0.25)$.

Note that in the notation introduced above, this means that, in turn, we take H_1, H_2, and H_3 to be the correct hypotheses; the resulting distributions of log odds ratio for single sources as well as catalogs of sources will be our foreground distributions. We also consider the case where the no-hair theorem holds, *i.e.* $\mathcal{H}_{\rm GR}$ is the correct hypothesis and the injections have $\delta\hat{\omega}_{22} = \delta\hat{\omega}_{33} = \delta\hat{\tau}_{22} = 0$. The log odds ratio distributions resulting from the latter, again for single sources and catalogs of sources, will be our backgrounds.

2.4 Testing the no-hair theorem with TIGER

Let us first focus on background and foreground distributions for the case where only a single ringdown detection is ever made by ET, *i.e.* $\mathcal{N} = 1$. Results for ~ 1300 sources are shown in Fig. 2.1. We see that in all of the cases, there is significant overlap between background and foreground, so that with a single source one has no guarantee that violations of the no-hair theorem at these levels will be picked up. With a maximum tolerable false alarm probability of $\beta = 0.05$, the efficiency ζ for a 10% shift in ω_{22} is 0.47, and for a 10% shift in ω_{33} it is 0.46. Note how the efficiencies for deviations in ω_{22} and ω_{33} are comparable; with our choice for the injected range of mass ratios ($0.3 < q < 1$, or $0.18 \lesssim \nu < 0.25$) there will be sources with $A_{33} > A_{22}$ as well as sources with $A_{33} < A_{22}$; see Fig. 1 of [74]. For a 10% shift in τ_{22} we find $\zeta = 0.05$. Thus, even with only a single detection, one will have a reasonable chance of finding a GR violation of the given size in ω_{22} and ω_{33}; however, the same shift in τ_{22} will be essentially unobservable.

At least for anomalies in ω_{22} and ω_{33}, the situation changes dramatically if information from multiple detections can be combined. This is shown in Fig. 2.2, for catalogs of 10 sources each. For the same maximum false alarm probability and the given shifts in ω_{22}, ω_{33}, and τ_{22}, the efficiencies become, respectively, 0.98, 0.98, and 0.13. Thus, there is a significant improvement in the first two cases, but the shift in τ_{22} remains hard to observe.

It is also of interest to see how the efficiencies grow with an increasing number of sources per catalog. This is shown in Fig. 2.3, for two choices of maximum tolerable false alarm probability: $\beta = 0.05$ and $\beta = 0.01$. Due to the finite number of catalogs considered, inevitably the numbers we quote for efficiencies are not exact; in the plot we show medians and 95% confidence intervals obtained for ζ when randomly combining the available simulated sources into catalogs of a given size in 1000 different ways. For the cases $\delta\hat{\omega}_{22} = 0.1$ and $\delta\hat{\omega}_{33} = 0.1$, we see that for either value of β, the efficiency reaches essentially 100% for ~ 20 sources per catalog. However, for a GR violation with $\delta\hat{\tau}_{22} = 0.1$ and as many as 50 sources per catalog, even with $\beta = 0.05$ the median efficiency is only ~ 0.2, with a large spread.

One may then wonder how large a deviation in τ_{22} needs to be before it becomes detectable with good efficiency, still assuming a few tens of sources per catalog. In the bottom right panel of Fig. 2.3, we show the evolution of median efficiencies and 95% confidence intervals for the case where $\delta\hat{\tau}_{22} = 0.25$. Here the efficiencies rise more steeply with the number of detections available, with the median efficiency for $\beta = 0.05$ reaching $\sim 50\%$, albeit still with a considerable spread.

We see that, by combining information from multiple sources, we greatly improve our ability to use ringdown signals observed by ET in testing the no-hair theorem. However, the advantages of TIGER are not limited to this. The use of multiple sub-hypotheses $H_{i_1 i_2 \ldots i_k}$ also has a significant impact in finding a deviation from GR, as illustrated in Fig. 2.4. Here we arrange simulated sources in order of increasing SNR, and we consider the Bayes factors $B^{i_1 i_2 \ldots i_k}_{\rm noise}$ and $B^{\rm GR}_{\rm noise}$ for the hypotheses $H_{i_1 i_2 \ldots i_k}$ and $\mathcal{H}_{\rm GR}$ against the noise-only hypothesis, respectively. In particular, what is plotted is the cumulative number of times that the Bayes factor against noise for a particular

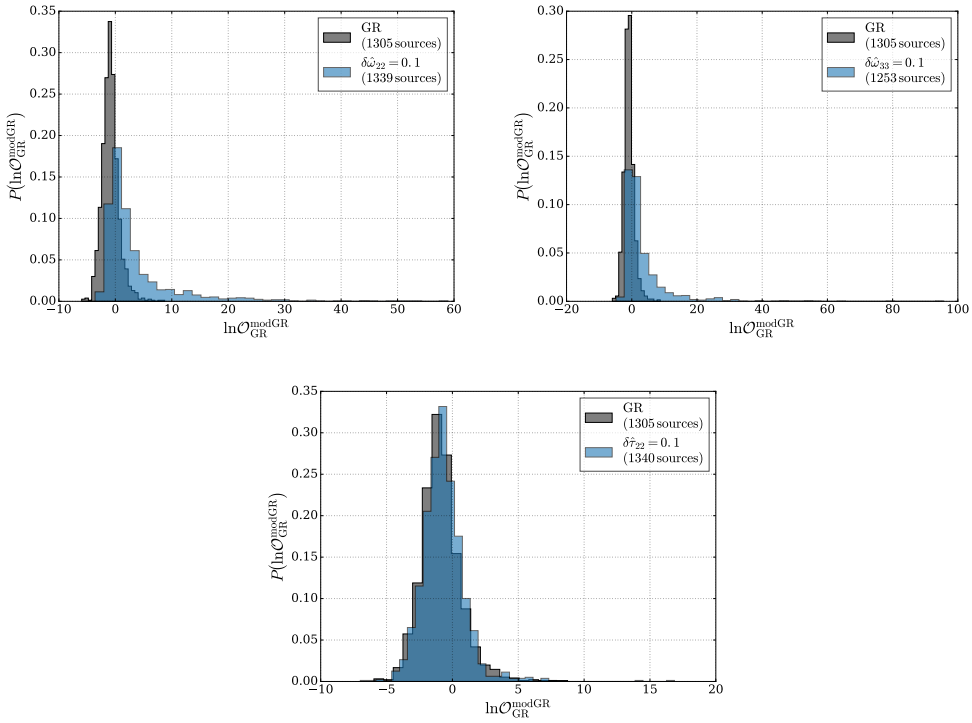


Figure 2.1: Single-source GR background distributions (dark gray) and foreground distributions (light gray), for a 10% deviation in ω_{22} (left), a 10% deviation in ω_{33} (middle), and a 10% deviation in τ_{22} (right). In all three cases there is significant overlap between background and foreground; for a maximum tolerable false alarm probability of $\beta = 0.05$, the efficiencies are, respectively, 47%, 46%, and 5%. The figures demonstrate that the chances of spotting a violation of GR using single sources are slim for ω_{22} and ω_{33} to nearly none in τ_{22}.

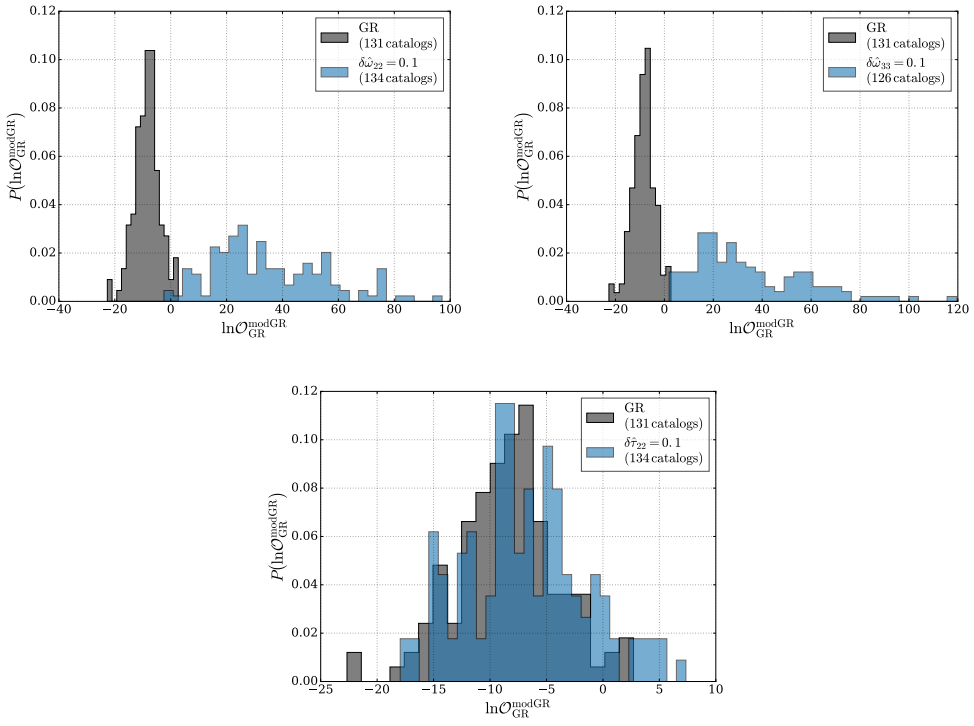


Figure 2.2: GR background distributions (dark gray) and foreground distributions (light gray), for a 10% deviation in ω_{22} (left), a 10% deviation in ω_{33} (middle), and a 10% deviation in τ_{22} (right). This time we considered *catalogs* of 10 sources each. Again with $\beta = 0.05$, this time efficiencies of 98% are attained for the two mode frequencies. On the other hand, the deviation in τ_{22} remains hard to detect, with an efficiency of only 14%. Here, as opposed to the single-source case in Fig. 2.1, we find that the foreground and background are almost completely separated for ω_{22} and ω_{33}, making it unlikely to miss the violation.

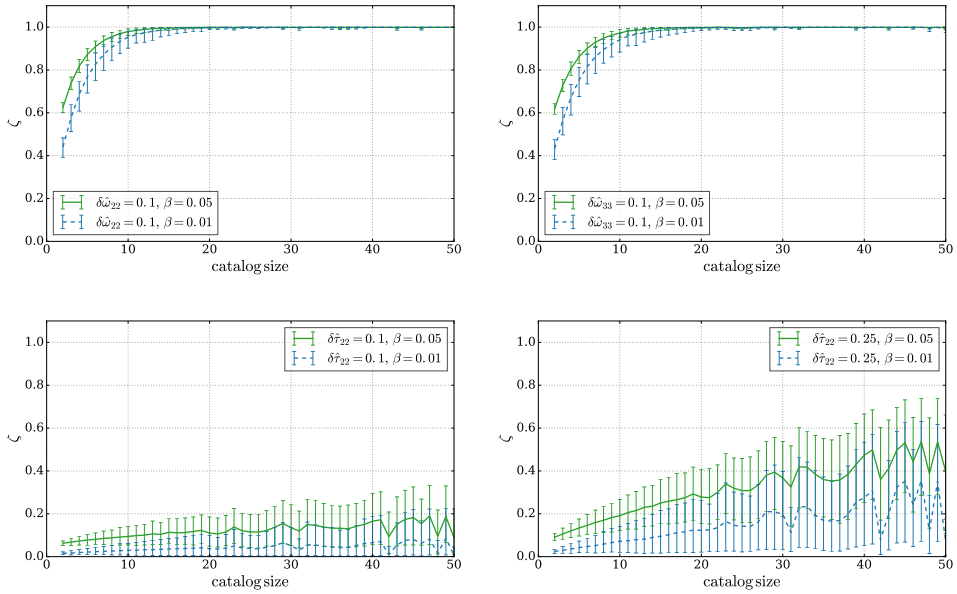


Figure 2.3: Growth of the efficiency ζ by which a violation may be spotted, with the number of sources per catalog, for a 10% deviation in ω_{22} (top left), ω_{33} (top right), τ_{22} (bottom left) and a 25% deviation in τ_{22} (bottom right), for maximum tolerable false alarm probabilities $\beta = 0.05$ and $\beta = 0.01$, respectively. In order to understand uncertainties in ζ due to having a finite number of catalogs, the available simulated sources were randomly combined into catalogs to obtain 1000 different realizations. Shown are the median efficiencies (solid and dashed lines) and 95% confidence intervals. Especially for ω_{22} and ω_{33}, the gain in ability to spot the 10% deviation as more sources are combined is evident, resulting in near 100% efficieny after only 20 combined sources.

hypothesis is the largest. We do this for the case where $\delta\hat{\omega}_{22} = 0.1$, so that the correct hypothesis is H_1. For SNRs up to ~ 18, we see that the GR hypothesis dominates. Going to higher SNRs, the correct hypothesis comes out on top the largest number of times. Even so, *incorrect hypotheses often dominate*. For example, the number of times that the incorrect hypothesis H_{12} has the largest Bayes factor against noise is not significantly lower than the number of times that H_1 has the largest Bayes factor. As the right hand panel in the Figure shows, at SNRs between 8 and 12, the hypothesis H_3 tends to be the most dominant after $\mathcal{H}_{\rm GR}$, yet it does not even involve $\delta\hat{\omega}_{22}$, where the GR violation occurs! Note also that in the latter SNR range, the least parsimonious hypothesis, H_{123}, does particularly badly.

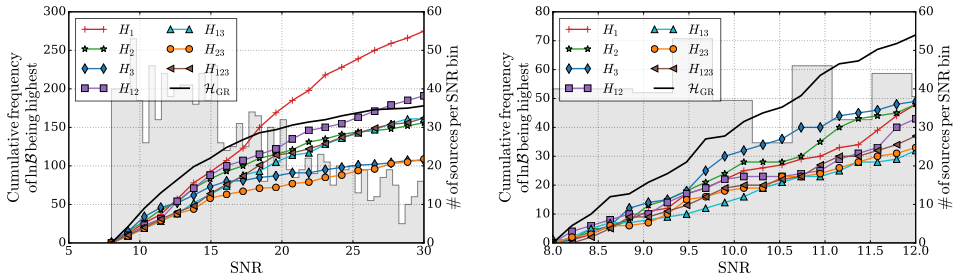


Figure 2.4: Left: The cumulative number of times that a given hypothesis (any of the $H_{i_1 i_2 \ldots i_k}$, or $\mathcal{H}_{\rm GR}$) has the largest Bayes factor against noise ($B^{i_1 i_2 \ldots i_k}_{\rm noise}$, or $B^{\rm GR}_{\rm noise}$), for single sources up to an SNR of 30. Right: The same, but focusing on SNRs up to 12. Note that the correct hypothesis H_1 is not necessarily always on top nor is the most inclusive one H_{123}, illustrating the importance of including all sub-hypotheses.

The above pertained to single sources. In Fig. 2.5 we consider, for an example *catalog* of sources, the evolution of the *combined* Bayes factors

$$\mathcal{B}^{i_1 i_2 \ldots i_k}_{\rm GR} = \prod_{A=1}^{\mathcal{N}} {}^{(A)}B^{i_1 i_2 \ldots i_k}_{\rm GR}, \tag{2.29}$$

as well as $\ln \mathcal{O}^{\rm modGR}_{\rm GR}$, as information from more and more detections is added; the sources are arranged in order of increasing SNR. We can make two observations:

- The hypothesis H_{123} where all the parameters $\{\delta\hat{\omega}_{22}, \delta\hat{\omega}_{33}, \delta\hat{\tau}_{22},\}$ are left free does not dominate the log odds ratio, and indeed is deprecated compared with some of the other sub-hypotheses. This illustrates how one can suffer significant loss in discriminatory power if the non-GR model is insufficiently parsimonious, *i.e.* has more free parameters compared with the number of parameters that is actually needed. TIGER does not have this problem.

- The correct hypothesis, in this case H_1, is *also* not necessarily the dominant one. Indeed, it can happen that detector noise obscures the true nature of the GR violation so that some other hypothesis (in this example H_{12}) ends up on

top. However, what is unlikely to happen is that the noise makes a non-GR signal look like a GR one. In a situation where most signals are weak, one should use TIGER with as many testing parameters $\{\delta\xi_1, \delta\xi_2, \ldots, \delta\xi_{N_T}\}$ as is computationally feasible.

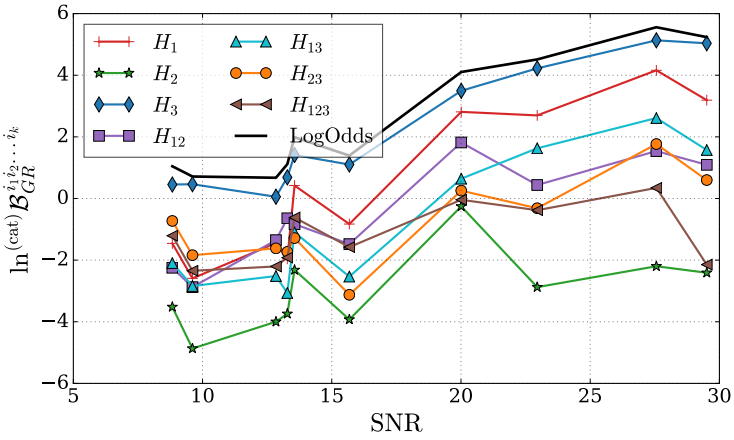


Figure 2.5: The progression of *combined* log Bayes factors within an example *catalog* of 20 sources. Shown are $\ln \mathcal{B}^{i_1 i_2 \ldots i_k}_{\mathrm{GR}}$, as well as the log odds ratio $\ln \mathcal{O}^{\mathrm{modGR}}_{\mathrm{GR}}$, with an increasing number of sources (sorted by SNR), for the case where the injections have $\delta\hat{\omega}_{22} = 0.1$. Here we demonstrate that the inclusion of all sub-hypotheses is important as the most inclusive hypothesis H_{123} contributes relatively little to the final value for $\ln \mathcal{O}^{\mathrm{modGR}}_{\mathrm{GR}}$ and even the correct hypothesis H_1 is not the most dominant one.

So far we have considered situations where GR violations are present, and we have studied how well one would be able to find them with TIGER, depending on the size of the violations and the number of detections available. In the next section we consider a scenario where the measured log odds ratio is consistent with GR.

2.5 Constraining the free parameters

If, when comparing the measured log odds ratio with some reasonable threshold, there turns out to be no reason to doubt the validity of GR, then one can consider *measuring* the parameters $\delta\hat{\omega}_{22}$, $\delta\hat{\omega}_{33}$, and $\delta\hat{\tau}_{22}$ in order to see what constraints can be put on them. Indeed, within the Bayesian parameter estimation framework implemented by Veitch and Vecchio [103, 104, 63] that we use here (see also [105] for comparisons with other Bayesian methods), given a waveform model corresponding to a hypothesis $\mathcal{H}$ with parameters $\boldsymbol{\lambda}$, the joint *posterior density function* (PDF) is obtained through

$$p(\boldsymbol{\lambda}|\mathcal{H}, d, I) = \frac{p(\boldsymbol{\lambda}|\mathcal{H}, I)\, p(d|\mathcal{H}, \boldsymbol{\lambda}, I)}{p(d|\mathcal{H}, I)}, \tag{2.30}$$

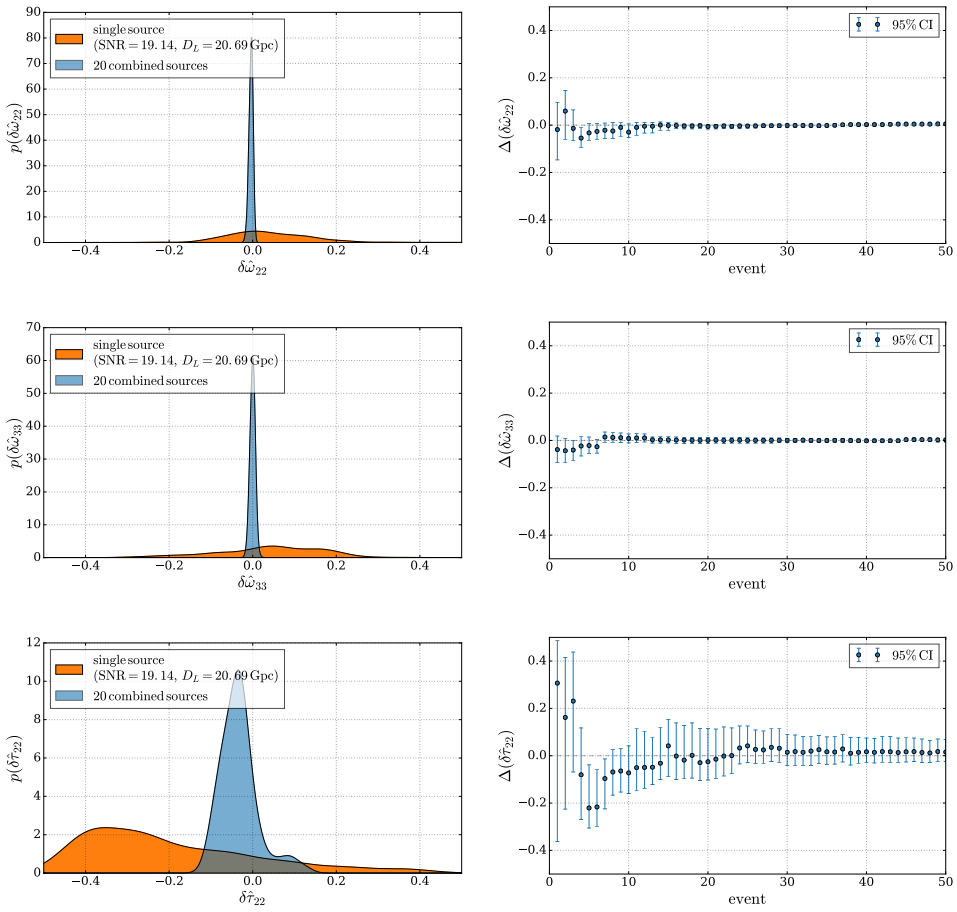


Figure 2.6: In this figure we demonstrate the evolution a combined posterior density function as more sources are included. Left panels: Posterior density functions for $\delta\hat{\omega}_{22}$ (top), $\delta\hat{\omega}_{33}$ (middle), and $\delta\hat{\tau}_{22}$ (bottom), both for a single source at a distance of 20.69 Gpc ($z = 2.47$) with an SNR of 19.14, and for a catalog of 20 sources. Right: Evolution of medians and 95% confidence intervals of PDFs as more and more sources are included.

where $p(\boldsymbol{\lambda}|\mathcal{H}, I)$ is the prior distribution of parameters before any measurement has been made, $p(d|\mathcal{H}, \boldsymbol{\lambda}, I)$ is the likelihood function (*i.e.* the probability of obtaining the data d given $\mathcal{H}$ and parameter values $\boldsymbol{\lambda}$), and $p(d|\mathcal{H}, I)$ is the prior probability of the data, which can be absorbed in a normalization factor. The probability density function for an individual component λ_1 of the vector $\boldsymbol{\lambda}$ is obtained straightforwardly by marginalizing over all the other parameters:

$$p(\lambda_1|\mathcal{H}, d, I) = \int \mathrm{d}\lambda_2 \mathrm{d}\lambda_3 \ldots \mathrm{d}\lambda_N \, p(\boldsymbol{\lambda}|\mathcal{H}, d, I). \tag{2.31}$$

If one anticipates λ_1 to be the same for all detections $d_1, d_2, \ldots, d_{\mathcal{N}}$ then one can combine PDFs across sources through

$$\begin{aligned} &p(\lambda_1|\mathcal{H}, d_1, d_2, \ldots, d_{\mathcal{N}}, I) \\ &= p(\lambda_1|\mathcal{H}, I)^{1-\mathcal{N}} \prod_{A=1}^{\mathcal{N}} p(\lambda_1|\mathcal{H}, d_A, I), \end{aligned} \tag{2.32}$$

as was done in a different context in e.g. [106].

In the present context one can obtain PDFs for $\delta\hat{\omega}_{22}$, $\delta\hat{\omega}_{33}$, and $\delta\hat{\tau}_{22}$ by e.g. letting $\mathcal{H}$ be H_1, H_2, and H_3, respectively. Now, if there is evidence that GR is violated (because of the measured $\ln \mathcal{O}^{\mathrm{modGR}}_{\mathrm{GR}}$ being above threshold), then there is no *a priori* reason to assume that the $\{\delta\hat{\omega}_{22}, \delta\hat{\omega}_{33}, \delta\hat{\tau}_{22}\}$ will be the same for all sources. Note that although this is the choice we made for injections in the previous section, even if these additional parameters had been non-constant it would not have been a problem to do *model selection* with multiple sources, since each of the Bayes factors ${}^{(A)}B^{i_1 i_2 \ldots i_k}_{\mathrm{GR}}$ only gauge whether the hypothesis $H_{i_1 i_2 \ldots i_k}$ is more probable than $\mathcal{H}_{\mathrm{GR}}$. In doing *parameter estimation* one has to be more careful as will be discussed in more detail in Sec. 4.1.

On the other hand, suppose that there is no evidence of GR being incorrect; *i.e.*, the measured $\ln \mathcal{O}^{\mathrm{modGR}}_{\mathrm{GR}}$ is well below threshold. Then one can expect that $\{\delta\hat{\omega}_{22}, \delta\hat{\omega}_{33}, \delta\hat{\tau}_{22}\}$ are all constant, namely $\delta\hat{\omega}_{22} = \delta\hat{\omega}_{33} = \delta\hat{\tau}_{22} = 0$, and it makes sense to combine PDFs from multiple sources as in Eq. (2.32). In turn, we let $\mathcal{H}$ be H_1, H_2, and H_3, and compute marginalized PDFs for $\delta\hat{\omega}_{22}$, $\delta\hat{\omega}_{33}$, and $\delta\hat{\tau}_{22}$, respectively.

Results are shown in Fig. 2.6. In the top panels, we consider PDFs both for an example single source at $D_{\mathrm{L}} = 20.69\,\mathrm{Gpc}$ ($z = 2.47$), and for a catalog of 20 sources. For the single source, the spreads of the PDFs are roughly consistent with an extrapolation of the results of Gossan *et al.* to the given luminosity distance. (For injections and templates with non-spinning progenitors and $D_{\mathrm{L}} < 6\,\mathrm{Gpc}$, we get uncertainties that are in close agreement with theirs.) As expected, the single-source PDFs are quite wide and uninformative, with standard deviations of 0.10, 0.13, and 0.21, respectively. For $\delta\hat{\omega}_{22}$ and $\delta\hat{\omega}_{33}$, with 20 sources the PDFs become strongly peaked (with widths of 0.0051 and 0.0066, respectively), and there is little bias. For $\delta\hat{\tau}_{22}$ the combined PDF is not only wide (with a standard deviation of 0.048), it also shows some bias (although the correct value of zero is within its support). In the bottom panels of the figure, we show the evolution of medians and 95% confidence intervals for the combined PDF as more and more detections are added. We see that even for $\delta\hat{\tau}_{22}$, the 95% confidence interval shrinks to ~ 0.1 when ~ 40 sources are at

our disposal. Hence there is a clear advantage in computing PDFs using all available detections.

2.6 Conclusions

We have revisited the problem of testing the no-hair theorem by using ringdown signals that will be seen by Einstein Telescope. In previous work [74], it was shown how deviations of up to 10% in the ringdown mode frequencies ω_{22}, ω_{33} and the damping time τ_{22} could be observed out to distances of $\sim 6\,\mathrm{Gpc}$, both through parameter estimation and model selection. Here we used the TIGER framework that was originally developed to test general relativity with stellar mass binary inspiral signals in second-generation detectors [66, 67, 93, 94, 68]. In this model selection scheme, parameterized deviations are introduced in the waveforms, and multiple auxiliary hypotheses are tested corresponding to all subsets of the extra free parameters. Information from multiple sources can trivially be combined. A log odds ratio $\ln \mathcal{O}^{\mathrm{modGR}}_{\mathrm{GR}}$ is computed, which compares the probability that one or more of the auxiliary hypotheses are correct with the probability that GR is the right theory. Given the expected distribution of $\ln \mathcal{O}^{\mathrm{modGR}}_{\mathrm{GR}}$ in the case that GR is correct, violations of GR are searched for by checking whether the measured log odds ratio is above a threshold set by a predetermined maximum false alarm probability. If this is not the case then there is no reason to doubt GR, and one can calculate bounds on the free parameters, again combining information from all available sources.

Ringdown signals from black holes with masses in the range $500 - 1000\,\mathrm{M}_{\odot}$ can result from coalescences of intermediate-mass binary black holes, but such events may be rare [88, 89, 90, 91, 92]. On the other hand, they can be seen with ET up to distances of 50 Gpc ($z \sim 5$). We have shown that with $\mathcal{O}(10)$ sources and using the TIGER framework, deviations of the same size as the ones considered in Gossan *et al.* [74] (10% in ω_{22}, ω_{33}) can be seen up to 50 Gpc instead of 6 Gpc.

Our work illustrates how TIGER is not tied to any particular waveform model (nor even any particular type of source). It is well suited to the regime of low signal-to-noise ratios due to its use of multiple sub-hypotheses, which increases the chance of finding a GR violation. Because of detector noise, the correct hypothesis (or for that matter, the most inclusive hypothesis) may not yield the largest contribution to the log odds ratio, but it is unlikely that noise will make a GR-violating signal look like one that is in accordance with GR. For concreteness we only considered possible deviations in $\{\omega_{22}, \omega_{33}, \tau_{22}\}$ (as was also done in [74]), leading to seven auxiliary hypotheses in the TIGER framework, but in reality one should include as many sub-hypotheses as is feasible and modeling allows.

TIGER offers an effective way of finding generic violations of GR. As shown in the context of compact binary coalescence, it can uncover deviations that are not included in any of the waveform models associated with the sub-hypotheses $\mathcal{H}_{i_1 i_2 \ldots i_k}$ [66, 67]. We fully expect the same to be true for ringdown; an example could be the appearance of modes with spin weights different from -2, as in the case of a black hole in certain $f(R)$ theories that are dynamically equivalent to Einstein-Proca theory [107, 108, 109]. An explicit demonstration is left for future work. On the other hand,

if the signal waveform is not among any of the template models, then fundamental bias can make it difficult to reliably pinpoint the underlying nature of the violation [110, 111, 66, 112, 113]. Also for ringdown, this remains an open problem.

When the log odds ratio does not indicate a violation of GR, upper limits can be put on deviations in the extra free parameters. A single source at large distance ($> 10\,\mathrm{Gpc}$) may only give weak bounds and could show considerable bias. On the other hand, with $\mathcal{O}(10)$ sources, deviations are well-constrained even for the parameter τ_{22}, for which no meaningful bounds can be obtained with a single source at SNR ~ 20.

As found in [74] for the case of *single* systems with $M \sim 10^6\,\mathrm{M}_\odot$, LISA will be able to perform tests of the no-hair theorem at a comparable level of accuracy as ET with $M \sim 10^3\,\mathrm{M}_\odot$ (Fig. 6 in [74]). Since the detection rate for such sources with LISA may be in the order of tens per year [114] (*i.e.* what we assumed for ET in this paper), results from TIGER, including the combining of information from multiple sources, should also be similar. Detailed investigations are left for future work.

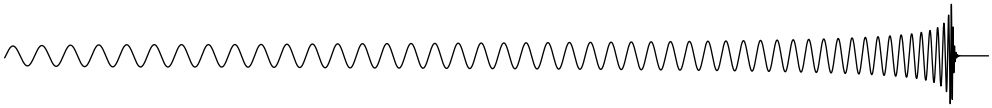
In light of the two BBH detections made with Advanced LIGO, we already expect to be able to place stringent constraints on GR violations well before ET will start taking data. However, we have demonstrated that even when we only have access to the ringdown, interesting tests of the strong-field dynamics of GR can still be done. In this study we deliberately restricted attention to a black hole mass range for which the preceding inspiral signal can not be seen, but the dominant ringdown mode is visible. The ringdown itself is of particular interest when investigating the black hole nature of the remnant.

Already attempts have been made to measure the properties of the least-damped QNM of GW150914. For example in [73] the mode frequency and damping time were found to be consistent with GR, but no strong constraint could be placed yet. As the QNMs are only a valid description of the waveform from about 3 ms after the merger takes place, the SNR is likely to be low and parameter estimation with a single source proves difficult. To achieve a reasonable SNR from only this part of the waveform, the source would need to be either 1) reasonably close by or 2) massive enough so that the ringdown falls within the most sensitive part of the Advanced LIGO and Advanced Virgo frequency bandwidth, or a combination of both. In conclusion, the method of testing GR presented in this chapter will allow us to make use of sources where the SNR is not ideal by way of combining information from multiple sources.

The tests described here can not directly be applied to sources like GW150914, *i.e.* sources that also contain parts of the late-inspiral and merger. Currently we are using a time domain waveform which needs to be Fourier transformed into the frequency domain to perform parameter estimation. Such a transformation always causes "spectral leaking", meaning that due to abrupt cut-off in time domain, power will be leaked into parts of the frequency domain that will not be present in the detector signal. This will likely cause severe bias in parameter estimation especially for low SNR signals. In conclusion, the *method* of testing GR presented in this chapter would allow us to test the no-hair theorem using sources where the SNR is not ideal by way of combining information from multiple sources. The next step is to apply this method with a frequency domain waveform, which is currently ongoing work.

Chapter 3

TIGER extended to light black hole binaries



The TIGER framework was introduced in Sec. 1.6 and has passed internal review by the collaboration for BNS systems. The goal of the work presented in this chapter is to extend the applicability of the TIGER pipeline to more massive systems, in particular to light binary black holes. There is no strict definition as to what "light" black hole binaries are, but in this chapter we restrict ourselves to component masses of up to $15\,\mathrm{M}_\odot$ [1]. This extension is not a trivial one; there are a few features of BBH systems that need to be considered carefully:

- Black holes in binary systems may have dimensionless spin magnitudes up to the maximum of 1. Furthermore, these spins may not be aligned and orbital precession becomes an important factor.
- Due to their heavier mass as opposed to BNS systems, BBH systems merge at lower frequencies, making the merger and ringdown become an important feature of the waveform since it will occur closer to the most sensitive part of the detector's frequency bandwidth.
- Also due to the fact that BBH systems are more massive, their signals will be shorter, i.e. the time they spend within the detector's sensitivity band will be shorter. This means that in our analysis we become more vulnerable to transient bursts of noise, called *glitches*, in the data.

A defining factor of how successful a TIGER analysis can be is the shape of the GR background, i.e. the distribution of odds ratios Eq. (1.161) obtained from simulated pure GR sources injected into a representative or actual detector data stream. In this work we aim to investigate what happens to the GR background when we inject light BBH sources. For example, glitches in the data could cause outliers in the odds ratio distribution (odds ratios strongly favoring the mod-GR hypothesis), making the distinction between actual violations of GR or background difficult and in case of large outliers unfeasible. If outliers do appear that can not easily be explained by large glitches, then we want to see if an observation-based cut can be applied to avoid these; in practice we would expect this to be a cut based on the chirp mass, as all other parameters will be hard to measure.

In order to perform background calculation studies, many simulated sources, typically on the order of thousands need to be analyzed to get a decent sense of the

[1] We limited the range for this investigation based on the fact that the LSO for a system with $m_1 = m_2 = 15\,\mathrm{M}_\odot$ is approximately 150 Hz, which is the most sensitive region in the advanced detectors bandwith. This will leave us with sources that still have a bit of inspiral left in band.

statistics involved. To make this computationally feasible, it is best to use a waveform model that is a closed-form analytical description of the frequency domain waveform. This omits the need for performing Fourier transforms and any additional evaluation of differential equations. Since we are trying to perform a study of the feasibility of testing general relativity using BBH systems, it is important to include as much of the relevant physics as possible, i.e. we need the full inspiral-merger-ringdown waveform including precession. At the time this study was performed, the most suitable waveform model was a phenomenological one called IMRPhenomPv1, which accurately models precessing BBH sources up to mass ratios of ~ 4 and to moderate spins with a magnitude of around 0.85.

This chapter is structured as follows: Sec. 3.1 introduces IMRPhenomPv1+, which is the model IMRPhenomPv1 extended to allow for parameterization of non-GR effects and we investigate how this parameterization affects the waveforms. We generate the GR background distribution in Sec. 3.2 and investigate any anomalies or outliers we find. The chapter is closed with some concluding remarks in Sec. 3.3.

3.1 Non-GR parameterization of IMRPhenomPv1

The precessing inspiral-merger-ringdown (IMR) phenomenological waveform IMRPhenomPv1 is based on the non-precessing IMR waveform IMRPhenomC. The precession is achieved by rotating the aligned-spin waveform in a similar fashion as is done for IMRPhenomPv2, described in Sec. 1.4.2-D. The difference is in the underlying aligned-spin waveform, which is also the model into which we will introduce our non-GR parameterization.

Since IMRPhenomPv1 is based on IMRPhenomC, we can separate the model into three dynamical stages that characterize how the binary coalesces. The first stage, the *inspiral*, is where the analytical standard phase approximation and post-Newtonian expansion still give an accurate description of the waveform dynamics. This stage is determined by PN coefficients φ_j, where $j = 0, \ldots, 7$ and the logarithmic PN coefficients $\varphi_{5l}, \varphi_{6l}$ defined in Eq. (1.108). This stage of the phase evolution is known *analytically* up to $(v/c)^7$. The second stage is called the *pre-merger* and is described by the phenomenological parameters ξ_n, where $n = 1, \ldots, 6$. The final stage, the *post-merger*, is a phenomenological representation of the ringdown, parameterized by ζ_1 and ζ_2.

We build the non-GR waveform model IMRPhenomPv1+ by introducing a set of testing coefficients $\delta\hat{p}_i$, which take the form of *fractional* deviations in the phasing parameters p_i:

$$p_i \to (1 + \delta\hat{p}_i) p_i^{\mathrm{GR}}. \tag{3.1}$$

In principle we are free to introduce such deviations to all stages of the waveform model. We omit the two post-merger parameters as they are determined from the set of $\{\xi_n\}$. The ξ_n themselves are all mapped to physical coordinates and have particular values that are determined by GR. We therefore consider deviations on the complete set of pre-merger coefficients ξ_n and the PN coefficients. The full set of testing coefficients that make up IMRPhenomPv1+ are then

inspiral $\{\delta\hat{\varphi}_0, \ldots, \delta\hat{\varphi}_7\}$ and $\{\delta\hat{\varphi}_{5l}, \delta\hat{\varphi}_{6l}\}$

pre-merger $\{\delta\hat{\xi}_1, \delta\hat{\xi}_2, \delta\hat{\xi}_3, \delta\hat{\xi}_4, \delta\hat{\xi}_5, \delta\hat{\xi}_6\}$

Note that $\{\delta\hat{\varphi}_i\}$ includes a 0.5PN testing parameter $\delta\hat{\varphi}_1$; since φ_1 is identically zero in general relativity, we let $\delta\hat{\varphi}_1$ be an absolute rather than a relative deviation. The frequency-dependence of each PN term is listed under the early-inspiral part of Table 1.1. The frequency-dependence of ξ_n is given in Eq. (1.115).

3.1.1 Effects of non-GR parameterization on the waveform

In the tests of GR performed with the TIGER pipeline, we are most sensitive to the phase and we leave the amplitude unperturbed. It is therefore fruitful to investigate the impact a particular fractional change in any of the $\delta\hat{p}_i$ has on the phase evolution of the waveform. Generating a background is computationally expensive and the number of runs that need to be performed scales strongly with the number of testing coefficients N_T that are considered ($\propto 2^{N_T} - 1$). Thus, more insight into the effects of the non-GR parameters may help us construct an ideal subset from all $\delta\hat{p}_i$. Moreover, a better understanding of the impact certain parameters have on the phase evolution may help us determine what a reasonable prior on $\delta\hat{p}_i$ may be.

In this chapter we aim to investigate the effect on the background when we include BBH systems. In doing so, we now include the merger and ringdown part of the waveform and corresponding phenomenological coefficients. As we increase the mass of the system, we expect to become less sensitive to the inspiral parameters as the merger-ringdown part moves closer to the most sensitive part of the detectors' frequency bandwidth and we lose GW cycles from the inspiral part.

To see to what degree this effect takes place, we can compare the phase calculated with IMRPhenomPv1 and that of IMRPhenomPv1+: $\Delta\Psi(f_{\rm ref}) = \Psi_{\text{non-GR}}(f_{\rm ref}) - \Psi_{\rm GR}(f_{\rm ref})$ at a certain reference frequency $f_{\rm ref}$ for a range in chirp masses $\mathcal{M}_c$. To test this effect we chose $f_{\rm ref} = 150\,\text{Hz}$; roughly the point where the detectors' sensitivity is at its best. In Fig. 3.1 we plot $\Delta\Psi(150\,\text{Hz})$ for a wide range in chirp masses $3\,\text{M}_\odot \leq \mathcal{M}_c \leq 30\,\text{M}_\odot$. We consider a symmetric system with $\eta = 0.25$ which, with the choice of chirp masses, corresponds to component masses in the range $3.4\,\text{M}_\odot \leq m_a \leq 17.2\,\text{M}_\odot$.

As expected, we see that for increasing chirp masses the pre-merger non-GR parameters become more important, whereas a relative shift in the PN coefficients becomes less distinguishable. When one needs to limit the number of testing coefficients used in a TIGER analysis using IMRPhenomPv1+, it is best to focus on the ones that have a relatively large impact on the phasing of the waveform and using Fig. 3.1 as guideline, the testing coefficients can be ranked accordingly. For the pre-merger parameters this ranking would be $\{\delta\hat{\xi}_3, \delta\hat{\xi}_4, \delta\hat{\xi}_5, \delta\hat{\xi}_2, \delta\hat{\xi}_1, \delta\hat{\xi}_6\}$ from largest $\Delta\Psi(150\,\text{Hz})$ to smallest. Indeed the relative size of $\Delta\Psi(150\,\text{Hz})$ depends on the chirp mass, but the general trend provides a reasonable suggestion as to a ranking in effect size. For the PN coefficients there are four that stand out from the others: $\delta\hat{\varphi}_0$, $\delta\hat{\varphi}_1$, $\delta\hat{\varphi}_2$ and $\delta\hat{\varphi}_3$. The most significant phase difference is seen when introducing a shift in $\delta\hat{\varphi}_0$, since $\Delta\Psi$ is a factor of 5 higher than the second largest it is omitted from the plot to better distinguish between the remaining parameters. When considering the

magnitude of $\Delta\Psi$, the other three, $\delta\hat{\varphi}_1$, $\delta\hat{\varphi}_2$ and $\delta\hat{\varphi}_3$ show similar effects. When we again consider the magnitude of $\Delta\Psi$ we note that the higher order PN terms might be ranked as follows (in order of decreasing effect size): $\delta\hat{\varphi}_4$, $\delta\hat{\varphi}_6$, $\delta\hat{\varphi}_{6l}$, $\delta\hat{\varphi}_7$, $\delta\hat{\varphi}_{5l}$. These rankings are mere suggestions for selecting testing parameters to use in TIGER, only a full TIGER analysis including the generation of a foreground will demonstrate what is truly the optimal set. However, since such an investigation is computationally costly, the sub-set one chooses to use could be based on the more basic principle demonstrated here.

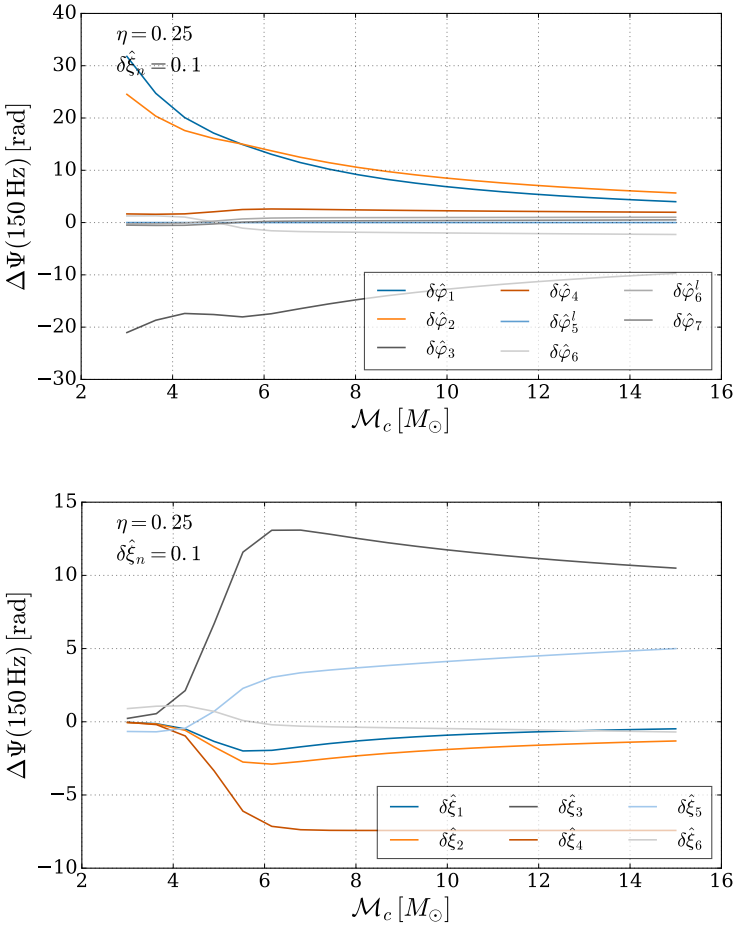


Figure 3.1: Phase difference $\Delta\Psi(150\,\mathrm{Hz})$ between IMRPhenomPv1 and IMRPhenomPv1+ at 150 Hz for the two sets of non-GR parameters. Parameters considered are those of the inspiral (left), and pre-merger (right) stages. The 0PN coefficient is omitted to aid distinguishabiliy between the other curves as $\delta\hat{\varphi}_0$ rises up 5 times as much as the others at $\mathcal{M}_c = 3.4$.

When using non-GR parameters, it may also be of interest to see what impact a particular parameter may have across the whole frequency range. To further investigate this, we also generated $\Delta\Psi(f)$ shown in Fig. 3.2, where f runs from 20 Hz to 1024 Hz, for a range in parameter values. Four parameters are considered: $\delta\hat{\varphi}_3$ (top left), $\delta\hat{\varphi}_4$ (top right), $\delta\hat{\xi}_3$ (bottom left) and $\delta\hat{\xi}_4$ (bottom right). The reasoning behind this choice is based on the ranking suggested by Fig. 3.1 and what parameters might be interesting to use in a background study. The PN parameter $\delta\hat{\varphi}_3$ is chosen because it is physically interesting: 1.5PN is the lowest PN order where dynamical non-linearities take place and it is the lowest order in which spin-orbit interactions enter. The next PN parameter $\delta\hat{\varphi}_4$ is chosen simply because it appears to have the strongest effect after $\delta\hat{\varphi}_3$. Aside from the physically interesting PN orders, one would also want to include parameters representing the genuinely strong-field dynamics when performing any test of GR with gravitational waves. To this end we decided on the most influential of the $\delta\hat{\xi}_n$ parameters: $\delta\hat{\xi}_3$ and $\delta\hat{\xi}_4$.

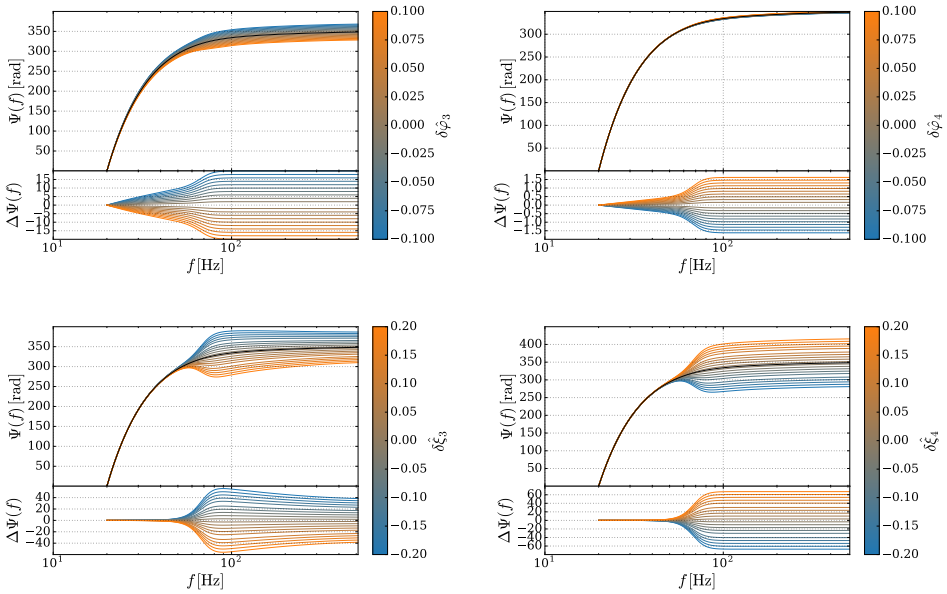


Figure 3.2: Phase evolution $\Psi(f)$ and difference $\Delta\Psi(f) = \Psi_{\text{non-GR}}(f) - \Psi_{\text{GR}}(f)$ as a function of frequency. Each curve corresponds to a particular shift of the non-GR parameter presented in the plot, except for the black curve, which represents the GR case i.e. $\delta\hat{p}_i = 0$. The non-GR parameters considered are $\delta\hat{\varphi}_3$ (top left), $\delta\hat{\varphi}_4$ (top right), $\delta\hat{\xi}_3$ (bottom left) and $\delta\hat{\xi}_4$ (bottom right), i.e. the phenomenological parameters that were ranked as having the strongest influence on the phasing and the lowest PN order to exhibit dynamical non-linearities and spin-orbit coupling ($\delta\hat{\varphi}_3$) and the next highest ranked PN parameter after that, $\delta\hat{\varphi}_4$. The source parameters in this case are similar to GW151226: $\mathcal{M}_c = 8.9$, $\eta = 0.22$ and spin magnitudes $|\mathbf{S}_1| \approx |\mathbf{S}_2| = 0.3$.

3.2 Background distribution

In TIGER, a violation from GR can be detected by comparing the odds ratio $\mathcal{O}^{\rm modGR}_{\rm GR}$ (the relative evidence between the modGR hypothesis $\mathcal{H}_{\rm modGR}$ and the GR hypothesis $\mathcal{H}_{\rm GR}$ resp., see Eq. (1.147)) of the detected signal with a large distribution of odds ratios calculated from simulated GR signals injected into detector noise. This distribution is called the *background* $P(\ln\mathcal{O}|\mathcal{H}_{\rm GR},\kappa_{\rm GR},I)$, where $\kappa_{\rm GR}$ is a large set of injections with GR waveforms, for each of which one computes $\ln\mathcal{O}$. The background distribution determines what value $\ln\mathcal{O}^{\rm modGR}_{\rm GR}$ needs to have in order to have confidence that a violation is present in the data. This value is called the threshold $\ln\mathcal{O}_\beta$ and is defined such that β is the fraction of the background distribution that is above $\ln\mathcal{O}_\beta$:

$$\beta = \int_{\ln\mathcal{O}_\beta}^{\infty} \mathrm{d}\ln\mathcal{O}\, P(\ln\mathcal{O}|\mathcal{H}_{\rm GR},\kappa_{\rm GR},I), \tag{3.2}$$

where β itself is the maximum false alarm probability that one is willing to tolerate. In other words, one is confident of a violation when the measured $\ln\mathcal{O}^{\rm modGR}_{\rm GR} > \ln\mathcal{O}_\beta$.

In the absence of noise one would expect all $\ln\mathcal{O}$ from injected GR sources to be below 0, however, noise distorts the waveforms and can mimic violations of GR causing the distribution to shift toward and into positive values or even create a longer positive tail, pushing the threshold $\ln\mathcal{O}_\beta$ into more positive values as well. Examples of this can be seen in Sec. 1.6.3. The length of the tail is determined by the nature of the detector noise and the length of the signals; it would only take a small glitch in the data to distort a short signal to the point where the odds ratio strongly favors $\mathcal{H}_{\rm modGR}$. Now that we will consider heavier sources, *e.g.* BBH, we expect to see a longer tail in the background distribution, since the signals will be shorter than those of BNS.

To investigate how the background behaves as we move to higher mass systems, we consider a set of 2500 BBH sources with a uniform distribution in component masses $5\,\mathrm{M}_\odot \leq m_a \leq 15\,\mathrm{M}_\odot$ and spins $0.5 \leq |\mathbf{S}_a| \leq 1$. The sources are uniformly distributed in volume between 50 Mpc and 150 Mpc. As we wish to test the performance of the TIGER pipeline, we exclude any of the sources that have a network SNR greater than 30 or less than 8. Because we are particularly interested in how realistic non-Gaussian and non-stationary noise affects the background, we inject these sources into recolored S6 noise, which combines a noise power spectral density that corresponds to that of the early-advanced detectors and true data from the 6th science run S6 (see also Sec. 1.3).

The GR background depends as well on the number of testing parameters that are considered; after all, one is also calculating evidence for the mod-GR hypothesis $\mathcal{H}_{\rm modGR}$, which consists of $2^{N_T}-1$ sub-hypotheses, where N_T is the number of testing parameters. The number of parameter estimation runs that need to be performed scales steeply with the number of testing parameters, so according to computational limitations one needs to decide on a smaller sub-set of $\delta\hat{p}_n$. In this investigation we consider 3 testing parameters: $\delta\hat{\varphi}_3$, $\delta\hat{\xi}_3$ and $\delta\hat{\xi}_4$. When one generates a GR background it should reflect the type of tests of GR one wants to perform. In our

case, we want to test the genuinely strong-field non-linear dynamics and we wish to include as many parameters that represent this regime. The lowest PN order that contains dynamical non-linearities is 1.5PN with the corresponding non-GR parameter $\delta\hat{\varphi}_3$. Furthermore, 1.5PN is the lowest order in which spin-orbit effects enter; it is therefore physically an interesting term to include in the analysis. We need to limit our choices in parameters due to computational constraints and since we are particularly interested in the dynamical strong-field regime, the next two parameters we include are the phenomenological $\delta\hat{\xi}_3$ and $\delta\hat{\xi}_4$; the two that were ranked to have the highest influence on the waveform phasing as shown in Fig. 3.1.

The chirp mass distribution that corresponds to the 2500 injected sources is shown in the left panel of Fig. 3.3. Note that we have injected uniformly in component masses, which results in a non-uniform distribution for chirp masses. Since the length of the signals are expected to have such an important effect on the background, we also plot the distribution of chirp times in the right panel of Fig. 3.3. The chirp time is the time for an inspiral signal to evolve from a lower frequency of $f_{\rm low}$ to the frequency at last stable orbit[2] $f_{\rm LSO}$, assuming the 0PN expression as derived in Eq. (1.90):

$$t_{\rm chirp} \simeq 2.18 \left(\frac{1.21 \mathrm{M}_\odot}{\mathcal{M}_c}\right)^{5/3} \left(\frac{100}{f_{\rm low}}\right)^{8/3} . \tag{3.3}$$

Note the strong dependence on the lower cut-off frequency $f_{\rm low}$. In the S6 BBH search [115], $f_{\rm low}$ was taken to be 40 Hz for LIGO and 30 Hz for Virgo; in our analysis we chose it to be 40 Hz for both.

The signals considered here are much shorter than those used for TIGER-BNS, for which the chirp times range between roughly 15 s and 40 s. For the BBH signals we see that a large quantity of sources have a chirp time of less than a second. To understand that such short signals are more likely to form a tail in the background one needs to realize that glitches in detector noise can have similar time spans. If a gravitational wave signal overlaps largely with a strong burst of energy that is caused by *e.g.* noise from the detector environment, the signal will be distorted in a way that does not correspond to GR and $\mathcal{H}_{\rm modGR}$ will be strongly favored.

The three parameters we consider in generating the background result in a mod-GR hypothesis that consists of $2^3 - 1 = 7$ sub-hypotheses (see *e.g.* Eq. (1.160)). When we plot the resulting distribution we only take into account sources for which $\ln B^{\rm GR}_{\rm noise} \geq 32$, which acts as a detection threshold (see Sec. 1.6.2), i.e. any source with a Bayes factor below that is not considered a confident detection. This leaves us with a background containing 1474 sources plotted in the left panel of Fig. 3.4. The right panel shows the distribution of the chirp masses versus the log odds ratio.

The bulk of the $\ln \mathcal{O}^{\rm modGR}_{\rm GR}$ distribution is similar to what we are used to seeing for TIGER-BNS and is almost Gaussian. However, we notice a much longer tail forming in the distribution as we have seen earlier for the BNS background in Fig. 1.12. Generating enough sources to see the full extent of the tail is costly, but at first glance it appears to go out to $\ln \mathcal{O}^{\rm modGR}_{\rm GR} \sim 25$. There are however sources that

[2]Note that in the frequency domain, waveforms can go a factor of a few times higher than the LSO frequency, but pre-merger/ringdown will take comparatively little time.

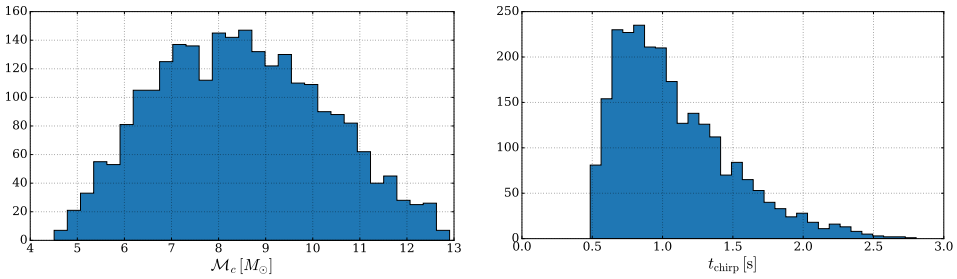


Figure 3.3: Histograms showing the chirp masses $\mathcal{M}_c$ (left) and chirp times $t_{\rm chirp}$ (right) of the 2500 injected BBH sources. Note that the chirp masses and chirp times for BNS injections are roughly $0.87\,\mathrm{M}_\odot \lesssim \mathcal{M}_c \lesssim 1.74\,\mathrm{M}_\odot$ and $15\,\mathrm{sec} \lesssim t_{\rm chirp} \lesssim 40\,\mathrm{sec}$ resp. The sources are uniformly distributed in component masses $5\,\mathrm{M}_\odot \leq m_a \leq 15\,\mathrm{M}_\odot$, resulting in a non-uniform distribution of the corresponding chirp masses and chirp times.

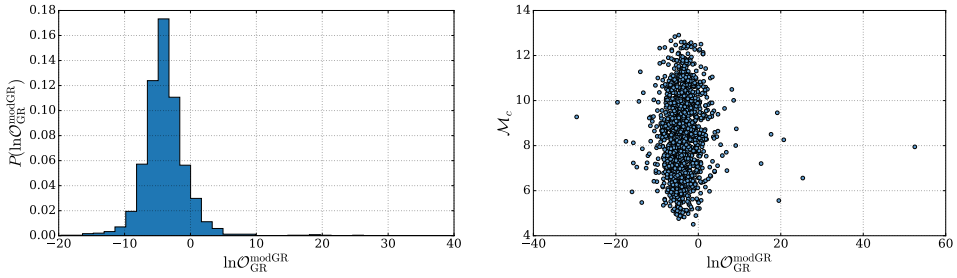


Figure 3.4: The GR background distribution for 1474 single sources (left). The right panel shows how the chirp masses are distributed with respect to $\ln\mathcal{O}^{\rm modGR}_{\rm GR}$. There are a few outliers with $\ln\mathcal{O}^{\rm modGR}_{\rm GR} > 50$ that are not shown here. These outliers are caused by glitches in the data and are listed in Table 3.1.

$\ln \mathcal{O}_{\mathrm{GR}}^{\mathrm{modGR}}$	$\mathcal{M}_c\,[\mathrm{M}_\odot]$	$t_{\mathrm{chirp}}\,[s]$	GPS injection time		
			Hanford	Livingston	Virgo
903.08	7.86	1.11	967152570	970658098	970918054
353.99	10.22	0.72	968938823	968732728	966718227
209.36	7.49	1.20	967876799	970459695	968605267
52.50	7.95	1.09	969182998	971500274	966849271
25.34	6.56	1.50	970804812	969606743	967692706
20.73	8.26	1.02	970674635	967698399	970859860
19.52	5.56	1.98	969196878	970129214	968601343
19.15	9.46	0.82	968906658	967217170	969932691
17.64	8.50	0.97	970814254	970751075	971496782
15.23	7.20	1.28	968149977	967678490	968950380

Table 3.1: List of background events with $\ln \mathcal{O}_{\mathrm{GR}}^{\mathrm{modGR}} > 10$. The first five events show glitches on or near the injection time in one of the detectors. The glitches of the first five entries can be viewed in Fig. 3.6.

exceed even $\ln \mathcal{O}_{\mathrm{GR}}^{\mathrm{modGR}} = 25$. We refer to these values as outliers: single sources that appear to have a disproportionately large log odds. Such outliers are generally caused by glitches in the data that have not been picked up by data quality investigations that veto certain parts of detector output (see Sec. 1.3). Four such outliers with $\ln \mathcal{O}_{\mathrm{GR}}^{\mathrm{modGR}} > 50$ are not included in the plots. Not only values that are obvious outliers are caused by glitches, some sources that appear in the tail of the distribution may also be caused by undiscovered glitches in the data. If that should be the case, such sources may be removed from the background and the tail might get shorter, or even disappear.

Table 3.1 lists the 10 events that have $\ln \mathcal{O}_{\mathrm{GR}}^{\mathrm{modGR}} > 10$. Listed here are the chirp masses, chirp times, and GPS times at which the source was injected in each detector[3]. The top five events are clear outliers, the rest may be a genuine tail that can not be explained by bad data quality.

We can empirically confirm whether or not the outliers and possibly some tail events are caused by glitches, by observing time-frequency scans of the detector data at and around the time of injection. These scans, also called Omega spectrograms are produced with the Omega Pipeline [116, 117] available on the LIGO Data Grid clusters. The scans are 2D density plots, essentially showing normalized energy output in each time-frequency bin within a certain range. Glitches are typically short in duration and widely spread over frequency.

Five events in the tail that have $\ln \mathcal{O}_{\mathrm{GR}}^{\mathrm{modGR}} > 20$ can be discarded due to the presence of a glitch; these are the top five events in Table 3.1. Fig. 3.6 shows the time-frequency scans corresponding to the times at which these events were injected.

[3]The injection times for the different observatories are not the same, since time slides as explained in Sec. 1.6.2 were used.

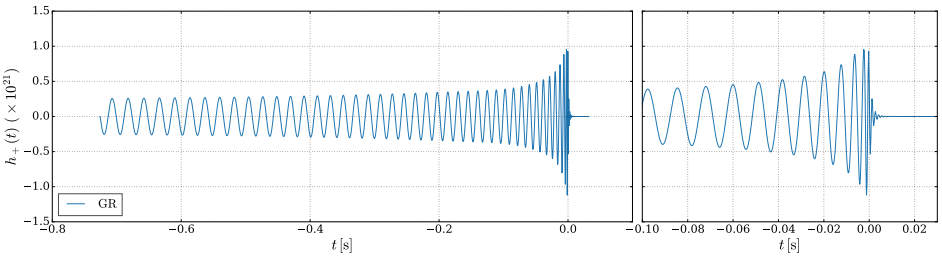


Figure 3.5: An inspiral-merger-ringdown waveform for a source with $\mathcal{M}_c = 8.9\,\mathrm{M}_\odot$ and $\eta = 0.22$ at a distance of 200 Mpc. The waveform is shown from approximately 57 Hz and has a chirp time $t_{\mathrm{chirp}} = 0.9\,\mathrm{s}$ (calculated with $f_{\mathrm{min}} = 40\,\mathrm{Hz}$). The figure demonstrates the amount of time the waveform spends up to the injection time $t = 0$, relative to the amount of time the merger-ringdown takes.

The time $t = 0$ corresponds to the GPS time where the event in question was injected into the corresponding detector data stream. This time also corresponds to $t = 0$ in Fig. 3.5. The top row shows the time-frequency plots around the first (left) and second (right) events in Table 3.1. The second row shows the third (left) and fourth (right) events and the bottom plot corresponds to the injection time of the fifth event. We note that in most cases the time duration of the glitch is of similar size as t_{chirp}. Note also that most of the signal comes before $t = 0$. After $t = 0$, only the merger and ringdown are left, which take only a fraction of the total duration of the signal. This is shown for an example waveform in Fig. 3.5 for a source with $\mathcal{M}_c = 8.9\,\mathrm{M}_\odot$ and $\eta = 0.22$ starting from 40 Hz, where $t = 0$ coincides with the injection time. Still, when we spot an outlier in the background and a glitch occurs near the injection time, we treat the frame as suspect and omit it from the analysis. All top 5 events are either suspect cases or have a clear glitch overlapping with the signal.

As was demonstrated, some events appearing in the tail of the background can be removed as they were either caused by glitches or were injected in suspect regions, however a tail to at least $\ln \mathcal{O}^{\mathrm{modGR}}_{\mathrm{GR}} = 20$ still remains. When the tail still appears to be undesirably large, one could decide to exclude sources based on a detection motivated cut. For example, since the chirp mass $\mathcal{M}_c$ is typically well measured it could be used to define an upper limit to the chirp mass we include in generating a background. Note that this also means that detections with $\mathcal{M}_c$ exceeding that value cannot be used in a TIGER analysis based on the corresponding background. The data set used to build the background is at this time too sparse to make a definitive decision for a cut in $\mathcal{M}_c$. We can however make a suggestion based on the current data: We note that when we omit any sources that have $\mathcal{M}_c > 8\,\mathrm{M}_\odot$, together with the events that were injected in data containing obvious glitches, the tail of the background will not exceed $\ln \mathcal{O}^{\mathrm{modGR}}_{\mathrm{GR}} \sim 20$, which is demonstrated in Fig. 3.7, where the full background and that without events with $\mathcal{M}_c > 8\,\mathrm{M}_\odot$ are shown and can be seen from the chirp mass distribution in the right panel of Fig. 3.4.

The choice of testing parameters can also have an effect on the shape of the background. If it turns out that certain parameters have a stronger effect on the

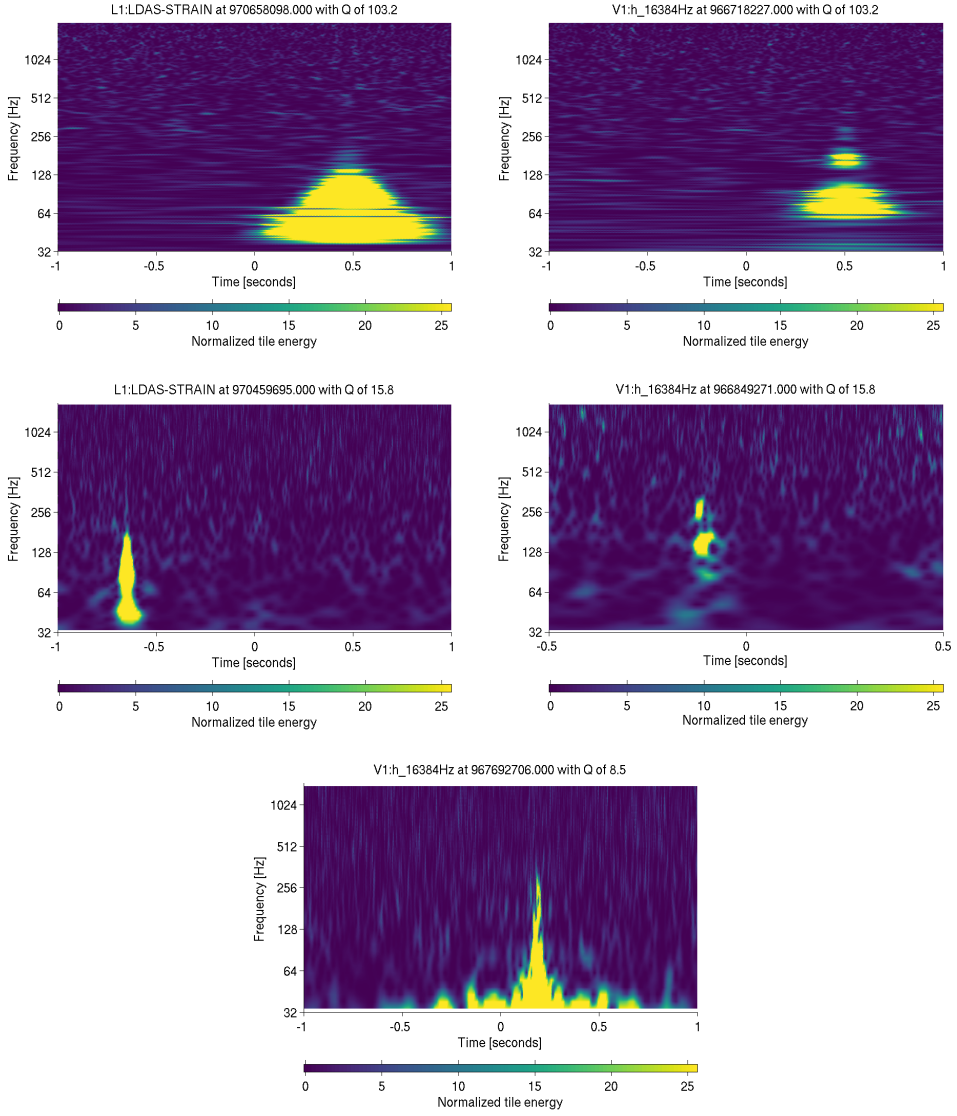


Figure 3.6: Omega scans around the injection times of the top five events in Table 3.1. The top left panel shows a time-frequency scan around the injection time of the first event ($\ln \mathcal{O}_{\rm GR}^{\rm modGR} \sim 903$) in the Livingston detector. The top right shows the scan for the second event ($\ln \mathcal{O}_{\rm GR}^{\rm modGR} \sim 354$) in the Virgo detector. Middle left: third event ($\ln \mathcal{O}_{\rm GR}^{\rm modGR} \sim 209$), Livingston detector. Middle right: Fourth event ($\ln \mathcal{O}_{\rm GR}^{\rm modGR} \sim 53$), Virgo detector. Bottom: Fifth event ($\ln \mathcal{O}_{\rm GR}^{\rm modGR} \sim 25$), Virgo detector.

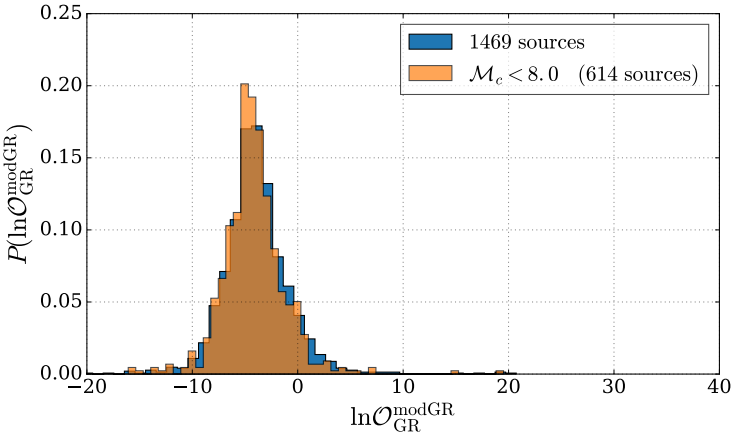


Figure 3.7: The full GR background (blue) and the background where any source with $\mathcal{M}_c > 8\,\mathrm{M}_\odot$ is left out (orange). Note that the injections were done for chirp masses up to $\sim 13\,\mathrm{M}_\odot$, hence the considerable decrease in sample size. The blue background does not contain the top 5 sources from Table 3.1 anymore, so there are 1474-5=1469 sources left.

shape of the background than others, it may be that certain outliers disappear when considering different choices of non-GR parameter sets. For example, in Fig. 3.8 we only calculate $\ln \mathcal{O}^{\mathrm{modGR}}_{\mathrm{GR}}$ for the subsets of parameters $\{\delta\hat{\xi}_3, \delta\hat{\xi}_4\}$, $\{\delta\hat{\varphi}_3, \delta\hat{\xi}_3\}$ and $\{\delta\hat{\varphi}_3, \delta\hat{\xi}_4\}$, i.e. we now have $N_T = 2$ instead of 3 and the number of sub-hypotheses $(2^{N_T} - 1)$ decreases from 7 to 3. For better visibility we only show the tail of the distribution where $\ln \mathcal{O}^{\mathrm{modGR}}_{\mathrm{GR}} > 0$. It seems, at least when considering the particular parameters in this example and assuming only a single detection, that the tail will not disappear no matter what subset of parameters we consider. Again a larger scale investigation might reveal more, but considering we are now using the parameters that seem to have the largest impact on the phasing of the waveform, the effect of omitting certain parameters is not likely to reduce the extent of the tail.

So far we have only generated a background corresponding to single detections. It is also interesting to see what happens if we consider the case where one has multiple detections to combine. As information from multiple sources is gathered, effects from noise become less significant and the tail of the background distribution we see for the single detection scenario is not relevant anymore. To see what effect combining sources might have on the background distribution, we take the 1469 sources of the current sample and separate those into 147 random groups (catalogs) of 10 sources each (9 in the last catalog). Now instead of calculating the log odds for a single source we calculate it for each catalog of 10 sources according to Eq. (1.172). Using these 147 values we generate the background distribution shown in Fig. 3.9. The resulting background is entirely below $\ln \mathcal{O}^{\mathrm{modGR}}_{\mathrm{GR}} = 0$ and there is no longer any evidence of a tail forming. Comparing the log odds of a group of detections happens in precisely

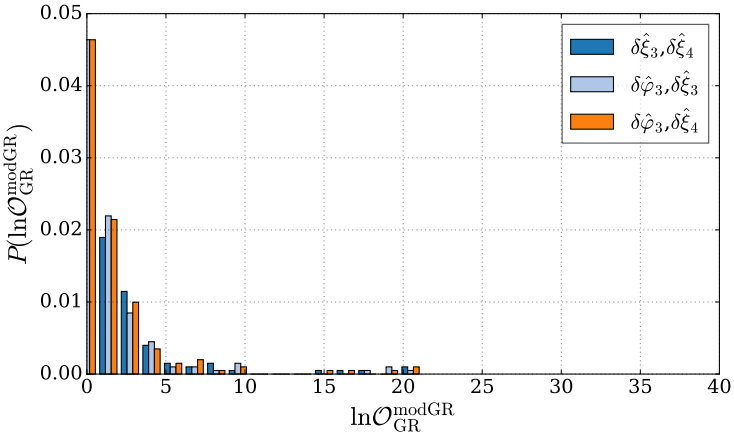


Figure 3.8: The GR background for several choices of parameters. Here $\ln\mathcal{O}_{\rm GR}^{\rm modGR}$ is calculated considering only these 3 subsets of parameters with $N_T = 2$ instead of 3 and the number of sub-hypotheses $(2^{N_T} - 1)$ decreases from 7 to 3. For better visibility we only show the tail of the distribution where $\ln\mathcal{O}_{\rm GR}^{\rm modGR} > 0$.

the same way as for a single detection, only in this case the *combined* odds ratio of the 10 detections must be above the value of $\ln\mathcal{O}_\beta$ determined with the new combined background. In addition to the background becoming more negative, the combined odds ratio of the detections in case of a violation is expected to become more positive (see *e.g.* [66]), making it deviate further from the background as more sources are added. This means that the TIGER pipeline is still effective in the case of multiple detections, even for more massive sources.

3.3 Discussion

In this chapter we presented the first steps taken to include light BBH sources in a TIGER analysis. A non-GR parameterization to the fully precessing inspiral-merger-ringdown waveform model IMRPhenomPv1 was introduced and its non-GR parameter effects were studied to arrive at a choice of parameters used to generate a GR background for BBH sources.

The main difficulty in extending the pipeline's functionality to include BBH sources is in the shape of the GR background. We generated a GR background containing 1474 BBH sources with component masses between 5 and 15 $\mathrm{M}_\odot$. We found that the background has a tendency to extend to larger positive $\ln\mathcal{O}_{\rm GR}^{\rm modGR}$ as opposed to the case where one considers only BNS sources. Outliers and even some events in the tail of the distribution could be discarded due to poor noise quality, however we expect the tail of the log odds distribution to extend to at least $\ln\mathcal{O}_{\rm GR}^{\rm modGR} = 20$, which should become more pronounced when a larger sample of sources is used to

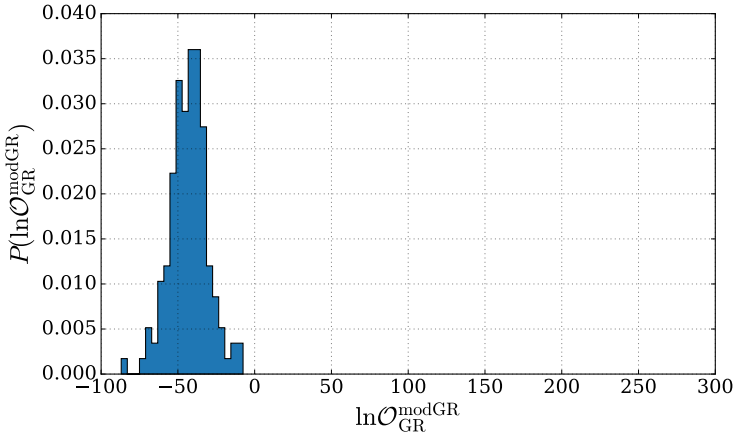


Figure 3.9: The GR background distribution for 147 groups (catalogs) of 10 random sources each (with the exception of the last one that has 9 sources) constructed from 1469 sources. Note that the entire distribution now lies below 0 and there is no evidence for tails forming.

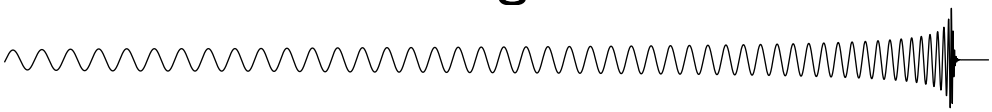
construct the background.

The appearance of a tail in the distribution can make a detection of a violation of GR difficult as the threshold odds ratio $\mathcal{O}_\beta$ that violations will need to overcome will increase. Should one wish to use TIGER on a single BBH detection, it can probably not be done for chirp masses exceeding $8\,\mathrm{M}_\odot$, unless the violation is a prominent one. To investigate how restrictive the tail of the single-source background will be in spotting violations, a foreground study needs to be performed as well, i.e. one will need to inject sources with a certain GR violation (one or more $\delta\hat{p}_i \neq 0$) and see how that distribution of $\ln\mathcal{O}_{\rm GR}^{\rm modGR}$ compares to the background. However as we combine information from multiple detections, we demonstrate that the corresponding combined background distribution will lose its tail and shift more toward negative values of $\ln\mathcal{O}_{\rm GR}^{\rm modGR}$, in fact, $\ln\mathcal{O}_{\rm GR}^{\rm modGR} < 0$ for the entire background when 10 sources are combined. This means that even for BBH sources, TIGER has the potential to efficiently pick up violations from GR as long as we can combine information from multiple detections.

For the second half of the second observing run O2, we intend to employ a new method to speed up likelihood calculations that I will present in Chapter 5. Using this method we do not need to restrict ourselves to a lower cut-off frequency of 40 Hz, but can instead go all the way down to 20 Hz, which in turn means that the chirp times $t_{\rm chirp}$ will become much larger and we become less sensitive to noise. We will need to perform new background studies using the fast likelihood evaluations to ascertain how this improves our ability to test GR using (light) binary black hole sources.

Chapter 4

Tests of general relativity with binary black holes in the first observing run



The first observing run (O1) with the Advanced LIGO detectors took place from September 12, 2015, to January 19, 2016 [11]. Two binary black hole systems were detected during O1. These BBH systems allow us access to a highly dynamical and strong-field regime of general relativity that has thus far remained unexplored by observations. The TIGER framework (Sec. 1.6) was developed specifically for testing the genuinely strong-field regime of relativity using gravitational wave observations. The pipeline provides us with a powerful generic test of GR that allows evidence for GR violations to build up when combining odds-ratios, even if there is only marginal evidence for individual sources.

Before the start of O1, the TIGER infrastructure had already been tested and reviewed for binary neutron star systems. However, the TIGER-BBH variant that extends the pipeline to be used on more massive systems was not reviewed by the time O1 began. Since we can therefore not employ TIGER-BBH for tests of GR with O1, we will instead use part of its functionality that has been reviewed for O1. This less extensive but nevertheless very effective test we refer to as the *parameterized test of GR*, comes in the form of parameter estimation and combination of posteriors from multiple sources.

For the parameterized test we start from the phenomenological waveform model IMRPhenomPv2 which is capable of describing inspiral, merger and ringdown, and partly accounts for precession (Sec. 1.4.2-D). The phase of this waveform model is described by phasing coefficients $\{p_i\}$, which include PN coefficients as well as phenomenological coefficients describing the merger and ringdown. All phasing parameters multiply specific functions of f as listed in Table 1.1 and are either well-determined functions of source parameters (PN coefficients) or determined by calibration against numerical waveforms (the phenomenological parameters describing merger and ringdown). In this waveform model we allow for possible departures from GR, parameterized by a set of testing coefficients $\delta\hat{p}_i$, which take the form of *fractional* deviations in the p_i [66, 68]. Thus, we replace $p_i \to (1 + \delta\hat{p}_i)p_i$ and let one ore more of the $\delta\hat{p}_i$ vary freely in addition to the source parameters that also appear in pure general relativity waveforms, using the GR expressions in terms of masses and spins for the p_i themselves.

The parameterized test allows us to still find even small GR violations, on condition that for a given source, parameter deviations occur preferentially in one direction so that the GR violation will not average out when combining information from multiple sources, *i.e.* $\delta\hat{p}_i > 0$ or $\delta\hat{p}_i < 0$ for each i and all sources. That this is a requirement for combining posteriors is demonstrated in Fig. 4.1: In the top panel

the combined posterior displays a significant deviation from GR in the positive direction as is also seen in the individual ones, whereas in the bottom panel the combined posterior suggests no violation was present, while the individual posteriors do show prominent shifts in opposite directions.

In the parameterized test one can choose to let any of $\delta\hat{p}_i$ vary freely along with the source parameters, but the more testing parameters are allowed to vary simultaneously, the less informative the posteriors will be. For both events we perform the single-parameter variant of the parameterized test, *i.e.* we let each one of the $\delta\hat{p}_i$ in turn vary freely while all others are fixed to their general relativity values, $\delta\hat{p}_j = 0$ for $j \neq i$. These tests model general relativity violations that would occur predominantly at a particular PN order (or in the case of the phenomenological parameters, a specific function of frequency in the relevant regime), although together they can capture deviations that are measurably present at more than one order, which we demonstrate using non-GR injections in section 4.1.2-B.

Moreover, as demonstrated in [118] and specific to the task at hand in Sec. 4.1.2-B, checking for a deviation from zero in a single testing parameter is an efficient way to uncover GR violations that occur at multiple PN orders. Furthermore, one can even find violations at powers of frequency that are distinct from the one the testing parameter is associated with ([66, 67] and Sec. 4.1.2-B). Hence such analyses are well suited to search for generic departures from GR, though it should be stressed that if a violation is present, the measured values of the $\delta\hat{p}_i$ will not necessarily reflect the predicted values of the correct alternative theory. To reliably constrain theory-specific quantities such as coupling constants or extra charges, one should directly apply full inspiral-merger-ringdown waveform models from specific modified gravity theories [1], but in most cases these are not yet available. However, in the present work the focus is on model-independent tests of general relativity itself.

The parameterized test will be explained in more detail in Sec. 4.1. Sec. 4.1.1 explains how the waveform model IMRPhenomPv2 is extended with additional testing parameters to be used in our framework to test GR. Sec. 4.1.2 demonstrates the impact certain deviations in various non-GR parameters have in the waveform dynamics as well as on parameter estimation results. In Sec. 4.1.3 we investigate the effect detector noise has on non-GR parameter posteriors.

With the final reviewed parameterized test using the extended IMRPhenomPv2+ we perform the first ever tests of the dynamical strong field regime of general relativity and are able to place constraints on possible deviations from the GR-predicted waveform coefficients. Section 4.2 presents our findings using the two detections made during O1. This chapter is closed in section 4.3 by discussing the implications of the results and providing an outlook on what may be required and achieved in O2 and beyond.

4.1 Parameterized test of GR

At the time of the first detections, the full TIGER framework (hypothesis testing in addition to parameter estimation) was only fully tested and reviewed for BNS sources. What we decided on doing instead is what we now call the *parameterized test of GR*.

The idea follows the TIGER framework closely; the waveforms allow for parameterized deformations with respect to GR in a model independent generic fashion. The difference is that we do not construct a background to compare evidence of a violation of GR, but instead perform parameter estimation on each of the non-GR parameters which are identically zero in case GR is correct. Similar to Refs. [66, 68], we only consider deviations from GR in the GW phase, while we leave the GW amplitude unperturbed. Indeed, at the SNR of GW150914 (i.e., SNR ~ 24) and especially that of GW151226 (SNR ~ 13), we expect to extract much more information from the GW phase rather than its amplitude, since the former does not involve poorly measurable extrinsic parameters such as sky position and orientation of the binary.

In the parameterized test, the posteriors on non-GR parameters provide a quantitative constraint on possible deviations from GR in that particular part of the waveform model. We have already touched upon this method in one of the TIGER publications [66] and the no-hair test presented in Chapter 2. The non-GR parameters $\delta\hat{p}_i$ are relative shifts with respect to the corresponding GR phasing coefficients p_i, i.e.

$$p_i^{\rm GR}(m_1, m_2, \mathbf{S}_1, \mathbf{S}_2) \to (1 + \delta\hat{p}_i) p_i^{\rm GR}(m_1, m_2, \mathbf{S}_1, \mathbf{S}_2), \tag{4.1}$$

where m_1, m_2, $\mathbf{S}_1$ and $\mathbf{S}_2$ are component masses and spins respectively.

In this parameterization, GR is uniquely defined as the locus in the parameter space where each of the non-GR parameters $\delta\hat{p}_i$ is zero. One can perform parameter estimation on any number of non-GR parameters simultaneously, but resulting posteriors will become less informative as more non-GR parameters are allowed to vary at the same time (We explored this in [73]). For each choice of testing parameters that are allowed to vary freely, we perform a separate LALInference analysis, where in concert with the full set of GR parameters [119] we also explore the posterior distributions for the specified set of testing parameters. In doing so we can verify that the posterior probability distributions for the coefficients include their GR values.

The choice of non-GR parameters depends on the waveform model that is used. For BBH we can typically not use any of the PN waveform families in the parameterized test for two reasons: 1) the strong-field dynamical regime of GR is at its richest in the late inspiral and merger-ringdown regimes, 2) the heavier the system, the lower the merger frequency, meaning that if one is limited to the PN formalism, only a limited number of cycles can be analyzed. For the parameterized tests on BBH we use IMRPhenomPv2 (Sec. 1.4.2-D), which is in good agreement with numerical waveforms for most of the BBH parameter space, is fast to evaluate, and extensible to include a set of non-GR parameters as explained in Sec. 4.1.1.

A useful feature of the parameterized test of GR is that it can be trivially extended to combine information from multiple sources. We have already demonstrated this concept in Chapter 2: Given posteriors $p(\delta\hat{p}_i|d_A, I)$ for given data streams d_A, prior information I and detections $A = 1, 2, \ldots, \mathcal{N}$ we can construct a combined posterior by way of Eq. (2.32), provided $\delta\hat{p}_i$ has the same sign for all detections d_A. For completeness, the equation for combining posteriors within the current context is

$$p(\delta\hat{p}_i|d_1, d_2, \ldots, d_{\mathcal{N}}, I) = p(\delta\hat{p}_i|I)^{1-\mathcal{N}} \prod_{A=1}^{\mathcal{N}} p(\delta\hat{p}_i|d_A, I). \tag{4.2}$$

To better understand the impact of combining different kinds of posteriors, Fig. 4.1 visualizes two different cases. The top panel demonstrates that when we combine posteriors from two sources in which the deviation in $\delta\hat{p}_i$ occurs in the positive direction, we arrive at a stronger result. It could however be that the violation depends on the source parameters, resulting in a very different shift in each source. The bottom panel shows what would happen in such a case if we combine the posteriors. Here the deviation from GR happens in the positive direction for the first detection, but in the negative direction for the second. This results in a combined posterior that suggests there is no violation of GR present in the data, which of course is incorrect as the individual posteriors show.

The non-GR parameters used in Eq. (4.1) are not source-dependent in the case GR is correct (since then $\delta\hat{p}_i = 0\,\forall\, i$). In that case one is indeed allowed to combine the posteriors across multiple detections and an increasingly narrow prior centered around zero is constructed. In that case one can construct upper bounds on possible violations on each of the $\delta\hat{p}_i$ which will improve roughly with the square root of the number of detections. However, in case a violation of GR is present, there is no a-priori reason to assume $\delta\hat{p}_i$ do not depend on the source parameters. In this case, combining posteriors could result in broad uninformative distributions or even one that suggests a false agreement with GR as shown in the bottom panel of Fig. 4.1. The far more general method to test for violations of GR is the TIGER pipeline in which we are not dependent on the nature of the GR violation and we do not combine posteriors of individual $\delta\hat{p}_i$, instead we combine evidences either supporting or rejecting consistency with general relativity.

4.1.1 IMRPhenomPv2+

The precessing inspiral-merger-ringdown phenomenological waveform IMRPhenomPv2 was introduced in Sec. 1.4.2-D. The waveform can be separated into three dynamical stages that characterize the coalescence process, which are depicted in Fig. 4.2. Recall that we refer to the *early-inspiral stage* as the PN part of the GW phase and is parameterized in terms of the PN coefficients φ_j, where $j = 0, \ldots, 7$ and the logarithmic PN terms $\varphi_{5l}, \varphi_{6l}$. This stage of the phase evolution is known *analytically* up to $(v/c)^7$. The *late-inspiral stage*, parameterized in terms of σ_j, where $j = 1, \ldots, 4$, is defined as the phenomenological extension of the PN series to $(v/c)^{11}$. The *intermediate stage* that models the transition between the inspiral and merger-ringdown phase is parameterized in terms of the phenomenological coefficients β_j, where $j = 1, 2, 3$. Finally, the *merger-ringdown* phase is parameterized in terms of the phenomenological coefficients α_j, where $j = 1, 2, 3, 4$.

The non-GR version of IMRPhenomPv2 – which will be referred to as IMRPhenomPv2+ – is parameterized by a set of testing coefficients $\delta\hat{p}_i$, which take the form of *fractional* deviations in the phasing parameters p_i listed above as defined in Eq. (4.1). During parameter estimation we let one or more of the $\delta\hat{p}_i$ vary freely, in addition to the source parameters that also appear in pure general relativity waveforms, using the general relativistic expressions for p_i in terms of masses and spins. In principle the non-GR testing parameters $\delta\hat{p}_i$ can consist of the full set where we have one $\delta\hat{p}_i$ for each p_i that appears in the model. However, due to the fitting proce-

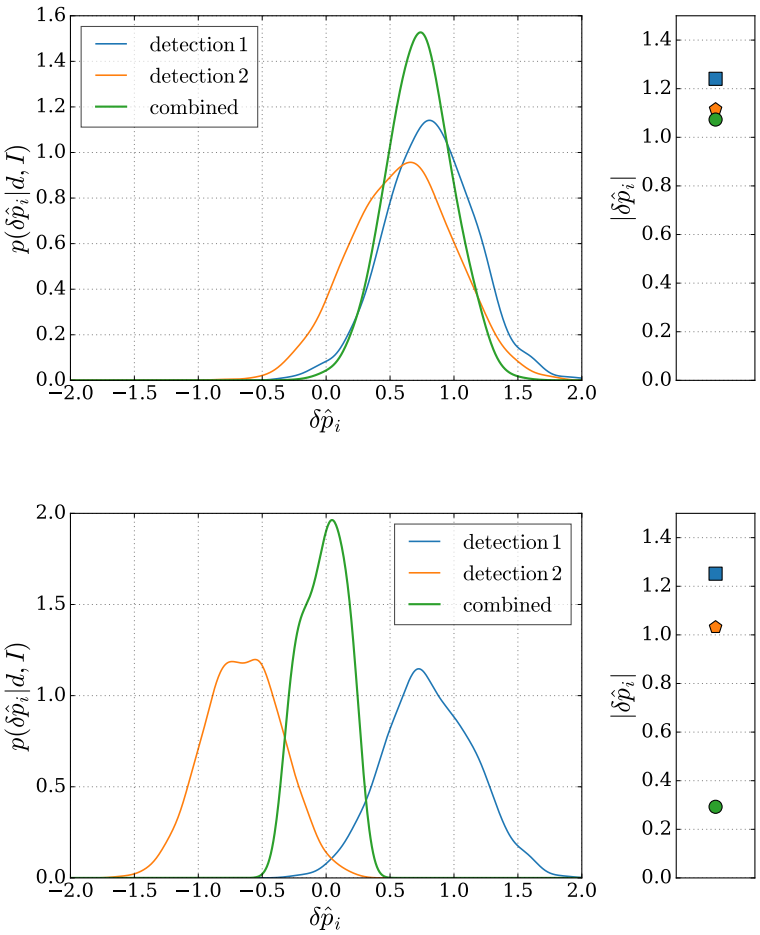


Figure 4.1: A schematic representation of the effect on the final combined posterior from two distinct situations. The figures demonstrate why it is only informative to combine posteriors when a deviation from GR in one particular parameter acts in either the positive or negative direction for all detections. Also shown are the 90% credible intervals of $|\delta\hat{p}_i|$ (detection 1 as a square, detection 2 as a diamond and the combined as a circle). The top panel demonstrates a scenario where a violation of GR is present in p_i and results in a positive shift in $\delta\hat{p}_i$ for both detections. The bottom panel demonstrates what would happen with similar measurements if the violation were to result in $\delta\hat{p}_i$ having a shift in opposite directions. Clearly, in the last case combining information from sources is not the right course of action as it results in a posterior suggesting there is no violation, whereas the individual posteriors show strong disagreement.

dure through which the model is constructed, there is an intrinsic uncertainty in the values of the phenomenological parameters of IMRPhenomPv2. The parameters we decided to use on O1 detections (see Sec. 4.2) were determined by the size of the uncertainties arising from calibration against numerical waveforms with respect to the corresponding statistical uncertainties for GW150914. For the merger-ringdown regime we verified these intrinsic uncertainties to be sufficiently small. In the late-inspiral case, the uncertainties associated with the calibration of the σ_j parameters are large and almost comparable with results from GW150914. Therefore, the results for the σ_j parameters are not included in Sec. 4.2. I will however consider this set of parameters as part of the IMRPhenomPv2+ waveform for future reference. Parameters that are degenerate with the reference time t_c or the reference phase ϕ_c, *i.e.* coefficients that appear in phasing terms that evolve with f or that are constant in frequency respectively, will not be considered either. Finally, we omit one additional phenomenological parameter $\delta\hat{\alpha}_5$, since α_5 appears in the same term as α_4 for which we do include a relative shift. Taking this into account, the full set of testing coefficients that make up IMRPhenomPv2+ are then

early-inspiral $\{\delta\hat{\varphi}_0, \ldots, \delta\hat{\varphi}_7\}$ and $\{\delta\hat{\varphi}_{5l}, \delta\hat{\varphi}_{6l}\}$

late-inspiral $\{\delta\hat{\sigma}_2, \delta\hat{\sigma}_3, \delta\hat{\sigma}_4\}$

intermediate $\{\delta\hat{\beta}_2, \delta\hat{\beta}_3\}$

merger-ringdown $\{\delta\hat{\alpha}_2, \delta\hat{\alpha}_3, \delta\hat{\alpha}_4\}$

Note that $\{\delta\hat{\varphi}_i\}$ includes a 0.5PN testing parameter $\delta\hat{\varphi}_1$; since φ_1 is identically zero in general relativity, we let $\delta\hat{\varphi}_1$ be an absolute rather than a relative deviation. The functional forms of the PN parameters are provided in Sec. 1.4.2-B, and the frequency-dependence of each term both PN and phenomenological are listed in Table 1.1.

4.1.2 Significance of GR violations in IMRPhenomPv2+

In the previous section a set of possible fractional changes $\delta\hat{p}_i$ with respect to the GR phasing functions p_i was introduced. The goal of this section is to investigate the impact a fractional change has on the phase evolution of the waveform and parameter estimation. More insight into the effects of the non-GR parameters may aid us in determining what a reasonable prior on $\delta\hat{p}_i$ may be and also which parameters are most likely to be measured well with a certain event. In Sec. 4.1.2-A we investigate several effects non-GR parameters have on the dynamics of the waveform by itself. Specific types of events are considered in which we monitor what kind of phase difference – with respect to GR – as a function of frequency $\Delta\Psi(f) = \Psi_{\text{non-GR}}(f) - \Psi_{\text{GR}}(f)$ appear in the waveform when giving one of $\delta\hat{p}_i$ a non-zero value. We also consider the impact of parameter deviations by plotting the accumulated phase difference at a certain test-frequency. This way we can study what type of effects are expected in the most sensitive part of the detectors' frequency bandwidth, corresponding to a frequency of around 150 Hz (Sec. 1.3). Furthermore, the latter allows us to monitor the impact as a function of one of the source parameters, rather than considering the

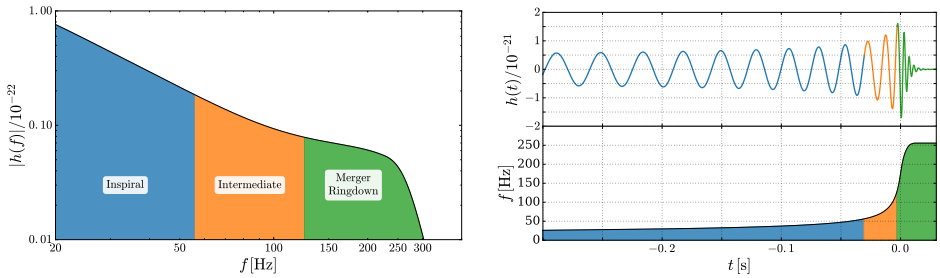


Figure 4.2: Frequency regions of the parameterized waveform model as based on the definitions in Sec. 1.4.2-D and [32]. **Left:** The absolute value of the most-probable waveform from GW150914 [119] in the frequency domain. The inspiral region (blue) from 20 Hz to ∼56 Hz corresponds to the early and late inspiral regimes. The intermediate region (orange) goes from ∼56 Hz to ∼126 Hz. Finally, the merger-ringdown (MR) region (green) goes from ∼126 Hz to the end of the waveform. As mentioned in Sec. 1.4.2-D, there is some ambiguity in defining these regions. Since we only consider the phasing parameters, only the boundaries as used in constructing the full IMR phase are indicated. **Right:** The same regions as in the left panel but now highlighted in the time-domain waveform (top). The bottom panel of the right figure shows the relation between time and frequency.

entire frequency evolution for one particular source. This gives us insight into what different types of detections might contribute to the parameterized test. In particular, we will show the effects as function of the chirp mass.

We are most sensitive to the GW phasing, but in some cases a shift in one of the $\{\delta\hat{p}_i\}$ can have a significant effect on the shape of the waveform as a whole, which demands inspection of the time-domain waveform $h(t)$. We evaluate $h(t)$ for a few particular source parameters and shifts in $\delta\hat{p}_i$ in order to verify that the size of non-GR effects on the waveforms do not break the time-domain evolution, *e.g.* we should always see an inspiral-merger-ringdown behavior.

A more in-depth investigation on the effects of particular GR violations is presented in Sec. 4.1.2-B in the way of a full parameter estimation exercise. Based on the detections currently made with Advanced LIGO and the parameter estimation results obtained from these, we can see how posteriors might change if a non-GR effect were to be present a few standard deviations away from the current measurements. Sec. 4.1.2-B addresses questions such as whether we can detect violations appearing in parameters that we do not include in parameter estimation and what would happen if some violation would influence more than one non-GR parameter.

4.1.2-A Effects on waveform dynamics

We first consider the general effect any of the non-GR parameters may have on the phase evolution as a function of frequency $\Psi(f)$. We do so by comparing the phase calculated with IMRPhenomPv2 and that of IMRPhenomPv2+: $\Delta\Psi(f) =$

$\Psi_{\text{non-GR}}(f) - \Psi_{\text{GR}}(f)$. This phase difference allows one to gauge the detectability of any violation when employing the TIGER pipeline or the parameterized test. Although we can not relate any particular phase shift directly to a quantitative statement of how much it may affect resulting posteriors or hypothesis testing – after all, detector noise can have have a large effect on measurements as will be discussed in Sec. 4.1.3 –, we can consider a rule of thumb that was introduced in [67]. It was found that for BNS systems, a violation causing an accumulated phase difference at 150 Hz of around five radians would (almost) certainly be detectable after combining 15 sources by using the TIGER pipeline. These five radians correspond to about one full GW cycle. A phase difference with respect to GR of a full cycle at the most sensitive region of the detectors' bandwidth is expected to be detectable. In the tests below, these five radians may be used as a guideline of detectability. However, note that with BBH sources we have three different regions (inspiral, intermediate and merger-ringdown) of which only one will fall within the detectors' most sensitive frequency window, and thus we expect only that region to contribute the most to parameter estimation.

In the following, two sources are considered of which the parameters are based on GW150914 and GW151226. For each of the non-GR parameters, the evolution of the phase difference $\Delta\Psi(f)$ as well as the phase $\Psi(f)$ itself are presented in Fig. 4.3 for a range in fractional shift sizes (color scale). The reference phase ϕ_c is defined such that $\Psi(f)$ is zero at $f = 20$ Hz. The waveforms were evaluated using the `lalsim-inspiral` module from LALSUITE with a lower frequency of 15 Hz. The left column in Fig. 4.3 corresponds to a source consistent with GW150914 while on the right we show results of a source consistent with GW151226. At the time this analysis was performed, there were no posteriors yet; however, to aid in interpreting the results presented in Sec. 4.2, the choice of parameter ranges is such that they cover at least the full extent of the posteriors in Fig. 4.13.

Three non-GR parameters are considered in Fig. 4.3, namely, $\delta\hat{\varphi}_3$ in the top row demonstrating the effect of one of the PN early-inspiral parameters, $\delta\hat{\beta}_2$ in the middle row for the intermediate stage and $\delta\hat{\alpha}_2$ in the bottom row showing effects in the merger-ringdown regime. These three parameters are a reasonable representation of the general behavior of the non-GR parameters in each respective region, the only exception being $\delta\hat{\alpha}_4$ of which the phase difference keeps increasing until the end of the waveform as shown in Fig. 4.4. Note that the inspiral parameter $\delta\hat{\varphi}_3$, accumulates a much larger phase difference for GW151226 as opposed to GW150914 by the time 150 Hz is reached. This effect is observed for most of the early-inspiral parameters and helps us understand why these parameters are measured much more accurately in GW151226, for which the inspiral part is much longer than GW150914.

The merger-ringdown parameter $\delta\hat{\alpha}_2$ follows a different trend. The accumulated phase difference increases, but reaches a maximum at some point after which it decreases again until the evolution stops. This maximum phase difference occurs in the most sensitive part of the detectors' bandwidth in case of a source such as GW150914, whereas for GW151226 this happens at much higher frequencies. Once more this explains why the posteriors on the ringdown parameters are not very informative for the second detection.

The intermediate stage parameter $\delta\hat{\beta}_2$ exhibits a relatively slowly increasing phase

accumulation and levels out between 100 and 200 Hz for both events, which is exactly where the detector is most sensitive. Indeed for both events, this parameter was measured quite accurately.

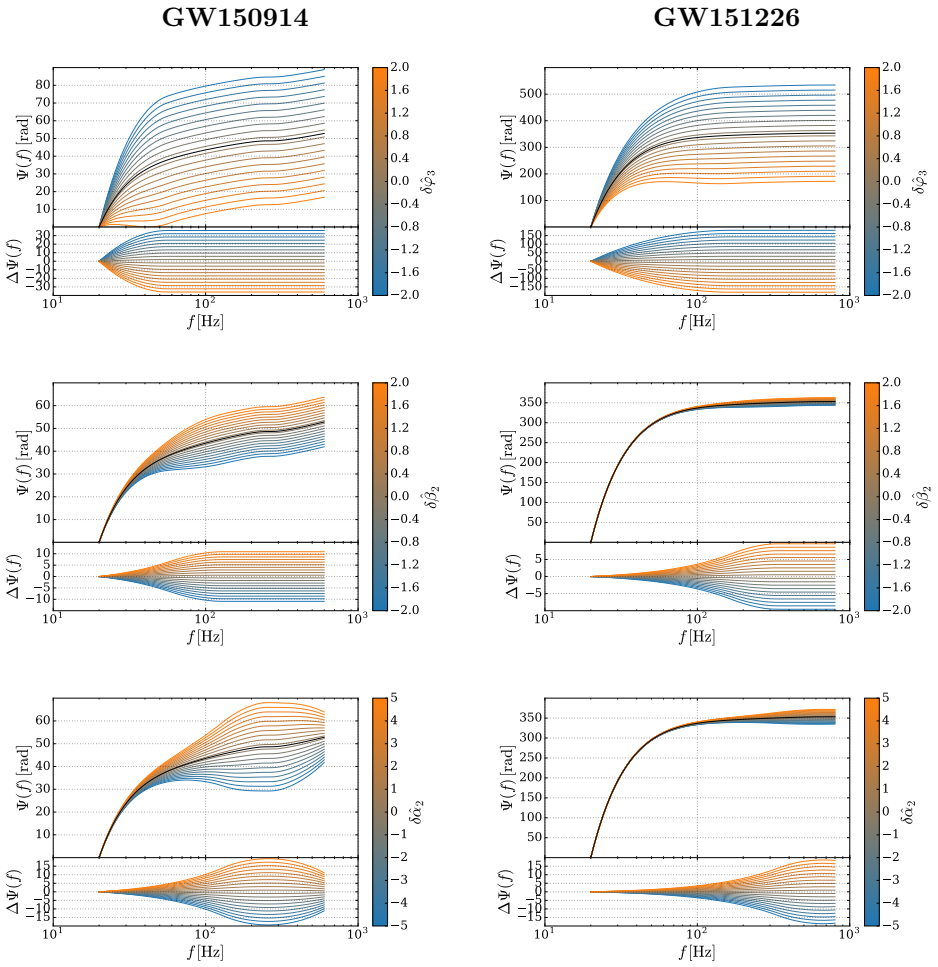


Figure 4.3: Phase evolution $\Psi(f)$ and difference $\Delta\Psi(f) = \Psi_{\text{non-GR}}(f) - \Psi_{\text{GR}}(f)$ as a function of frequency. The left column corresponds to source parameters that are consistent with GW150914, while the right column represents a source that is consistent with GW151226. Each curve corresponds to a particular shift of the non-GR parameter presented in the plot, except for the black curve, which represents the GR case *i.e.* $\delta\hat{p}_i = 0$. The non-GR parameters considered are $\delta\hat{\varphi}_3$, $\delta\hat{\beta}_2$ and $\delta\hat{\alpha}_2$ in the top, middle and bottom rows respectively.

The phase and amplitude of IMRPhenomPv2 are by construction guaranteed to be smooth functions of f at the transition points between the various regions. However, as we introduce physically unmotivated deviations in the phasing coefficients,

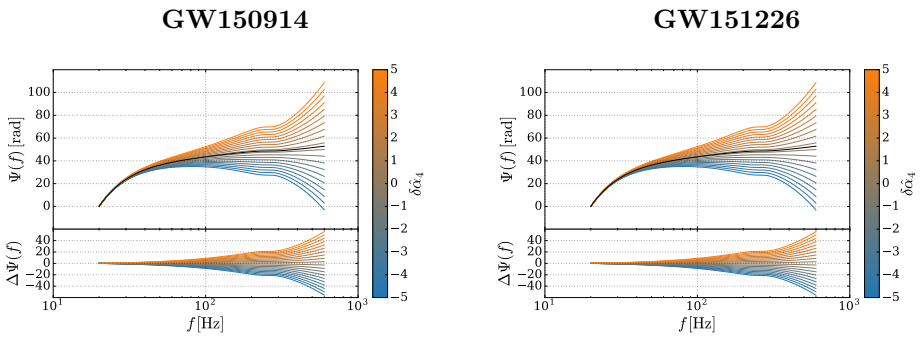


Figure 4.4: Same as in Fig. 4.3, but now only considering $\delta\hat{\alpha}_4$, which shows different behavior from $\delta\hat{\alpha}_2$ and $\delta\hat{\alpha}_3$. As opposed to these, the phase difference for $\delta\hat{\alpha}_4$ keeps increasing until the waveform terminates.

the smooth connections between the various regions may result in strange behavior in the time-domain waveform. We expect that the effects will not be extreme, since the different parts of the waveform are matched by adjusting essentially only the reference phases and times, and the PN coefficients and phenomenological ones encoding the dynamics of the waveform are left as they are, *i.e.* φ_i, β_i and α_i where $i \neq \{0, 1\}$ are left to their original values.

Here, we demonstrate some of the effects in the time-domain waveform $h(t)$ after introducing non-GR parameters into IMRPhenomPv2. We consider even reasonably large fractional shifts in some of the parameters that are already excluded by observations to get a feel for the extremes.

This analysis was performed on only one type of source resembling GW150914 and considering only a few parameters, mainly serving as a sanity check on the time-domain evolution. A large scale study to investigate this for a broad range in source parameters and non-GR parameters is left for future work. We do want to note here that although we are interested in seeing what happens in the time-domain, both TIGER and the parameterized test are most sensitive to the phase evolution rather than the amplitude. Fig. 4.5 shows $h_+(t)$ for three non-GR parameters[1], one from each region of the waveform model: $\delta\hat{\varphi}_3, \delta\hat{\beta}_2$ and $\delta\hat{\alpha}_2$. Plotted are the GR waveform (represented by the solid black curve) and two shift sizes for each parameter. The relative shift sizes are $\delta\hat{\varphi}_3 = \{0.1, 2.0\}$, $\delta\hat{\beta}_2 = \{1.0, 2.0\}$ and $\delta\hat{\alpha}_2 = \{1.0, 5.0\}$. We note that even for relative shift sizes that fall outside the upper bounds presented in Sec. 4.2, the waveforms still look reasonable. The largest impact occurs when $\delta\hat{\varphi}_3$ is increased to well outside the 90% credible interval from Fig. 4.13. The resulting waveform shows an unusual ringdown signal, but still features the usual inspiral-merger-ringdown progression at the end of which it decays to zero amplitude.

For the final part of the investigation into the waveform dynamics, the phase difference $\Delta\Psi(f)$ is once again considered, but in this instance we are interested in

[1]Plotting only the + polarization is sufficient, since the difference between + and × polarization is only a constant phase shift.

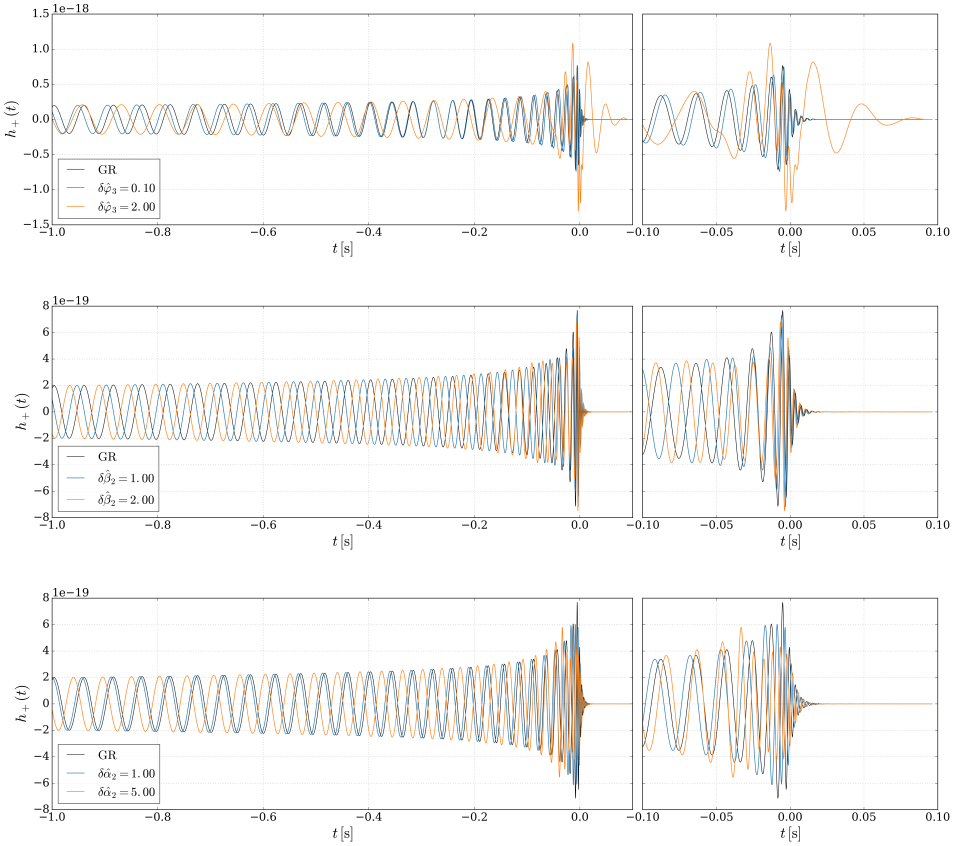


Figure 4.5: The effect of deviations from GR in the three regions of the IMRPhenomPv2 waveform on the time-domain +-polarization $h_+(t)$. **Top:** $\delta\hat{\varphi}_3 = 0.1$ in blue and $\delta\hat{\varphi}_3 = 2.0$ in orange. Note that even at this massive violation compared to the 90% credible intervals shown in Fig. 4.13, the waveform still looks smooth. **Middle:** $\delta\hat{\beta}_2 = 1.0$ in blue and $\delta\hat{\beta}_2 = 2.0$ in orange. **Bottom:** $\delta\hat{\alpha}_2 = 1.0$ in blue and $\delta\hat{\alpha}_2 = 5.0$ in orange. The GR waveform is shown for comparison in each case as a solid black curve. The right panels show the same waveforms, but zoomed in to the final tenths of a second.

what kind of effect we see in only the most sensitive part of the detectors' bandwidth, *i.e.* $f \sim 150\,\mathrm{Hz}$. In doing so the effect as a function of the source's chirp mass $\mathcal{M}_c$ can be monitored. The total mass and mass ratio of the source have a large influence on the time it spends within the detectors' sensitive band. We consider a slightly non-symmetric system ($\eta = 0.2$) with spin-magnitudes $|S_i| = 0.9$ turned on and off. The chirp mass is varied between $2\,\mathrm{M}_\odot$ and $40\,\mathrm{M}_\odot$, which with $\eta = 0.2$ corresponds to component masses in the range $1.45\,\mathrm{M}_\odot$ to $76.02\,\mathrm{M}_\odot$.

Fig. 4.6 shows the phase differences at 150 Hz for three sets of non-GR parameters as function of $\mathcal{M}_c$. As in Fig. 4.3, here too the waveforms were evaluated starting from 15 Hz and phases were set to zero at 20 Hz. The plots shown in Fig. 4.6 consider two spin cases, one in which the spins are zero (solid curves) and the other case in which the spins are set to be very large at 0.9 (dashed curves). The difference can not be observed for the inspiral PN parameters, but is considerable for the intermediate parameters $\delta\hat{\beta}_i$ and visible in $\delta\hat{\alpha}_i$. The values of the constant non-GR shifts in each parameter are based on the 90% credible upper limits as set by the posteriors from GW150914. These choices allow one to observe the effect of realistic shifts in each individual parameter.

We note that as expected, phase shifts increase drastically for PN parameters with decreasing mass due to the number of GW cycles that are in band up to 150 Hz. For the intermediate and merger-ringdown parameters the opposite happens as these phase shifts become increasingly more pronounced as the mass of the system increases and the merger part shifts more towards 150 Hz. We can thus see that for light systems, shifts in the PN coefficients are more pronounced, for intermediate systems the $\delta\hat{\beta}_i$ have a greater measurability and for the heaviest systems the most relevant parameters are $\delta\hat{\alpha}_i$.

4.1.2-B Effects on parameter estimation

The parameterized test as presented in [73] considers two types of parameter estimation efforts: (i) We allow only one of the whole set of non-GR parameters to vary at any one time and repeat this for all the parameters. (ii) We split the parameters into three "batches", in which case we allow all parameters in one batch to vary while the parameters in the other batches are set to zero. In the O1 BBH paper [11] and Sec. 4.2, we only perform the first test because the latter is unable to place reasonable constraints on any of the non-GR parameters. Moreover, when a GR violation occurs at some part in the waveform phasing, *e.g.* $\delta\hat{\varphi}_3$, but we recover using any of the other $\delta\hat{p}_i$, where $i \neq 3$, it is expected that these will also deviate from zero. The reason is that during parameter estimation, the discrepancy at the 1.5PN phasing term must somehow be accommodated by other parameters. Depending on the strength of the violation, this effect will present itself as an offset in some, if not all of the non-GR parameters we test with. This section is intended to demonstrate this effect quantitatively by performing full parameter estimation runs on non-GR waveforms using LALInference as would be done in a real detection scenario.

To simulate realistic non-Gaussian noise effects that are also present in O1 detections, we use the so-called recolored S6 noise curves (see also Sec. 1.3). The recolored noise combines a noise power spectral density curve that corresponds to that of the

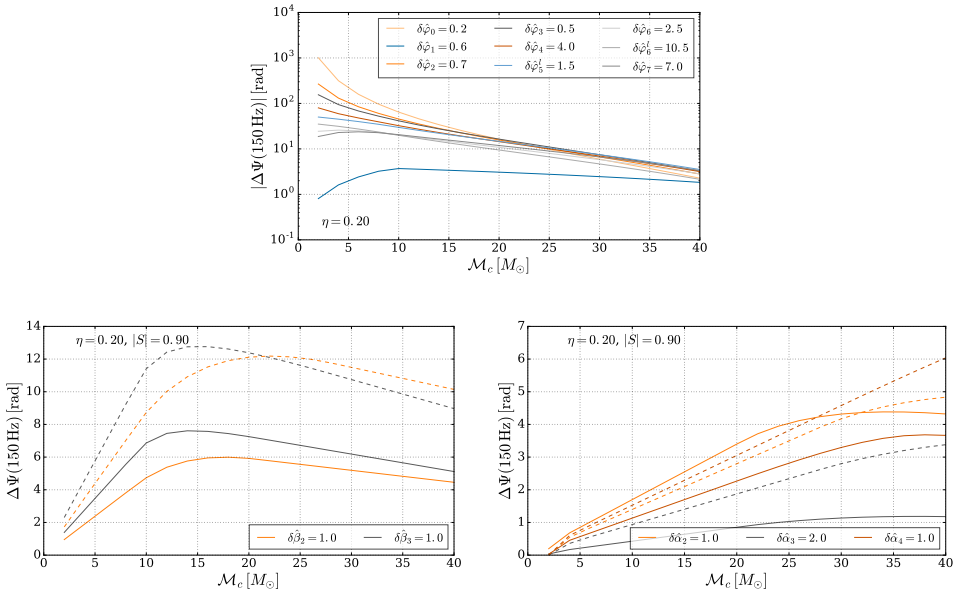


Figure 4.6: Phase difference $\Delta\Psi(150\,\mathrm{Hz})$ for three sets of non-GR parameters. Spins are turned on with $|S| = 0.9$ in the dashed curves and off in the solid curves. For the inspiral PN parameters only the curves without spin are shown as they are indistinguishable from the curves that include spin. We can conclude that at 150 Hz, spins do not have a significant effect on the inspiral PN parameters at any mass. Parameters considered are those of the early-inspiral (top), intermediate (bottom left) and merger-ringdown (bottom right) stages. Note that the top panel shows absolute values to better fit the wide range in $\Delta\Psi(150\,\mathrm{Hz})$ on a logarithmic scale.

non-GR parameter	mean	stdev	injection ($\sim 5 \times$ stdev)
$\delta\hat{\varphi}_3$	0.075	0.086	± 0.4
$\delta\hat{\varphi}_4$	-0.702	0.660	± 3.3
$\delta\hat{\beta}_2$	0.115	0.140	± 0.7
$\delta\hat{\beta}_3$	0.084	0.159	± 0.8
$\delta\hat{\alpha}_2$	-0.071	0.265	± 1.3
$\delta\hat{\alpha}_4$	-0.079	0.317	± 1.6

Table 4.1: The parameters considered for non-GR injections where we only allow a single injected parameter to deviate from zero. Shown are the mean and standard deviation (stdev) as measured from GW150914 and the injected values considered in Figs. 4.7, 4.8 and 4.9.

early-advanced detectors and true data from the 6th science run S6. The resulting noise realization represents what we see in O1. We simulate the first event GW150914 using IMRPhenomPv2 for a few different deviations in a selection of the non-GR parameters $\delta\hat{p}_i$. These simulated non-GR waveforms are then injected into the recolored S6 data-stream on which the parameterized test is applied as we would do on a real detection.

Ideally one would perform various types of injections on all the non-GR parameters that are considered for recovery; however, such an extended investigation would be computationally too costly. Instead we restrict the analysis to a small subset of parameters; two in each region of the waveform. First we consider deviations from GR in one parameter at a time, namely, $\delta\hat{\varphi}_3$ and $\delta\hat{\varphi}_4$ for the early-inspiral, $\delta\hat{\beta}_2$ and $\delta\hat{\beta}_3$ for the intermediate regime and $\delta\hat{\alpha}_2$ and $\delta\hat{\alpha}_4$ for the merger-ringdown part of the waveform. Choices for the magnitudes of the deviations in the simulated signals are based on the accuracy with which these parameters were recovered in the case of GW150914. The size of the relative shifts, *i.e.* the values of $\delta\hat{p}_i$, are taken to be approximately five times the standard deviation as presented in Table 4.1. The mean, standard deviation (stdev) and $5 \times$ stdev values are listed in the same table for easy reference.

In each case, we recover with the same choice of non-GR parameters used in Sec. 4.2 *i.e.* we recover one-by-one with the PN parameters $\delta\hat{\varphi}_0, \ldots, \delta\hat{\varphi}_7$ and $\delta\hat{\varphi}_{5l}, \delta\hat{\varphi}_{6l}$, the intermediate $\delta\hat{\beta}_2, \delta\hat{\beta}_3$ and merger-ringdown parameters $\delta\hat{\alpha}_2, \delta\hat{\alpha}_3, \delta\hat{\alpha}_4$. Posteriors on $\delta\hat{p}_i$ resulting from the parameterized test are presented in the form of a violin plot: For each parameter the x-axis is a symmetric representation of the posterior "counts" while the y-axis shows the parameter values. The violin plots in Fig. 4.7, Fig. 4.8 and Fig. 4.9 show the posteriors and 95% confidence intervals for each of the recovered non-GR parameters. In each case we injected signals with a positive (orange) and negative (blue) shift corresponding to the 5σ values of the GW150914 posteriors.

Indeed we can observe that if one of the phasing coefficients is not as predicted by GR – at least when the deviation is large enough – this will have an impact on most, if not all of the other non-GR parameters.

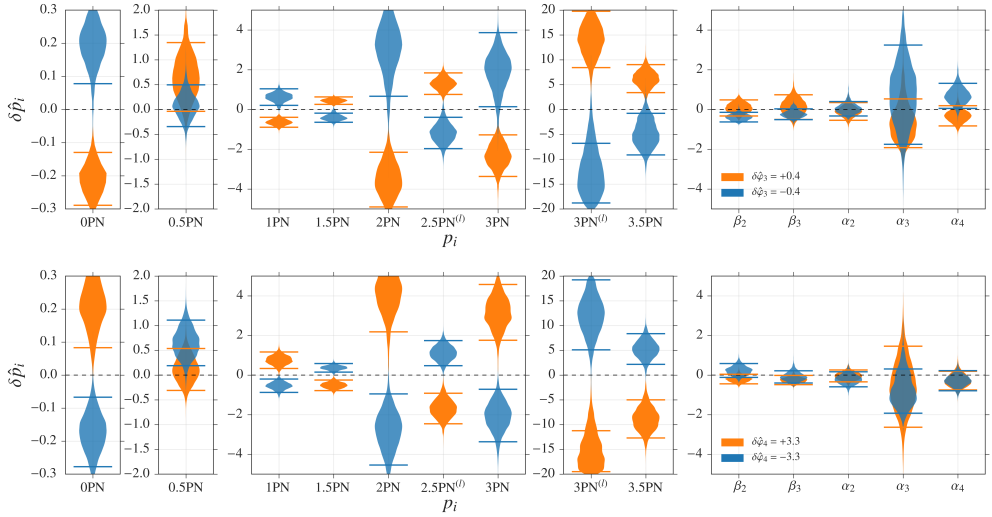


Figure 4.7: Posteriors for a full parameterized test on various non-GR sources. Positive injected shifts are represented by the orange curves, negative by the blue. The top panel shows results for the non-GR injections $\delta\hat{\varphi}_3 = \pm 0.4$ and the bottom panel that of $\delta\hat{\varphi}_4 = \pm 3.3$.

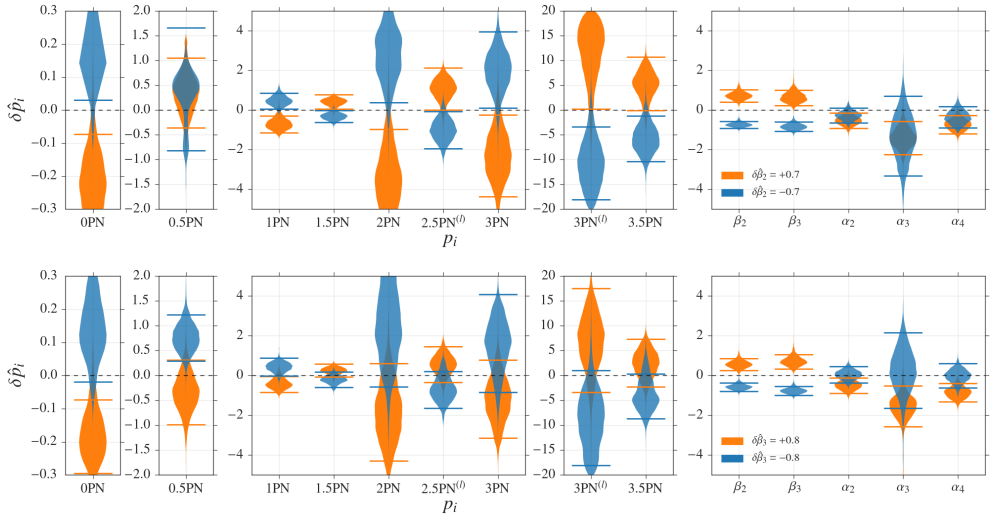


Figure 4.8: Same as in Fig. 4.7, but for injections $\delta\hat{\beta}_2 = \pm 0.7$ (top) and $\delta\hat{\beta}_3 = \pm 0.8$ (bottom).

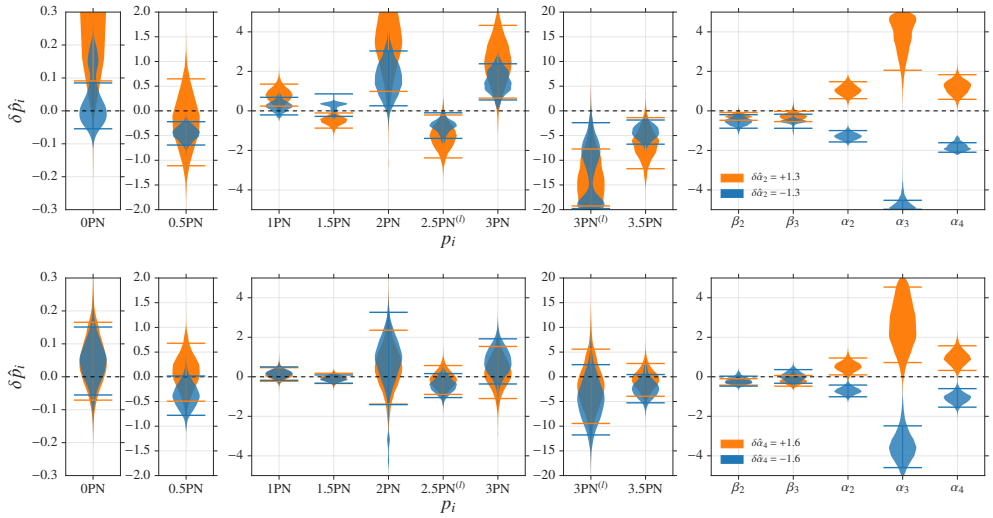


Figure 4.9: Same as in Fig. 4.7, but for injections $\delta\hat{\alpha}_2 = \pm 1.3$ (top) and $\delta\hat{\alpha}_4 = \pm 1.6$ (bottom).

In most known alternative theories of gravity, the corrections to GR extend to all PN orders even if in most cases they have been computed only at leading PN order [120, 121, 122]. We therefore also consider what happens to our results if that is indeed the case. To demonstrate that such violations are easily picked up by the parameterized test we have set up another set of non-GR injections. In this case two scenarios are considered:

1. All non-GR parameters starting from 1.5PN $\delta\hat{\varphi}_{\{3,4,5l,6,6l,7\}}$, $\delta\hat{\beta}_{\{2,3\}}$ and $\delta\hat{\alpha}_{\{2,3,4\}}$ have the same fractional shift $\delta\hat{p}_i = 0.5$.

2. All non-GR parameters starting from 1.5PN have a shift that varies in sign from one coefficient to the other, according to the way they are correlated: $\delta\hat{\varphi}_{\{3,5l,6l,7\}} = 0.5$ and $\delta\hat{\varphi}_{\{4,6\}} = -0.5$. For the intermediate and merger-ringdown parameters we simply choose to continue the alternating sign convention, *i.e.* $\delta\hat{\beta}_2 = \delta\hat{\alpha}_2 = \delta\hat{\alpha}_4 = 0.5$ and $\delta\hat{\beta}_3 = \delta\hat{\alpha}_3 = -0.5$.

As before in Figs. 4.7-4.9, we recover with all the individual non-GR parameters to observe the effect on each one. The results are presented in Fig. 4.10. All recovered parameters show very convincing offsets from their GR values and we can indeed conclude that should a violation of GR present itself in multiple phasing coefficients, the parameterized test of GR has no difficulty detecting it.

The waveform tests performed in this section demonstrate that when the non-GR parameters are varied within the upper bounds set by current observations, we witness no effects on the resulting time-domain evolution that should raise concern. In some cases the signal can exhibit an unusual ringdown signature in the case of relatively large deviations in the non-GR parameters. One example of such an effect is given

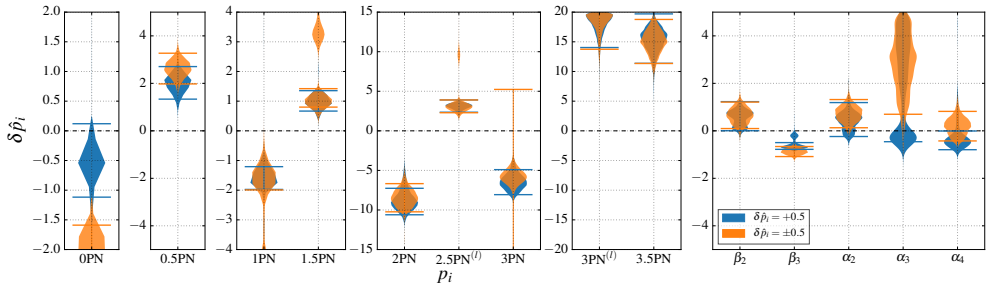


Figure 4.10: Posteriors for a full parameterized test on two non-GR sources that have violations present at all coefficients. The orange distributions represent the case where the non-GR parameters $\delta\hat{\varphi}_{\{3,4,5l,6,6l,7\}}$, $\delta\hat{\beta}_{\{2,3\}}$ and $\delta\hat{\alpha}_{\{2,3,4\}}$ have the same shift of $\delta\hat{p}_i = 0.5$. The blue distributions represent the case where all the non-GR parameters have shifts of magnitude 0.5, but with alternating sign that follows the correlation between the PN parameters and is continued all the way up to $\delta\hat{\alpha}_4$.

by increasing the relative shift in $\delta\hat{\varphi}_3$ to 200% which is already excluded by current observations. By investigating the effect of parameter shifts on the phase evolution we have gained some insight into which type of sources are most suitable to detect small deviations in particular phasing parameters and that indeed the PN parameters contribute most for light systems, whereas the influence of intermediate and merger-ringdown parameters are much more apparent in heavier systems, where the relevant regime appears in the detectors' most sensitive frequency band.

A quantitative investigation into the effects of GR violations on the parameterized test in a parameter estimation scenario has demonstrated the strength of the test, even when recovery is done using one non-GR parameter at a time. Furthermore we have demonstrated that when a violation appears in multiple phasing coefficients at the same time, the parameterized test has no difficulty distinguishing it from GR.

4.1.3 Impact of noise on the parameterized test

It is not unlikely that instrumental noise fluctuations would cause the degree of apparent deviation from GR found to occur in the single-parameter measurements for GW150914 (See Sec. 4.2), even in the absence of an actual deviation from GR. However, we cannot fully exclude a systematic origin from inaccuracies or even missing physics in our waveform models or the way the pipeline is built. In this section we would like to demonstrate that the systematic inaccuracies are either very small or non-existent by presenting the results from numerical injections. We will inject an NR source (SXS:BBH:0307 [123]) resembling GW150914 in recolored S6 noise, which is publicly available. In the case there are no systematic errors in either the waveform model or the pipeline, one expects the GR quantiles defined as

$$Q_{\rm GR} = \int_{-\infty}^{0} \mathrm{d}x\, p(x|d, I), \tag{4.3}$$

to be distributed uniformly on the interval $[0, 1]$. This follows from the way the quantile is defined and is independent from the shape of the posterior $p(x|d, I)$. The GR quantile is close to zero if there is a significant positive offset in the posterior, it will be close to unity in the case of a large negative offset and 0.5 in case the median is at the GR value of zero.

When we look at the combined results from GW150914 and GW151226 presented in Sec. 4.2 (Fig. 4.13), we see that the offset observed in the inspiral non-GR parameters has significantly decreased due to the accurate measurements from GW151226. Because of the many cycles observed in GW151226, any noise effects are more easily canceled out resulting in the nicely centered posteriors we see now. This does however not prove that there are no systematic errors. We injected an NR waveform into 20 different noise realizations from recolored S6 data that showed no obvious glitch features[2] that might interfere with the signal. In the parameter recovery we only considered the early-inspiral PN coefficients to demonstrate the robustness of the parameterized test. Indeed a set of 20 samples is not large enough to prove the resulting quantiles are indeed uniformly distributed, but it does give us a compelling argument in favor of the pipeline being robust. The sparseness of the number of data sets is due to computational expense. Fig. 4.11 shows the resulting posteriors for the NR injections. By looking at the posteriors one can already see that in some noise realizations, similar offsets away from zero as observed in [73] and Fig. 4.13 are present. Evidently, there is at least a one in twenty chance that the posteriors measured for one event exhibit a 2σ offset toward positive or negative deviations from zero. Note that if there is an offset present toward either positive or negative values, the other parameters show an alternating deviation according to their correlations. This is exactly what we expect to see based on the nature of the PN parameters which have alternating signs (see Eq. (1.108)). Fig. 4.12 provides a histogram of the GR quantiles of all the posteriors. We expect that for a large enough set of NR injections, the quantiles will be uniformly distributed. Indeed from the histograms we can observe that no deviations from a uniform distribution that should raise concern are evident. Only $\delta\hat{\varphi}_1$ seems to show a preference toward low values, however we need to keep in mind that we are dealing with low number statistics. Also note that there are similar large or small quantiles (*i.e.* far from the ideal 0.5) as there were in the first GW detection [73], demonstrating that noise alone is indeed capable of mimicking a violation in the PN parameters for such short signals.

4

4.2 O1 results

The previous sections demonstrated the robustness and strength of the parameterized test. All these tests were performed in a somewhat more concise manner for the internal review of the parameterized test. This section will summarize the results of the parameterized test of GR for both GW150914 and GW151226.

Because GW150914 was emitted by a binary black hole in its final phase of rapid orbital evolution, its gravitational phasing (or phase evolution) encodes non linear

[2]This was done in the same way as in Sec. 3.2; by checking spectrograms made from the stretches of time surrounding the signal for sharp features.

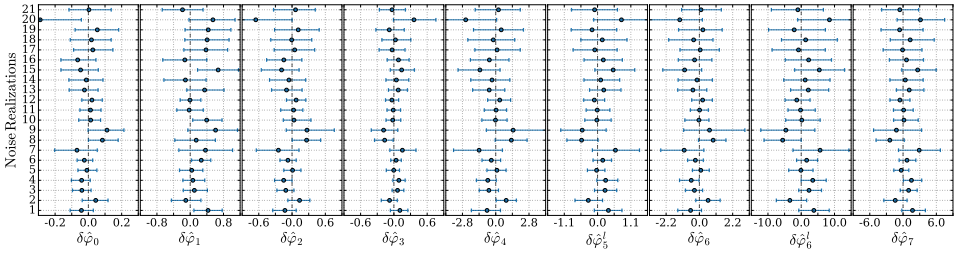


Figure 4.11: The 95% confidence intervals and medians of all the early-inspiral PN non-GR parameters recovered from the same numerical injection in 21 different noise realisations. The injected waveform is a numerically evaluated source that resembles GW150914. For each noise realization, spectrograms were checked by eye for glitches within the analyzed stretch of data.

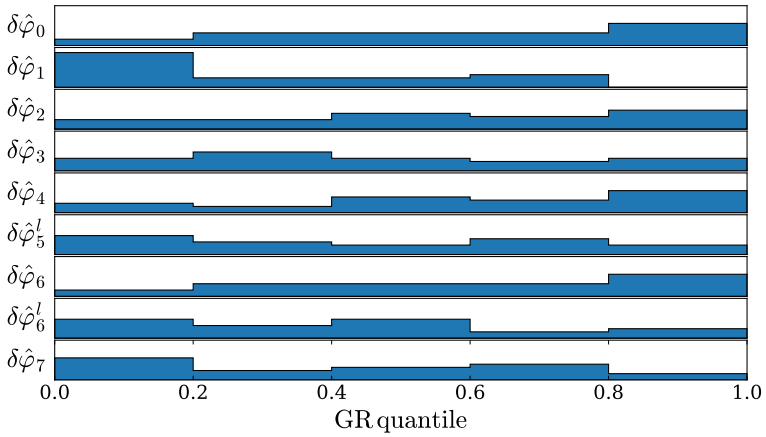


Figure 4.12: Histograms of the GR quantiles as calculated with Eq. (4.3) for the non-GR posteriors on the 21 NR injections presented in Fig. 4.11. From the histograms we can observe that no deviations from a uniform distribution that should raise concern are evident. Only $\delta\hat{\varphi}_1$ seems to show a preference toward low values; however, we must keep in mind that we are dealing with low number statistics. The sparseness of the number of data sets is due to computational expense.

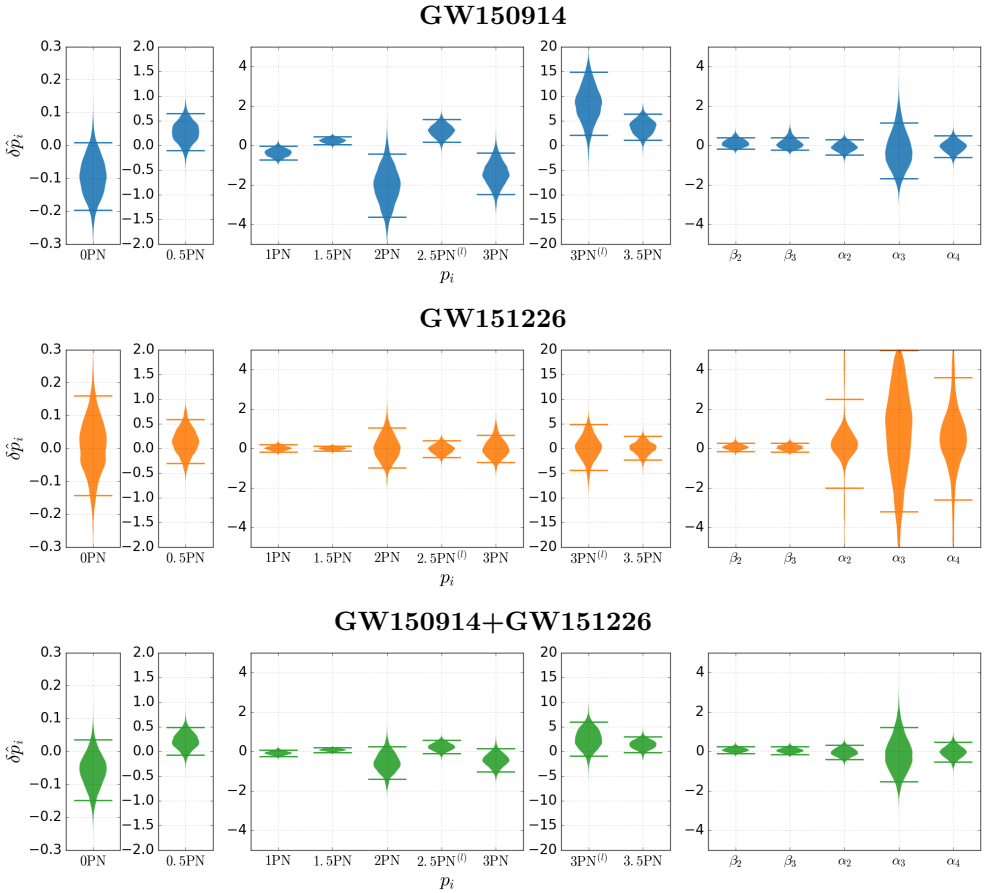


Figure 4.13: Posterior density distributions and 90% credible intervals for relative deviations $\delta\hat{p}_i$ in the PN parameters p_i (where $^{(l)}$ denotes the logarithmic correction), as well as intermediate parameters β_i and merger-ringdown parameters α_i. The top panel is for GW150914 by itself and the middle one for GW151226 by itself, while the bottom panel shows *combined* posteriors from GW150914 and GW151226. While the posteriors for deviations in PN coefficients from GW150914 show large offsets, the ones from GW151226 are well-centered on zero as well as being more tight, causing the combined posteriors to similarly improve over those of GW150914 alone. For deviations in the β_i, the combined posteriors improve over those of either event individually. For the α_i, the joint posteriors are mostly set by the posteriors from GW150914, whose merger-ringdown occurred at frequencies where the detectors are the most sensitive.

conservative and dissipative effects that are not observable in binary pulsars, whose orbital period changes at an approximately constant rate. Those effects include tails of radiation due to backscattering of GWs by the curved background around the coalescing black holes [124], non linear tails (i.e., tails of tails) [125], couplings between black-hole spins and the binary's orbital angular momentum, interactions between the spins of the two bodies [126, 127, 54], and excitations of quasi-normal modes [128, 129, 130] as the remnant black hole settles into the stationary configuration. In other words GW150914 provided us with the first empirical access to the genuinely strong-field dynamics of gravity. The waveform amplitude peaked at a frequency that was well aligned with the best instrument sensitivity, which allowed us to study the late inspiral and merger-ringdown regime in great detail as described in Sec. 4.1.

Various tests were performed among which are the measurement of the least damped quasi-normal mode frequency, the consistency check between masses and spins estimated from different portions of the waveform [131] and our own parameterized tests of the waveform as a whole. Even though not much of the early inspiral of GW150914 was in the detector's sensitivity band, interesting bounds have been placed on the PN coefficients up to 3.5PN. An example of how these bounds relate to previous electromagnetic observations is given in Fig. 4.14, where we plot the 95% upper bounds on each of the PN parameters measured with the parameterized test together with the bounds obtained from the measured orbital-period derivative $\dot{P}_{\rm orb}$ of the double pulsar J0737-3039 [132].

The second detection GW151226 merged much later, at around 450 Hz. This event was therefore a perfect addition concerning tests of GR to what we could already do with GW150914: The signal provides the opportunity to probe the PN inspiral with many more waveform cycles. The SNR of GW151226 was rather low at ~ 13, especially compared to GW150914 which had an SNR of ~ 24, but by combining posteriors of both detections, we were capable of placing even tighter constraints on the non-GR parameters, especially in the PN regime.

We have for both events performed the single-parameter variant of the parameterized test, meaning we let each of the $\delta\hat{p}_i$ in turn vary freely while all others are fixed to their general relativity values, $\delta\hat{p}_j = 0$ for $j \neq i$. The single-parameter scenario models GR violations that would occur predominantly at a particular PN order (or in the case of the intermediate and merger-ringdown parameters, a specific power of frequency in the relevant regime). However, as we demonstrated in Sec. 4.1.2-B by injecting various non-GR waveforms, the single-parameter variant of the parameterized test is perfectly capable of capturing deviations that are measurably present at more than one order.

We choose the prior probability distributions of all non-GR parameters to be uniform and wide enough to encompass the full posterior probability density function. This ensures we do not introduce any bias by a priori assuming certain relative shifts are excluded. In particular we set $\delta\hat{\varphi}_i \in [-20, 20]$; $\delta\hat{\beta}_i \in [-3, 3]$; $\delta\hat{\alpha}_i \in [-5, 5]$. In all cases we obtain estimates of the physical parameters – e.g., masses and spins – that are in agreement with those reported in the parameter estimation companion paper [119], in which we do not perform tests of GR.

In [73], for completeness we had also shown results from analyses where the parameters in each of the regimes early-inspiral, intermediate and merger-ringdown are

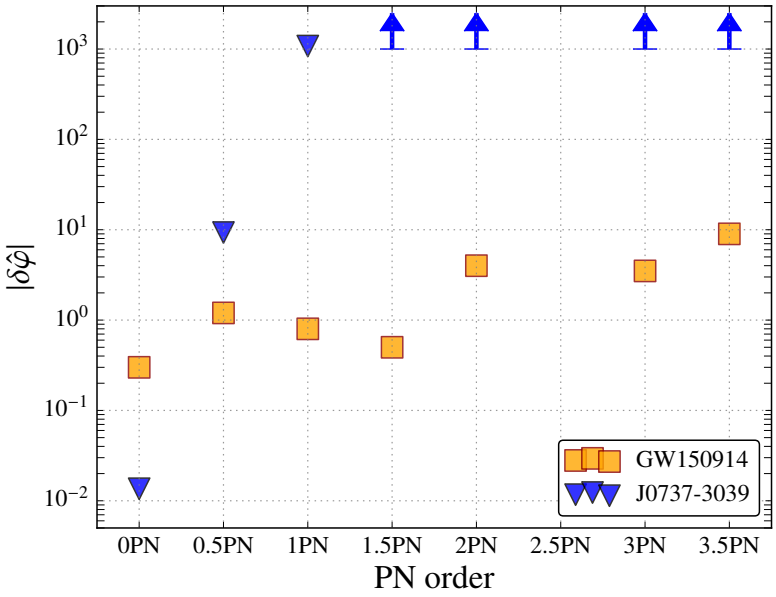


Figure 4.14: 90% upper bounds on the fractional variations of the known PN coefficients with respect to their GR values. The orange squares are the 90% upper bounds obtained from the single-parameter analysis of GW150914. As a comparison, the blue triangles show the 90% upper bounds extrapolated exclusively from the measured orbital-period derivative $\dot{P}_{\rm orb}$ of the double pulsar J0737-3039 [133, 132]. The GW phase deduced from an almost constant $\dot{P}_{\rm orb}$ cannot provide significant information as the PN order is increased, so we show the bounds for the latter only up to 1PN order. We do not report on the deviation of the 2.5PN coefficient, which is unmeasurable because it is degenerate with the reference phase. We also do not report on the deviations of the logarithmic terms in the PN series at 2.5PN and 3PN order, which can be found in Fig. 4.15.

4

allowed to vary simultaneously, but these tests return wide and uninformative posteriors. By contrast, analyses where the testing parameters $\delta\hat{p}_i$ are varied one at a time have much smaller statistical uncertainties. As we mentioned earlier, testing for deviations in a single non-GR parameter is an efficient way to expose GR violations that occur at multiple PN orders and we have shown in Sec. 4.1.2-B that violations at different powers of frequency as the parameter is associated with can be seen. These important points show that the analyses considered here are indeed well suited in the search for generic deviations from GR. Finally, we repeat that it is important to realize that if a violation is present, the measured values of the $\delta\hat{p}_i$ will not necessarily reflect the predicted values of the correct alternative theory. In order to place reliable constraints on theory-specific quantities such as coupling constants or extra charges, waveform models need to be used that describe the full inspiral-merger-ringdown for a specific modified theory of gravity [1]. In most cases, such models are not yet avail-

able. However, in the present work the focus is on model-independent tests of general relativity itself.

Given the observation of more than one BBH merger, posterior distributions for the $\delta\hat{p}_i$ can be combined to yield stronger constraints. The way posteriors are combined is explained in section 2.5 and also given in Eq. (4.2). In Fig. 4.13 we show the posteriors from GW150914 and GW151226 by themselves, generated with final instrumental calibration, as well as joint posteriors from the two events together. We do not present similar results for the candidate LVT151012 since it is not as confident a detection as the others; furthermore, its smaller detection SNR means that its contribution to the overall posteriors is insignificant.

For GW150914, the testing parameters for the PN coefficients, $\delta\hat{\varphi}_i$ and $\delta\hat{\varphi}_{il}$, showed moderately significant (2–2.5σ) deviations from their general relativity values of zero (Fig. 4.13). In Sec. 4.1.3 we have seen that such deviations – at least for a source as heavy as GW150914 – occurs due to the influence of noise alone for at least one in twenty detections. Another notable feature is the sign-changes the shifts display. It is however not surprising this should happen for the majority of the early-inspiral parameters since we find that these parameters have a substantial degree of correlation. Thus, if a particular noise realization causes the posterior distribution of one parameter to be off-centered with respect to zero, we expect that the posteriors of all the other parameters will also be off-centered. This is indeed what we observe as the medians of the early-inspiral single-parameter posteriors show opposite sign shifts that closely follow the sign pattern found in the PN series (see Fig. 4.13). By contrast, the posteriors of GW151226 tend to be centered on the general relativity value. As a result, the offsets of the combined posteriors are smaller. Moreover, the joint posteriors are considerably tighter, with a 1-σ spread as small as 0.07 for deviations in the 1.5PN parameter φ_3, which encapsulates the leading-order effects of the dynamical self-interaction of space-time geometry (the "tail" effect) [134, 135, 124, 136] as well as spin-orbit interaction [126, 127, 54].

In Fig. 4.15, we show the 90% credible upper bounds on the magnitude of the fractional deviations in PN and phenomenological coefficients $|\delta\hat{\varphi}_i|$ (except for $\delta\hat{\varphi}_5$, which is degenerate with the reference phase), which are affected by both the offsets and widths of the posterior density functions for the $\delta\hat{\varphi}_i$ parameters. We show bounds for GW150914 and GW151226 individually, as well as the joint upper bounds resulting from the combined posterior density functions of the two events. Not surprisingly, the quality of the joint bounds is mainly due to GW151226 in the PN coefficients, because of the larger number of inspiral cycles in the sensitive frequency band of the detectors. Note how at high PN order the combined bounds are slightly looser than the ones from GW151226 alone; this is because of the large offsets in the posteriors from GW150914. As opposed to the PN parameters, GW150914 has the largest effect on the bounds in the merger-ringdown parameters, because for this event, it was the merger-ringdown regime that fell in the most sensitive part of the frequency band of the detectors.

Following [133], we also show the bounds obtained from the measured orbital-period derivative $\dot{P}_{\rm orb}$ of the double pulsar J0737-3039 [132] in Fig. 4.14. Not surprisingly, since in binary pulsars the orbital period changes at essentially a constant rate, the corresponding bounds quickly become rather loose as the PN order is increased.

As a consequence, the double-pulsar bounds are significantly less informative than GW150914, except at 0PN order, where the double-pulsar bound is better thanks to the long observation time ($\sim$ 10 years against $\sim$ 0.4 s for GW150914)[3]. Thus, the combined posteriors from GW150914 and GW151226 allow us for the first time to constrain the coefficients in the PN series of the phasing up to 3.5PN order.

Next we consider the intermediate-regime coefficients $\delta\hat{\beta}_i$, which pertain to the transition between inspiral and merger-ringdown. For both GW150914 and GW151226, this stage is well inside the sensitive part of the detectors' frequency band. Returning to Fig. 4.13, we see that the measurements for GW151226 are of comparable quality to GW150914, and the combined posteriors improve on the ones from either detection by itself.

Last, we look at the merger-ringdown parameters $\delta\hat{\alpha}_i$. For GW150914, this regime corresponds to frequencies of $f \in [130, 300]$ Hz, while for GW151226 it occurred at $f \gtrsim 400\,\mathrm{Hz}$. As expected, the posteriors from GW151226 are not very informative for these parameters, and the combined posteriors are essentially determined by those of GW150914.

In summary, GW151226 makes its most important contribution to the combined posteriors in the PN inspiral regime, where both offsets and statistical uncertainties have significantly decreased over the ones from GW150914, in some cases to the $\sim$ 10% level. In the combined parameterized test, the GR value is usually found to be very close to the median of the marginalized distributions. This is mostly due to the fact that when combining the two events we have accurate measurements of both the early-inspiral and merger-ringdown regimes.

4.3 Tests of GR in O2 and beyond

The two detections that the first GW observing run has brought us have allowed us to test general relativity in the genuinely strong-field dynamical regime. For the first time we have been able to place constraints on possible deviations from GR in all the PN coefficients of the early-inspiral part of the GW waveform. Some of these constraints, like the 1.5PN non-GR coefficient restrict violations to the 10% level. The value of combining posteriors from multiple sources was demonstrated and bounds placed on the non-GR parameters will continue to improve as more sources are combined, roughly with the square root of the number of detections.

In the current chapter we demonstrated what the parameterized test is capable of, by way of investigating waveform dynamics, simulating detection scenarios and application to detected gravitational waves. The parameterized test does not suffer from measurable systematic errors coming from either the pipeline or the waveform model. The method is not restricted to picking up violations occurring at only one of the GR coefficients that go into the frequency evolution of the GW phasing as we have demonstrated by injecting various types of non-GR waveforms into simulated realistic data.

[3]Note that when computing the upper bounds with the binary-pulsar observations, we include the effect of eccentricity only in the 0PN parameter. For the higher PN parameters, the effect is not essential considering that the bounds are not very tight.

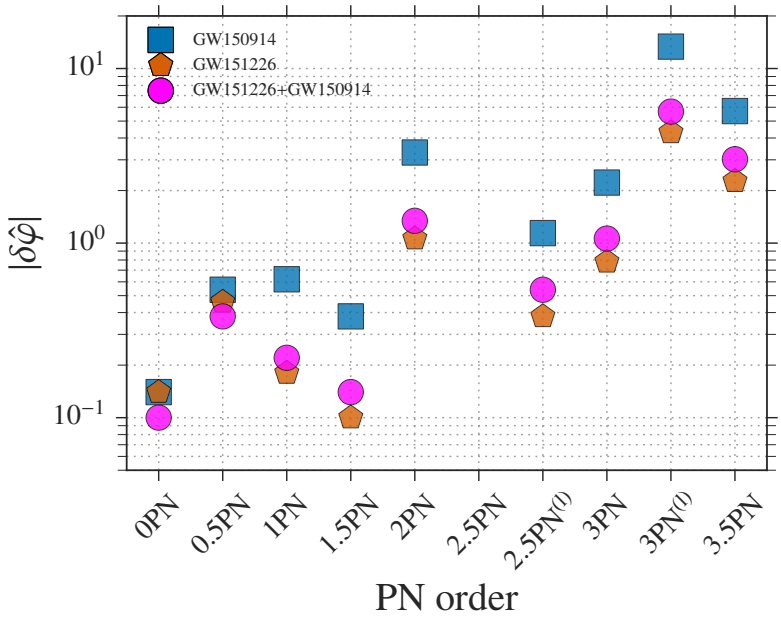
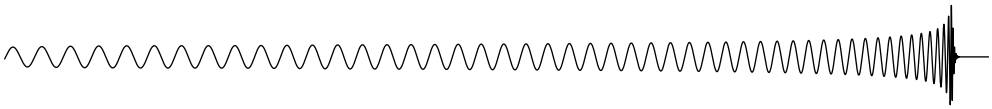


Figure 4.15: The 90% credible upper bounds on the magnitude of deviations in the PN coefficients. Also shown are joint upper bounds from the two detections; the main contributor on $\delta\hat{\varphi}_i$ is GW151226, which had many more inspiral cycles in band than GW150914. At 1PN order and higher the joint bounds are slightly looser than the ones from GW151226 alone; this is due to the large offsets in the posteriors for GW150914.

A very strong aspect of the parameterized test is the ability to combine posteriors from multiple sources as was already evident by combining the posteriors of the relative shifts in the waveform coefficients using only two detections. In the second observing run O2 and certainly the third O3, we expect to see many more sources [11] and combining posteriors will quickly result in very stringent bounds on the various non-GR parameters.

Chapter 5

Accelerated waveform generation and rapid likelihood evaluation



The LIGO-Virgo Collaboration (LVC) as a whole requires substantial computing power. Computational resources are needed for all facets and at all stages of detection and parameter estimation. A single parameterized test on an event like GW151226, requires 56 LALInference executions (chains)[1] running continuously for roughly 30 days. The rate estimates in [11] for O2 and O3 suggest we may observe many more BBH. In fact it is estimated that the chance of observing more than 30 significant events over the course of O3 is more than 80%. The example here only considers events that are as massive as GW151226, but O2 and O3 may detect much lighter sources, meaning much longer signals in the advanced detector bandwidth, scaling up the computation time considerably. Next to the parameterized tests of GR we plan to generate a TIGER background to perform hypothesis tests for the second part of O2 and certainly for O3. A proper background would have to consist of at least 1000 sources and if we decide on considering even only 4 out of 14 non-GR parameters, the previous estimate would scale by a factor of 4000. Needless to say, this quickly becomes computationally too expensive.

The major bottleneck in parameter estimation is the evaluation of waveforms and the computation of the likelihood. In a typical LALInference job, on the order of $10^7 - 10^8$ likelihood evaluations are performed [63]. A lot of effort over the past few years has been put into faster waveform evaluation with either surrogate models [137, 138, 139] or phenomenological models [59, 140, 32], speeding up the computation of the likelihood [64, 141] and in more recent cases a combination of both [142, 143]. The most impressive speed-up can be achieved using the class of techniques called *reduced order modeling* (see *e.g.* [137, 138, 139]), in particular so-called *reduced order quadratures* prove to be effective [142, 143]. The goal of reduced order models is to create a sparse basis that encapsulates a particular waveform model across a broad range in parameter space. This means that any waveform within this parameter space can be approximated as an expansion in the basis components.

In section 5.1 I will explain the concept of a reduced order model (ROM) and reduced order quadrature (ROQ) in more detail and provide the reader with the mathematical background that goes into it. Section 5.2 discusses the more practical side of building a ROM and ROQ. In section 5.3 I will describe the application of the concepts to significantly reduce computational requirements for the TIGER pipeline applied to BNS sources. Finally in section 5.4 I will present an ROQ for

[1] For a production parameter estimation run, we use 4 LALInference chains executed in parallel of which the results are combined afterwards. This is multiplied by 14 non-GR parameters, resulting in 56 chains in total.

TIGER applied to BBH sources and provide a realistic example by performing the parameterized test of GR for three coefficients on an injection similar to GW151226. The results and future work is discussed in the closing section 5.5.

5.1 ROMs and ROQs

Reduced order modeling is a class of techniques that can be used to construct a so-called surrogate model that provides fast and accurate approximations of a waveform model. Several methods with different applications have been proposed in the literature. One notable method is singular value decomposition (SVD). SVD-based methods have been used to calibrate phenomenological models with NR waveforms [144] and also for interpolating time-domain waveforms [145, 146, 147, 137, 138]. An SVD is a method to find a basis that accurately represents a space of functions such as GW waveforms. An SVD typically takes a dense set of waveforms (often called a *training set*) across parameter space and decomposes it into a set of orthogonal basis functions and corresponding projection coefficients, using a noise-weighted inner product as defined in Eq. (1.76). The number of these basis functions are equal to the number of waveforms in the training set. Not all the basis functions are required to accurately represent all the waveforms in the set and the SVD can be truncated to reach whatever representation accuracy is required. Any waveform in the set can then be represented by an expansion in that sub set of basis functions, weighted by the projection coefficients. For parameter estimation we need a model that represents the continuum rather than a discrete set of waveforms, hence some form of interpolation is always required. In order to build a basis for the continuum, the space between the coefficients can be interpolated using for example Chebyshev polynomials [147, 137]. Any waveform in the domain of the training set is then approximately represented by a linear combination of the basis functions weighted by the interpolated projection coefficients.

One disadvantage of using an SVD method can be the interpolation; the representation errors of the interpolants depend on the density of the training set. As is demonstrated throughout this chapter, the size of the training set is a major constraint in what can feasibly be achieved with current computing resources. What has proven to be an effective method while keeping training sets small, is the *greedy reduced basis* (RB) method [148, 149, 150]. Computing the SVD scales with the size of the training set squared, while the greedy algorithm we use for the reduced basis (RB) method [151] scales linearly. Furthermore, the greedy RB method has good scalability due to its embarrassingly parallel nature. It can easily be scaled up to use many computing cores [152, 153].

The greedy RB method is capable of generating a globally optimal basis (*i.e.* where the representation errors are minimized everywhere over the function space) even when sparse training sets are used. The method tends to be used in conjunction with the empirical interpolation method (EIM) [154, 155, 156] to build a final interpolant that accurately represents the entire space of gravitational waveforms.

A reduced order quadrature (ROQ) is essentially the discrete overlap calculation split into a data-dependent sum which only needs to be evaluated once for each

detection or source to be analyzed and a much shorter sum that takes care of the parameter dependent part of the overlap calculation. This can be achieved for closed-form analytical models for which the waveform interpolant can be evaluated at any frequency. Substituting the interpolant into the expression for the discrete overlap then constitutes the ROQ.

In this thesis I will focus on using a greedy RB method together with empirical interpolation to build ROQs for IMRPhenomPv2+ and TaylorF2+. These methods have seen considerable development over the past few years, are computationally efficient and allow for trivial parallelization. The computational efficiency and their numerical stability make these methods ideal candidates for tackling high dimensional models such as IMRPhenomPv2+.

5.2 Building process for reduced order quadratures

This section will focus on the details of how a reduced basis and finally an ROQ is built to be used in *e.g.* LALInference. Although the method is widely applicable, I will focus on the application to two specific gravitational wave models designed for the parameterized tests of GR and the full TIGER framework. The goal is to accurately represent a d-dimensional space of waveforms with as few basis elements as possible.

For gravitational waves, the space of waveforms is a continuous manifold where each point is a gravitational waveform $h(\boldsymbol{\theta})$ evaluated at a set of parameter values $\boldsymbol{\theta}$. This collection of waveforms do not themselves form a vector space, since the sum of two GW waveforms $h(\boldsymbol{\theta}_1) + (\boldsymbol{\theta}_2)$ is not a GW waveform. However, the waveforms can still be viewed as vectors in an infinite-dimensional space of functions. Within this space one can try to find a subspace spanned by a finite number of basis vectors $\hat{e}_i$ such that all GWs in the relevant part of parameter space can be well-approximated by a linear combination of the $\hat{e}_i$. Finding such a basis is discussed in the following section.

5.2.1 Constructing a basis

The method of choice in this work is the *greedy algorithm* (algorithm (5.1)). If the waveforms $h(\boldsymbol{\theta}; f)$ to be represented are known at a set of training points $\mathcal{T}_N = \{\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \ldots, \boldsymbol{\lambda}_N\}$, where $\boldsymbol{\lambda}_k$ is a set of parameter values, the greedy algorithm will construct a set of n orthonormal basis functions $\mathcal{E}_n = \{\hat{e}_i(f)\}_{i=1}^n$ such that

$$h(\boldsymbol{\lambda}; f) \approx \sum_{i=1}^{n} c_i(\boldsymbol{\lambda}) \hat{e}_i(f) \quad \text{for} \quad \boldsymbol{\lambda} \in \mathcal{T}_N, \tag{5.1}$$

where $n \leq N$. To be able to use the basis to achieve a proper performance increase in waveform calculations and likelihood evaluations one must have $n \ll N$, where the final value of n depends on the desired accuracy by which $h(\boldsymbol{\lambda}; f)$ needs to be approximated. For gravitational waves, $\boldsymbol{\lambda}$ represents the intrinsic parameters and f runs from some lower frequency $f_{\min}$ to an upper frequency $f_{\max}$, sampled at L points. I will refer to these frequency points as a *frequency series*. The number of

parameters in $\boldsymbol{\lambda}$ is referred to as the parameter dimension and L is the number of frequency bins. The basis constructed by a greedy algorithm will always satisfy

$$\sigma_n \equiv \max_{\boldsymbol{\lambda}} \min_{c_i} \| h(\boldsymbol{\lambda}; f) - \sum_{i=1}^{n} c_i(\boldsymbol{\lambda}) \hat{e}_i(f) \|^2 \leq \epsilon, \tag{5.2}$$

for any waveform $h(\boldsymbol{\lambda}; f)$ in the training set. The user-defined bound ϵ for the error is here taken to be 10^{-12}, which roughly corresponds to roundoff errors in numerical calculations. The minimization part is over the coefficients $c_i(\boldsymbol{\lambda})$, meaning that the $c_i(\boldsymbol{\lambda})$ are chosen such that the approximant is optimized for each training point $\boldsymbol{\lambda}_i$. The maximization is then over $\boldsymbol{\lambda}$, resulting in σ_n being the largest error of the best approximant. In other words, σ_n quantifies the "worst best" approximation by the basis. This ensures that the basis will represent all waveforms from the training set with errors of at most 10^{-12}.

From linear algebra we learn that the optimal representation (or best approximant) of a function by a set of orthonormal vectors is the projection onto its span.

$$\mathcal{P}_{\mathcal{E}_n}(h(\boldsymbol{\lambda}; f)) \equiv \sum_{i=1}^{n} \langle \hat{e}_i | h(\boldsymbol{\lambda}) \rangle \hat{e}_i(f), \tag{5.3}$$

where the inner product $\langle . | . \rangle$ is defined as

$$\langle f | g \rangle \equiv \int_{x_{\min}}^{x_{\max}} \mathrm{d}x \, f^*(x) g(x), \tag{5.4}$$

with * indicating the complex conjugate.

The basis $\mathcal{E}_n$ will not precisely represent every waveform in the training set $\mathcal{T}_N$, but the approximation in Eq. (5.1) minimizes the error

$$\| h(\boldsymbol{\lambda}) - \sum_{i=1}^{n} c_i(\boldsymbol{\lambda}) \hat{e}_i(f) \|^2, \tag{5.5}$$

when the coefficients $c_i(\boldsymbol{\lambda})$ are chosen such that

$$\langle \sum_{i=1}^{n} c_i(\boldsymbol{\lambda}) \hat{e}_i - \mathcal{P}_{\mathcal{E}_n}(h(\boldsymbol{\lambda})) | \hat{e}_j \rangle = 0 \quad \forall \quad \hat{e}_j \in \mathcal{E}_n. \tag{5.6}$$

Solving Eq. (5.6) for $c_i(\boldsymbol{\lambda})$ gives

$$c_i(\boldsymbol{\lambda}) = \langle \hat{e}_i | h(\boldsymbol{\lambda}) \rangle, \tag{5.7}$$

and Eq. (5.1) becomes

$$h(\boldsymbol{\lambda}; f) \approx \sum_{i=1}^{n} \langle \hat{e}_i | h(\boldsymbol{\lambda}) \rangle \hat{e}_i(f). \tag{5.8}$$

Initially the training set is just a large collection of waveforms and will of course not be orthonormal. In order to build an orthonormal basis the greedy algorithm (5.1)

selects waveforms from the training set that contribute most to the basis. These waveforms are orthonormalized using the Modified Gram-Schmidt procedure and added to the set of $\{\hat{e}_i\}$ until the approximation in Eq. (5.8) reaches the desired tolerance ϵ, i.e.

$$\|h(\boldsymbol{\lambda}; f) - \mathcal{P}_{\mathcal{E}_n}(h(\boldsymbol{\lambda}; f))\|^2 \leq \epsilon \quad \text{for} \quad \boldsymbol{\lambda} \in \mathcal{T}_N. \tag{5.9}$$

The evolution of the approximation error, also called *greedy error*, is shown in Fig. 5.7 for a few basis-building cases that are discussed in Sec. 5.4. The exponential decay is a typical feature of the greedy process [142, 143]. The set of orthonormalized functions $\mathcal{E}_n = \{\hat{e}_i(f)\}_{i=1}^n$ is called the reduced basis and will be the output of the greedy algorithm, together with the corresponding set of *greedy points* $\{\boldsymbol{\lambda}_i\}$ that were selected at each iteration. The process by which the RB algorithm (5.1) selects basis points is demonstrated graphically in Fig. 5.1. The gradient with which the points are colored visualizes which points were chosen early in the process and which in the end. The optimal approximation is given by

$$h(\boldsymbol{\lambda}; f) \approx \mathcal{P}_{\mathcal{E}_n}(h(\boldsymbol{\lambda}; f)) \equiv \sum_{i=1}^{n} \langle \hat{e}_i | h(\boldsymbol{\lambda}) \rangle \hat{e}_i. \tag{5.10}$$

Note that the resulting basis can only represent waveforms with the desired maximum error ϵ, if they are part of the original training set. However, provided the training set is dense enough and the waveforms have a smooth dependence on the values in $\boldsymbol{\lambda}$, the basis will approximate any waveform on the entire manifold bounded by the parameter ranges in $\mathcal{T}_N$ with the desired maximum representation error ϵ. Typically, for a high dimensional parameter space it is difficult to produce a dense training set such that the approximation in Eq. (5.10) is satisfactory and special techniques are necessary to achieve a good representation across the desired parameter space. These techniques will be discussed in sections 5.3 and 5.4, where the RB method will be applied to specific waveform models.

As an addition to finding a projection that covers the parameter space, we can also build interpolants that reduce the size of the frequency space required to represent the waveforms by using a sparse set of well chosen frequency points. The following section explains how we build such interpolants.

5.2.2 Empirical interpolation method

Each basis vector $\hat{e}_i(f)$ in the basis constructed with the greedy algorithm is represented by L frequency bins. The number of bins L is based on the duration T of longest waveform in the training set and depends on the maximum frequency $f_{\max}$ and the minimum frequency $f_{\min}$. At a constant sampling at the Nyquist rate $\Delta f = 1/T$, the length will be

$$L = (f_{\max} - f_{\min})T + 1. \tag{5.11}$$

We can now build interpolants for the waveforms, using the basis and the *empirical interpolation method* (EIM), which will reduce L to be the same length as the number

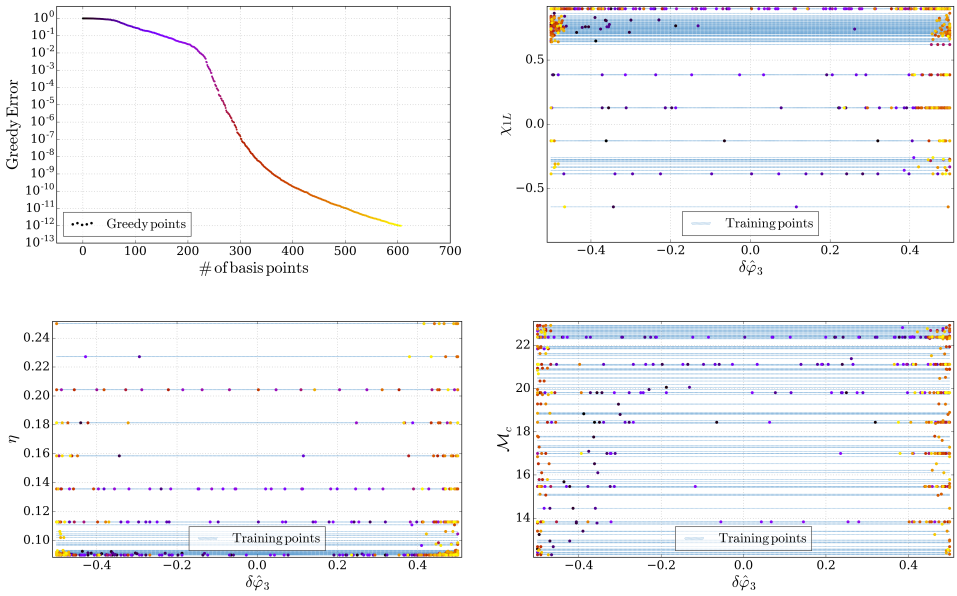


Figure 5.1: A demonstration of the greedy algorithm selecting points (gradient colored) from a training set (blue) while converging to a user-defined maximum tolerance error $\epsilon \leq 10^{-12}$. This example comes from an actual greedy sweep performed for building the IMRPhenomPv2+ basis where the model includes the non-GR parameter $\delta\hat{\varphi}_3$ (Sec. 5.4). The top right panel shows a 2D slice in parameter space for χ_{1L} and $\delta\hat{\varphi}_3$. The bottom left panel shows the same process for η and $\delta\hat{\varphi}_3$ and the bottom right panel for $\mathcal{M}_c$ and $\delta\hat{\varphi}_3$. The gradient visualizes which points were chosen early in the process (dark) and which in the end (light). The top left panel shows the decreasing greedy error as more basis points are added.

Algorithm 5.1 The greedy algorithm in its simplest form

Input: $\mathcal{T}_N = \{h(\boldsymbol{\lambda}_i; f)\}_{i=1}^N$ and $\{\boldsymbol{\lambda}_i\}_{i=1}^N$
Seed choice: $\boldsymbol{\lambda}_1$ (arbitrary)
$\mathcal{E}_1 = \{\mathbf{e}_1(f)\}$, where $\mathbf{e}_1(f) = h(\boldsymbol{\lambda}_1; f)/\|h(\boldsymbol{\lambda}_1; f)\|$
$i \leftarrow 1$
$\sigma_1 \leftarrow 1$
while $\sigma_i \geq \epsilon$ **do**
  $i \leftarrow i + 1$
  $\sigma_i \leftarrow \max_{\boldsymbol{\lambda}\in\mathcal{T}} \|h(\boldsymbol{\lambda}; f) - \mathcal{P}_{\mathcal{E}_{i-1}}(h(\boldsymbol{\lambda}; f))\|$
  $\boldsymbol{\lambda}_i \leftarrow \text{argmax} \|h(\boldsymbol{\lambda}; f) - \mathcal{P}_{\mathcal{E}_{i-1}}(h(\boldsymbol{\lambda}; f))\|$
  Perform orthonormalization with Gram-Schmidt
    on the set $\mathcal{E}_{i-1} \cup h(\boldsymbol{\lambda}_i)$ to get $\mathbf{e}_i(f)$
  $\mathcal{E}_i = \mathcal{E}_{i-1} \cup \mathbf{e}_i(f)$
end while
Output: basis $\mathcal{E}_i$ and greedy points $\mathcal{G}_i = \{\boldsymbol{\lambda}_k\}_{k=1}^i$

of basis elements n. Not only will this speed up the calculation of Eq. (5.10), the empirical interpolant will be required to build ROQs (Sec. 5.2.3) which in turn are used to speed up the calculation of the likelihood. The EIM was proposed in 2004 [154] and has since been applied in the effort to speed up parameterized inner product (overlap) calculations [152, 142, 143].

The goal is to build an *empirical interpolant* $\mathcal{I}_n[h](\boldsymbol{\lambda}; f)$ of a waveform function $h(\boldsymbol{\lambda}; f)$ spanned by a basis $\{\hat{e}_i(f)\}_{i=1}^n$. Note that this method does not rely on how the basis was built. In general, a well-posed interpolation problem for n basis functions requires n interpolation points $\{\mathcal{F}_i\}_{i=1}^n$. These points must ensure an accurate representation. Finding these points can be done in more than one way, but we will use the EIM algorithm. The interpolation points are selected as a subset of the full L samples. Before I explain how the algorithm selects these points I will first explain how the interpolant is built from an existing set of $\mathcal{F}_i$.

The empirical (*i.e.* problem-dependent) interpolant we seek can be written as

$$\mathcal{I}_n[h](\boldsymbol{\lambda}; f) \equiv \sum_{i=1}^{n} x_i(\boldsymbol{\lambda})\hat{e}_i(f), \tag{5.12}$$

where the coefficients x_i are defined as the solution to

$$\mathcal{I}_n[h](\boldsymbol{\lambda}; \mathcal{F}_k) \equiv h(\boldsymbol{\lambda}; \mathcal{F}_k) \quad \forall \quad k = 1, \ldots, n. \tag{5.13}$$

Solving for x_i means solving an n-by-n system of linear equations (this is why we need at least n frequency nodes), i.e.

$$\begin{aligned} \sum_{i=1}^{n} x_i(\boldsymbol{\lambda})\hat{e}_i(\mathcal{F}_k) &\equiv h(\boldsymbol{\lambda}; \mathcal{F}_k) \\ &\Rightarrow \\ (e_1(\mathcal{F}_k)\, e_2(\mathcal{F}_k)\, \ldots\, e_n(\mathcal{F}_k)) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} &= h(\mathcal{F}_k) \\ &\Rightarrow \\ A\mathbf{x} &= \mathbf{h}, \end{aligned} \tag{5.14}$$

where

$$\mathbf{h} \equiv \begin{pmatrix} h(\mathcal{F}_1) \\ h(\mathcal{F}_2) \\ \vdots \\ h(\mathcal{F}_n) \end{pmatrix}, \tag{5.15}$$

and

$$A \equiv \begin{pmatrix} e_1(\mathcal{F}_1) & e_2(\mathcal{F}_1) & \cdots & e_n(\mathcal{F}_1) \\ e_1(\mathcal{F}_2) & e_2(\mathcal{F}_2) & \cdots & e_n(\mathcal{F}_2) \\ e_1(\mathcal{F}_3) & e_2(\mathcal{F}_3) & \cdots & e_n(\mathcal{F}_3) \\ \vdots & \vdots & \ddots & \vdots \\ e_1(\mathcal{F}_n) & e_2(\mathcal{F}_n) & \cdots & e_n(\mathcal{F}_n) \end{pmatrix}. \tag{5.16}$$

Since the column vectors of A are linearly independent (they are the basis vectors), the matrix is guaranteed to be invertible. The unique solution to Eq. (5.13) is therefore $\mathbf{x} = A^{-1}\mathbf{h}$. We can now substitute the expression for x_i into Eq. (5.12) to find

$$\mathcal{I}_n[h](\boldsymbol{\lambda}; f) = \sum_{i=1}^{n}\sum_{k=1}^{n} \left(A^{-1}\right)_{ik} h(\boldsymbol{\lambda}; \mathcal{F}_k)\hat{e}_i(f) = \sum_{k=1}^{n} B_k(f) h(\boldsymbol{\lambda}; \mathcal{F}_k), \tag{5.17}$$

where

$$B_k(f) = \sum_{k=1}^{n} \left(A^{-1}\right)_{ik} \hat{e}_i(f). \tag{5.18}$$

Note that in the evaluation of $\mathcal{I}_n[h](\boldsymbol{\lambda}; f)$, the waveform function $h(\boldsymbol{\lambda}; \mathcal{F}_k)$ enters explicitly. This means that in order to evaluate the interpolant, the waveform function needs to be known at the nodes $\{\mathcal{F}_k\}_{k=1}^n$. For the models considered in the next sections, this is not a problem, since they are both closed-form expressions, already in the frequency domain. The empirical interpolant is of course not exact and will introduce an additional error. There are many options to choose from when trying to interpolate a function, all will have some manner of representation errors. One of the advantages of using the EIM is that, unlike most other forms of interpolation, this method is optimal in a precise sense. Namely the interpolant always satisfies

$$\max_{\boldsymbol{\lambda}} \|h(\boldsymbol{\lambda}; f) - \mathcal{I}_n[h](\boldsymbol{\lambda}; f)\|^2 \leq \Lambda_n^2 \sigma_n, \tag{5.19}$$

where σ_n is the maximum projection error of the basis ($\leq \epsilon$) and Λ_n^2 is a Lebesgue constant that can be computed as is shown in section 4.2 of [152]. Errors may often turn out to be much lower than the upper bound given in Eq. (5.19). The Lebesgue constant in the current context is around 100, suggesting that with a maximum projection error ϵ, the interpolants will never exceed an error larger than $\sim 100\epsilon$.

We have demonstrated how to calculate the empirical interpolant *if* we know $\mathcal{F}_k$. In the following I will explain how these frequency nodes are determined by the EIM algorithm. The input for the algorithm is the basis set $\{\hat{e}_i(f)\}_{i=1}^n$ we built earlier and an arbitrary number of frequency samples $f = \{f_i\}_{i=1}^L$. Note that this set can be chosen arbitrarily as opposed to using a frequency sampling at a constant Δf, as long as all the waveforms in the original training set are sufficiently sampled for this choice. This measure of freedom will be important when building ROQs for long waveforms. The empirical interpolation nodes $\{\mathcal{F}_k\}_{k=1}^n$ are selected from $\{f_i\}_{i=1}^L$, where $n \ll L$. The EIM algorithm 5.2 is not as intuitive as the greedy algorithm (5.1), but it can

be explained in a qualitative manner. At the first iteration the algorithm simply selects the frequency point that maximizes the amplitude of the first basis vector *i.e.* $|e_1(\mathcal{F}_1)| \geq |e_1(f_i)|$ for all samples. The following iteration will then build an interpolant of $e_2(f)$ using only e_1 and $\mathcal{F}_1$. The frequency point that maximizes the point-wise interpolation error *i.e.* $|\mathcal{I}_1[e_2](\mathcal{F}_2) - e_2(\mathcal{F}_2)| \geq |\mathcal{I}_1[e_2](f_i) - e_2(f_i)|$ will be the second interpolation point $\mathcal{F}_2$. At each iteration one frequency point will be added, which is in turn used to build up the interpolant. In the final iterations, the maximum point-wise interpolation error will get small, since more and more of the optimal basis is used to construct the interpolant. In principle this is another greedy process that selects points that "add the most information" to the interpolant. Where the greedy RB algorithm choses waveforms that point the most outside of the basis at the current iteration, the EIM algorithm selects frequency nodes at which the amplitude is most different from the interpolant at the current iteration. Fig. 5.2 shows a few steps of the process.

Algorithm 5.2 The EIM algorithm

Input: $\mathcal{E}_n = \{\mathbf{e}_i(f)\}_{i=1}^n$ and $\{f_i\}_{i=1}^L$
$i \leftarrow \text{argmax}(|\mathbf{e}_1(f)|)$ (*i.e.* index of largest entry in $\mathbf{e}_1(f)$)
$\mathcal{F}_1 \leftarrow f_i$
for all $j \in \{2, 3, \ldots, n-1\}$ **do**
  Calculate $\mathcal{I}_{j-1}[\mathbf{e}_j](f)$
  Calculate point-wise error $\mathbf{d}(f) = \mathcal{I}_{j-1}[\mathbf{e}_j](f) - \mathbf{e}_j(f)$
  $i \leftarrow \text{argmax}(|\mathbf{d}(f)|)$
  $\mathcal{F}_j \leftarrow f_i$
end for
output: $\{\mathcal{F}_k\}_{k=1}^n$

5.2.3 Fast likelihood evaluations with reduced order quadratures

The interpolants constructed in the previous section allow us to evaluate waveforms for any value of the intrinsic parameters $\boldsymbol{\lambda}$ within the bounds on which the $\{\hat{e}_i(f)\}_{i=1}^n$ is valid. We will now discuss how the interpolants can be used to speed up the likelihood evaluations in *e.g.* a nested sampling algorithm. The likelihood can be expressed in terms of weighted inner products as defined in Eq. (1.76). Assume that the detector data stream in the frequency domain is given by $d(f) = h(\boldsymbol{\theta}; f) + n(f)$, where $h(\boldsymbol{\theta}; f)$ is the GW strain depending on intrinsic and extrinsic parameters and $n(f)$ is instrumental noise. When we consider Gaussian noise, the likelihood can then be written as

$$\mathcal{L}(d|\boldsymbol{\theta}) \propto \exp(-\chi^2/2), \tag{5.20}$$

where

$$\chi^2 \equiv \langle n|n\rangle = (d - h(\boldsymbol{\theta})|d - h(\boldsymbol{\theta}))\,, \tag{5.21}$$

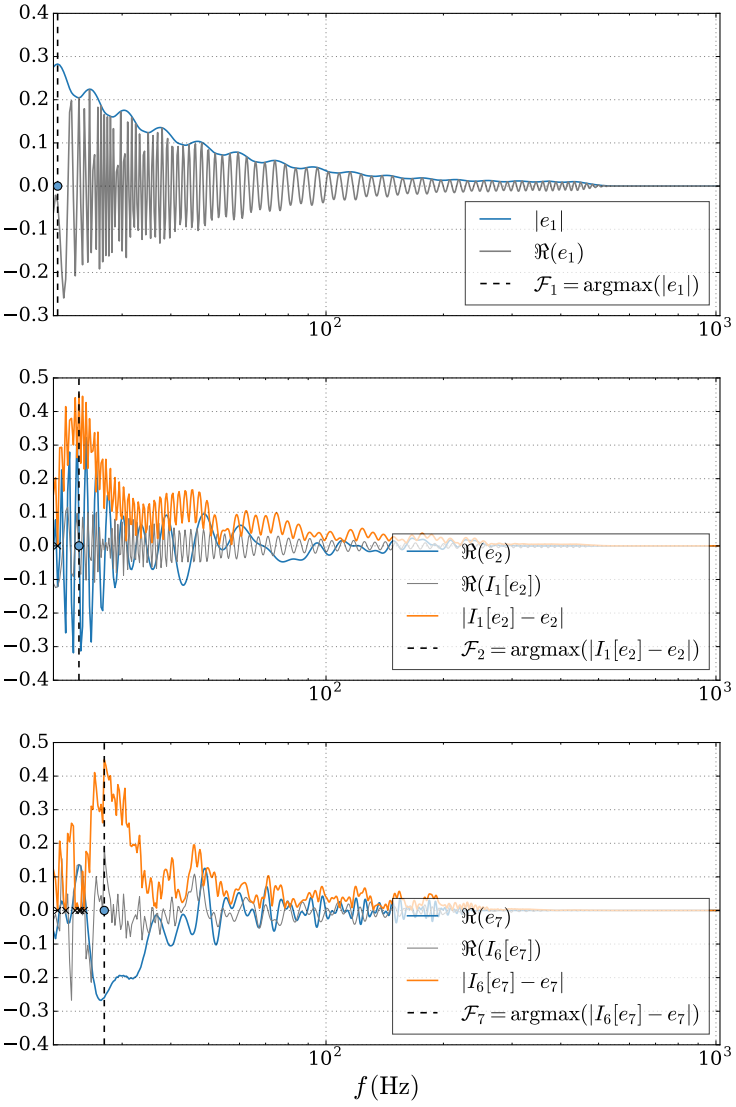


Figure 5.2: A few interpolants and basis vectors as they are used in the EIM algorithm. The 0th 1st and 6th iteration in the algorithm are shown here. The vertical dashed lines indicate where the EIM node $\mathcal{F}_n$ was found, *i.e.* where $|\mathcal{I}_1[e_2](\mathcal{F}_n) - e_2(\mathcal{F}_n)| \geq |\mathcal{I}_1[e_2](f_i) - e_2(f_i)| \quad \forall \quad i$. The black crosses indicate the previously found nodes. The basis functions and interpolants shown here are those of the IMRPhenomPv2+ model for case $\delta\hat{\varphi}_3^A$ (see Table 5.3).

and we note that we use the PSD-weighted inner product

$$(a|b) \equiv 4\Re \int_{f_{\min}}^{f_{\max}} \mathrm{d}f \, \frac{\tilde{a}^*(f)\tilde{b}(f)}{\tilde{S}_n(f)}. \tag{5.22}$$

The log likelihood is then written as

$$\log \mathcal{L} = \frac{1}{2}\left[2\,(d|h(\boldsymbol{\theta})) - (h(\boldsymbol{\theta})|h(\boldsymbol{\theta})) - (d|d)\right]. \tag{5.23}$$

Before going into the details, we can already demonstrate how an ROQ can in principle be built. When we look at Eq. (5.23), we notice inner products of the form $(h|h)$ and $(d|h)$. The most obvious power of the ROQ representation presents itself when we substitute the interpolant given in Eq. (5.17) into the discretized expression for $(d|h)$, we find

$$\begin{aligned}
(d|h(\boldsymbol{\lambda})) &\equiv 4\Delta f \Re \sum_{i=1}^{L} \frac{d^*(f_i)h(\boldsymbol{\lambda}; f_i)}{S_n(f_i)} \\
&\approx 4\Delta f \Re \sum_{i=1}^{L}\sum_{k=1}^{n} \mathcal{B}_k(f_i)h(\boldsymbol{\lambda}; \mathcal{F}_k)\frac{d^*(f_i)}{S_n(f_i)} \\
&= 4\Delta f \Re \sum_{k=1}^{n}\left[\sum_{i=1}^{L} \mathcal{B}_k(f_i)\frac{d^*(f_i)}{S_n(f_i)}\right] h(\boldsymbol{\lambda}; \mathcal{F}_k) \\
&\Rightarrow \\
(d|h(\boldsymbol{\lambda}))_{\mathrm{ROQ}} &\equiv \sum_{k=1}^{n} w_k h(\boldsymbol{\lambda}; \mathcal{F}_k). \qquad (5.24)
\end{aligned}$$

(5.25)

Note that we have now decoupled the parameter-dependent part from the much larger sum over all L frequency bins. The latter has now been written as a set of n ROQ weights $\{w_k\}_{k=1}^{n}$ where

$$w_k \equiv 4\Delta f \Re \sum_{i=1}^{L} \mathcal{B}_k(f_i)\frac{d^*(f_i)}{S_n(f_i)}, \tag{5.26}$$

which only need to be calculated once for each detection, since they only depend on the data stream and the parameter-independent part of the interpolants. Once the weights are calculated, only Eq. (5.25) needs to be evaluated, which is a sum over n elements as opposed to L. This means that with the ROQ a speedup of L/n can be achieved at runtime, which in practice can be anywhere between a factor of a few and several hundreds depending on the complexity of the model and the length of the signals (see Sec. 5.4).

We have now demonstrated the potential of using ROQs, but from Sec. 1.3 we know that $h(\boldsymbol{\theta}; f)$ is actually a combination of $+$ and $\times$ polarization states which do require some attention. The polarization states are projected into the detector's frame with

the antenna pattern functions $F_+(\mathrm{ra}, \mathrm{dec}, \psi, r)$ and $F_\times(\mathrm{ra}, \mathrm{dec}, \psi, r)$. Substituting the full expression for $h(\boldsymbol{\theta}; f)$ (Eq. (1.129)) into Eq. (5.23) we get the full expression of the log likelihood in terms of various inner products:

$$\begin{aligned} 2\log\mathcal{L} =& 2F_+ \left(d|h_+\right) + 2F_\times \left(d|h_\times\right) - |F_+|^2 \left(h_+|h_+\right) - |F_\times|^2 \left(h_\times|h_\times\right) \\ & - 2F_+F_\times \left(h_+|h_\times\right) - \left(d|d\right). \end{aligned} \tag{5.27}$$

In this expression we note that h_+ and $h_\times$ enter linearly into terms of the form $(d|h_A)$ and quadratically in $(h_A|h_B)$, where A and B may take the values $+$ or $\times$. In this work therefore, we will build ROQs for the linear part

$$h_A(\boldsymbol{\lambda}; f_i) \approx \sum_{j=1}^{N_\mathrm{L}} \mathcal{B}^\mathrm{L}_j(f_i) h_A(\boldsymbol{\lambda}; \mathcal{F}^\mathrm{L}_j), \tag{5.28}$$

and the quadratic part

$$\Re\left[h_A(\boldsymbol{\lambda}; f_i) h^*_B(\boldsymbol{\lambda}; f_i)\right] \approx \sum_{k=1}^{N_\mathrm{Q}} \mathcal{B}^\mathrm{Q}_k(f_i) \Re\left[h_A(\boldsymbol{\lambda}; \mathcal{F}^\mathrm{Q}_k) h^*_B(\boldsymbol{\lambda}; \mathcal{F}^\mathrm{Q}_k)\right], \tag{5.29}$$

so that we have an accurate representation of both polarization states and all its products appearing in the likelihood. The vectors $\{\mathcal{B}^\mathrm{L}_j(f)\}_{j=1}^{N_\mathrm{L}}$ contain the basis for the polarizations and $\{\mathcal{B}^\mathrm{Q}_k(f)\}_{k=1}^{N_\mathrm{Q}}$ contain the basis for all the products between the polarizations. As Eq. (5.28) suggests, both polarization states h_+ and $h_\times$ can be represented with a single basis $\mathcal{B}^\mathrm{L}_j$. In Eq. (5.29) we expect to require only $\mathcal{B}^\mathrm{Q}_k$ to represent all products $h_+h^*_+$, $h_\times h^*_\times$ and $\Re h_+h^*_\times$. That one only needs these two bases to represent the various polarizations and their products was empirically demonstrated in [143] and we will follow their approach in this work as well. Whether the basis will be accurate in a parameter estimation setting depends on the full process with which we will construct $\mathcal{B}^\mathrm{L}_k$ and $\mathcal{B}^\mathrm{Q}_k$. As the polarizations share a basis, they also share the same interpolation points $\{\mathcal{F}^\mathrm{L}_j(f)\}_{j=1}^{N_\mathrm{L}}$ and so do the products with their own set of points $\{\mathcal{F}^\mathrm{Q}_k(f)\}_{k=1}^{N_\mathrm{Q}}$.

We will now return to the likelihood in Eq. (5.27) and substitute the approximants defined in Eq. (5.28) and (5.29) to construct the reduced order quadratures

$$\begin{aligned} 2\log\mathcal{L} \approx& 2F_+ \left(d|h_+\right)_\mathrm{ROQ} + 2F_\times \left(d|h_\times\right)_\mathrm{ROQ} \\ & - |F_+|^2 \left(h_+|h_+\right)_\mathrm{ROQ} - |F_\times|^2 \left(h_\times|h_\times\right)_\mathrm{ROQ} \\ & - 2F_+F_\times \left(h_+|h_\times\right)_\mathrm{ROQ} - \left(d|d\right) \\ =& 2\log\mathcal{L}_\mathrm{ROQ}, \end{aligned} \tag{5.30}$$

in which we have the *linear ROQ*

$$\left(d|h_A(\boldsymbol{\lambda})\right)_\mathrm{ROQ} = \sum_{j=1}^{N_\mathrm{L}} w^\mathrm{L}_j(t_c) h_A(\boldsymbol{\lambda}; \mathcal{F}^\mathrm{L}_k), \tag{5.31a}$$

$$w^\mathrm{L}_j(t_c) = 4\Re\Delta f \sum_{i=1}^{L} \frac{d^*(f_i)\mathcal{B}^\mathrm{L}_j(f_i)}{S_n(f_i)} e^{-2\pi\mathrm{i}t_c f_i} \tag{5.31b}$$

and the *quadratic*

$$(h_A(\boldsymbol{\lambda})|h_B(\boldsymbol{\lambda}))_{\mathrm{ROQ}} = \sum_{j=1}^{N_{\mathrm{Q}}} w_j^{\mathrm{Q}} h_A(\boldsymbol{\lambda}; \mathcal{F}_k^{\mathrm{Q}}) h_B(\boldsymbol{\lambda}; \mathcal{F}_k^{\mathrm{Q}}), \tag{5.32a}$$

$$w_j^{\mathrm{Q}} = 4\Re\Delta f \sum_{i=1}^{L} \frac{\mathcal{B}_j^{\mathrm{Q}}(f_i)}{S_n(f_i)}. \tag{5.32b}$$

Note in Eq. (5.28) we temporarily set t_c and ϕ_c to zero. The waveform ends at t_c, so setting t_c to zero means that the waveform ends at $t = 0$. We implement an arbitrary non-zero t_c in the expression of the weights, simply by adding the offset $-t_c$ to the waveform, which in the frequency domain corresponds to multiplying by $e^{-2\pi \mathrm{i} f t_c}$. The coalescence phase ϕ_c is an overall phase shift that can be written as a multiplication by a complex constant and keeps the waveform in the RB space, but t_c requires a special approach introduced in [142].

A discrete set of ROQ weights is generated for n_c values of t_c, equally spaced on an interval $[t_{\mathrm{trigger}} - W, t_{\mathrm{trigger}} + W]$:

$$w_j^{\mathrm{L}}(t_k) = w_j^{\mathrm{L}}(t_{\mathrm{trigger}} - W + k\Delta t) \quad \text{for} \quad k \in \{0, 2, \ldots, n_c - 1\}, \tag{5.33}$$

where the width of the time window W is estimated by the search pipeline and $\Delta t = 2W/n_c$. This somewhat brute force method is feasible thanks to the high accuracy with which we can typically measure t_c (see *e.g.* [157]). The accurate measurement on t_c allows us to consider a narrow window W and achieve a dense sampling with few points n_c. Finally we can interpolate the discrete set of n_c weights $\{w_j^{\mathrm{L}}(t_k)\}_{k=0}^{n_c-1}$ we calculated and end up with weights as a continuous function of t_c: $w_j^{\mathrm{L}}(t_c)$. Calculating the weights is a one-time startup cost. This will need to be repeated for each detection or in the case of generating a TIGER background for the analysis of each injected waveform, but the major cost of a LALInference run is in evaluating the likelihood millions of times. As we have seen from the general definition of the ROQ in Eq. (5.25), the time it takes to evaluate the likelihood can be decreased by a factor L/n. In this case n depends on the length of the sums in both the linear and quadratic ROQs (Eqns. (5.31a) and (5.32a)) and the speedup becomes $L/(N_{\mathrm{L}} + N_{\mathrm{Q}})$. From [143] we know that this factor can be between a few and a few hundred for precessing BBH. Such speedups can make analyses on light systems starting at low frequencies feasible as running times of one year can be reduced to a few days.

5.2.4 The complete work-flow

The preceding sections have laid out the principles, mathematics and algorithms that go into first constructing a reduced basis, then building an empirical interpolant to finally construct the reduced order quadratures. All these steps fit into a generic work-flow. In Sec. 5.2.1 I briefly mentioned the purpose of a training set and how a sufficiently dense training set can produce a basis that covers the parameter space continuum. Even though I skipped over the details of generating such a set, it is a crucial step that can have an impact on the final efficiency of the ROQs. There are many ways to sample the points that go in the training set and not all will yield an

accurate basis across the entire parameter space. Building a training set is problem-dependent and for more complex models there is no single universal approach to find an optimal training set. In this thesis I present two methods, one for a relatively simple waveform model in Sec. 5.3 and another method that relies on an already existing basis for a much more complex model in Sec. 5.4. The training set can never be blindly trusted and therefore the resulting basis will be verified for accuracy for many random points spread out across the parameter space. In this section I will discuss the work-flow of how we can achieve an accurate basis starting with some initial training set $\check{\mathcal{T}}_M$ of size M. The outline of the work-flow can be sketched out in a few steps:

1. Generate an initial training set $\check{\mathcal{T}}_M$ of size M that provides reasonable coverage of the parameter space $\boldsymbol{\lambda}$, while still fitting within the memory capacity of one's computing resources.

2. Run a greedy algorithm to produce a first basis $\mathcal{E}^0_{n_0} = \{\mathbf{e}_i\}_{i=1}^{n_0}$, where each choice of $\mathbf{e}_i$ corresponds to a parameter value $\boldsymbol{\lambda}_i$. We will call the set of these parameters $\mathcal{G}^0_{n_0}$ the *greedy points*.

3. Construct the interpolants $\mathcal{I}_{n_0}[h](\boldsymbol{\lambda}; f)$ using an EIM algorithm.

4. Generate a large set of N random points $\Lambda^1_N = \{\boldsymbol{\lambda}_i\}_{i=1}^N$ covering the entire parameter space and test the interpolants by evaluating their errors $\epsilon_i = \|h(\boldsymbol{\lambda}_i; f) - \mathcal{I}_{n_0}[h](\boldsymbol{\lambda}_i; f)\|^2$ for all $\boldsymbol{\lambda}_i \in \Lambda^1_N$.

5. Collect all $\boldsymbol{\lambda}_i$ where the error ϵ_i exceeds some threshold β and call these "bad points" $\Lambda^{\mathrm{b},1}_{n_{\mathrm{b},1}}$.

6. Finally generate a new training set $\mathcal{T}^1_{n_1+n_{\mathrm{b},1}}$ that is the combination of the greedy points and the bad points, *i.e.* $\mathcal{T}^1_{n_1+n_{\mathrm{b},1}} = \mathcal{G}^0_{n_0} \cup \Lambda^{\mathrm{b},1}_{n_{\mathrm{b},1}}$ and repeat steps 2-5 until $\epsilon_i \leq \beta$ for all i.

The last three steps in this work-flow are called the *enrichment* process which was introduced in [143] and ensures that the final basis is accurate across the entire parameter space, even if the original training set was not sufficiently dense. Step 4 is what I refer to as the *validation* and is part of the enrichment process. In many cases n_0 can still be rather large, which demonstrates how poorly the original training set represents the continuum. In most cases, only one or two enrichment steps are required to make the basis converge to a complete coverage (*i.e.* $\Lambda^{\mathrm{b},i}_{n_{\mathrm{b},i}} = \emptyset$). In general I will omit the subscripts indicating the size of the various sets for brevity, but will mention the size if necessary. In section 5.4 we will see that for more complex[2] waveforms it might not be possible to find a sparse basis when including certain parts of parameter space. In this last case, the enrichment process will continue to add many points to the basis at each iteration and $\Lambda^{\mathrm{b},i}$ will remain large. In such a situation, the points are typically densely clustered in a small region of parameter space. This is in fact a useful feature of the basis building process, since it helps identify regions in parameter space where the waveform model exhibits strong dependence on parameter

[2]Waveforms that show rapid changes with respect to parameter variation.

variations and possibly unphysical peaks in amplitude and/or phase. We will see such an example in section 5.4.

The details of step 1 are left to sections 5.3 and 5.4, as each will use a different specialized method. I have developed a few Python modules that help setting up and walking through the work-flow presented above. The modules generate files and scripts which rely on algorithms developed by Field *et al.* [152, 153] to run, called GREEDYCPP and EIM respectively. This is publicly accessible software that can be downloaded from Bitbucket [151]. The Python code consists of 3 parts. The first part is a script (`generate_trainingpoints`) that generates a training set and is now part of the official GREEDYCPP code, the second part is a script I called ROQ_PIPE[3] which sets up everything needed to run steps 2 and 3. The third part is the script ROQ_VALIDATE, which takes care of steps 4-6. All scripts and modules required to walk through the work-flow can be downloaded from my Bitbucket page [158] which also includes some documentation. I have drawn a flowchart of the process and indicated the codes involved in Fig. 5.3. In principle it is possible to construct a front-to-back pipeline that builds ROQs for any model using this work-flow, if everything can be run on a single computing resource. Algorithm (5.3) shows what such an algorithm might look like.

Algorithm 5.3 Complete ROQ-building work flow in algorithmic representation

$\mathcal{T}^0 \leftarrow \check{\mathcal{T}}_M$
$i \leftarrow 0$
repeat
  $\mathcal{T}^i \rightarrow$ `greedycpp` $\rightarrow \mathcal{E}^i, \mathcal{G}^i$
  $\mathcal{E}^i \rightarrow$ `eim` $\rightarrow \{\mathcal{F}_m\}, \{\mathcal{B}_m\}$
  generate validation set Λ^{i+1}
  for all $\boldsymbol{\lambda}_n \in \Lambda^{i+1}$ **do**
    $\epsilon_n \leftarrow \|h(\boldsymbol{\lambda}_n; f) - \mathcal{I}_{\mathcal{E}^i}[h](\boldsymbol{\lambda}_n; f)\|^2$
    if $\epsilon_n > \beta$ **then**
      $\Lambda^{\mathrm{bad},i+1} \leftarrow \boldsymbol{\lambda}_n$
    end if
  end for
  $\mathcal{T}^{i+1} \leftarrow \mathcal{G}^i \cup \Lambda^{\mathrm{bad},i+1}$
  $i \leftarrow i + 1$
until $\Lambda^{\mathrm{bad},i} \neq \emptyset$
Calculate weights

In the following sections I will employ the reduced basis and interpolation methods discussed so far to two specific applications, namely the inspiral part of BNS systems (Sec. 5.3) and full inspiral-merger-ringdown BBH/NSBH systems (Sec. 5.4).

[3]Note that this name is rather optimistic since this part of the code only sets up scripts to run GREEDYCPP and EIM. The validation is done separately.

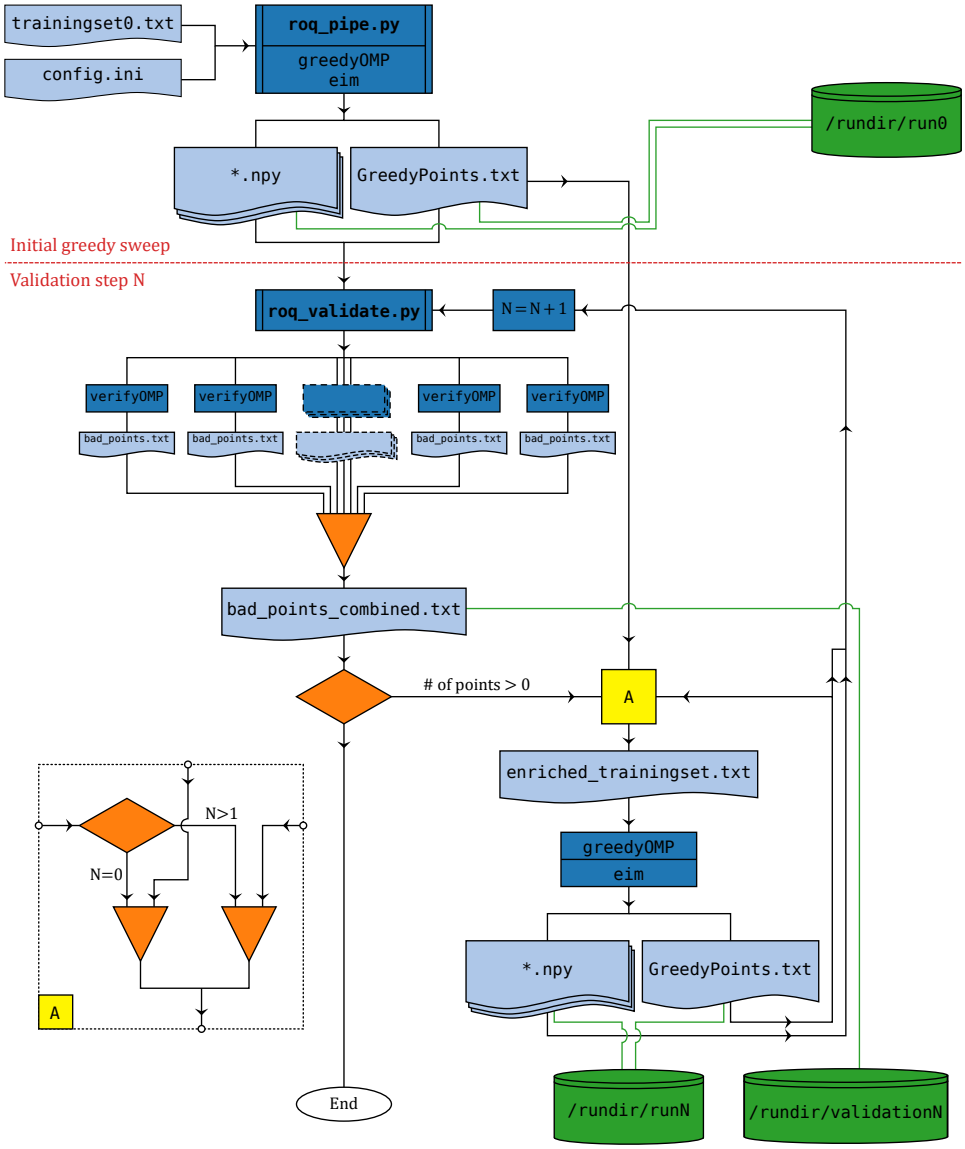


Figure 5.3: Flowchart representing the work-flow using my pipeline modules [158] (`roq_pipe.py` and `roq_validate.py`) and Field's GREEDYCPP software [151] (`greedyOMP`, `eim` and `verifyOMP`). `roq_pipe.py` takes care of the setup for the first greedy sweep, while `roq_validate.py` sets up the validation runs and enrichment steps. The validation process starts at $N = 1$ and once there are no more bad points, the final basis and EIM nodes will be located in the run folder $(N - 1)$.

5.3 ROQs for TIGER-BNS

To perform comprehensive tests of GR in O2 and O3, we need to make full use of all detections, which means we also need to be prepared to deal with signals originating from BNS mergers. Even though development is ongoing to apply the full TIGER framework to BBH systems, we should therefore not neglect the BNS implementation of TIGER.

Even though the waveforms in TIGER-BNS are cut off at a maximum frequency of 400 Hz to avoid influences of unknown tidal effects, the full parameter estimation runs still require much computing resources. Regardless of the fact that the ROQ-building project for these inspiral waveforms was already ongoing by the time IMRPhenomPv2 ROQs became available, there is good motivation to continue building the inspiral-only ROQs: IMRPhenomPv2 is an exceptionally accurate waveform for BBH systems, but we can not trust the merger-ringdown part for BNS- or even NSBH-mass systems. Since we need to cut off the waveforms at around 400 Hz anyway, it does not make sense to try and build an ROQ for these systems using the fully precessing 7 dimensional IMRPhenomPv2 model. After all, for BNS systems the small component spins will make precession effects much less important because they are difficult to measure. For TIGER-BNS this is demonstrated in [68] and it should therefore be sufficient to build ROQs for the inspiral-only TaylorF2 waveform.

In this section I will discuss work that I have done on generating ROQs meant for TIGER-BNS applications, using the TaylorF2 waveform model in a frequency range of 30 to 400 Hz. I will demonstrate the feasibility of generating a full 4+1 dimensional basis from scratch using a template bank as training set. Template banks have been used as a training set before to perform an SVD in *e.g.* [137]. I will also propose a method to more quickly build ROQs for any inspiral non-GR parameter based on my work with building the IMRPhenomPv2+ ROQs (see Sec. 5.4) that does not require a template bank.

As explained in the previous section, the initial training set $\check{\mathcal{T}}$ does not necessarily have to represent the entire continuum due to the validation steps that later enrich the basis built with $\check{\mathcal{T}}$. Regardless, finding an optimal training set, or one that is nearly optimal prevents one from having to perform potentially computationally expensive enrichment steps. A dense enough sampling in one parameter might require as much as 100 points. For TIGER-BNS, the parameter space consists of at least 5 parameters; the component masses (m_1, m_2), the component spins in the z-direction (S_{1z}, S_{2z}) and the additional non-GR parameter $\delta\hat{p}_i$, making the full set of parameters $\boldsymbol{\lambda} = \{m_1, m_2, S_{1z}, S_{2z}, \delta\hat{p}_i\}$. A naive dense sampling would then result in $100^5 = 10^{10}$ points. In order for the greedy algorithm to be feasible within a reasonable time, all the waveforms of the training set must be stored in memory. Storing the waveforms in memory is a trade-off between either spending considerable amounts of computing time during the greedy algorithm or requiring a lot of memory. The computing time would not only be determined by othogonalization and projection (which by itself can take a long time in case of large training sets and long waveforms), but most of the time would be spent evaluating all the waveforms in the training set in each iteration of the greedy algorithm (5.1). This quickly becomes problematic, especially in the

case of waveforms that take long to evaluate as is for example the case for low mass BNS sources. The memory storage required in the naive sampling example would be $10^{10} \times L \times 2 \times 8\,\mathrm{B} \approx L \times 160\,\mathrm{TB}$, where L is the number of frequency samples. Clearly, this naive sampling is not the way to do it, but there are ways to decrease memory requirements considerably by creating more intelligent training sets.

The enrichment method allows us to use an initial training set where each parameter dimension may only need to be sampled at around 30 points or even less. One solution to building a compact, yet rich training set comes from the way gravitational waves are detected. As described in section 1.3, some of the detection pipelines use template banks. Rather than evaluating waveforms on random points in the parameter space continuum to perform matched filtering, template banks are set up off-line which are a dense grid covering the entire set of sources expected to be detectable (and for which we have proper waveform models). Such a template bank can be used as a training set for the greedy algorithm, since it has identified the most important parts of the parameter space already, *i.e.* some parts of parameter space are sampled much more densely than others, where the density of template placement is determined by how much two neighbouring templates are alike. A template bank comes with a certain efficiency, *e.g.* an efficiency of 90% corresponds to a maximum mismatch of 10% between any neighboring templates. The template banks used in detection pipelines need to be accurate, and are large at an efficiency of 97% [159]. For building a basis such a dense template bank is probably not necessary and we use instead an efficiency of 90%. As we shall see shortly in section 5.3.1, the resulting template bank may still be rather large, in the order 10^6 points.

The memory estimate that I gave before also scales with the number of frequency bins L in the waveforms. The number of frequency bins, using a constant Δf and a length of 32 seconds, needs to be at least $128 \times (512 - 30) + 1 = 61697$ points when sampled at the Nyquist rate. Taking into account the number of training points, this would amount to a memory requirement of around $2 \times 8 \times 61697 \times 10^6 \approx 987\,\mathrm{GB}$, which is still rather large. Thankfully, there is a way to considerably reduce the number of frequency bins. To this end, we apply so-called multibanding, or adaptive sampling. The idea is that essentially the waveform is severely over-sampled at high frequencies when choosing a constant $\Delta f = 1/T$. This over-sampling can be remedied by dividing the waveform into multiple frequency bands with corresponding Δf, ensuring that each band is appropriately sampled (at least the Nyquist rate). The multiband approach is rather straight-forward for inspiral-only signals since the frequency is monotonically increasing:

1. Split the frequency range into a set of bands *e.g.* $[20, 32]\,\mathrm{Hz}$, $[32, 64]\,\mathrm{Hz}$, $[64, 128]\,\mathrm{Hz}$, $[128, 256]\,\mathrm{Hz}$, $[256, 512]\,\mathrm{Hz}$ and $[512, 1024]\,\mathrm{Hz}$.

2. For each band, calculate the time the waveform spends within this frequency range, *e.g.* $T_{[20,32]} \approx 213\,\mathrm{s}$

3. Calculate Δf for this band as the inverse of the next-power-of-two T: $\Delta f_{[20,32]} = 1/256\,\mathrm{Hz}$

4. Repeat steps 2 and 3 for all bands to construct the adaptive frequency samples f_i and weights Δf_i

The band-time $T_{[f_0,f_1]}$ is calculated using an aligned spin EOB function from LALSUITE called `XLALSimInspiralChirpTimeBound`. This function calculates the time until coalescence from a starting frequency $f_{\rm start}$, taking into account the mass, mass ratio and two spin components S_{1z} and S_{2z}. The band-time $T_{[f_0,f_1]}$ is then $T_{[f_0,f_c]} - T_{[f_1,f_c]}$, with f_c the frequency at coalescence. To make sure we have sufficient frequency samples for all waveforms in $\check{\mathcal{T}}$, the frequency samples and weights are calculated for the longest waveform in the set. In this case the longest waveform should be the one with the lowest chirp mass $\mathcal{M}_c = 0.87\,\mathrm{M}_\odot$, equal masses $\eta = 0.25$ corresponding to $m_1 = m_2 = 1.0\,\mathrm{M}_\odot$, and maximum aligned spins, which I took to be 0.99 even though the basis will be built for spins up to 0.5. Table 5.1 lists the frequency bands chosen and the corresponding band-times $T_{[f_0,f_1]}$ and frequency weights Δf_i. The resulting number of frequency samples is now $L = 4001$ as opposed to $(512-30)\times 128 + 1 = 61697$ when we use constant Δf. Rounding the band-times up to the next-power-of-two ensures that the sampling in that band is at or above the Nyquist rate. There are two things to notice when looking at the numbers: 1) The power-of-two steps I used in selecting the bands are just a guideline. In one case I chose to split the band at 50 Hz instead of 56 Hz, because it decreased the final number of points by quite a bit: $5729 \to 4001$. This is because $T_{[32,56]} \approx 64.01\,\mathrm{s}$, only just above 64 s, resulting in $\Delta f = 1/128\,\mathrm{Hz}$ for that band. 2) In the last band $T_{[256,512]} < 1\,\mathrm{s}$, the next-power-of-two will then always be one, which results in a $\Delta f = 1\,\mathrm{Hz}$ that is too large in most cases. Here I choose to round the band-time to the inverse of the next-power-of-two; 1/4 s in this case.

There is a small caveat in using this multiband approach for TAYLORF2 when including non-GR parameters. The way we construct a multiband as described above assumes we know how long the longest waveform in the set will be. The fact is that we cannot calculate this with `XLALSimInspiralChirpTimeBound`, because the underlying waveform model for this function (spin-aligned EOB) does not have non-GR parameters. The longest waveform when taking $\delta\hat{p}_i$ into account may be slightly longer than predicted by the GR waveform. A quick workaround to take into account this uncertainty, is to implement an augmentation factor. In practice, for TAYLORF2, I increase the band-times calculated with `XLALSimInspiralChirpTimeBound` by 10%. This augmentation factor increases the number of samples a bit ($4001 \to 4353$), but ensures a dense enough frequency sampling for all waveforms in the set, assuming the effect of $\delta\hat{p}_i$ on the waveform length is no larger than 10%.

Using multiband sampling, and assuming we can manage to reduce the training set to about 10^6 points, we require roughly $2 \times 8 \times 10^6 \times 4000 = 64\,\mathrm{GB}$, which is reasonable. In the next section I will demonstrate how a near optimal training set of around 10^6 points can be achieved using the template bank approach.

5.3.1 Generating a specialized training set using template banks

As mentioned before, generating a training set can in principle be trivial. Take for example a waveform that is only dependent on component masses m_1 and m_2. It is possible to simply sample the $m_1 - m_2$ plane with an $N \times N$ grid. Make N large enough and one will not miss any information. The question is of course how large

f-range (Hz)	band-time $T_{[f_0,f_1]}$	T^{NextPow2}	$\Delta f = 1/T^{\mathrm{NextPow2}}$
$30 \leq f < 32$	15.39 s	16 s	0.062 500 Hz
$32 \leq f < 50$	57.39 s	64 s	0.015 625 Hz
$50 \leq f < 128$	23.59 s	32 s	0.031 250 Hz
$128 \leq f < 256$	1.88 s	2 s	0.500 000 Hz
$256 \leq f < 512$	0.32 s	1/4 s	4.000 000 Hz

Table 5.1: The frequency bands chosen to sample the TaylorF2 waveforms in $\check{\mathcal{T}}$. The total length of the frequency vector using these bands is 4001 (4353 when using an augmentation factor of 10%). Note that the last band has a band-time of less than of less than a second. In this case I took the next-power-of-two length to be the inverse of a power of two.

should N be? Instead of assuming every part of parameter space is equally important (as is done in this naive method), it is fruitful to think about what kind of effect a change in parameters has on the waveform model. For example, a small change in component masses has a more substantial impact on the waveform in one part of the parameter space than others. This depends greatly on the chirp mass of the system. So instead of using the naive grid, we should try and use a more dense sampling in the regions of parameter space that matter most and thus minimize the number of points while maximizing the coverage of the parameter space. Finding this sampling has been the area of extensive studies in the field of template bank generation.

In [160] a stochastic method for template bank generation was presented. The general idea is that the density of waveforms in any part of the parameter space considered is determined by ensuring that the mismatch between any two neighbouring points is no less that a certain value. This mismatch is related to what is called the distance between two points, which in turn can be related to a metric. The nature of the metric used in the template placement algorithm is explained in detail in [161] for aligned spin BNS sources. The only difference for the current situation is that we will introduce at least one shift $\delta\hat{\varphi}_n$ in one of the PN phasing parameters that we wish to treat as a free parameter. I have adapted the PyCBC code [162] that follows this method called pycbc_aligned_stoch_bank.

The complete parameter space will consist of the component masses, the spins along the z-direction and some set of additional non-GR parameters $\Phi = \{\delta\hat{\varphi}_n\}$ (see Sec. 1.6), *i.e.* $\boldsymbol{\theta} = \{m_1, m_2, S_{1z}, S_{2z}, \Phi\}$. In pycbc_aligned_stoch_bank, one uses TaylorF2 accurate up to 3.5PN order in orbital phase, with leading order spin-orbit and spin-spin terms. The phase $\Psi(\bar{x})$ is given by

$$\Psi(\bar{x}) = 2\pi v_0^3 \bar{x}^3 t_c - \phi_c + \bar{x}^{-5} \sum_{n=0}^{7} \left(\bar{\lambda}_n + 3\bar{\lambda}_n^{(l)} \ln \bar{x}\right) \bar{x}^n, \tag{5.34}$$

where the frequency f used in [161] is converted to orbital velocity $v = (\pi M f)^{-1/3}$. To write everything in a consistent manner I defined the fiducial orbital velocity $v_0 \equiv (\pi M f_0)^{-1/3}$, where f_0 is the fiducial frequency as used in [161] and the parameter

$\bar{x} \equiv v/v_0$. The parameters t_c and ϕ_c are the usual coalescence time and constant phase offset respectively. The $\bar{\lambda}_n$ can be expressed in terms of the familiar TaylorF2 phasing terms presented in Sec. 1.4.2-B, i.e.

$$\begin{aligned} \bar{\lambda}_n &= \psi_n \, v_0^{n-5} \\ \bar{\lambda}_n^{(l)} &= \psi_n^{(l)} \, v_0^{n-5}. \end{aligned} \tag{5.35}$$

To incorporate possible deviations from GR as discussed in Sec. 1.6, the phasing parameters $\bar{\lambda}_n$ undergo the same relative shift i.e.

$$\begin{aligned} \bar{\lambda}_n &\to \begin{cases} \psi_n(1+\delta\hat{\varphi}_n)\, v_0^{n-5} & \text{if} \quad n \neq 1 \\ \delta\hat{\varphi}_1 \, v_0^{-4} & \text{if} \quad n = 1 \end{cases} \\ \bar{\lambda}_n^{(l)} &\to \psi_n^{(l)}(1+\delta\hat{\varphi}_n^{(l)})\, v_0^{n-5}. \end{aligned} \tag{5.36}$$

The goal is to construct a template bank with as few waveforms as possible for which any plausible BNS signal has a certain minimum required overlap. The higher the demand on the minimum overlap, the larger the template bank will be. The distance between two infinitesimally separated normalized waveforms on the parameter space is given by

$$\begin{aligned} \|h(\boldsymbol{\theta}+\delta\boldsymbol{\theta}) - h(\boldsymbol{\theta})\|^2 &= \|h(\boldsymbol{\theta}) + \frac{\partial h(\boldsymbol{\theta})}{\partial\theta^\mu}\delta\theta^\mu + \cdots - h(\boldsymbol{\theta})\|^2 \\ &\simeq \|\frac{\partial h(\boldsymbol{\theta})}{\partial\theta^\mu}\delta\theta^\mu\|^2 \\ &= \left(\frac{\partial h(\boldsymbol{\theta})}{\partial\theta^\mu}\Big|\frac{\partial h(\boldsymbol{\theta})}{\partial\theta^\nu}\right)\delta\theta^\mu\delta\theta^\nu \\ &= g_{\mu\nu}\delta\theta^\mu\delta\theta^\nu, \end{aligned} \tag{5.37}$$

where $\boldsymbol{\theta}$ contains the parameters of the signal, in this case the masses, spins and non-GR parameters and Einstein's summation convention is used (summation over repeated indices). The overlap $\mathcal{O}(\boldsymbol{\theta}, \boldsymbol{\theta}+\delta\boldsymbol{\theta})$ between two infinitesimally separated vectors $h(\boldsymbol{\theta})$ and $h(\boldsymbol{\theta}+\delta\boldsymbol{\theta})$ on the waveform parameter space is given by

$$\mathcal{O}(\boldsymbol{\theta}, \boldsymbol{\theta}+\delta\boldsymbol{\theta}) \equiv (h(\boldsymbol{\theta})|h(\boldsymbol{\theta}+\delta\boldsymbol{\theta}))\,. \tag{5.38}$$

The waveforms in the template bank are normalized so $\mathcal{O}(\boldsymbol{\theta}, \boldsymbol{\theta}) = 1$. This means that for vectors that are infinitesimally close to each other, the overlap can be expanded about its maximum value of 1:

$$\mathcal{O}(\boldsymbol{\theta}, \boldsymbol{\theta}+\delta\boldsymbol{\theta}) = 1 + \frac{1}{2}\frac{\partial^2\mathcal{O}}{\partial\theta^\mu\partial\theta^\nu}\delta\theta^\mu\delta\theta^\nu + \ldots. \tag{5.39}$$

The *mismatch* between the two vectors is $1 - \mathcal{O} = g_{\mu\nu}\delta\theta^\mu\delta\theta^\nu$, so we can write

$$\mathcal{O}(\boldsymbol{\theta}, \boldsymbol{\theta}+\delta\boldsymbol{\theta}) \simeq 1 + \frac{1}{2}\frac{\partial^2\mathcal{O}}{\partial\theta^\mu\partial\theta^\nu}\delta\theta^\mu\delta\theta^\nu = 1 - g_{\mu\nu}\delta\theta^\mu\delta\theta^\nu, \tag{5.40}$$

which means that the metric on this parameter space may also be written as

$$g_{\mu\nu}(\boldsymbol{\theta}) = -\frac{1}{2}\frac{\partial^2\mathcal{O}}{\partial\theta^\mu\partial\theta^\nu}, \tag{5.41}$$

which is the metric used in [161] and [163]. To construct the template bank, the $\bar{\lambda}_n$ and $\bar{\lambda}_n^{(l)}$ are treated as eight independent parameters. Following [161] and [163] the metric in Eq. (5.41) can be evaluated when maximized over ϕ_c and the resulting metric can be written in terms of a 9 dimensional space. The parameter t_c and the phasing parameters $\bar{\lambda}_n$ and $\bar{\lambda}_n^{(l)}$ are the coordinates in the 9 dimensional space. Finally, when an upper and lower frequency cut-off is determined, the metric can be maximized over t_c as well, leaving us with a metric that can be expressed in terms of only $\bar{\lambda}_n$ and $\bar{\lambda}_n^{(l)}$. For our particular goal it is interesting to note that the parameter space metric, in the $\bar{\lambda}_n$ coordinate system, has no dependence on the values of $\bar{\lambda}_n$ themselves. This means that the additional $\delta\hat{\varphi}_n$ parameters do not require the construction of a different metric. The PYCBC_ALIGNED_STOCH_BANK algorithm presented in [160] can therefore be trivially implemented to construct a training set for our purpose.

The initial training set $\check{\mathcal{T}}$ we aim to generate here, depends on 5 parameters; the component masses m_1 and m_2 both ranging between $1.0\,\mathrm{M}_\odot$ and $2.5\,\mathrm{M}_\odot$, the aligned component spins S_{1z} and S_{2z} ranging between ± 0.5 and one additional non-GR parameter $\delta\hat{\varphi}_3$ in the 1.5PN coefficient, ranging between ± 1.5. When constructing $\check{\mathcal{T}}$, I wanted a reasonably accurate template bank and set the efficiency to 90%. The frequency range was set such that it would extend beyond what is needed for a TIGER-BNS ROQ: $15\,\mathrm{Hz} \leq f \leq 500\,\mathrm{Hz}$. The final template bank has a size $N = 1,069,048$ and is plotted for some parameters in Fig. 5.4. Note that there are clearly parts in parameter space that are more densely populated, indicating a stronger dependence on these parameters in those regions. For example, $\delta\hat{\varphi}_3$ has a stronger impact on the waveform for lower $\mathcal{M}_c$ and systems that are closer to equal mass. The bottom row in Fig. 5.4 shows there is only a weak dependence of the impact of $\delta\hat{\varphi}_3$ on the component spins, but it is certainly not a homogeneous distribution, *e.g.* for strong positive shifts in the 1.5PN coefficient combined with strong negative component spins, the parameter space is more densely sampled.

In the next section this template bank is used as the initial training set $\check{\mathcal{T}}$ to build the ROQs.

5.3.2 TaylorF2 ROQ

Section 5.2.3 shows what is needed to compute the ROQs for the likelihood given in Eq. (5.25). The goal is to construct the linear and quadratic bases $\mathcal{B}^{\mathrm{L}}$ and $\mathcal{B}^{\mathrm{Q}}$, with corresponding frequency nodes $\mathcal{F}^{\mathrm{L}}$ and $\mathcal{F}^{\mathrm{Q}}$. For the linear part we follow the process outlined in Fig. 5.3 and algorithm (5.3). The quadratic term, thanks to the relatively simple nature of TaylorF2, we can evaluate analytically. Recall that the quadratic basis is built for any $h_A(\boldsymbol{\lambda}; f)h_B^*(\boldsymbol{\lambda}; f)$ and should work for all products of h_A and h_B, where A and B can be $+$ or $\times$. For TaylorF2, this follows analytically. We start by defining $H(\boldsymbol{\lambda}; f) \equiv h_A(\boldsymbol{\lambda}; f)h_B^*(\boldsymbol{\lambda}; f)$ and write down the interpolant as in Eq. (5.17)

$$\mathcal{I}_n[H](\boldsymbol{\lambda}; f) \equiv \sum_{j=1}^{n} \mathcal{B}_j^{\mathrm{Q}}(f) H(\boldsymbol{\lambda}; \mathcal{F}_j^{\mathrm{Q}}). \tag{5.42}$$

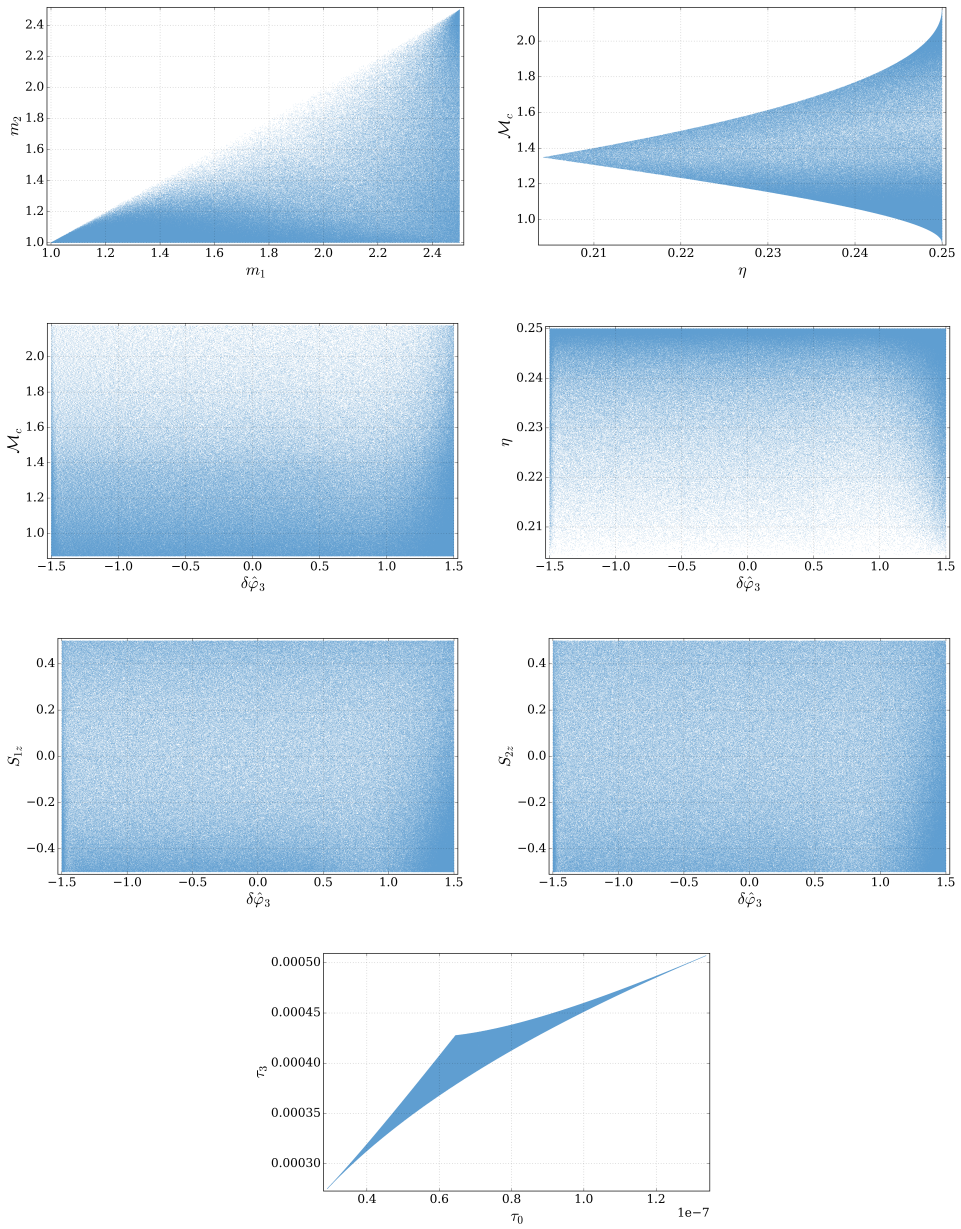


Figure 5.4: 2D sub-selections of the 5D template bank with a size of 1,069,048 points. The first row shows the parameter distribution in the mass parameters $m_1, m_2, \mathcal{M}_c$ and η. The second and third row shows the distribution of the non-GR parameter $\delta\hat{\varphi}_3$ as a function of the other 4 parameters.

We then write down the quadratic term as a discrete sum over frequencies

$$\begin{aligned}\langle h_A(\boldsymbol{\lambda};f)|h_B(\boldsymbol{\lambda};f)\rangle &= \Re\Delta f\sum_{i=1}^{L} h_A(f_{\min}+i\Delta f)h_B^*(f_{\min}+i\Delta f)\\ &= \Re\Delta f\sum_{i=1}^{L} H(\boldsymbol{\lambda};f_i).\end{aligned} \tag{5.43}$$

Here we observed that the product $h_A(\boldsymbol{\lambda};f_i)h_B^*(\boldsymbol{\lambda};f_i)$ is what we have defined as $H(\boldsymbol{\lambda};f_i)$. If we then substitute the interpolant Eq. (5.42) in its place, we get the quadratic part in terms of the basis $\mathcal{B}^{\mathrm{Q}}$ and EIM nodes $\mathcal{F}$:

$$\langle h_A(\boldsymbol{\lambda})|h_B(\boldsymbol{\lambda})\rangle = \Re\Delta f\sum_{i=1}^{L}\sum_{j=1}^{n}\mathcal{B}_j^{\mathrm{Q}}(f_i)H(\boldsymbol{\lambda};\mathcal{F}_j^{\mathrm{Q}}). \tag{5.44}$$

A polarization of the TaylorF2 waveform can be written as $h_A(\boldsymbol{\lambda};f) = \mathcal{A}_A f^{-7/6}e^{\mathrm{i}\Psi(\boldsymbol{\lambda};f)}$ so that $H(\boldsymbol{\lambda};f) = \mathcal{A}_A\mathcal{A}_B f^{-14/6}$. We can now substitute this expression into both 5.43 and 5.44 and solve for values for $\mathcal{F}_j^{\mathrm{Q}}$ and $\mathcal{B}_j^{\mathrm{Q}}$:

$$\Re\Delta f\sum_{i=1}^{L}H(\boldsymbol{\lambda};f_i) = \Re\Delta f\sum_{i=1}^{L}\sum_{j=1}^{n}\mathcal{B}_j^{\mathrm{Q}}(f_i)H(\boldsymbol{\lambda};\mathcal{F}_j^{\mathrm{Q}})$$

$$\Re\Delta f\mathcal{A}_A\mathcal{A}_B f_i^{-14/6} = \Re\Delta f\sum_{j=1}^{n}\mathcal{B}_j^{\mathrm{Q}}(f_i)\mathcal{A}_A\mathcal{A}_B(\mathcal{F}_j^{\mathrm{Q}})^{-14/6} \tag{5.45}$$

$$f_i^{-14/6} = \sum_{j=1}^{n}\mathcal{B}_j^{\mathrm{Q}}(f_i)(\mathcal{F}_j^{\mathrm{Q}})^{-14/6}. \tag{5.46}$$

We can solve this for $n=1$ when we set

$$\mathcal{B}_1^{\mathrm{Q}}(f) = \left(\frac{f}{\mathcal{F}_1^{\mathrm{Q}}}\right)^{-14/6}, \tag{5.47}$$

where $\mathcal{F}_1^{\mathrm{Q}}$ is allowed to be anything, but for practical reasons I use $f_{\min}$. We can substitute this into the definition of the ROQ (Eqns. (5.25) and (5.26)) and find

$$\langle h_A(\boldsymbol{\lambda})|h_B(\boldsymbol{\lambda})\rangle_{\mathrm{ROQ}} = w_1 H(F_1;\boldsymbol{\lambda}), \tag{5.48}$$

where we have a single weight

$$w_1 = 4\Delta f\,\Re\sum_{i}^{L}\frac{\mathcal{B}_1^{\mathrm{Q}}(f_i)}{S_n(f_i)}. \tag{5.49}$$

5.3.3 Performance

Constructing the basis for TaylorF2 with one additional non-GR parameter $\delta\hat{\varphi}_3$ required the use of the multiband given in Table 5.1 and the template bank described

in section 5.3.1 and displayed in Fig. 5.3.1 functioning as initial training set $\check{\mathcal{T}}_N$, where $N = 1069048$. The final memory footprint of the first greedy sweep was around 165 GB which was calculated with the memory estimator provided in the GREEDYCPP repository[4]. The first iteration, *i.e.* the generation of $\mathcal{E}^0_{n_0}$, resulted in $n_0 = 1466$ basis points. Enriching the basis took quite a few iterations, which was surprising considering the simplicity of the model. After each iteration a single bad point or at most 4 were found. All of these points lie in a small part of parameter space all the way in the bottom left of the m_1-m_2 plane. After closer inspection of the template bank in the top left panel of Fig. 5.4, we see why this behavior is not entirely unexpected. The region of small masses is densely populated, if we then generate the validation set by sampling *uniformly* in the m_1-m_2 plane, we can expect to miss some points in this corner of parameter space. A better way to generate the validation set for a waveform such as this is to sample in the so-called *chirp times* [164] instead of component masses or η and $\mathcal{M}_c$: The nice feature of the chirp times, defined as

$$\tau_0 = \frac{5}{256\pi f_{\min}}(\pi \mathcal{M}_c f_{\min})^{-5/3}, \quad \tau_3 = \frac{\eta^{-3/5}}{8 f_{\min}}(\pi \mathcal{M}_c f_{\min})^{-2/3}, \tag{5.50}$$

is that points are much more uniformly distributed in this parameterization, which can be seen in the bottom panel of Fig. 5.4. When we sample uniformly on this plane the chance of having to do multiple enrichment steps is much lower.

The enrichment steps added only a few more points resulting in a final basis $\mathcal{E}^3_{n_3}$, where $n_3 = 1468$. Fig. 5.5 shows the overlap errors after each validation step using 10^7 points randomly and drawn from the 5 dimensional parameter space $\{m_1, m_2, S_{1z}, S_{2z}, \delta\hat{\varphi}_3\}$, where $1.0 \leq m_a \leq 2.5$, $-0.5 \leq S_{az} \leq 0.5$ and $-1.5 \leq \delta\hat{\varphi}_3 \leq 1.5$, which are the same used in the template bank and thus $\check{\mathcal{T}}_N$.

After validation is complete, we are left with a reliable basis of size $n_3 = 1468$ from which we can construct the ROQs. However, in this case we need to perform one additional step: up-sampling. The final basis as it is now consists of $\{\mathcal{B}^{\text{L}}(f)\}_{i=0}^{n_3}$ and $\mathcal{B}^{\text{Q}}_1(f)$, where f is the multibanded frequency series we constructed earlier in section 5.3 (see Table 5.1). In practice, when computing the ROQ weights in Eqns. (5.31b) and (5.32b), the basis needs to be sampled in the same way as the detector output. The detector output in LALINFERENCE is always sampled at a constant Δf at the Nyquist rate: $L = (f_{\max} - f_{\min})T + 1$, where $T = 128\,\text{s}$ in this case. To achieve this, we need to up-sample the basis. Since we sampled the basis in such a way that no information is lost, we can up-sample the basis by direct evaluation of the waveform at the n_3 greedy points and re-orthogonalize. In other words, we perform a new greedy sweep, but in this case setting the training set $\check{\mathcal{T}}_N$ to the greedy points of the final enriched set $\mathcal{G}^3_{n_3}$. In general, this up-sampling technique can not be trusted blindly and one should perform another validation run. Recall that in order to incorporate the unknown effect of $\delta\hat{\varphi}_i$ on the length of the longest waveform in the set we introduced an augmentation factor. This means that if we underestimated the effect of $\delta\hat{\varphi}_i$, the final up-sampled set might result in approximation errors that are larger than the ones we have seen in the multiband case. The purpose of the work presented in this thesis

[4]The actual size in memory also depends on the a priori estimated maximum size of the reduced basis and the number of parameters.

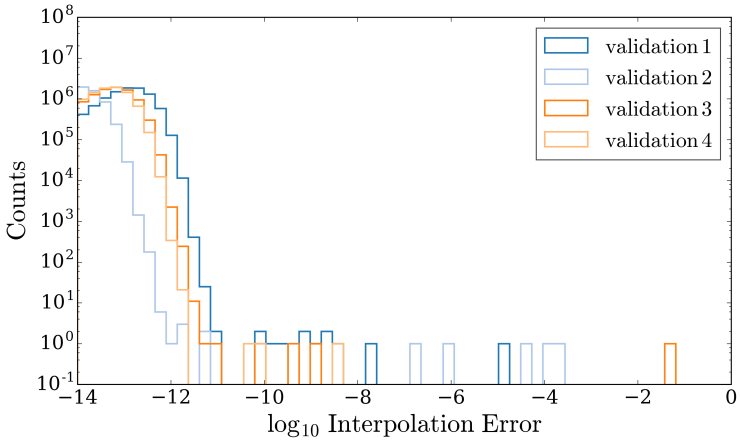


Figure 5.5: The interpolation errors for all points in Λ_N^i, where $N = 10^7$ and $i = 1, 2, 3, 4$. After 4 validation runs no error exceeds the threshold $\beta = 10^{-6}$ and the basis is at that point finished.

is mostly a proof of principle and I have left either the validation of the up-sampled basis or better determination of the longest waveform in the set to future work.

Note that the final up-sampled basis is actually a bit smaller with $n_f = 1467$ points than the one generated using adaptive sampling, which has $n_3 = 1468$ points. This is to be expected since the basis itself was generated for a frequency range up to $f_{\max} = 512\,\mathrm{Hz}$, while the final up-sampled basis has the usual TIGER-BNS cut-off of $f_{\max} = 400\,\mathrm{Hz}$ and thus less complexity in the model. For constructing the initial basis I deliberately chose a frequency that is a power of two to leave some room in case we would prefer to increase the cut-off frequency. Furthermore, preliminary tests with the LALInference ROQ implementation show that only powers of two in $f_{\max}$ are compatible with the software.

Fig. 5.6 shows the final up-sampled greedy points plotted over the original template bank for a few choices of parameters. Most of the points are clustered in regions that were also more densely represented in the original template bank. What is not seen in the template bank is the horizontal feature in the $\delta\hat{\varphi}_3$-$\mathcal{M}_c$ plot (middle-left panel) around $\mathcal{M}_c = 1.3\,\mathrm{M}_\odot$ for $\delta\hat{\varphi}_3 > 0.5$. Note that the chirp mass determines the leading-order behavior of inspiral amplitude and phase, so that a periodic feature related to this parameter is not surprising.

After being satisfied with the accuracy of the basis that has been demonstrated with validation runs, one should apply it to a full parameter estimation run to confirm there are no significant differences in the resulting parameter posteriors. I will however leave this final test for future work as it is at this time computationally too demanding: The length of the longest signals that need to be generated with the current mass priors is $\sim 98.5\,\mathrm{sec}$. In the near future we will probably be able to perform such analyses using a new adaptive sampling feature in LALInference that is currently

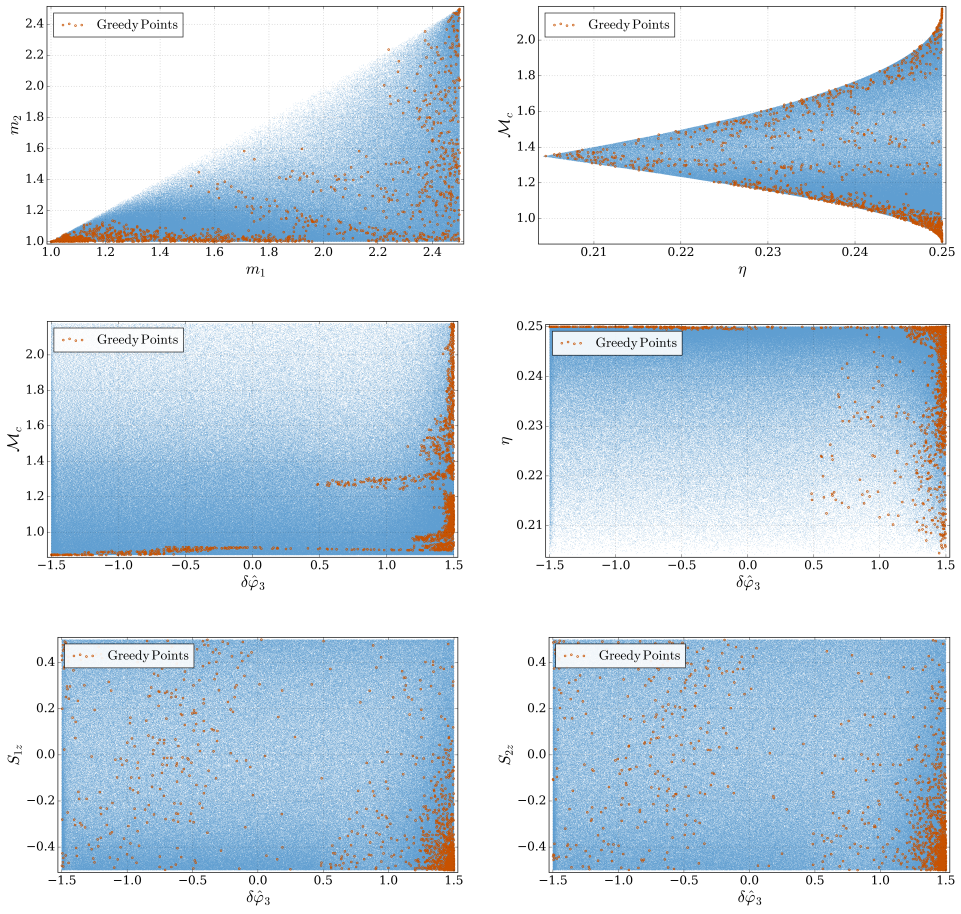


Figure 5.6: Final greedy points for the up-sampled TaylorF2 basis with a size $n_f = 1467$. The greedy points are represented as red circles. The blue dots are the original template bank with a size of 1,069,048 points. The first row shows the parameter distribution in the mass parameters $m_1, m_2, \mathcal{M}_c$ and η. The second and third rows show the distribution of the non-GR parameter $\delta\hat{\varphi}_3$ as a function of $\mathcal{M}_c, \eta, S_{1z}$ and S_{2z}.

in development [141].

For TIGER-BNS production runs we might wish to go down to a lower frequency $f_{\rm min} = 20\,{\rm Hz}$. To get an idea of the kind of theoretical speedup we may achieve in that case, I generated one final basis that goes all the way down to $f_{\rm min} = 20\,{\rm Hz}$. One must realize that in that case the waveforms become much longer: For a source of component masses $m_1 = m_2 = 1.0\,{\rm M}_\odot$ and spin magnitudes of 0.5, the band-time $T_{[20,400]} \sim 286\,{\rm s}$ for which the next power of two is 512 s which is used for calculating T. The final size of the basis (after up-sampling) is 2907, giving a theoretical speedup of $T/(N_{\rm L} + N_{\rm Q}) = 194561/2908 \approx 67$. This is a factor of 2 more than the theoretical speedup when starting at 30 Hz ($T/(N_{\rm L} + N_{\rm Q}) = 47361/1468 \approx 32$) even though the length L of the longest waveform has increased by approximately a factor of 4. The fact that the basis size does not grow as fast as the length L has been noticed in previous ROQ studies [143] and we will see this again shortly for the BBH ROQs.

5.4 ROQs for BBH and NSBH

Building ROQs and thus a basis for BBH sources introduces entirely different challenges than the previously discussed aligned-spin inspiral-only TaylorF2. BBH or NSBH sources have a wide range in masses and can show considerable precession due to strong non-aligned component spins. The models currently used for parameter estimation on BBH or NSBH sources are SEOBNRv2/v3 and SEOBNRv2_ROM, but for the parameterized test of GR these models are less suitable and we instead use the frequency domain waveform IMRPhenomPv2, which is in excellent agreement with numerical simulations and SEOBNRv3 across a wide range in parameter space. The main challenge, aside from the fact that we are now dealing with a much more complex model than before (TaylorF2), is that due to the random spins we are now dealing with a 7D parameter space instead of 5D. The challenge in this might not immediately seem obvious, but recall how the size of a training set typically scales. If we initially required on the order of 64 GB to store the training set in memory, we would now need $64 \times n_1 \times n_2$, where n_i is however many samples one needs for each additional parameter. Even with a single additional parameter, the memory requirements quickly increase by an order of magnitude. The 7D case is however only the GR version of IMRPhenomPv2. We will need ROQs for IMRPhenomPv2+, which has at least one additional parameter, increasing the parameter space to at least an 8 dimensional one. It is clear that this is not a trivial task. Thankfully, only recently, bases have been generated for IMRPhenomPv2 [143] and in this chapter I will demonstrate how we can adapt these to construct a basis for at least one additional non-GR parameter, *i.e.* a basis for IMRPhenomPv2+.

5.4.1 IMRPhenomPv2 ROQ

Since the method introduced here to build a basis for IMRPhenomPv2+ greatly depends on the already available bases for IMRPhenomPv2, I will first present the latter. The ROQs for IMRPhenomPv2 that have recently been developed by Smith *et al.* [143] are divided into bins of overlapping chirp mass. These overlapping regions

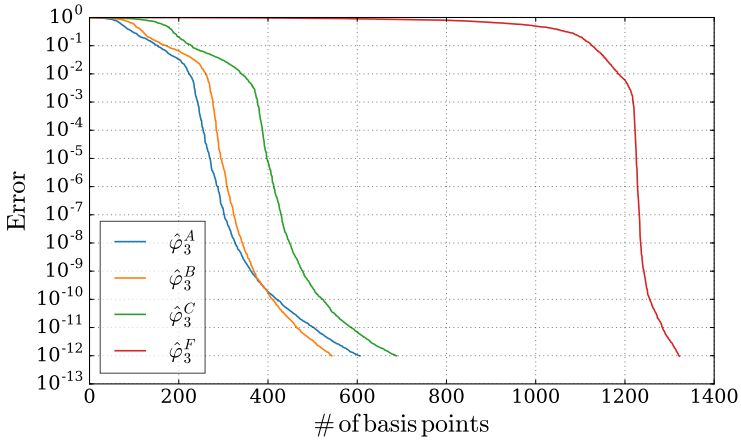


Figure 5.7: Exponential decay of the greedy error as more basis functions are added. The greedy algorithm is set to terminate when the error reaches the threshold $\epsilon \leq 10^{-12}$.

are defined by

$$\mathcal{M}_c(T = \tfrac{3}{8}2^n\,\mathrm{s}) \leq \mathcal{M}_c \leq 1.2\mathcal{M}_c(T = 2^n\,\mathrm{s}), \tag{5.51}$$

where T is the waveform duration from 20 Hz. The function $\mathcal{M}_c(T)$ is an interpolation to obtain the waveform's chirp mass as a function of duration. In Eq. (5.51), the following powers of 2 are considered: $n = 2, 3, 4, 5, 6, 7$. These choices correspond to regions in $\mathcal{M}_c$-space describing signals with durations $1.5\,\mathrm{s} \leq T \leq 4\,\mathrm{s}$, $3\,\mathrm{s} \leq T \leq 8\,\mathrm{s}$, $6\,\mathrm{s} \leq T \leq 16\,\mathrm{s}$, $12\,\mathrm{s} \leq T \leq 32\,\mathrm{s}$, $24\,\mathrm{s} \leq T \leq 64\,\mathrm{s}$ and $48\,\mathrm{s} \leq T \leq 128\,\mathrm{s}$. So the complete set covers binary systems with waveform durations between 1.5 s and 128 s starting from 20 Hz and masses ranging from heavy BNS systems to BBH systems with $\mathcal{M}_c \leq 23\,\mathrm{M}_\odot$. Table 5.2 lists the 6 cases that were covered in [143]. The upper frequencies in Table 5.2 correspond to what Smith *et al.* call the maximum-over-configuration ringdown frequency, rounded to the next-highest-power-of-two. In all cases the ROQs are valid in the mass ratio interval $1 \leq q \leq 9$, which is within IMRPhenomPv2's calibration range [32]. Such a wide range in q allows the model to represent BNS, BBH and NSBH systems, depending on the total mass. The component masses are restricted to be $m_1 \geq m_2 \geq 1\,\mathrm{M}_\odot$. The ranges in magnitudes of the spin-related parameters are $(-0.9, -0.9, 0) \leq (\chi_{1L}, \chi_{2L}, \chi_p) \leq (0.9, 0.9, 0.9)$. Finally, the full range in spin angles is used: $(0, 0) \leq (\theta_J, \alpha_0) \leq (\pi, 2\pi)$.

The work done in [143] is essential to realizing an ROQ for IMRPhenomPv2+. As mentioned before, a training set quickly scales in size when adding only a single parameter to the model. Consider for example that we want to create a non-GR ROQ for case B. If one would start from scratch and sample the extra non-GR parameter with only 8 points, it means generating a training set of $(32^2) \times (8^5) \times 8$ waveforms, based on [143], where they used 32 points for η and $\mathcal{M}_c$. With a sampling of $\Delta f = 1/8$ Hz and a bandwidth of 1024 Hz, storing this training set would already require as

Case	$f_{\max}$ (Hz)	Signal duration (s)	$\mathcal{M}_c\,(\mathrm{M}_\odot)$	N^L	N^Q	Speedup
A	1024	$1.5 \leq T \leq 4$	$12.3 \leq \mathcal{M}_c \leq 23$	300	197	4
B	1024	$3 \leq T \leq 8$	$7.9 \leq \mathcal{M}_c \leq 14.8$	388	278	12
C	2048	$6 \leq T \leq 16$	$5.2 \leq \mathcal{M}_c \leq 9.5$	360	233	54
D	2048	$12 \leq T \leq 32$	$3.4 \leq \mathcal{M}_c \leq 6.2$	524	254	83
E	2048	$24 \leq T \leq 64$	$2.2 \leq \mathcal{M}_c \leq 4.2$	749	270	127
F	4096	$48 \leq T \leq 128$	$1.4 \leq \mathcal{M}_c \leq 2.6$	1253	487	300

Table 5.2: Details of the original IMRPhenomPv2 bases as presented in [143]. All waveforms start at a lower frequency $f_{\min} = 20\,\mathrm{Hz}$ and in all cases $\Delta f = 1/\max(T)$.

much as 40 TB of memory. Even if we were to use adaptive sampling to shorten the training set waveforms, we would still require at least 20 TB. Note that this example considers one of the shortest waveforms in Table 5.2, meaning that covering the whole range of chirp masses would require increasingly vast amounts of memory. Instead of starting from scratch, I propose to use the bases built by Smith *et al.* and extend those with an additional non-GR parameter to arrive at a new training set. Fig. 5.8 sketches out the concept of extending the existing basis with an additional parameter. In doing so, we have room to sample the additional non-GR parameter more densely, putting less strain on the enrichment part of building the basis. If we again consider case B and sample the non-GR parameter densely with 500 points, we only require about 20 GB, which is reasonable. Note that there is no guarantee that this approach will work. It could well be that the correlation between the non-GR parameter and any other intrinsic parameter is too strong. For example, it could be that some non-GR parameter $\delta\hat{p}_i$ is strongly correlated with the chirp mass. Since we have only a limited number of points in $\mathcal{M}_c$ taken from the IMRPhenomPv2 basis, this would mean that we under-sample the chirp mass dimension for certain values of $\delta\hat{p}_i$. This could be an issue when basing the training set on the IMRPhenomPv2 basis alone, instead of using a denser sampling in all dimensions. In principle the enrichment steps would record any points in parameter space that were not accurately represented and add them to the basis. There is a reason however to put some thought into constructing an accurate training set: If there is a large under-sampling in the training set for a certain parameter, the resulting set of bad points could be so large that we run into memory issues again. The goal of this work is to demonstrate the feasibility of the basis extension approach.

5.4.2 IMRPhenomPv2+ ROQ

Building the IMRPhenomPv2+ ROQ follows the same process as described in section 5.2.3 (see Fig. 5.3 and algorithm (5.3)). To generate the training set $\check{\mathcal{T}}$, the basis extension approach explained in the previous section is employed; given an original linear (Eq. (5.28)) GR basis $\mathcal{B}^{\mathrm{L}}_{\mathrm{GR}}$ with size N_{L} and a set of additional non-GR parameters $\{\delta\hat{p}_i\}$, extend each original parameter dimension with $N_{\delta\hat{p}_i}$ samples in the $\delta\hat{p}_i$

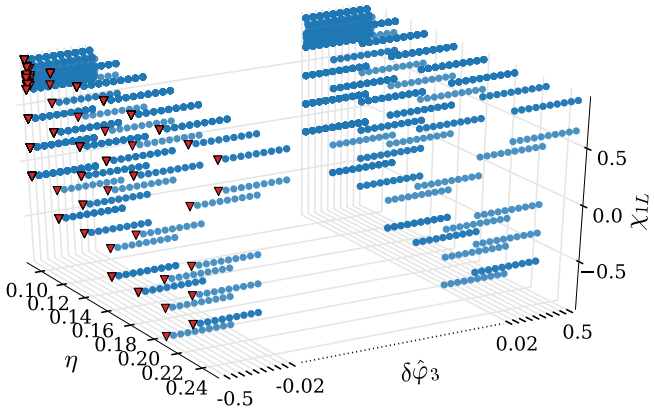


Figure 5.8: Illustration of how the original basis (triangles) is extended in an additional parameter dimension ($\delta\hat{\varphi}_3$ in this example) to form a new training set. The illustration only considers a three dimensional slice of the full parameter space, but the concept illustrated in this figure is extensible to any selection of two IMRPhenomPv2 parameters. Note that points with $-0.02 \leq \delta\hat{\varphi}_3 \leq 0.02$ are not shown to aid the visualization.

direction. An example is sketched out in Fig. 5.8 for a 2D slice from the full IMRPhenomPv2 parameter space. If the GR basis consists of 2 parameters m_1 and m_2, each with N_m points and we wish to extend it with one additional non-GR parameter $\delta\hat{p}_i$ at $N_{\delta\hat{p}_i}$ samples, we will have $N_{\delta\hat{p}_i}$ for each value of (m_1, m_2), resulting in a training set of $N_m^2 \times N_{\delta\hat{p}_i}$ points. For generating the bases for IMRPhenomPv2+, it appeared sufficient to use a uniform sampling in the extra non-GR dimension, but it might even be possible to capture a non-trivial dependence on other parameters. For example, instead of sampling uniformly in $\delta\hat{p}_i$, one could use a more dense sampling for higher values of $\delta\hat{p}_i$. This method will not capture any particular functional dependence on any one of the IMRPhenomPv2 parameters, but it might capture the expectation that stronger violations of GR would result in a more significant effect on the general shape of the waveform, requiring denser sampling in the training set.

This approach might be useful in the future when we wish to build bases for wider ranges in the non-GR parameters, or when we wish to include more than one non-GR parameter, so I will demonstrate the concept here. Even though enrichment steps can adjust an initial basis with missed points, the goal is to set up a training set that is as densely sampled as possible in regions where one expects faster changes in the waveform as a function of the considered parameters. In particular we wish to avoid ending up with too many bad points after a validation run, since that could once again result in exceeding the memory resources. For non-GR waveforms, one way to probe the effect a parameter $\delta\hat{p}_i$ has on the waveform as a function of intrinsic GR parameters, is to plot $\max_i |\Delta\Psi(f_i)|$, where $\Delta\Psi(f_i) = \Psi_{\mathrm{GR}}(f_i) - \Psi_{\mathrm{non-GR}}(f_i)$ is

the maximum phase difference between the GR and non-GR waveform. The reason I choose to investigate the phase only is because this is what we are most sensitive to. Of course this phase difference is highly dependent on all the parameters that go into the model, not just $\delta\hat{p}_i$, so what I have plotted in the left panel of Fig. 5.9, are the medians and 95% confidence intervals of $\max_i |\Delta\Psi(f_i)|$ calculated for 1000 random waveforms evaluations. The histogram in the background shows how the parameter would be sampled for the training set using a fitted curve as sampling function. Of course the fitted curve is only an indication of a more optimal sampling, but may help us out to more efficiently sample the non-GR parameters for higher dimensional models or wider ranges. Already for one non-GR parameter, namely $\delta\hat{\varphi}_1$, too many bad points after the first validation were collected. Constructing a training set for this parameter using the method described here, might help. The right panel of Fig. 5.9 is a plot of one of the waveforms with the largest phase difference, together with the GR waveform at these parameter values.

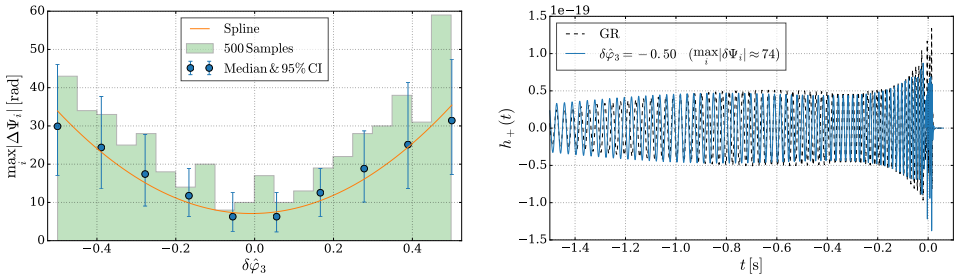


Figure 5.9: left: The maximum phase differences for 1000 waveforms randomly drawn from the training set parameter space as a function of non-GR parameter shift. Shown here are the 95% confidence intervals and medians. **right:** One of the waveforms with the largest phase difference together with its GR counterpart.

The full range of non-GR parameters for IMRPhenomPv2+ consists of 14 relative phase-shifts if we stick to the ones we used in the analysis of GW150914 and GW151226:

$$\begin{aligned}\hat{P} =&\{\delta\hat{\varphi}_0, \delta\hat{\varphi}_1, \delta\hat{\varphi}_2, \delta\hat{\varphi}_3, \delta\hat{\varphi}_4, \delta\hat{\varphi}_{5l}, \delta\hat{\varphi}_6, \delta\hat{\varphi}_{6l}, \delta\hat{\varphi}_7,\\ &\delta\hat{\beta}_2, \delta\hat{\beta}_3,\\ &\delta\hat{\alpha}_2, \delta\hat{\alpha}_3, \delta\hat{\alpha}_4\}. \end{aligned} \tag{5.52}$$

For production TIGER runs or parameterized tests of GR as described in section 4.1 we need ROQs for all these parameters and for all chirp mass bins listed in Table 5.2. This is a large amount of work and the results presented in this thesis are not yet the complete set. Instead, I will consider only case B in Table 5.2 for $\delta\hat{\varphi}_3$, $\delta\hat{\alpha}_2$ and $\delta\hat{\beta}_2$ and cases A,B,C and F for $\delta\hat{\varphi}_3$. These cases should provide a good example of the feasibility of the basis extension method for different types of non-GR parameters as well as for the longest waveforms that require the adaptive sampling technique. Table 5.3 lists the cases that are discussed here.

Case	$f_{\rm max}$ (Hz)	Signal duration (s)	$\mathcal{M}_c\,(\mathrm{M}_\odot)$	N^L	N^Q	Speedup
$\delta\hat{\varphi}_3^A$	1024	$1.5 \leq T \leq 4$	$12.3 \leq \mathcal{M}_c \leq 23$	607	281	4.52
$\delta\hat{\varphi}_3^B$	1024	$3 \leq T \leq 8$	$7.9 \leq \mathcal{M}_c \leq 14.8$	553	253	9.97
$\delta\hat{\varphi}_3^C$	2048	$6 \leq T \leq 16$	$5.2 \leq \mathcal{M}_c \leq 9.5$	689	265	34.01
$\delta\hat{\varphi}_3^F$	4096	$48 \leq T \leq 128$	$1.4 \leq \mathcal{M}_c \leq 2.6$	1352	350	306.54
$\delta\hat{\beta}_2^B$	1024	$3 \leq T \leq 8$	$7.9 \leq \mathcal{M}_c \leq 14.8$	610	252	9.32
$\delta\hat{\alpha}_2^B$	1024	$3 \leq T \leq 8$	$7.9 \leq \mathcal{M}_c \leq 14.8$	543	244	10.21

Table 5.3: The set of cases for which a non-GR basis has been built. The parameter ranges are as follows $-0.5 \leq \delta\hat{\varphi}_3 \leq 0.5$, $-5.0 \leq \delta\hat{\alpha}_2 \leq 5.0$ and $-2.0 \leq \delta\hat{\beta}_2 \leq 2.0$. All waveforms start at a lower frequency $f_{\rm min} = 20\,\mathrm{Hz}$ and in all cases $\Delta f = 1/\max(T)$.

The non-GR parameter ranges are based on the posteriors we found in the parameterized tests of GR performed on O1 data. The range of $\delta\hat{\varphi}_3$ is at this time rather narrow with $-0.5 \leq \delta\hat{\varphi}_3 \leq 0.5$ and serves as a proof of concept. For production runs we will probably want to get it closer to ± 1.0 to fit the posteriors of both events in O1. For $\delta\hat{\beta}_2$ and $\delta\hat{\alpha}_2$ I tried for a wider range encompassing most of the posteriors as seen in Fig. 4.13: $-0.5 \leq \delta\hat{\beta}_2 \leq 0.5$ and $-0.5 \leq \delta\hat{\alpha}_2 \leq 0.5$. For the other intrinsic parameters appearing in IMRPhenomPv2, I used the same ranges as was done in [143]: $1 \leq q \leq 9$, $(-0.9, -0.9, 0) \leq (\chi_{1L}, \chi_{2L}, \chi_p) \leq (0.9, 0.9, 0.9)$, $(0, 0) \leq (\theta_J, \alpha_0) \leq (\pi, 2\pi)$ and $m_1 \geq m_2 \geq 1\,\mathrm{M}_\odot$. In the enrichment steps, when validating the basis using a large set of random waveforms Λ_N, I also implement the bound Smith *et al.* imposed: $\chi_{1L} \geq 0.4 - 7\eta$. This last bound is not one that follows directly from the physics of the model, but was found empirically. Before this bound was implemented, bad points would cluster in this region of parameter space indicating a problem in the model. Indeed, it seems that the IMRPhenomPv2 model exhibits sharp features in the amplitude and does not agree well with SEOBNRv3, suggesting that both models are pathological in this region, which was also demonstrated in [143]. Such features are difficult to capture with a sparse basis and result in a clustering of bad points. In the same way I have discovered an additional problematic region in parameter space when dealing with non-GR parameters, or at least $\delta\hat{\varphi}_3$. The extra parameter appears to introduce more complications for IMRPhenomPv2 at high component spins χ_{1L} and large mass ratios q. The left panel of Fig. 5.10 shows clustering of bad points in the top left corner, corresponding to large spins and mass ratios. The right panel shows this cluster plotted in the χ_{1L}-χ_p plane where the spin-bound $\sqrt{\chi_{1L}^2 + \chi_p^2} \leq 1$ is indicated with a black dashed curve. Note how most of the points lie along this curve and close to it. Along the spin-bound, the extra dimension in the $\delta\hat{\varphi}_3$ direction gets filled up with bad points after each validation. Apparently this extra degree of freedom has brought to light an additional problem in the model: Note that in the bottom panel of Fig. 5.10 the agreement between the waveform models is bad, regardless of the fact that they are GR (*i.e.* $\delta\hat{\varphi}_3 = 0$).

This problematic part of parameter space can be omitted by slightly tightening

the spin-bound to $\sqrt{\chi_{1L}^2 + \chi_p^2} \leq 0.99$. In practice, this gets rid of the problem and I can build compact bases for all the cases discussed in this thesis. However, this bound is just a first guess as to what might work. It is possible that setting the spin-bound to 0.999 will be sufficient and this method is not guaranteed to work for all non-GR parameter cases. This feature is certainly worth looking into in future work. At some point one can expect that not including certain parts in parameter space might harm parameter estimation efforts. This is definitely something that needs to be taken into consideration when constructing an ROQ for production use. In this case we have omitted only extreme corners of parameter space, *e.g.* large component spins, combined with large mass ratios. When we happen to stumble across such a source in nature, we can of course not use the ROQs, but the same can be said for the full IMRPhenomPv2 waveform, and possibly even SEOBNRv3 considering their poor agreement and spiky characteristics. The dynamics in these extreme parts of parameter space are too complex to capture with the current waveform models. As development in this field progresses, new ROQ models will need to be produced. Indeed an algorithm such as LALInference will touch upon the abovementioned parts of parameter space where the ROQ is not valid, if one does not impose the specialized ROQ bounds in this algorithm. However, these points will have a poor likelihood and will quickly be excluded as the algorithm climbs up the likelihood surface. The definitive test of the ROQ will therefore be to implement it in a parameter estimation run, which is done in section 5.4.3.

The sizes of the final bases with the bounds as given above are listed in Table 5.3. Regardless of the additional non-GR parameter, the theoretical speedups achieved are quite close to those in Table 5.2. In case $\delta\hat{\varphi}_3^F$ the speedup is even larger, which is likely due to the additional constraint placed on $\chi_{1L}^2 + \chi_p^2$. In addition to generating the linear basis $\mathcal{B}^{\text{L}}$, IMRPhenomPv2+ also requires a non-trivial quadratic part $\mathcal{B}^{\text{Q}}$. Constructing the quadratic part also required a few enrichment steps in most cases listed in Table 5.3. Fig. 5.11 shows the final selected basis points for case $\delta\hat{\varphi}_3^F$ for a few slices in parameter space and Fig. 5.12 shows the errors after the various enrichment steps for $\delta\hat{\varphi}_3^A$, $\delta\hat{\varphi}_3^B$ and $\delta\hat{\varphi}_3^F$. The errors after enrichment steps for cases $\delta\hat{\beta}_2^B$ and $\delta\hat{\alpha}_2^B$ are shown in Fig. 5.13 and Fig. 5.14 respectively. The basis points for $\delta\hat{\varphi}_3^F$ show a few features: Especially the top left and bottom left plots show grid-like structures. These are simply the points left over from the original IMRPhenomPv2 basis constructed in [143], which used a grid-like training set. Any points scattered in between grid lines are a result of the enrichment process. The middle and top right plots show that mainly the quadratic part of the basis still requires many points in the regions of high spin and low mass ratio. This is not a problem, since after only two enrichment steps we already had a good basis. The interpolation errors in Fig. 5.12, Fig. 5.13 and Fig. 5.14 also demonstrate different situations one can come across when constructing a basis. For example, in case $\delta\hat{\varphi}_3^A$ we see that the original basis was already spot on (validation 2 was performed for some measure of security) as opposed to case $\delta\hat{\varphi}_3^F$, where the original basis was apparently poor with the error tail extending all the way to unity. It is also interesting to see that although the initial *linear* basis of case $\delta\hat{\varphi}_3^A$ was spot on, the *quadratic* part required three enrichment steps to converge to a basis with the desired accuracy. For the cases considered in

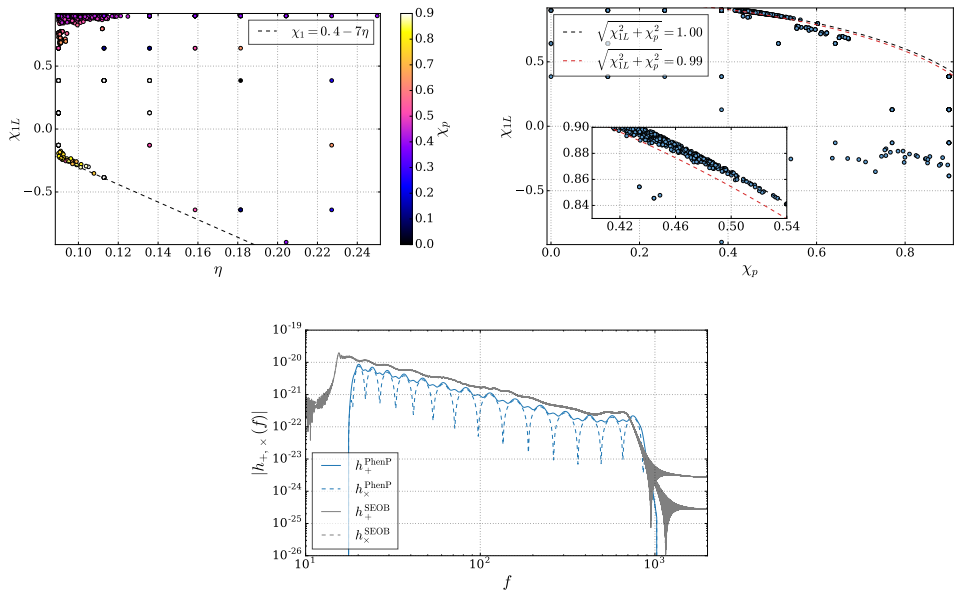


Figure 5.10: This figure shows how points are clustered in a difficult part of parameter space. Points shown here are those collected after 4 enrichment runs. **left:** Clustering of points is visible in the corner of parameter space where component spins and mass ratio are large. **right:** Clustering of points along between the original spin-bound $\sqrt{\chi_{1L}^2 + \chi_p^2} \leq 1$ (black dashed) and the revised spin-bound $\sqrt{\chi_{1L}^2 + \chi_p^2} \leq 0.99$ (red dashed). The inset zooms in on the bulk of the cluster. **bottom:** Example waveforms at one of the points in the cluster. Both IMRPhenomPv2 and SEOBNRv3 are plotted. The non-GR parameter $\delta\hat{\varphi}_3$ is set to 0. This source is close to the spin-bound with $\sqrt{\chi_{1L}^2 + \chi_p^2} \sim 0.999$. The full set of source-parameters are $\chi_{1L} = 0.90$, $\chi_p = 0.44$, $\chi_{2L} = 0.00$, $\theta_j = 0.96$, $\alpha_0 = 5.88$ and has a high mass ratio: $m_1 = 29.27\,\mathrm{M}_\odot$ and $m_2 = 3.60\,\mathrm{M}_\odot$.

Table 5.3, these are about the most extreme examples. That said, these cases are only a small fraction of the cases we need to cover in order to have production quality ROQs for IMRPhenomPv2+, and one can expect to run into unforeseen features.

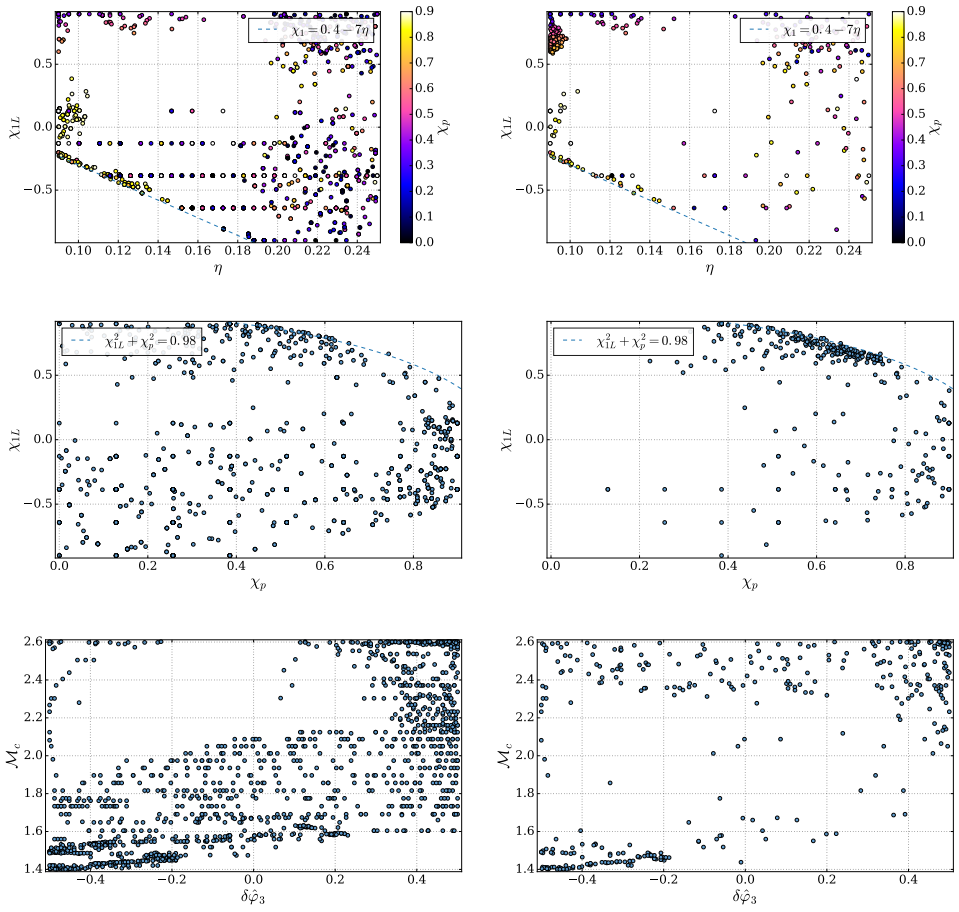


Figure 5.11: Final selected basis points (linear on the left, quadratic on the right) for case $\delta\hat{\varphi}_3^F$ for a few slices in parameter space.

5.4.3 PE tests and performance

The ultimate test of an ROQ is applying it in a realistic parameter estimation setting. The previous section has summarized the results for a few different cases in Table 5.3. Due to time constraints the ROQ has only been tested for the cases $\delta\hat{\varphi}_3^B$, $\delta\hat{\alpha}_2^B$, $\delta\hat{\beta}_2^B$. I deliberately started with the the B-cases since the chirp mass bin nicely accommodates the second detection GW151226 and we already experienced the parameter estimation to be painfully slow due to the length of the signal. These cases make the perfect

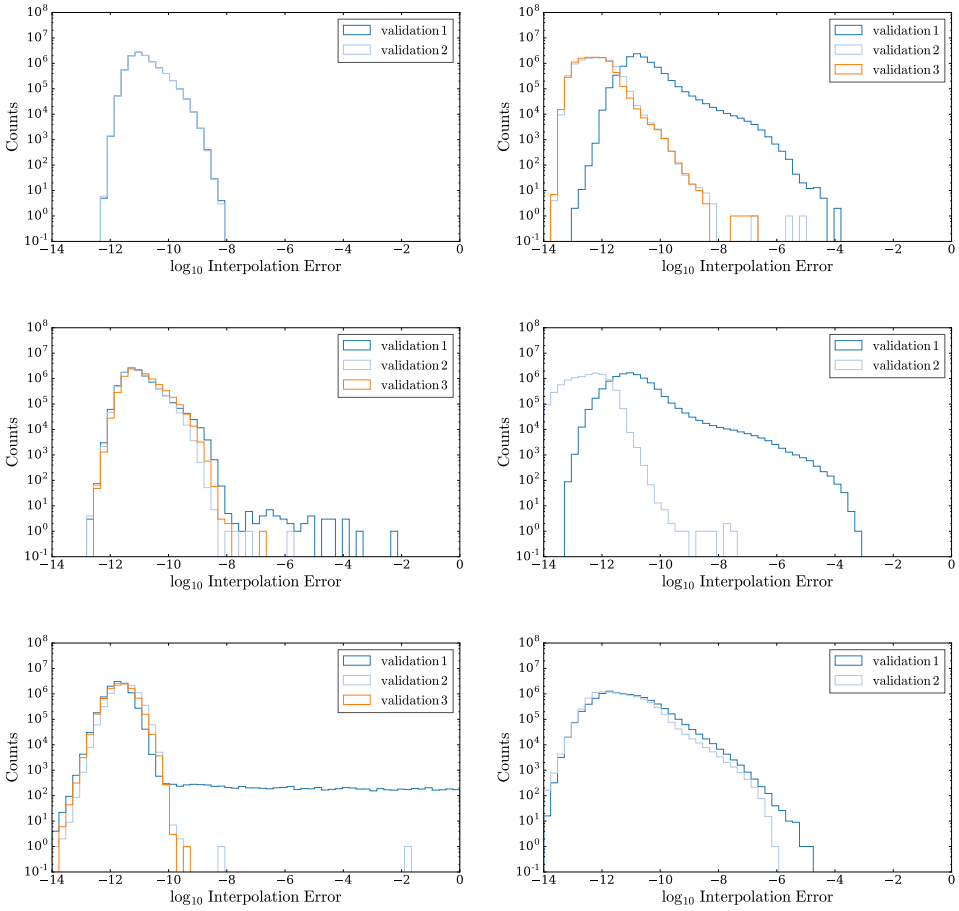


Figure 5.12: The interpolation errors for all points in the validation sets Λ_N^i (validation i), where $N = 10^7$ for the linear (left) and quadratic (right) basis. The last iteration is where all the waveforms in the validation set have an interpolation error of less than $\beta = 10^{-6}$ which is the desired tolerance. The rows display the cases $\delta\hat{\varphi}_3^A$, $\delta\hat{\varphi}_3^B$ and $\delta\hat{\varphi}_3^F$ from top to bottom.

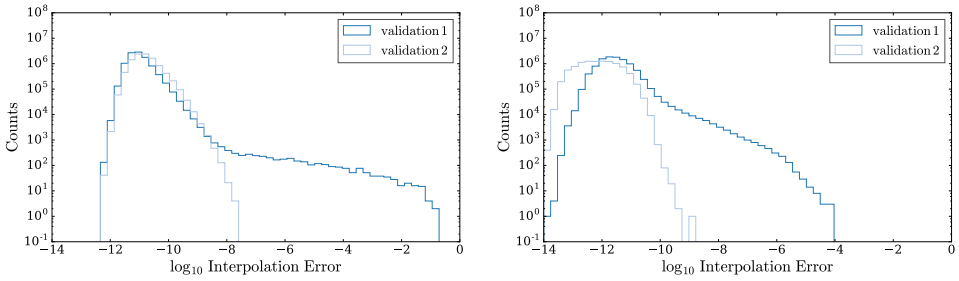


Figure 5.13: Same as Fig. 5.12, but for case $\delta\hat{\beta}_2^B$.

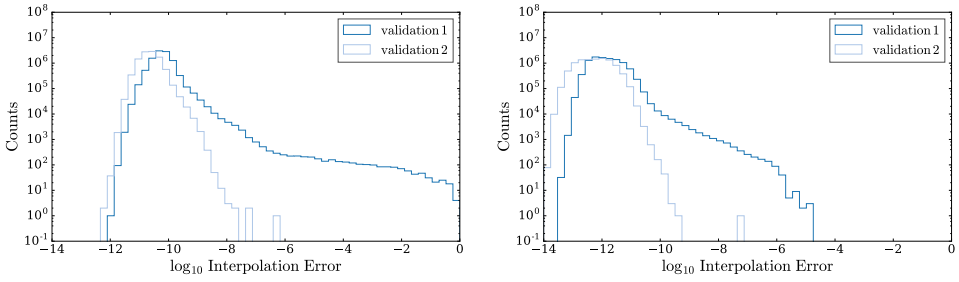


Figure 5.14: Same as Fig. 5.12, but for case $\delta\hat{\alpha}_2^B$.

candidates for demonstrating the effectiveness of the IMRPhenomPv2+ROQs in a realistic parameter estimation application. In this section 3 LALInference comparison runs are set up where the same priors as in the O1 parameterized tests of GR listed in chapter 4 are used. The only differences are the prior for the chirp mass, mass ratio and those for the non-GR parameters. The chirp mass prior is taken to be close to the width of chirp mass bin B $[7.93\,\mathrm{M}_\odot, 14.76\,\mathrm{M}_\odot]$, the mass ratio is close to what is used in official parameter estimation runs with $1 \leq q \leq 8$ and the priors for the non-GR parameters are reduced to what the current ROQs allow, namely $[-0.499, 0.499]$, $[-1.99, 1.99]$ and $[-4.99, 4.99]$ for $\delta\hat{\varphi}_3$, $\delta\hat{\beta}_2$ and $\delta\hat{\alpha}_2$ respectively. All the priors are slightly narrower than what the ROQ was built for to avoid any edge-effects. This precaution might not be necessary, so it would be worth running the same tests while setting the prior to the full width of the ROQ range. Again, this is left for future studies. The source to be analyzed is an injected waveform with source parameters that are within the posterior distributions of GW151226 as presented in [72]: $\mathcal{M}_c \approx 8.91\,\mathrm{M}_\odot$, $q \approx 1.99$ and distance $D_L = 200\,\mathrm{Mpc}$. The source has an SNR of approximately 18 and is GR, *i.e.* $\delta\hat{p}_i = 0$. This source has been injected into an EARLY_RECOLORED_S6 (see section 1.3.3) data frame. The injection GPS time was chosen such that for the full length of the waveform plus some buffer ($\simeq 10\,\mathrm{s}$), the data are free of glitches.

The comparison is made between a LALInference run that uses the regular IMRPhenomPv2+ without ROQs and a run with the exact same parameters and source, but using ROQs instead. Fig. 5.15 shows the posteriors of the three cases for the chirp mass $\mathcal{M}_c$ (right column) and the non-GR parameter (left column). The results are not precisely alike, but we do not expect them to be since every LALInference run has some measure of randomness; the sampling process is randomized. Aside from small differences the posteriors agree well. The chirp mass is always the best recovered intrinsic parameter because it enters in the amplitude as well as the phase of the waveform at lowest PN order. Accurately recovering this parameter in particular is therefore a good test of the model.

In this parameter estimation example we use a source that falls nicely within one of the chirp mass bins listed in Tables 5.2 and 5.3, but what if the detected source happens to be right on the edge of one of the bins? Even though this might not be the case most of the time, it is certainly something we need to consider. Most of the times when a source passes the detection threshold and is indeed treated as a detection, the SNR is sufficiently high to get an accurate measurement of the chirp mass. For example, GW151226 was not a loud observation such as GW150914, but its chirp mass measurement was still accurate with $\mathcal{M}_c = 8.9^{+0.3}_{-0.3}\,\mathrm{M}_\odot$ [72]. The chirp mass bins have overlaps ranging from $0.4\,\mathrm{M}_\odot$ for the longest waveforms to $2.5\,\mathrm{M}_\odot$ for the shortest. These overlaps are thus in most cases large enough to accommodate the entire posterior of the chirp mass in one of the bins. Note that the smallest overlap between bins is for the case of the longest waveforms. These waveforms have a large number of GW cycles in band and the chirp mass posterior will likely be extremely narrow. In the few cases we are dealing with a detection that shows a posterior leaking into a neighboring bin, we can run the analysis on two bins and combine the posteriors. We would like to avoid that of course, since it would require roughly twice as many CPU hours.

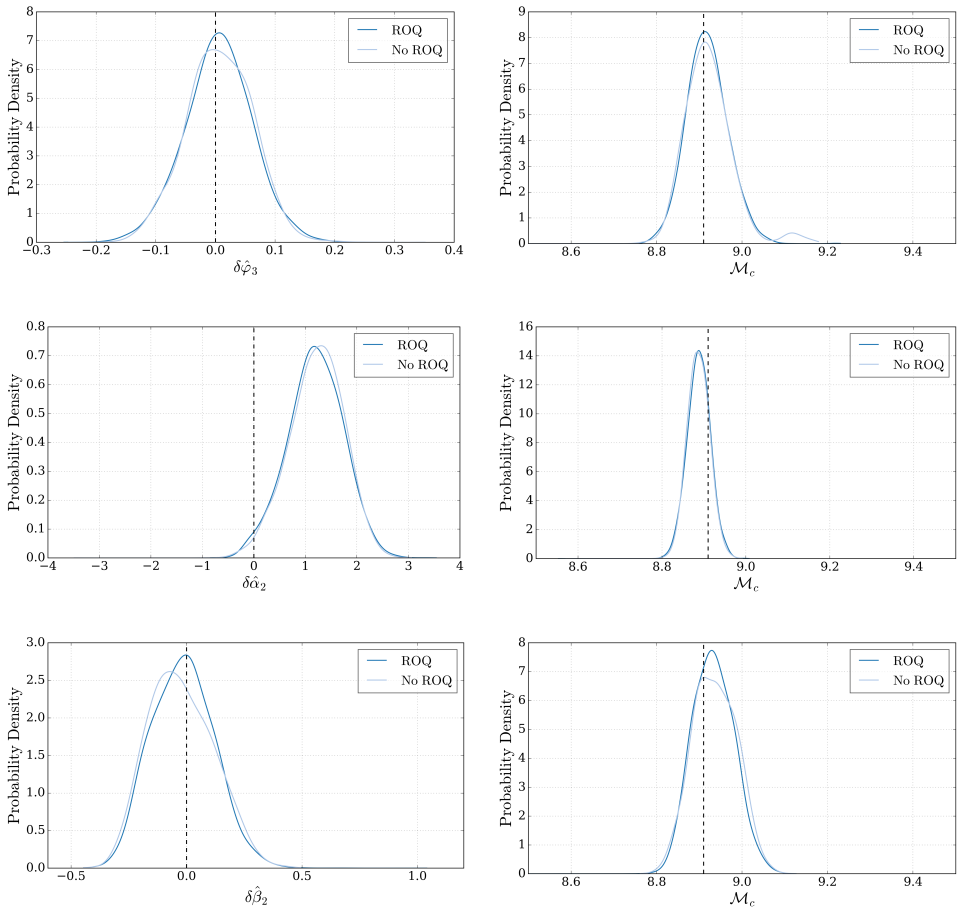


Figure 5.15: Comparison plots of parameter estimation results on a source consistent with GW151226. Results are shown for the cases $\delta\hat{\varphi}_3^B$, $\delta\hat{\alpha}_2^B$ and $\delta\hat{\beta}_2^B$ in the top, middle and bottom row respectively. The left column show the posteriors for the non-GR parameter and the right column shows results for the chirp mass. The injected values are $\mathcal{M}_c \simeq 8.9\,\mathrm{M}_\odot$, $q \approx 1.99$, $D_L = 200\,\mathrm{Mpc}$, $\delta\hat{p}_i = 0$ (dashed vertical lines) and SNR $\rho \simeq 18$.

Case	ROQ run times		Standard run times		Theoretical	Actual speedup	
	min	max	min	max	speedup	min	max
$\delta\hat{\varphi}_3^B$	1d : 9h	1d : 18h	9d : 1h	10d : 1h	9.97	5.22	7.31
$\delta\hat{\beta}_2^B$	2d : 4h	2d : 8h	9d : 3h	9d : 22h	9.32	3.95	4.40
$\delta\hat{\alpha}_2^B$	1d : 8h	1d : 10h	16d : 7h	17d : 5h	10.21	11.59	12.91

Table 5.4: Actual running times (roughly) and speedups recorded from the LALInference output. The times cited are from the main LALInference algorithm. Calculation of ROQ weights took about 15 s in all cases.

The speedups recorded by the parameter estimation runs considered in this chapter vary quite a bit. Table 5.4 lists the times the LALInference algorithm used. I have recorded the maximum and minimum times of the 4 chains that each LALInference job ran. We can see that the actual speedup for $\delta\hat{\beta}_2$ is quite a bit less than the theoretical prediction based on the basis size in Table 5.3. On the other hand $\delta\hat{\alpha}_2$ appears to exceed expectations which is likely due to the fact that we are unable to do a proper one-to-one comparison: In parameter estimation runs, the runtime is influenced by factors such as the status of the computing cluster, queue times and possible differences in computing nodes. We see though that when we compare actual speedups with the theoretical ones that they are ordered in the same way, *i.e.* $\delta\hat{\beta}_2$ has the lowest speedup and $\delta\hat{\alpha}_2$ the highest. Of course, not much can be concluded from these few examples. To get a more accurate idea of the typical speedup of an ROQ one would need to perform many of these comparisons and look for instance at the mean speedup to be able to average out unpredictable effects such as computing cluster related scheduling. All we can learn from this small data set is that the ROQs give accurate results and generate speedups between factors of roughly 5 to 10, making a run that would normally take a week, likely to finish within a day. The times recorded in the Table are for the main LALInference algorithm only, since the additional startup time for computing the ROQ weights is, as we expected, negligible at around 15 seconds for each case.

5.5 The future of parameterized tests of GR using ROQs

In the parameterized tests of GR (see [73] and section 4.1), we use the flagship frequency domain waveform model IMRPhenomPv2 in the case of BBH and NSBH sources and TaylorF2 for BNS. These models are relatively quick to evaluate since they are closed form analytical representations already in the frequency domain in which the likelihood is evaluated. Although these waveforms are the fastest models while at the same time perfectly suited for their respective applications, a typical parameter estimation run on a single source costs days (a week for GW150914) to months (about 30 days for GW151226) and more. This in itself asks for faster evaluations of the models and in particular the overlap calculations in the likelihood. ROQs are cur-

rently the most promising technique for achieving large speedups and have been produced for IMRPhenomPv2. The implementation of ROQs on IMRPhenomPv2+ is even more pressing, since for each parameterized test, not one analysis, but at least 14 need to be run to test the various non-GR parameters. I have employed the bases built in [143] to build a basis for IMRPhenomPv2+. This work has demonstrated the feasibility of the basis extension method, where an existing basis is extended to one or more non-GR parameters to create a sparse but suitable training set. The ROQs for IMRPhenomPv2+ reproduce posterior distributions on the non-GR parameters and chirp masses well and much quicker than the full model. Speedups have been recorded ranging from a factor 5 to 12. Building ROQs to implement on the full parameterized test of GR is a process that is currently underway. There is however more investigation required into the various degrees of freedom in generating an ROQ. For example, currently a small part of parameter space is excluded by setting an upper limit to the squared sum of χ_{1L} and χ_p to be less than 0.99. It may be worth investigating what effect it has on the basis and parameter estimation results if this bound would be relaxed a bit more. The current version of IMRPhenomPv2 is not reliable at extreme spins, and even though building a basis where $\sqrt{\chi_{1L}^2 + \chi_p^2} \leq 0.99$ works, it might be sufficient to set it to 0.98 or even less, possibly resulting in a much smaller basis that works just as well in a realistic parameter estimation setting. That said, care must be taken when considering to reduce basis size at the cost of losing parameter space coverage. Another thing that needs more attention is the ranges of the non-GR parameters. Currently, $\delta\hat{\varphi}_3$ is possibly too narrow and the current approach of using the same ranges for all chirp mass bins might not be optimal especially for long signals. One can expect for example to reach much narrower posteriors on $\delta\hat{\varphi}_3$ for long signals, due to the large difference in GW cycles in the detector bandwidth. Finally, for IMRPhenomPv2+, I would recommend to use more validation points in the enrichment steps than I have done now. A guideline could be to use 10^n points, where n is the number of parameters. Currently, I have used 10^7 validation points, which is in principle equivalent to sampling each parameter dimension (of which there are 8) with less than 8 points ($x = 10^{7/8} \lesssim 7.5$). With such a sampling bad points might be missed, especially in parts of parameter space where the complexity of the waveform does not vary uniformly with the parameter that we draw from.

Even though quite a bit of work is still left to be done for constructing IMRPhenomPv2+ ROQs that we can use preferably in the second half of O2, the ROQs for TIGER-BNS are also still to be finalized. Even when the ROQ work for IMRPhenomPv2+ is finished, we will not be able to use these for typical BNS systems, as the chirp mass bins considered only go down to masses corresponding to the heaviest BNS systems. To cover the complete range, BNS ROQs are required, but for TIGER we only require the simple inspiral-only TaylorF2 model up to a frequency of 400 Hz. I have demonstrated how we can use template bank generation to construct an effective non-GR training set. In the process of building the TaylorF2 basis, I identified some points that need more attention in the future. First of all, the setup used in this work was meant to be a proof of principle and waveforms started from 30 Hz as opposed to the 20 Hz lower limit that could be achieved with Advanced Virgo and LIGO. In this case the basis was built for an upper frequency of 512 Hz, which I do not think is really necessary, since we have already shown in previous

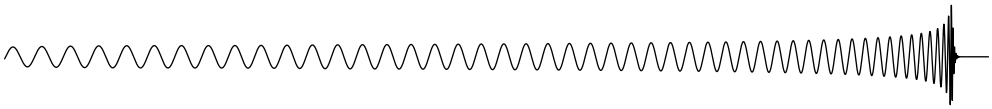
work [66, 67, 68] that a frequency cut-off of 400 Hz works perfectly well for hypothesis testing in TIGER-BNS. In future work I also suggest to perform validation studies on the up-sampled basis, before applying the ROQs to a parameter estimation scenario. For TaylorF2 it might be fruitful to give more attention to determining the longest waveform in the training set (including the effect of the non-GR parameter), in order to avoid the need to use the augmentation factor in generating an adaptively sampled frequency series. Finally, the basis extension method introduced in section 5.4 might be the way to a TaylorF2 ROQ using more than one non-GR parameter if we first build a basis for the GR TaylorF2 with the four parameters. For the latter, I would suggest using a template bank as training set.

For TaylorF2 it is difficult to predict what actual speedup can be achieved until a full parameter estimation study has been done. We may however expect to exceed a factor of 10 considering the size of the basis and length of the original waveforms. We have also demonstrated that a much larger speedup can be achieved if we consider waveforms starting at 20 Hz as opposed to 30 Hz.

I have demonstrated that ROQs can be built within a reasonable time frame for sources ranging from BNS to BBH sources. When ROQs for both TaylorF2 and IMRPhenomPv2+ are available, we can employ these to considerably increase the performance of the parameterized tests. Currently I have achieved theoretical speedups of around 10 for GW151226-like events and over 300 for the lightest BBH systems. These values may increase in the near future and background calculations for the TIGER framework that used to take months could be done in a matter of days. Background calculations for light BBH systems will be feasible for the first time, making the future of the TIGER pipeline an optimistic one.

Chapter 6

Conclusions and outlook



The two gravitational wave detections made by Advanced LIGO in September 2015 ushered in a new age for astronomy and fundamental physics. General relativity makes clear predictions about what such signals should look like and how the nature of the source imprints its properties onto the structure of the waveforms that pass by our detectors. This allows us to test whether the predictions made by Einstein one hundred years ago still stand.

In this thesis I presented our work on the first tests of general relativity in the strong-field dynamical regime using the first discovered gravitational wave signals and introduced a method that is capable of significantly increasing the speed by which the required gravitational waveforms can be generated and likelihoods are calculated.

Tests of general relativity – In this work we have presented two main model-independent tests of general relativity (GR) that are related to one another; one is to perform parameter estimation on the coefficients of the various GR-predicted functions of frequency entering the phase of the gravitational wave signals, the other is to perform model selection between two competing hypotheses, one stating that GR is correct, the other stating that one or more of the coefficients are not as predicted by GR. We stress that we did not pursue the verification or dismissal of any particular alternative theories of gravity, rather we are approaching the tests completely model-independent, allowing the possibility to spot violations of any kind even one that is not captured in any known alternative theory of gravity.

To perform model selection we employ the TIGER pipeline, which has been internally reviewed by the LIGO-Virgo Collaboration when using BNS sources. We have presented the first steps taken to extend TIGER's functionality to include light BBH systems. We found that the GR background distribution exhibits a tail toward positive odds ratios, making strong statements on either the preference for or against GR difficult. However, we have shown that this tail disappears as one combines information from multiple sources, meaning that in the case of multiple detections TIGER still constitutes an effective model-independent test of GR. An investigation of how to include heavier BBH systems is left for for future work.

Aside from using the TIGER pipeline to test GR, we introduced a parameterized test. Rather than performing model selection, we measure posteriors on individual non-GR parameters, *i.e.* relative deviations in functions or parameters that characterize the waveform phase. In doing so, for the first time we have been able to place constraints on possible deviations from GR in all the PN coefficients of the early-

inspiral part of the GW waveform. Some of these constraints, like the 1.5PN non-GR coefficient, restrict violations to the 10% level. In addition to doing this for both individual detections we have improved our bounds by combining information from the two signals and bounds placed on the non-GR parameters will continue to improve as more sources are combined, roughly with the square root of the number of detections.

Lastly, in addition to using the full inspiral-merger-ringdown signal to perform tests of GR we have demonstrated what may be accomplished when only the ringdown is visible, even when this would be at a low SNR. The TIGER pipeline proves flexible enough to allow one to test the no-hair theorem combining ringdown information from multiple sources. This was shown making use of the ET noise curves, but could in principle be applied to Advanced LIGO and Advanced Virgo, given appropriate frequency models and taking care with separation between inspiral-merger and ringdown parts of the signal. The latter is not a trivial task and is currently being investigated.

Increasing speed of likelihood evaluations – The above-mentioned tests of general relativity demand a lot of computing resources. Especially the construction of a GR background, not to mention the many foreground simulations required for method testing are time consuming operations. To increase the speed at which waveforms are generated and more importantly, likelihoods are evaluated I have applied the concept of reduced order quadratures (ROQs) to the waveform models used in the parameterized tests and possibly TIGER. By considering non-GR waveforms, the already large seven dimensional parameter space of precessing BBH signals grows even larger.

By extending the existing basis introduced in [143] into the direction of a non-GR parameter, we can overcome the memory storage constraints implied by the greedy algorithm used to generate a basis. This proof of principle is very promising for the parameterized test as here one typically only needs to consider a single additional non-GR parameter. However, if we wish to implement the non-GR ROQs in the TIGER pipeline, we ideally need to have ROQs for even higher dimensional models. This problem is left for future investigation, but the method presented in this thesis might work for additional parameters as well. To complete the non-GR ROQ work started here, all chirp mass bins and wider ranges in GR violations need to be included.

I have demonstrated that we can reach theoretical speedups of around 10 for signals much like GW151226 and over 300 for the lightest BBH systems. This implies that background calculations for the TIGER framework that used to take months could be done in a matter of days. Finally, due to the significant speedups for the longest waveforms, background calculations for light BBH systems will be feasible for the first time, making the future of the TIGER pipeline a optimistic one.

Bibliography

[1] N. Yunes, K. Yagi, and F. Pretorius. Theoretical Physics Implications of the Binary Black-Hole Mergers GW150914 and GW151226. *Phys. Rev.* **D94** (2016)(8), 084002. arXiv:1603.08955.

[2] M. Kramer *et al.* Tests of general relativity from timing the double pulsar. *Science* **314** (2006), 97. arXiv:astro-ph/0609417.

[3] D. Shoemaker. Advanced LIGO anticipated sensitivity curves (2015). https://dcc.ligo.org/LIGO-T0900288/public.

[4] S. Hild, S. Chelkowski, A. Freise, J. Franc, *et al.* A Xylophone Configuration for a third Generation Gravitational Wave Detector. *Class. Quant. Grav.* **27** (2010), 015003. arXiv:0906.2655.

[5] J. Slutsky *et al.* Methods for Reducing False Alarms in Searches for Compact Binary Coalescences in LIGO Data. *Class. Quant. Grav.* **27** (2010), 165023. arXiv:1004.0998.

[6] H. Luck *et al.* The upgrade of GEO600. *J. Phys. Conf. Ser.* **228** (2010), 012012. arXiv:1004.0339.

[7] C. Affeldt, K. Danzmann, K. L. Dooley, H. Grote, *et al.* Advanced techniques in GEO 600. *Class. Quant. Grav.* **31** (2014)(22), 224002.

[8] G. M. Harry. Advanced LIGO: The next generation of gravitational wave detectors. *Class. Quant. Grav.* **27** (2010), 084006.

[9] J. Aasi *et al.* Advanced LIGO. *Class. Quant. Grav.* **32** (2015), 074001. arXiv:1411.4547.

[10] F. Acernese *et al.* Advanced Virgo: a second-generation interferometric gravitational wave detector. *Class. Quant. Grav.* **32** (2015)(2), 024001. arXiv:1408.3978.

[11] B. P. Abbott *et al.* Binary Black Hole Mergers in the first Advanced LIGO Observing Run. *Phys. Rev.* **X6** (2016)(4), 041015. arXiv:1606.04856.

[12] B. P. Abbott *et al.* Prospects for Observing and Localizing Gravitational-Wave Transients with Advanced LIGO and Advanced Virgo. *Living Reviews in Relativity* **19** (2016)(1), 1. arXiv:1304.0670.

[13] K. Somiya. Detector configuration of KAGRA – the Japanese cryogenic gravitational-wave detector. *Class. Quant. Grav.* **29** (2012)(12), 124007.

[14] Y. Aso, Y. Michimura, K. Somiya, M. Ando, *et al.* Interferometer design of the KAGRA gravitational wave detector. *Phys. Rev.* **D88** (2013), 043007.

[15] C. S. Unnikrishnan. IndIGO and LIGO-India: Scope and plans for gravitational wave research and precision metrology in India. *Int. J. Mod. Phys.* **D22** (2013), 1341010.

[16] ET design study document. http://www.et-gw.eu/etdsdocument/.

[17] J. A. Faber and F. A. Rasio. Binary Neutron Star Mergers. *Living Rev. Rel.* **15** (2012), 8. arXiv:1204.3858.

[18] R. Ruffini and J. Wheeler. . *Physics Today* **24** (1971), 30.

[19] E. Leaver. An Analytic Representation for the Quasi-Normal Modes of Kerr Black Holes. *Proc. R. Soc. Lond.* **402** (1985), 285.

[20] K. Kokkotas and B. Schmidt. Quasi-Normal Modes of Stars and Black Holes. *Liv. Rev. Rel.* **2** (1999), 2.

[21] E. Berti, V. Cardoso, and A. O. Starinets. Quasinormal modes of black holes and black branes. *Class. Quant. Grav.* **26** (2009), 163001. arXiv:0905.2975.

[22] LIGO/LSC algorithms library, LALSuite. https://wiki.ligo.org/DASWG/LALSuite.

[23] B. Brügmann, J. A. González, M. Hannam, S. Husa, *et al.* Calibration of moving puncture simulations. *Phys. Rev.* **D77** (2008), 024027.

[24] S. Husa, J. A. Gonzalez, M. Hannam, B. Bruegmann, *et al.* Reducing phase error in long numerical binary black hole evolutions with sixth order finite differencing. *Class. Quant. Grav.* **25** (2008), 105006. arXiv:0706.0740.

[25] SXS Gravitational Waveform Database. https://www.black-holes.org/waveforms/.

[26] E. Newman and R. Penrose. An Approach to gravitational radiation by a method of spin coefficients. *J. Math. Phys.* **3** (1962), 566.

[27] N. T. Bishop and L. Rezzolla. Extraction of Gravitational Waves in Numerical Relativity. *Living Reviews in Relativity* **19** (2016)(1), 2. ISSN 1433-8351. arXiv:1606.02532.

[28] A. Taracchini, A. Buonanno, Y. Pan, T. Hinderer, *et al.* Effective-one-body model for black-hole binaries with generic mass ratios and spins. *Phys. Rev.* **D89** (2014), 061502.

[29] P. Kumar, K. Barkett, S. Bhagwat, N. Afshari, *et al.* Accuracy and precision of gravitational-wave models of inspiraling neutron star-black hole binaries with spin: Comparison with matter-free numerical relativity in the low-frequency regime. *Phys. Rev.* **D92** (2015), 102001.

[30] B. Szilágyi, J. Blackman, A. Buonanno, A. Taracchini, *et al.* Approaching the Post-Newtonian Regime with Numerical Relativity: A Compact-Object Binary Simulation Spanning 350 Gravitational-Wave Cycles. *Phys. Rev. Lett.* **115** (2015), 031102.

[31] L. Santamaría, F. Ohme, P. Ajith, B. Brügmann, *et al.* Matching post-Newtonian and numerical relativity waveforms: Systematic errors and a new phenomenological model for nonprecessing black hole binaries. *Phys. Rev.* **D82** (2010), 064016.

[32] S. Khan, S. Husa, M. Hannam, F. Ohme, *et al.* Frequency-domain gravitational waves from nonprecessing black-hole binaries. II. A phenomenological model for the advanced detector era. *Phys. Rev.* **D93** (2016)(4), 044007. arXiv:1508.07253.

[33] P. Schmidt, I. Harry, and H. Pfeiffer. Numerical Relativity Injection Infrastructure. https://dcc.ligo.org/cgi-bin/private/DocDB/ShowDocument?docid=T1500606&version=.

[34] C. R. Galley and P. Schmidt. Fast and efficient evaluation of gravitational waveforms via reduced-order spline interpolation (2016). arXiv:gr-qc/1611.07529.

[35] A. Buonanno, B. R. Iyer, E. Ochsner, Y. Pan, *et al.* Comparison of post-Newtonian templates for compact binary inspiral signals in gravitational-wave detectors. *Phys. Rev.* **D80** (2009)(8), 084043.

[36] T. Damour, P. Jaranowski, and G. Schaefer. Dimensional regularization of the gravitational interaction of point masses. *Phys. Lett.* **B513** (2001), 147. arXiv:gr-qc/0105038.

[37] L. Blanchet, T. Damour, and G. Esposito-Farèse. Dimensional regularization of the third post-Newtonian dynamics of point particles in harmonic coordinates. *Phys. Rev.* **D69** (2004), 124007.

[38] T. Damour, P. Jaranowski, and G. Schäfer. Equivalence between the ADM-Hamiltonian and the harmonic-coordinates approaches to the third post-Newtonian dynamics of compact binaries. *Phys. Rev.* **D63** (2001), 044021.

[39] V. C. de Andrade, L. Blanchet, and G. Faye. Third postNewtonian dynamics of compact binaries: Noetherian conserved quantities and equivalence between the harmonic coordinate and ADM Hamiltonian formalisms. *Class. Quant. Grav.* **18** (2001), 753. arXiv:gr-qc/0011063.

[40] L. Blanchet and B. R. Iyer. Third postNewtonian dynamics of compact binaries: Equations of motion in the center-of-mass frame. *Class. Quant. Grav.* **20** (2003), 755. arXiv:gr-qc/0209089.

[41] Y. Itoh and T. Futamase. New derivation of a third post-Newtonian equation of motion for relativistic compact binaries without ambiguity. *Phys. Rev.* **D68** (2003), 121501.

[42] L. Blanchet, T. Damour, G. Esposito-Farèse, and B. R. Iyer. Gravitational Radiation from Inspiralling Compact Binaries Completed at the Third Post-Newtonian Order. *Phys. Rev. Lett.* **93** (2004), 091101.

[43] L. Blanchet, T. Damour, G. Esposito-Farèse, and B. R. Iyer. Dimensional regularization of the third post-Newtonian gravitational wave generation from two point masses. *Phys. Rev.* **D71** (2005), 124004.

[44] L. Blanchet, B. R. Iyer, and B. Joguet. Gravitational waves from inspiraling compact binaries: Energy flux to third post-Newtonian order. *Phys. Rev.* **D65** (2002), 064005.

[45] L. Blanchet, G. Faye, B. R. Iyer, and B. Joguet. Gravitational-wave inspiral of compact binary systems to 7/2 post-Newtonian order. *Phys. Rev.* **D65** (2002), 061501.

[46] L. Blanchet and B. R. Iyer. Hadamard regularization of the third post-Newtonian gravitational wave generation of two point masses. *Phys. Rev.* **D71** (2005), 024004. arXiv:gr-qc/0409094.

[47] P. C. Peters and J. Mathews. Gravitational Radiation from Point Masses in a Keplerian Orbit. *Phys. Rev.* **131** (1963), 435.

[48] P. C. Peters. Gravitational Radiation and the Motion of Two Point Masses. *Phys. Rev.* **136** (1964), B1224.

[49] A. Buonanno and T. Damour. Effective one-body approach to general relativistic two-body dynamics. *Phys. Rev.* **D59** (1999), 084006.

[50] A. Buonanno, Y. Pan, J. G. Baker, J. Centrella, *et al.* Approaching faithful templates for nonspinning binary black holes using the effective-one-body approach. *Phys. Rev.* **D76** (2007)(10), 104049.

[51] 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.* **D62** (2000), 084011.

[52] 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.* **D84** (2011), 104027. arXiv:1107.2904.

[53] 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.* **D49** (1994), 6274.

[54] L. E. Kidder. Coalescing binary systems of compact objects to postNewtonian 5/2 order. 5. Spin effects. *Phys. Rev.* **D52** (1995), 821. arXiv:gr-qc/9506022.

[55] L. Blanchet, A. Buonanno, and G. Faye. Higher-order spin effects in the dynamics of compact binaries. II. Radiation field. *Phys. Rev.* **D74** (2006), 104034.

[56] T. Damour. Coalescence of two spinning black holes: An effective one-body approach. *Phys. Rev.* **D64** (2001), 124013.

[57] E. Poisson. Gravitational waves from inspiraling compact binaries: The quadrupole-moment term. *Phys. Rev.* **D57** (1998), 5287.

[58] S. Husa, S. Khan, M. Hannam, M. PÃ¼rrer, *et al.* Frequency-domain gravitational waves from nonprecessing black-hole binaries. I. New numerical waveforms and anatomy of the signal. *Phys. Rev.* **D93** (2016)(4), 044006. arXiv:1508.07250.

[59] P. Schmidt, M. Hannam, and S. Husa. Towards models of gravitational waveforms from generic binaries: A simple approximate mapping between precessing and non-precessing inspiral signals. *Phys. Rev.* **D86** (2012), 104063. arXiv:1207.3088.

[60] M. Hannam, P. Schmidt, A. Bohé, L. Haegel, *et al.* Simple Model of Complete Precessing Black-Hole-Binary Gravitational Waveforms. *Phys. Rev. Lett.* **113** (2014)(15), 151101. arXiv:1308.3271.

[61] P. Schmidt, F. Ohme, and M. Hannam. Towards models of gravitational waveforms from generic binaries: II. Modelling precession effects with a single effective precession parameter. *Phys. Rev.* **D91** (2015), 024043.

[62] P. Schmidt, M. Hannam, S. Husa, and P. Ajith. Tracking the precession of compact binaries from their gravitational-wave signal. *Phys. Rev.* **D84** (2011), 024046.

[63] J. Veitch and A. Vecchio. Bayesian coherent analysis of in-spiral gravitational wave signals with a detector network. *Phys. Rev.* **D81** (2010), 062003.

[64] J. Veitch *et al.* Parameter estimation for compact binaries with ground-based gravitational-wave observations using the LALInference software library. *Phys. Rev.* **D91** (2015)(4), 042003. arXiv:1409.7215.

[65] 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.

[66] T. G. F. Li, W. Del Pozzo, S. Vitale, C. Van Den Broeck, *et al.* Towards a generic test of the strong field dynamics of general relativity using compact binary coalescence. *Phys. Rev.* **D85** (2012), 082003.

[67] T. Li, W. Del Pozzo, S. Vitale, C. Van Den Broeck, *et al.* Towards a generic test of the strong field dynamics of general relativity using compact binary coalescence: Further investigations. *J. Phys. Conf. Ser.* **363** (2012), 012028. arXiv:1111.5274.

[68] M. Agathos, W. Del Pozzo, T. G. F. Li, C. Van Den Broeck, *et al.* TIGER: A data analysis pipeline for testing the strong-field dynamics of general relativity with gravitational wave signals from coalescing compact binaries. *Phys. Rev.* **D89** (2014), 082001.

[69] J. Meidam, M. Agathos, C. Van Den Broeck, J. Veitch, *et al.* Testing the no-hair theorem with black hole ringdowns using TIGER. *Phys. Rev.* **D90** (2014), 064009. arXiv:1406.3201.

[70] C. Cutler, T. A. Apostolatos, L. Bildsten, L. S. Finn, *et al.* The last three minutes: Issues in gravitational-wave measurements of coalescing compact binaries. *Phys. Rev. Lett.* **70** (1993), 2984.

[71] B. P. Abbott *et al.* Observation of Gravitational Waves from a Binary Black Hole Merger. *Phys. Rev. Lett.* **116** (2016)(6), 061102. arXiv:1602.03837.

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

[73] N. A., R. Robie, F. Robinet, A. Rocchi, *et al.* Tests of General Relativity with GW150914. *Phys. Rev. Lett.* **116** (2016), 221101.

[74] S. Gossan, J. Veitch, and B. S. Sathyaprakash. Bayesian model selection for testing the no-hair theorem with black hole ringdowns. *Phys. Rev.* **D85** (2012), 124056.

[75] W. Israel. Event horizons in static vacuum space-times. *Phys. Rev.* **164** (1967), 1776.

[76] W. Israel. Event horizons in static electrovac space-times. *Commun. Math. Phys.* **8** (1968), 245.

[77] S. Hawking. Gravitational radiation from colliding black holes. *Phys. Rev. Lett.* **26** (1971), 1344.

[78] S. Hawking. Black holes in general relativity. *Commun. Math. Phys.* **25** (1972), 152.

[79] B. Carter. Axisymmetric Black Hole Has Only Two Degrees of Freedom. *Phys. Rev. Lett.* **26** (1971), 331.

[80] P. Chrusciel. On completeness of orbits of Killing vector fields. *Class. Quant. Grav.* **10** (1994), 2091. arXiv:9304029.

[81] S. L. Detweiler. Black holes and gravitational waves. III. The resonant frequencies of rotating holes. *Astrophys. J.* **239** (1980), 292.

[82] O. Dreyer, B. J. Kelly, B. Krishnan, L. S. Finn, *et al.* Black hole spectroscopy: Testing general relativity through gravitational wave observations. *Class. Quant. Grav.* **21** (2004), 787. arXiv:gr-qc/0309007.

[83] E. Berti, V. Cardoso, and C. M. Will. Gravitational-wave spectroscopy of massive black holes with the space interferometer LISA. *Phys. Rev.* **D73** (2006)(6), 064030.

[84] E. Berti, J. Cardoso, V. Cardoso, and M. Cavaglia. Matched-filtering and parameter estimation of ringdown waveforms. *Phys. Rev.* **D76** (2007), 104044. arXiv:0707.1202.

[85] I. Kamaretsos, M. Hannam, S. Husa, and B. S. Sathyaprakash. Black-hole hair loss: Learning about binary progenitors from ringdown signals. *Phys. Rev.* **D85** (2012), 024018.

[86] eLISA Science. https://www.elisascience.org/.

[87] S. Yoshida, Y. Eriguchi, and T. Futamase. Quasinormal modes of boson stars. *Phys. Rev.* **D50** (1994), 6235.

[88] M. C. Miller and E. J. M. Colbert. Intermediate - mass black holes. *Int. J. Mod. Phys.* **D13** (2004), 1. arXiv:astro-ph/0308402.

[89] J. M. Fregeau, S. L. Larson, M. C. Miller, R. W. O'Shaughnessy, *et al.* Observing IMBH-IMBH Binary Coalescences via Gravitational Radiation. *Astrophys. J.* **646** (2006), L135. arXiv:astro-ph/0605732.

[90] P. Amaro-Seoane and M. Freitag. Intermediate-mass black holes in colliding clusters: Implications for lower-frequency gravitational-wave astronomy. *Astrophys. J.* **653** (2006), L53. arXiv:astro-ph/0610478.

[91] P. Amaro-Seoane and L. Santamaria. Detection of IMBHs with ground-based gravitational wave observatories: A biography of a binary of black holes, from birth to death. *Astrophys. J.* **722** (2010), 1197. arXiv:0910.0254.

[92] J. R. Gair, I. Mandel, M. C. Miller, and M. Volonteri. Exploring intermediate and massive black-hole binaries with the Einstein Telescope. *Gen. Rel. Grav.* **43** (2011), 485. arXiv:0907.5450.

[93] C. Van Den Broeck. Probing dynamical spacetimes with gravitational waves (2013). arXiv:1301.7291.

[94] M. Agathos, W. Del Pozzo, T. G. F. Li, C. Van Den Broeck, *et al.* Testing general relativity using gravitational waves from binary neutron stars: Effect of spins. In *Proceedings, 13th Marcel Grossmann Meeting on Recent Developments in Theoretical and Experimental General Relativity, Astrophysics, and Relativistic Field Theories (MG13)*, pages 1710–1712. 2015. arXiv:1305.2963.

[95] S. A. Teukolsky. Rotating Black Holes: Separable Wave Equations for Gravitational and Electromagnetic Perturbations. *Phys. Rev. Lett.* **29** (1972), 1114.

[96] S. A. Teukolsky. Perturbations of a Rotating Black Hole. I. Fundamental Equations for Gravitational, Electromagnetic, and Neutrino-Field Perturbations. *Astrophys. J.* **185** (1973), 635.

[97] I. Kamaretsos. From black holes to their progenitors: A full population study in measuring black hole binary parameters from ringdown signals. *Journal of Physics: Conference Series* **363** (2012)(1), 012047.

[98] L. London, D. Shoemaker, and J. Healy. Modeling ringdown: Beyond the fundamental quasinormal modes. *Phys. Rev.* **D90** (2014)(12), 124032. arXiv:1404.3197.

[99] I. Kamaretsos, M. Hannam, and B. S. Sathyaprakash. Is Black-Hole Ringdown a Memory of Its Progenitor? *Phys. Rev. Lett.* **109** (2012), 141102.

[100] L. Rezzolla, E. Barausse, E. N. Dorband, D. Pollney, *et al.* Final spin from the coalescence of two black holes. *Phys. Rev.* **D78** (2008), 044002.

[101] E. Barausse and L. Rezzolla. Predicting the direction of the final spin from the coalescence of two black holes. *Astrophys. J.* **704** (2009), L40. arXiv:0904.2577.

[102] E. Berti and M. Volonteri. Cosmological black hole spin evolution by mergers and accretion. *Astrophys. J.* **684** (2008), 822. arXiv:0802.0025.

[103] J. Veitch and A. Vecchio. Bayesian approach to the follow-up of candidate gravitational wave signals. *Phys. Rev.* **D78** (2008), 022001.

[104] J. Veitch and A. Vecchio. Assigning confidence to inspiral gravitational wave candidates with Bayesian model selection. *Class. Quant. Grav.* **25** (2008), 184010. arXiv:0807.4483.

[105] J. Aasi, J. Abadie, B. P. Abbott, R. Abbott, *et al.* Parameter estimation for compact binary coalescence signals with the first generation gravitational-wave detector network. *Phys. Rev.* **D88** (2013), 062001.

[106] W. Del Pozzo, T. G. F. Li, M. Agathos, C. Van Den Broeck, *et al.* Demonstrating the Feasibility of Probing the Neutron-Star Equation of State with Second-Generation Gravitational-Wave Detectors. *Phys. Rev. Lett.* **111** (2013), 071101.

[107] P. Pani, V. Cardoso, L. Gualtieri, E. Berti, *et al.* Perturbations of slowly rotating black holes: Massive vector fields in the Kerr metric. *Phys. Rev.* **D86** (2012), 104017.

[108] V. Vitagliano, T. P. Sotiriou, and S. Liberati. Dynamics of generalized Palatini theories of gravity. *Phys. Rev.* **D82** (2010), 084007.

[109] H. A. Buchdahl. Representation of the Einstein-Proca Field by an A(a)*. *J. Phys.* **A12** (1979), 1235.

[110] N. Yunes and F. Pretorius. Fundamental theoretical bias in gravitational wave astrophysics and the parametrized post-Einsteinian framework. *Phys. Rev.* **D80** (2009), 122003.

[111] W. Del Pozzo, J. Veitch, and A. Vecchio. Testing general relativity using Bayesian model selection: Applications to observations of gravitational waves from compact binary systems. *Phys. Rev.* **D83** (2011), 082002.

[112] S. Vitale and W. Del Pozzo. How serious can the stealth bias be in gravitational wave parameter estimation? *Phys. Rev.* **D89** (2014), 022002.

[113] L. Sampson, N. Cornish, and N. Yunes. Mismodeling in gravitational-wave astronomy: The trouble with templates. *Phys. Rev.* **D89** (2014), 064037.

[114] P. Amaro-Seoane, S. Aoudia, S. Babak, P. Binétruy, *et al.* eLISA: Astrophysics and cosmology in the millihertz regime. *GW Notes, Vol. 6, p. 4-110* **6** (2013), 4. arXiv:1201.3621.

[115] J. Aasi *et al.* Search for gravitational waves from binary black hole inspiral, merger, and ringdown in LIGO-Virgo data from 2009-2010. *Phys. Rev.* **D87** (2013)(2), 022002. arXiv:1209.6533.

[116] S. Chatterji, L. Blackburn, G. Martin, and E. Katsavounidis. Multiresolution techniques for the detection of gravitational-wave bursts. *Class. Quant. Grav.* **21** (2004), S1809. arXiv:gr-qc/0412119.

[117] S. Chatterji. The search for gravitational wave bursts in data from the second LIGO science run (2005).

[118] L. Sampson, N. Cornish, and N. Yunes. Gravitational Wave Tests of Strong Field General Relativity with Binary Inspirals: Realistic Injections and Optimal Model Selection. *Phys. Rev.* **D87** (2013)(10), 102001. arXiv:1303.1185.

[119] B. P. Abbott *et al.* Properties of the binary black hole merger GW150914 (2016). https://dcc.ligo.org/LIGO-P1500218/public/main, arXiv:1602.03840.

[120] C. M. Will. The Confrontation between General Relativity and Experiment. *Living Reviews in Relativity* **17** (2014), 4. arXiv:1403.7377.

[121] E. Berti *et al.* Testing General Relativity with Present and Future Astrophysical Observations. *Class. Quant. Grav.* **32** (2015), 243001. arXiv:1501.07274.

[122] N. Yunes and X. Siemens. Gravitational-Wave Tests of General Relativity with Ground-Based Detectors and Pulsar Timing-Arrays. *Living Rev. Rel.* **16** (2013), 9. arXiv:1304.3473.

[123] SXS Gravitational Waveform Database. https://www.black-holes.org/waveforms/.

[124] L. Blanchet and G. Schaefer. Gravitational wave tails and binary star systems. *Class. Quant. Grav.* **10** (1993), 2699.

[125] L. Blanchet. Gravitational wave tails of tails. *Class. Quant. Grav.* **15** (1998), 113. [Erratum: Class. Quant. Grav.22,3381(2005)], arXiv:gr-qc/9710038.

[126] J. Lense and H. Thirring. Über den Einfluß der Eigenrotation der Zentralkörper auf die Bewegung der Planeten und Monde nach der Einsteinschen Gravitationstheorie. *Physikalische Zeitschrift* **19** (1918), 156.

[127] B. M. Barker and R. F. O'Connell. Gravitational two-body problem with arbitrary masses, spins, and quadrupole moments. *Phys. Rev.* (1975), 329.

[128] C. V. Vishveshwara. Scattering of Gravitational Radiation by a Schwarzschild Black-hole. *Nature* **227** (1970), 936.

[129] W. H. Press. Long Wave Trains of Gravitational Waves from a Vibrating Black Hole. *Astrophys. J.* **170** (1971), L105.

[130] S. Chandrasekhar and S. Detweiler. The quasi-normal modes of the Schwarzschild black hole. *Proc. R. Soc. Lond.* **A344** (1975), 441.

[131] A. Ghosh *et al.* Testing general relativity using golden black-hole binaries. *Phys. Rev.* **D94** (2016)(2), 021101. arXiv:1602.02453.

[132] N. Wex. Testing Relativistic Gravity with Radio Pulsars. *ArXiv e-prints* (2014). arXiv:1402.5594.

[133] N. Yunes and S. A. Hughes. Binary pulsar constraints on the parametrized post-Einsteinian framework. *Phys. Rev.* (2010)(8), 082002. arXiv:1007.1995.

[134] L. Blanchet and T. Damour. Hereditary effects in gravitational radiation. *Phys. Rev.* **D46** (1992), 4304.

[135] E. Poisson. Gravitational radiation from a particle in circular orbit around a black hole. 1: Analytical results for the nonrotating case. *Phys. Rev.* **D47** (1993), 1497.

[136] A. G. Wiseman. Coalescing binary systems of compact objects to (post)Newtonian**5/2 order. 4V: The Gravitational wave tail. *Phys. Rev.* **D48** (1993), 4757.

[137] R. J. E. Smith, K. Cannon, C. Hanna, D. Keppel, *et al.* Towards Rapid Parameter Estimation on Gravitational Waves from Compact Binaries using Interpolated Waveforms. *Phys. Rev.* **D87** (2013)(12), 122002. arXiv:1211.1254.

[138] M. Puerrer. Frequency domain reduced order models for gravitational waves from aligned-spin compact binaries. *Class. Quant. Grav.* **31** (2014)(19), 195010. arXiv:1402.4146.

[139] B. D. Lackey, S. Bernuzzi, C. R. Galley, J. Meidam, *et al.* Effective-one-body waveforms for binary neutron stars using surrogate models (2016). arXiv:1610.04742.

[140] M. Hannam, P. Schmidt, A. Bohé, L. Haegel, *et al.* Simple Model of Complete Precessing Black-Hole-Binary Gravitational Waveforms. *Phys. Rev. Lett.* **113** (2014), 151101.

[141] I. M. S. Vinciguerra, J. Veitch. Accelerating gravitational wave parameter estimation with multi-band template interpolation. *In Preparation* (2017).

[142] P. Canizares, S. E. Field, J. R. Gair, and M. Tiglio. Gravitational wave parameter estimation with compressed likelihood evaluations. *Phys. Rev.* **D87** (2013)(12), 124005. arXiv:1304.0462.

[143] R. Smith, S. E. Field, K. Blackburn, C.-J. Haster, *et al.* Fast and accurate inference on gravitational waves from precessing compact binaries. *Phys. Rev.* **D94** (2016)(4), 044031. arXiv:1604.08253.

[144] K. Cannon, J. D. Emberson, C. Hanna, D. Keppel, *et al.* Interpolation in waveform space: enhancing the accuracy of gravitational waveform families using numerical relativity. *Phys. Rev.* **D87** (2013)(4), 044008. arXiv:1211.7095.

[145] K. Cannon, A. Chapman, C. Hanna, D. Keppel, *et al.* Singular value decomposition applied to compact binary coalescence gravitational-wave signals. *Phys. Rev.* **D82** (2010), 044025. arXiv:1005.0012.

[146] K. Cannon, C. Hanna, and D. Keppel. Efficiently enclosing the compact binary parameter space by singular-value decomposition. *Phys. Rev.* **D84** (2011), 084003. arXiv:1101.4939.

[147] K. Cannon, C. Hanna, and D. Keppel. Interpolating compact binary waveforms using the singular value decomposition. *Phys. Rev.* **D85** (2012), 081504. arXiv:1108.5618.

[148] S. E. Field, C. R. Galley, F. Herrmann, J. S. Hesthaven, *et al.* Reduced basis catalogs for gravitational wave templates. *Phys. Rev. Lett.* **106** (2011), 221102. arXiv:1101.3765.

[149] P. Binev, A. Cohen, W. Dahmen, R. A. DeVore, *et al.* Convergence Rates for Greedy Algorithms in Reduced Basis Methods. *SIAM J. Math. Analysis* **43** (2011)(3), 1457.

[150] R. DeVore, G. Petrova, and P. Wojtaszczyk. Greedy Algorithms for Reduced Bases in Banach Spaces. *Constructive Approximation* **37** (2013)(3), 455. ISSN 0176-4276.

[151] GREEDYCPP. https://bitbucket.org/sfield83/greedycpp/.

[152] H. Antil, S. Field, F. Herrmann, R. Nochetto, *et al.* Two-Step Greedy Algorithm for Reduced Order Quadratures. *J. Sci. Comput.* **57** (2013)(3), 604. ISSN 0885-7474.

[153] H. Antil, D. Chen, and S. Field. A note on QR-based model reduction. *In Preparation* (2017).

[154] M. Barrault, Y. Maday, N. C. Nguyen, and A. T. Patera. An 'empirical interpolation' method: application to efficient reduced-basis discretization of partial differential equations. *Comptes Rendus Mathematique* **339** (2004)(9), 667 . ISSN 1631-073X.

[155] Y. Maday, N. C. Nguyen, A. T. Patera, and S. H. Pau. A general multipurpose interpolation procedure: the magic points. *Communications on Pure and Applied Analysis* **8** (2009), 383.

[156] J. S. Hesthaven, B. Stamm, and S. Zhang. Efficient greedy algorithms for high-dimensional parameter spaces with applications to empirical interpolation and reduced basis methods. *ESAIM: Mathematical Modelling and Numerical Analysis* **48** (2014)(01), 259.

[157] S. Fairhurst. Triangulation of gravitational wave sources with a network of detectors. *New J. Phys.* **11** (2009), 123006. [Erratum: New J. Phys.13,069602(2011)], arXiv:0908.2356.

[158] GWTOOLBOX. https://bitbucket.org/jmeidam/gwtoolbox/.

[159] B. P. Abbott *et al.* GW150914: First results from the search for binary black hole coalescence with Advanced LIGO. *Phys. Rev.* **D93** (2016)(12), 122003. arXiv:1602.03839.

[160] I. W. Harry, B. Allen, and B. S. Sathyaprakash. Stochastic template placement algorithm for gravitational wave data analysis. *Phys. Rev.* **D80** (2009), 104014.

[161] D. A. Brown, I. Harry, A. Lundgren, and A. H. Nitz. Detecting binary neutron star systems with spin in advanced gravitational-wave detectors. *Phys. Rev.* **D86** (2012), 084017.

[162] PyCBC documentation. http://ligo-cbc.github.io/pycbc/latest/html/.

[163] B. J. Owen. Search templates for gravitational waves from inspiraling binaries: Choice of template spacing. *Phys. Rev.* **D53** (1996), 6749.

[164] B. S. Sathyaprakash. Filtering post-Newtonian gravitational waves from coalescing binaries. *Phys. Rev.* **D50** (1994), R7111.

Acronyms

BBH binary black hole . . . 24
BNS binary neutron star . . . 24
CBC compact binary coalescence . . . 21
EFE Einstein field equations . . . 3
EIM empirical interpolation method . . . 120
EOB effective one body . . . 26
ET Einstein Telescope . . . 21
GR general relativity . . . 1
GW gravitational wave . . . 13
IMR inspiral-merger-ringdown . . . 36
LSO last stable orbit . . . 22
LVC LIGO-Virgo Collaboration . . . 119
NR numerical relativity . . . 32
NSBH neutron star and black hole . . . 24
PN Post-Newtonian . . . 26
PSD power spectral density . . . 17
QNM quasi-normal modes . . . 59
RB reduced basis . . . 120
ROM reduced order model . . . 119
ROQ reduced order quadrature . . . 119
SNR signal-to-noise ratio . . . 16
SPA stationary phase approximation . . . 29
SVD singular value decomposition . . . 120

The common thread of the research presented in this thesis is performing tests of the strong-field dynamics of the theory of general relativity using gravitational waves. The main goal of this summary is to make sense of the previous sentence.

I will explain these concepts to the interested reader that does not have background in physics and perhaps for those who simply wish to refresh their memory. The theory of general relativity can be rather counter-intuitive and I will only be able to skim its 4-dimensional surface, so please forgive me for being brief on such a widely fascinating and complex topic.

Relativity

Before we get into general relativity (GR) we should first refresh our knowledge of the theory of relativity itself; a theory that everyone is familiar with whether they realize it or not.

First we will need to get some terminology out of way. When I speak of a frame of reference (sometimes I will just say "frame"), think of it as an observer performing his or her physics experiment. A frame of reference is always indicated by 3 spatial coordinates, a time coordinate and possibly some velocity or even acceleration. A frame of reference could for example be a moving car, an elevator, or a stationary room. Certain physical processes may appear differently when observed from different frames of reference. A simple example would be a balloon in a car: The balloon will behave differently in an accelerating car than it does in a parked car. An important special frame of reference is the inertial frame; a frame that is not accelerating. An inertial frame moving at some velocity is said to be in uniform motion. To make sure the laws of physics make sense to any observer, we need rules to transform from one frame to another. When some description of a physical process changes when we do this it is said to be invariant under this particular transformation.

The principle of relativity is not a novel concept in physics. Galileo Galilei already proposed that for the laws of mechanics there is no preferred state of motion for any particular inertial observer. This means that the laws of mechanics are the same for any inertial frame of reference regardless of its velocity. This can easily be made intuitive: Consider opening a bag of cookies upside down, surely you expect the cookies to fall to the ground approximately vertically. Now suppose you are standing in a high speed train travelling at a smooth constant speed of 300 km/h. Once again you clumsily open the bag of cookies upside down. Now, surely you do not expect the cookies to fly in your face at 300 km/h. This is the concept of relativity that is intuitive to us and it is often referred to as Galilean relativity.

In this principle of relativity, absolute velocities are not measurable. Consider once more that you are the observer in the train. If you are unable to look outside and thus unable to see the landscape moving past, you have no way of determining whether you are moving or not. Movement can only be measured relative to other frames of reference or when uniform motion is interrupted. When the train enters a sharp bend or when the tracks are no longer smooth, you are no longer in an inertial

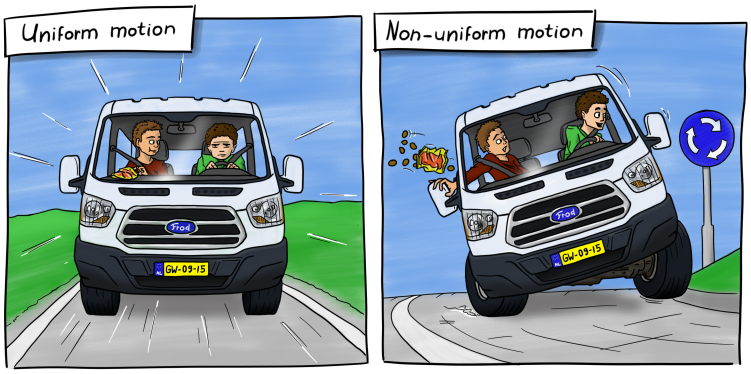


Figure 6.1: left: A car traveling toward us in uniform motion. **right:** The same car going around a bend, exhibiting noticeably non-uniform motion.

frame and this can be measured. Fig. 6.1 demonstrates the difference with a car, and for completeness, another bag of cookies. What you experience is force and Newton describes force to be a result of acceleration and not mere motion. Newton states that as long as an object is not accelerating, no matter how fast it moves, no force acts upon it.

So far we only considered the laws of mechanics (how stuff moves). Things get more interesting when electromagnetism is taken into account. Newton thought that light was made of particles moving at some velocity and following the same laws of mechanics as everything else. Early in the nineteenth century however it was shown that light behaves as waves. This lead to the assumption that light must travel through some medium like sound waves through the air or ripples on the surface of a pond. Another important observation was that James Clerk Maxwell's equations that accurately describe electromagnetism, were not invariant when going from one moving frame of reference to another. Even at low velocities the electric and magnetic fields become messy and nonphysical when we try to do this. Hendrik Lorentz devised a transformation under which Maxwell's equations do behave. His transformation works by deforming the medium through which light was thought to propagate.

The existence of a medium through which light propagates, called the Ether, was later disproved by an experiment performed by Michelson and Morley. This resulted in another significant discovery. Scientists were able to prove that the speed by which light propagates must be the same for all observers: The speed of light (in vacuum) is a constant!

This is where Einstein comes in. He was rethinking the way space and time behave starting with two assumptions: 1) The speed of light is constant for all observers and 2) the laws of physics must be the same for all inertial observers. The second can be rephrased a bit more technically by stating that the laws of physics must be invariant under a transformation between inertial frames. He already knew that the Lorentz transformation was required to achieve this for electromagnetism. Einstein then realized that the one transformation that worked with electromagnetism implied

that the laws of mechanics too must hold under the Lorentz transformation. What followed was what we now know as the special theory of relativity. In the special theory of relativity all the laws of physics hold under a single transformation by allowing not some Ether, but space and time itself to deform.

What makes the theory special is that it only applies to special frames of reference, namely inertial frames. It only applies to observers moving at constant velocities. Einstein realized this and set out to make his theory general. He came up with a brilliant thought experiment.

Consider the following thought experiment: You are standing on a high diving board, as anywhere else on Earth, you notice gravity is pulling you down. Now you step off the diving board and let yourself fall. Suddenly, the sensation of weight, the telltale sign that gravity is pulling you toward the Earth has vanished (forgetting for a moment that you can see you are falling). Of course, the Earth's gravitational pull is still present and people did not suddenly start floating as you stepped off the edge of the diving board (consider the alternative in Fig. 6.2).

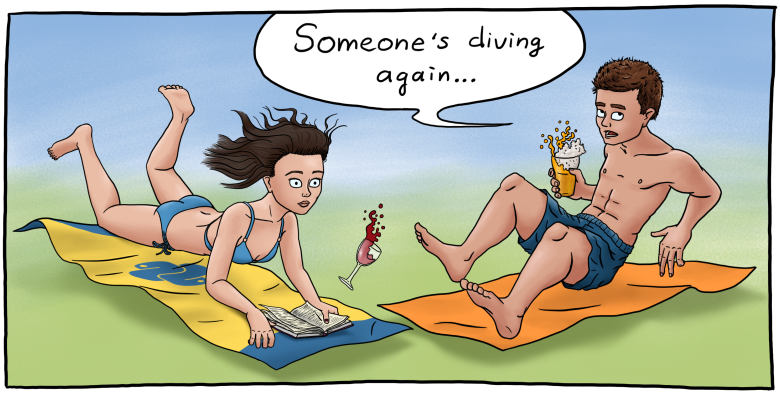


Figure 6.2: When you dive off a diving board you lose the sensation of weight, yet gravity remains. The alternative would result in a strange universe indeed.

Einstein realized how the world must work when he performed a similar thought experiment (his involved someone walking off a roof). He realized that the behaviour of objects in the presence of gravity is indistinguishable from them being in an accelerated frame of reference. This may be a big leap, but as illustrated in Fig. 6.3, it basically comes down to an astronaut not feeling the difference between standing still at the launch site or being accelerated though empty space at $9.8\,\mathrm{m/s^2}$ (roughly the gravitational acceleration at the Earth's surface).

Special relativity already tells us that space and time are connected and should really be described as space-time as a whole. This equivalence between accelerating frames of reference and the presence of gravity is know as the equivalence principle. Putting together the special theory of relativity and the equivalence principle allowed Einstein to formulate the general theory of relativity.

In the general theory of relativity, space-time can be considered as a fabric that twists and bends under the influence of matter and energy. All matter follows the

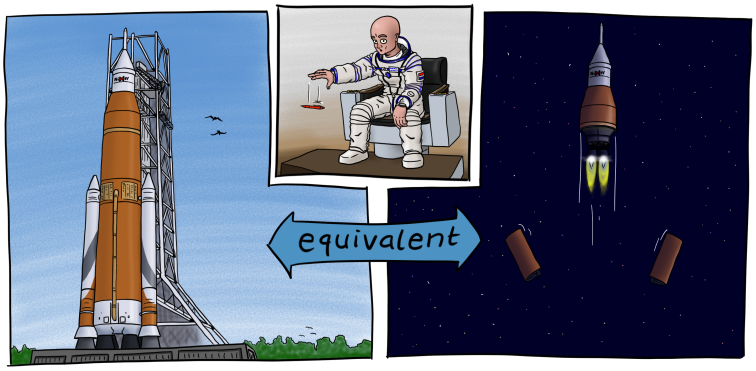


Figure 6.3: An astronaut dropping a pen in two equivalent situations. **left:** In a rocket waiting at the launch site. **right:** In the same rocket, but this time in empty space under acceleration $g = 9.8\,\mathrm{m/s^2}$. The astronaut will find that the pen falls to the floor in precisely the same way in both situations.

curvature of space-time and force is only experienced when something is prevented from doing so. What we experience as gravitational force on Earth is a result of space-time bending under the influence of the mass of the Earth. We follow the curvature but our path is obstructed by the earth itself and we experience force. When stepping off a diving board, we are in free fall and we can follow the curvature of spacetime unimpeded. As soon as we hit the water, our sensation of weight returns as the water now prevents us from following the curvature of spacetime.

Einstein's theory is remarkably elegant (everything it predicts is written in a single short equation) and especially in 1915 when he first presented it to the Prussian Academy of Science, it was unorthodox. Of course, as is the way of science, the theory had to be tested.

First tests of general relativity

General relativity predicts that everything in free-fall – only influenced by gravity and nothing else – follows the curvature of space. Let's use an often considered example to provide an idea of how this works: Imagine a trampoline with a heavy ball at the center. The fabric curves around the ball. This trampoline fabric can be thought of as the fabric of space-time curving under the influence of the mass of the sun, represented by the ball. Now roll a small marble into the pit without hitting the center ball. Its path will be curved. If you roll it just right you can even make it circle around the large ball. The marble can be thought of as the Earth revolving around the sun. Of course, unlike the marble, the Earth will not crash into the sun. In the case of the marble, friction reduces its velocity causing it to spiral into the pit. So Earth follows the curvature of space around the sun and because it has just the right velocity, it keeps going in circles.

Light is also just another form of matter obeying the curvature of space-time. Think back to the trampoline experiment. As long as we give the marble a high enough velocity it will not fall into the gravitational well. Instead, when going really fast, it will bend slightly and zoom past it. As light is no exception, it too will curve very slightly and change direction when passing the sun closely.

This effect – the fact that matter can curve the path of light – is the first prediction of GR that could be tested. Astronomers around the globe set out to be the first to confirm or disprove Einstein's prediction. They observed the apparent change in position of stars as their light bends under the influence of the sun.

Performing these observations was no easy task, but the principle of how they could prove Einstein's predictions can be demonstrated with a few sketches as in Fig. 6.4. As light from a distant star moves past the sun toward us, it bends slightly and the apparent position of the star shifts. Fig. 6.4 shows the positions of a few stars when the sun is not present in the left panel. The middle panel shows how light from a distant star would deviate under the influence of the sun and the right panel shows how this changes our observations of the same stars. The effect is tiny and was only measurable for stars that appeared very close to the edge of the sun. Indeed, under normal circumstances, distant stars cannot be seen right next to the blazing brightness of the sun itself and the observations had to be done during a total solar eclipse. Eddington was the first to provide observations that confirmed the predicted apparent deviation of the positions of the stars by observing them during a solar eclipse.

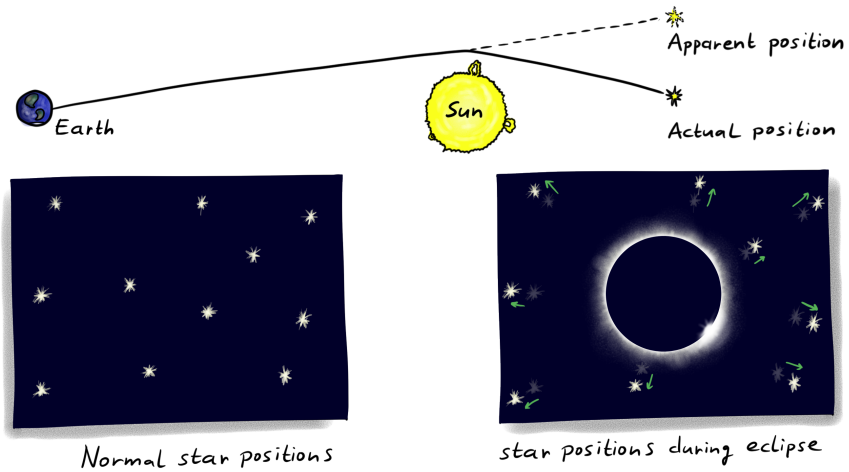


Figure 6.4: An illustration of how the bending of light around a massive object was observed. The top illustration illustrates the path of a light beam originating from a star far away. The bottom left panel shows the positions of a few stars in the night sky. The right panel shows how the position of the stars seem to have shifted when looking at them during a total solar eclipse.

And so general relativity passed its first test. The angle of deviation predicted

was also observed, many times over.

GR has been tested across a wide range of physical processes with great precision and has passed each and every one with flying colors. However, what has been covered so far is only the tip of the proverbial iceberg. Before we get to the part where that makes sense, we first need to take a small detour into the life of a star.

A bit of stellar evolution

As mentioned once or twice, space-time curves under the influence of matter. The more matter, the stronger the curvature. Let's consider that for a moment. Turning back to the trampoline example, we know that when we roll a marble past the central mass at a certain distance, it is its velocity that determines whether it will fly past, orbit, or plunge into the pit. At this point we can leave the trampoline behind and consider a mass forming a gravitational well at which we hurl all sorts of test objects. Fig. 6.5 illustrates a few situations.

If there is more mass at the center, the well will be steeper and an even larger velocity is required to make an object pass it at the same distance. We also know that the same goes for light; it too will follow the curvature of space. In fact, when the mass at the center is sufficiently large, even light, the fastest stuff in our universe can plunge toward the object and never come out. Such an object is called a black hole.

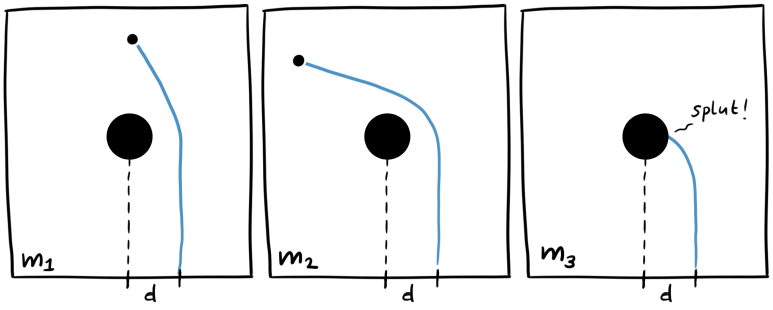


Figure 6.5: Demonstration of the effect of mass on an approaching object. The approaching object moves at the same initial velocity in each picture and passes at the same distance (d). The mass of the central object is different in each panel, getting heavier from left to right. At some point the approaching object does not pass fast enough to escape the central object's gravitational attraction.

Black holes are quite weird. Their masses range from a few times the mass of the sun, to billions of times as much. At the "surface" of the black hole, the escape velocity (the velocity it takes to fly away from the object) is the speed of light. This means that no light can be emitted from it, hence its blackness. To form a black hole, enough mass needs to be compressed into a small enough volume. This is where stellar evolution comes in. It tells us that if a star is massive enough, it will eventually collapse and form a black hole.

Our sun shines because the hydrogen it is predominantly made of, is compressed so much by gravity that the atoms are smashed into each other at the core, producing energy. The energy that is released by this fusion generates an outward pressure. In the sun, and any other star, there is a balance between inward pressure caused by gravity and outward pressure caused by the energy release of fusing atoms (see Fig. 6.6).

Fusion of light elements such as hydrogen, creates heavier elements such as helium and releases energy. These heavier elements can in turn be fused together to produce even heavier elements such as carbon. This chain of fusing elements into heavier elements goes all the way up to iron at which point fusion no longer releases energy, but requires energy. This fusion process therefore ends at this point; a star, such as the sun eventually runs out of fuel[1]. When the sun, or any star, runs out of fuel the outward pressure will be gone and only gravity remains: The star collapses.

There are different forms of collapse depending on the star's mass. A star such as the sun will not collapse all the way and form a compact ball of carbon and oxygen. Heavier stars, around twice the mass the sun will collapse in something that is even more compact: A neutron star. These are balls of matter roughly the size of Amsterdam, but twice as massive as the sun! A drop of this matter would weigh around one billion kilograms. So what about stars that are even more massive?

Given enough mass, in a small enough volume, the density will be so high that no form of matter we know of can counter the gravitational force it produces. Stars exceeding five times the mass of the sun inevitably collapse into a black hole. At this point light can no longer escape its surface. It is interesting to note that for such an object the exact nature of the central mass is hidden from us. Anything happening outside the black hole is described by pure general relativity: Black holes are the ideal laboratory to study GR as no complicated descriptions of matter are required to describe the surrounding space-time.

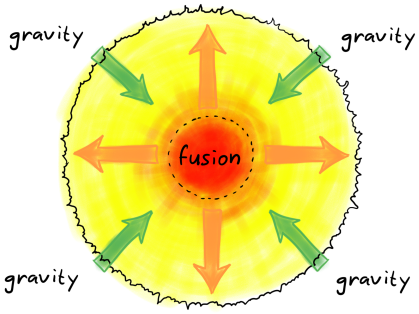


Figure 6.6: The two forces playing in a star: Nuclear fusion generates outward pressure while gravity causes inward pressure. The two balance each other as long as there is fuel for the fusion process.

Weird as they may be, there is strong evidence for the existence of black holes. For example, at the center of our galaxy we can see stars whizzing around some dark

[1]Don't worry we still have a reasonable 4 billion years left before the sun runs out of fuel.

center that must have the mass of a billion suns confined in a small volume.

The reason we only have indirect evidence for the existence of black holes is because these are objects that cannot directly produce electromagnetic radiation (radio, infrared, x-ray, etc...). Yet our ideal laboratory is not completely unobservable: It turns out that GR predicts a form of information that can escape even the deepest gravitational wells.

Gravitational waves

Space-time is dynamic, it can be stretched by a stationary mass, but an accelerating mass will create waves in space-time! We spark our imagination one last time with an analogy closer to home: The surface of a pond. We can also deform the surface of water by placing a little ball on it. Again, the ball represents a mass that is curving space-time, which in turn is represented by the surface of the water. Now move the ball around a bit and you create waves.

Waves on a surface happen when the surface needs time to adjust to changes. When we suddenly remove the ball out of the water, we see an expanding circular wave. The dip in the surface will not disappear instantly. It is said that the surface has a certain stiffness: The stiffer the surface, the faster the wave will travel and the faster the surface adjusts itself to changes.

Something similar happens in the fabric of space-time. When some mass excellerates, the gravitational well it produces follows it along. Of course the place where the object used to be no longer has a well since the mass is now gone. However, like the surface of the pond, space-time has a certain stiffness and its shape can not instantly change from curved to flat. So accelerating masses generate waves in space-time, which we call gravitational waves. Space-time turns out to be incredibly stiff. A lot of mass and energy is required to generate some decent waves and they travel at the speed of light.

Strong evidence supporting the existence of gravitational waves was already found in the 1970s by Russell Alan Hulse en Joseph Hooton Taylor. They observed two neutron stars revolving around each other for many years. They found that the separation between the two objects decreased precisely according to what general relativity predicts. The reason according to GR, is energy being carried away from the system through the emission of gravitational waves.

Recall that a marble on a trampoline could not sustain a perfect orbit due to friction. These two neutron stars revolving around each other are dragging their way through spacetime. So even in empty space such a system experiences something analogous to friction. At some point they will fall into each other. The Earth-Sun system moves much too slow for this to be a serious effect, but the neutron stars in the Hulse-Taylor binary orbit each other every eight hours! Even so, the gravitational waves leaving this system are still too weak to directly see with our gravitational wave detectors on Earth.

The strong-field dynamical regime of GR

The fate of this system of neutron stars – something we call a binary neutron star – is inevitable collision. So what does the future of such a system look like until that moment?

Gravitational waves carry energy away from the system causing the two objects to move closer to each other. This in turn increases their orbital velocity. Now that they are revolving around each other even faster, the system generates stronger waves and loses energy faster. So as a result the objects move into tighter orbits and their velocities increase even more, et cetera. In fact this happens exceedingly fast, resulting in a runaway effect until the two neutron stars merge into a single object. For a system like the Hulse-Taylor binary this will still take about three hundred million years and unfortunately the waves will only be strong enough to detect on Earth for the last minute or so (unless in three hundred million years we are able to just fly there and enjoy the show locally).

Even though we all hope to soon detect gravitational waves originating from such a system, I will focus here on an even more extreme binary system: A binary black hole (BBH), a system where two black holes revolve around each other.

For quite some time already BBH systems were believed to exist, but it was very difficult to estimate how many there would be as there is really no way of seeing them directly. It was an exciting surprise to detect gravitational waves originating from a merging BBH system on September 14th 2015.

Each black hole in this system had a mass of around thirty suns. We witnessed the last ten orbits the black holes made before plunging toward each other releasing an enormous amount of energy as they collided and formed a single black hole. The event that we were able to observe by detecting its gravitational waves has proven to be the most powerful event ever observed by mankind. Space-time was violently stretched, twisted and turned during these final ten orbits in only a fraction of a second. These final moments of such a massive system are unveiling the nature of GR under the most extreme circumstances.

As I mentioned before, black holes are the ideal laboratory to test general relativity and a binary black hole merger shows off all that GR has to offer. Never before have we had such an opportunity.

The first tests of the genuinely strong-field dynamics of general relativity

All tests of GR that we have been able to perform so far have been in either a regime where fields are weak (only subtle curvature in space-time) or where there are only small dynamical aspects (slowly moving and evolving systems).

Particles in close orbit around black holes could be described by approximating GR as the more complex effect in GR are too subtle to notice. Even though these particles are moving in a very strong gravitational field, the system is still static; the mass producing the field is just sitting there. The Hulse-Taylor binary is not static

and does provide a more interesting test of GR, but even in such a system only small corrections to Newtonian mechanics can be measured. The effects of GR in such systems are still subtle and we can not use them to confirm if all the complicated dynamic aspects of GR are also correct.

Before September 2015 we have never been able to observe the strong-field dynamics of GR. But there is a deeper reason behind wanting to observe and investigate these illusive waves: Nature is telling us there is something fishy going on with our current understanding of the universe. GR explains the large aspects of the universe to unprecedented precision, while quantum mechanics shines in explaining the very smallest of things. But here's the funny thing: These two theories are not compatible with each other. GR does not feel comfortable at the quantum scale and quantum mechanics has nothing to say at large scales. Something must be missing. If there is something missing in GR, our only hope to observe it is in its currently unexplored strong-field dynamics!

This is the main topic of this thesis as the subtitle – not so subtly – states. In the remainder of this section I will explain how we have accomplished this with the first detections of gravitational waves.

The shape of a signal

I shall skip the method of detecting gravitational waves as the focus of this thesis is not the detection itself, but rather what physics can be extracted from detected signals. What I do need to mention is that the shape of a wave passing a detector is never clean. Detecting a gravitational wave is done by measuring extremely tiny length differences (a fraction of the size of an atom) in two perpendicular vacuum tubes, typically a few kilometers long. Unfortunately not only the stretching of space-time results in measuring such length differences. Many other sources of vibrations such as tremors in the earth cause the detector's output signal to be messy. The resulting undesired signal that is not caused by gravitational waves is called noise. In Fig. 6.7 I illustrate what a short piece of data from a detector can look like.

We are capable of finding tiny signals buried in all this noise. When a signal is found buried in the noise, we can start working on the stretch of data containing the signal. As mentioned above, such a signal is buried in noise and gets distorted. In order to find it we need to have a pretty good idea of what a signal can look like.

General relativity predicts precisely what a gravitational wave (GW) signal passing our detectors should look like, depending on the type of source that created it. In Fig. 6.8 the shape of such a signal originating from two black holes is illustrated. The two black holes will revolve around each other, slightly getting closer each orbit while their velocity is increased. This increase in velocity means a higher GW frequency. This process continues to pick up the pace. The GW frequency will rise faster and faster as the black holes move toward each other. The GWs will gain strength so the amplitude increases as well. Finally they will plunge into each other forming a single deformed black hole. This remnant black hole will then ring for a while (very much like a bell does when hit with a hammer) until it has settled down and it no longer emits any waves.

The shape of the signal looks simple to the eye, but to accurately describe it,

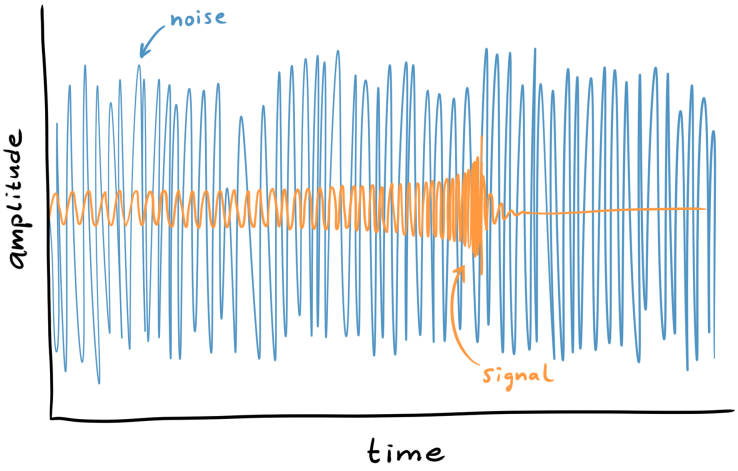


Figure 6.7: A small stretch of what data from a detector would look like containing a signal. The blue messy lines are noise resulting from all sorts of things like seismic vibrations and the tiny red wiggle is a gravitational wave.

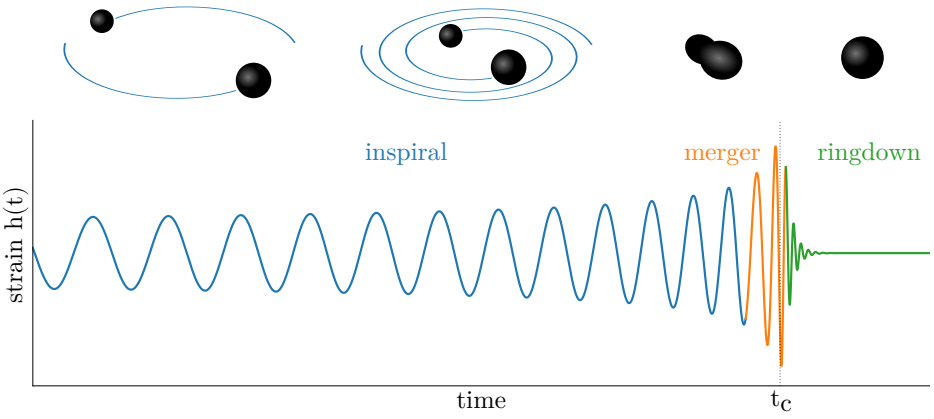


Figure 6.8: The shape of a gravitational wave resulting from two objects revolving around each other closer and closer. Highlighted in blue, orange and green are the three regions that are often referred to as inspiral, merger and ringdown. The time t_c, is roughly where the two objects merge into a single one.

the full complexity of GR is required. Also, the signal will be different for different sources. In fact, its shape depends – in the case of black holes – on the two masses of the black holes m_1 and m_2, and the way they spin around their own axis described by six spin-components that give the direction and how fast they are spinning. Change any of these numbers a bit and the signal will change as well.

Estimating the source parameters

When we analyze the signal and try to measure what parameters the source must have had, we keep tweaking the masses and spins until our tweaked waveform matches the signal in the data. Noise in the data will have distorted the signal, so we can never know precisely what the true masses and spins were, but we can find a range in values that fit the signal best. This range in possible values is a probability distribution for the parameters, based on the data. As this is rather a mouthful I will refer to these as "posteriors".

This range in possible values is called a posterior distribution and is illustrated in Fig. 6.9. The width of this posterior distribution tells us how certain we are about our best match. The wider the distribution, the larger the grain of salt we must take with the most likely value at the top.

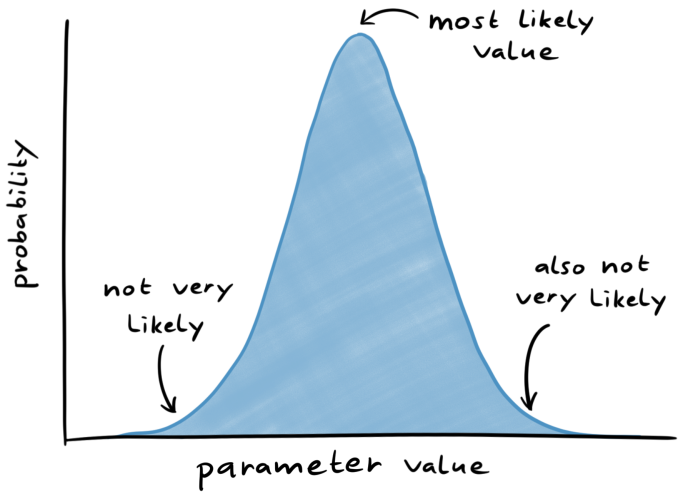


Figure 6.9: The distribution of possible values for a particular parameter after performing parameter estimation on the signal. The distribution, called a posterior, is highest at the most likely value. The width indicates our uncertainty about the most likely value. Values in the tails are unlikely to be the true value of the source.

So for each parameter we have a control knob, we turn them back and forth while we note at each setting how much the tweaked waveform matches the signal. Thankfully we do not have to do this by hand and can use clever algorithms and lots of computing power.

How we performed the tests of GR

The tweaked signal that we keep using is in fact calculated using GR. If GR is the true underlying theory, the only parameters that can change the signal are indeed the masses and spins. If GR is however not correct, it is the theory itself and not only the parameters of the source we should change to get a good match with the signal. There are quite a few theories that try to extend or change GR in some way or another and in principle we could repeat the tweaking process for each of these models and see which one works best. In this thesis however, we do not assume we have any idea where GR may or may not fail. We introduce testing parameters that work very much like the tuning knobs for the source parameters, only these will tweak certain aspects of GR itself.

So if GR is not correct, we should need to tweak it along with the source parameters to properly fit the signal. To do so we add a few testing parameters which we give some fancy labels: $\delta\hat{\varphi}_1$, $\delta\hat{\varphi}_2$, et cetera. These testing parameters are special, as they should all be zero if GR is the correct theory.

We now perform the parameter estimation exercise again, this time also tuning the testing parameters. As we did with the source parameters, we obtain posterior distributions.

These distributions are the main result of this thesis as their implications are significant. As with the posteriors of the source parameters, the width tells how certain we are of our measurements. There are two things we need to be mindful of here: 1) The posterior should overlap with zero for GR to be an acceptable theory to describe the signal and 2) the width of the posterior tells us how much room for error there still is in GR. In other words if the posterior is narrow for parameter $\delta\hat{\varphi}_3$ and its peak is roughly at zero, we are very confident that the part of GR that $\delta\hat{\varphi}_3$ represents is pretty accurate.

The testing parameters are cleverly chosen and in the tests we performed they roughly represent different physical processes that appear in GR. Let's take $\delta\hat{\varphi}_3$ for example. This parameter represents interactions between the spins of the black holes. Only the strong-field dynamics of GR are able to reveal its secrets in the way of gravitational waves. The width of our posterior on that parameter means that we were able to place a bound on its value: This particular effect that GR describes still stands strong. If an alternative theory for gravity is thought up, it has to fit between the bounds we place.

Finally, if we take the posteriors for the testing parameters, put them on their side and show them all next to each other we get Fig. 6.10: The actual results of the first ever tests of the genuinely strong-field dynamics of GR! From these figures alone we can conclude that once again general relativity prevails.

Fig. 6.10 shows some nice constraints for some of the parameters, but others still have a lot of wriggle room. A lot of the posteriors are still rather wide, but these will get narrower as more detections are made and more information is combined. GR is still within the range of acceptable models (all posteriors overlap 0), but this wriggle room could perhaps accommodate a slight change in GR that would make it work at the quantum scale as well.

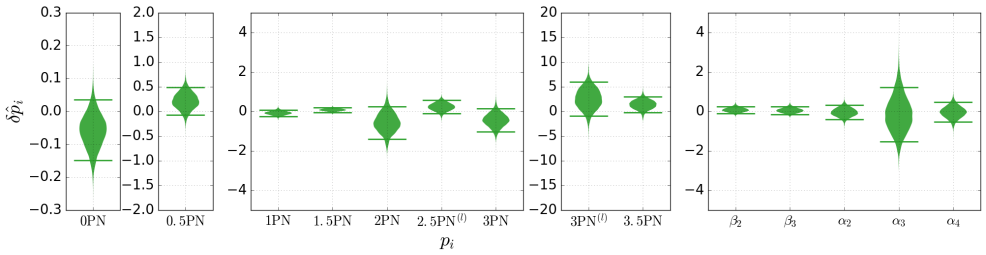


Figure 6.10: These are the posteriors (turned sideways and mirrored) we obtained from tweaking the various testing parameters. The numbers on the vertical axis show fractions; 0.1 means a 10% difference from GR and 0.0 is GR exactly. The results I show here are in fact the combined results from the first two detections. Each posterior shows how certain we can be about how much wriggle room there still is for changes in GR. For example the posterior labelled with 1.5PN is the parameter representing spin interactions between the two black holes as mentioned in the text. We show that this particular effect is pretty well described by GR, there is no more than about 3% room for error anymore. Note that *all* posteriors overlap with zero meaning that our measurements are still consistent with GR!

Speeding things up

In the previous sections I explained that we need to calculate many different waveforms for many values of source parameters and testing parameters to obtain posteriors. This requires a lot of computation. In some cases it took a month of continuous calculation on a super computer to obtain the results. This is a severe bottleneck in performing our tests of GR.

In my thesis I present a method to significantly increase the speed at which we calculate waveforms and evaluate how well they fit the signal. Although the core concept of how this can be done is not new in the field of gravitational waves, it is a challenging task to get it to work with extra testing parameters.

The process by which we can speed up the calculations is actually by being lazy; we only calculate what we absolutely have to. Pretend for a moment that a gravitational wave is determined by only the masses of the two objects that are generating the wave. If we want to find the masses that best fit the signal we see in the detector, it looks like we need to calculate a waveform for every imaginable combination of masses. Maybe the best value for one of the masses is 10, then again it could also be 10.1, or 10.0001. You can probably also imagine that if we calculate the waveform for a mass of 10.0001, we probably will not need to calculate it again for 10.0002. At some point, there is no way to tell the difference.

So clearly we do not need to calculate all possible waveforms, which is a good thing as there are infinitely many possible values for the masses. It turns out that we can cleverly choose a few values for the masses, evaluate their waveforms and use them as a basis for calculating a waveform for any other combination of masses. It is a little bit like a cooking recipe; a few ingredients can be combined into many different meals

depending on the amounts we use for example. The ingredients are the basis of the recipe.

When considering the gravitational waves we can form a set of combinations of masses and calculate their waveforms. These are the ingredients and we'll call them e_i, where i can be anything between 0 and however many ingredients we have. We can then create any other waveform by adding a dash of e_1, a spoon-full e_2, a pinch of e_3, etc.

Finding the proper ingredients, or more appropriately called basis functions for gravitational waves is challenging. It gets more challenging the more parameters we add to the waveform. For testing general relativity we need additional testing parameters on top of the already parameter-heavy waveforms. In this thesis I describe the method by which we can overcome some of the difficulties involved in finding the proper e_i and demonstrate that we can make our tests of GR up to 300 times as efficient!

To give a sense of how we find these e_i, pretend once more that gravitational waves are only described by the masses of the two black holes that generated them. We begin by collecting lots of pairs of mass values. So if we have ten values for the first mass and ten for the other, we have one hundred possible combinations. We calculate the waveform for one of these combinations and call it e_1. We then compare this e_1 with the waveforms corresponding to every other mass combination. The combination that generates the waveform that is most different to e_1, we remember and call e_2. These two waveforms are now a basis. To return to the recipe analogy: Using only two ingredients e_1 and e_2 is just not enough to generate all the recipes we want, so we continue the process of finding more ingredients until we are happy with how accurately we can reproduce any recipe.

This is how we continue building the basis for our waveforms. At each step we compare all the onehundred waveforms with our current basis and remember the one that has the most to add. This process of finding basis functions e_i, that contribute the most to our current basis continues until we are happy with how accurately we can reproduce any waveform we want.

In my thesis I introduce a way to beat the curse of dimensionality: Using ten values for each mass gives one hundred possible combinations, but when we wish to include all parameters (two masses, six spin parameters and one testing parameter), this amount becomes huge with 10^9 possible combinations. That's a ten digit number. All this needs to either be stored in a computer's memory which can add up to well over 200 GB (2 TB when adding only a single additional parameter) or waveforms would need to be calculated over and over again for all these combinations. The latter would quickly result in a computation that takes a lifetime. By making use of existing bases for pure GR waveforms and carefully choosing ranges in parameter values I was able to generate bases for waveforms that also included testing parameters.

The title explained

I hope I have been able to convince the reader that the celebrated theory of general relativity still needs to be put to the test and that the universe may still have some

secrets in store for us. What we have seen of GR is just the tip of the iceberg. Binary black holes spiralling toward each other and collapsing into a single object while emitting gravitational waves is the perfect laboratory to perform the ultimate test of GR.

Black holes can be described by pure GR. As far as the outside observer is concerned there is no matter there, only curved space-time: Black holes way as well be voids to us. We substantiated[2] this void and thereby GR, by performing our first tests of the genuinely strong-field regime and concluding that Einstein's predictions still stand strong.

[2]Substantiate, verb: Provide evidence to support or prove the truth of.

Nederlandse samenvatting

Het doel van het onderzoek dat in deze scriptie gepresenteerd wordt, is het testen van het dynamische sterkeveldenregime van de algemene relativiteitstheorie met behulp van zwaartekrachtsgolven. Het doel van deze samenvatting is de vorige zin bevattelijk te maken.

Relativiteit

Voordat we het kunnen hebben over de algemene relativiteitstheorie (ART), moeten we eerst onze kennis over het concept relativiteit verversen. Relativiteit is een concept waar iedereen mee bekend is, ook al realiseert niet iedereen zich dat.

Een sleutelbegrip van de relativiteitstheorie, is het zogeheten 'referentiekader'. Een referentiekader is het kader waarbinnen een persoon een experiment uitvoert. Een referentiekader wordt beschreven aan de hand van drie ruimtelijke coordinaten, een tijd en eventueel een snelheid of versnelling. Het referentiekader kan bijvoorbeeld een optrekkende auto zijn, een stijgende lift, of een niet-bewegende kamer. Sommige natuurkundige processen kunnen er anders uitzien in verschillende referentiekaders. Een eenvoudig voorbeeld is het gedrag van een ballon. In een optrekkende auto, lijkt een ballon zich anders te gedragen dan in een stilstaande kamer. Een belangrijk speciaal referentiekader is het zogenaamde intertiaalsysteem: een referentiekader dat niet versnelt. Een dergelijk referentiekader dat in beweging is (maar niet versnelt dus) bevindt zich in een staat van eenparige beweging. Omdat de natuurwetten niet mogen verschillen voor observeerders in verschillende referentiekaders, moeten we deze wetten kunnen vertalen van het ene naar het andere kader. Als het blijkt dat een natuurwet na een vertaling verandert, dan is deze niet 'invariant'. Dit zal verderop nog wat duidelijker worden.

Het concept 'relativiteit' is niet nieuw in de wereld van natuurkunde. Galileo Galilei kwam al met de suggestie dat de natuurwetten hetzelfde zijn voor elk inertiaal referentiekader, onafhankelijk van de snelheid waarmee het kader beweegt. Dit kunnen we inzichtelijker maken aan de hand van een dagelijks voorbeeld. Stel je voor dat je een zak met koekjes ondersteboven open maakt (je scheurt de bodem open). Uiteraard verwacht je dat de koekjes recht naar beneden op de grond vallen. Stel je nu voor dat je in een hogesnelheidstrein staat, die zich vloeiend over het spoor voortbeweegt met 300 km/u. Wederom maak je een zak koekjes onhanding ondersteboven open. Ik neem aan dat je nu niet verwacht dat de koekjes met 300 km/u in je gezicht vliegen. Ze vallen gewoon weer naar beneden. Dit is het concept van relativiteit die voor ons intuitief is.

Een gevolg hiervan, is dat absolute snelheden niet meetbaar zijn binnen het referentiekader. Beschouw nogmaals de situatie waar je je in de rijdende trein bevindt. Als je niet naar buiten kunt kijken, en het landschap dus niet voorbij ziet bewegen, is er geen manier om te bepalen of je beweegt of niet. Beweging kan alleen bepaald worden relatief aan andere referentiekaders, of wanneer uniforme beweging wordt onderbroken (zoals wanneer de trein plots remt, en je deze beweging voelt). Wanneer de trein bijvoorbeeld remt, een bocht in gaat, of wanneer het spoor niet langer vlak is, bevindt je je niet langer in een inertiaal referentiekader. Dat kan gemeten wor-

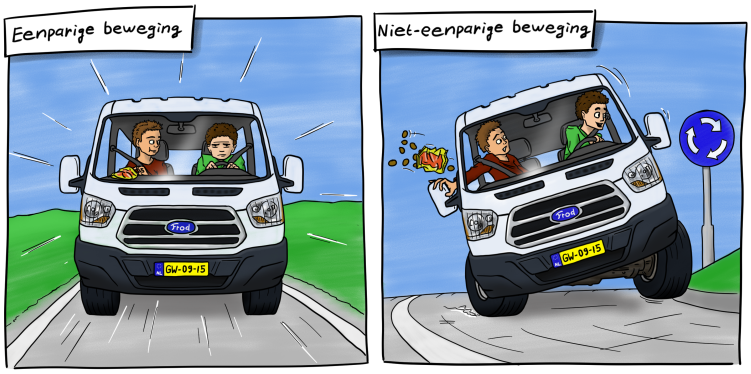


Figure 6.11: links: Een auto in eenparige beweging die op ons af komt. **rechts:** Dezelfde auto die een bocht om gaat, waarbij deze duidelijke niet-eenparige beweging demonstreert.

den. De illustratie in Fig. 6.11 demonstreert dit verschil tussen uniforme beweging en niet-uniforme beweging aan de hand van een rijdende auto en voor de zekerheid een zak met koekjes. Je ervaart in deze situatie een kracht. Newton beschrijft kracht als gevolg van versnelling, en niet enkel snelheid. Newton zegt dat zolang een object niet versnelt, hoe hoog de huidige snelheid ook mag zijn, er geen kracht op wordt uitgeoefend.

Tot nu toe hebben we het alleen gehad over de mechanische natuurwetten (hoe dingen bewegen). Het wordt een stuk interessanter wanneer we electromagnetisme in ons verhaal verwerken. Newton dacht dat licht bestond uit deeltjes die zich met een bepaalde snelheid voortbewegen en die zich net als alle andere objecten, aan de wetten van de mechanica houden. Aan het begin van de negentiende eeuw is echter aangetoond dat licht zich gedraagt als golven. Dit deed men denken dat licht zich door een bepaald medium moest voortbewegen net als geluid door de lucht of golven aan het oppervlak van een vijver.

Een andere belangrijke ontdekking was dat de vergelijkingen van James Clerk Maxwell, die zeer nauwkeurig de electromagnetisme beschrijven, niet invariant waren bij de verplaatsing van een bewegend referentiekader naar een ander. Zelfs bij lage snelheden worden de magnetische velden erg rommelig en onnatuurlijk wanneer wij dit proberen. Hendrik Lorentz heeft vervolgens een transformatie gevonden waarbij de maxwellvergelijkingen zich wel gedragen wanneer we van het ene intertiaalsysteem naar het andere gaan. Zijn transformatieregel werkt doormiddel van het vervormen van het medium, waar men dacht dat licht zich doorheen voortbeweegt.

Het bestaan van dit medium, wat men de ether noemde, werd later weerlegd met een experiment door Michelson en Morley. Dit resulteerde in wederom een significante ontdekking. Wetenschappers konden aantonen dat de snelheid waarmee licht zich voortbeweegt hetzelfde moet zijn voor alle observeerders: de snelheid van licht (in vacuum) is een constante!

Dit is waar Einstein zijn entree maakt. Hij had een hele andere kijk op hoe de

ruimte en tijd zich gedragen. Hij ging uit van twee aannames: 1) de snelheid van het licht is hetzelfde voor alle observeerders en 2) de natuurwetten moeten hetzelfde zijn voor alle inertiele referentiekaders. We kunnen het tweede punt op wat technischere wijze formuleren door te zeggen dat de natuurwetten invariant moeten zijn onder een transformatie tussen inertiaalsystemen.

Het was bekend dat de enige transformatieregel waarbij de wetten van de electromagnetisme invariant blijven, de lorentz-transformatie is. Einstein realiseerde zich dat de enige transformatieregel die werkt voor electromagnetisme ook moet gelden voor alle andere natuurwetten. Hieruit volgde wat wij nu de speciale relativiteitstheorie noemen. In de speciale relativiteitstheorie blijven alle natuurwetten invariant onder de lorentz-transformatie, niet doordat de 'ether' vervormt, maar doordat ruimte en tijd zelf vervormen.

Wat deze theorie 'speciaal' maakt, is dat deze enkel opgaat voor speciale referentiekaders, namelijk de inertiaalsystemen. De speciale relativiteitstheorie kan enkel worden toegepast op observeerders die zich met uniforme snelheid voortbewegen. Einstein realiseerde zich dit uiteraard ook en hij begon direct aan het generaliseren van zijn theorie. Hij bedacht een briljant gedachten-experiment.

Stel je voor: je staat op een duikplank en, net als overal op de aarde, voel je dat de zwaartekracht je naar beneden trekt. Nu stap je van de duikplank af en laat je jezelf vallen. Plotseling merk je dat je het gevoel van gewicht – het teken dat de zwaartekracht je naar de aarde trekt – kwijt bent. Uiteraard is de zwaartekracht niet plotseling uitgeschakeld. Mensen om je heen begonnen niet opeens te zweven toen je van de duikplank afstapte (beschouw het alternatief eens in Fig. 6.12).

Figure 6.12: Wanneer je van een duikplank afstapt voel je je gewichtloos, ook al blijft de zwaartekracht onveranderd. Het alternatief zou een vreemd universum vormen.

Einstein realiseerde zich hoe de wereld werkt dankzij een soortgelijk gedachten-experiment (zijn gedachte-experiment ging echter over iemand die van een dak afliep.) Hij besefte dat het gedrag van een object bij de aanwezigheid van zwaartekracht, niet te onderscheiden is van het gedrag van eenzelfde object dat zich in een versnellend referentiekader bevindt. Dit is misschien een grote denkstap, maar zoals getekend in Fig. 6.13 komt het er simpelweg op neer dat een astronaut het verschil niet kan voelen

tussen het stilstaan bij de lanceerplek op aarde, of het accelereren door de ruimte met $9.8\,\mathrm{m/s^2}$ (ongeveer de zwaartekrachtsversnelling aan het aardoppervlak).

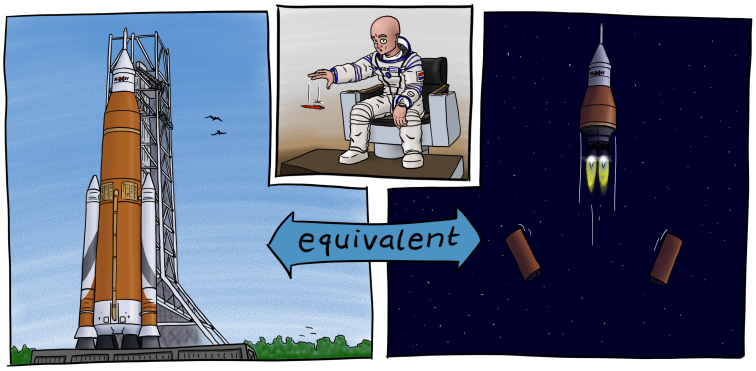


Figure 6.13: Een astronaut laat een pen vallen in twee gelijksoortige (equivalente) situaties. **links:** In een raket die rustig staat te wachten op een lanceerplatform. **rechts:** In dezelfde raket, maar nu in de lege ruimte met een versnelling van $9.8\,\mathrm{m/s^2}$. De astronaut zal de pen in beide situaties op exact dezelfde manier naar de vloer zien vallen.

De speciale relativiteitstheorie vertelt ons al dat ruimte en tijd gekoppeld zijn en dat ze eigenlijk beschreven moeten worden als één geheel: de ruimtetijd. De equivalentie tussen versnellende referentiekaders en de aanwezigheid van zwaartekracht staat bekend als het equivalentieprincipe. Doormiddel van het samenvoegen van deze twee – de speciale relativiteitstheorie en het equivalentieprincipe – kon Einstein zijn algemene relativiteitstheorie ontwikkelen.

Volgens de algemene relativiteitstheorie kun je ruimtetijd zien als een canvas dat kan rekken en buigen onder de invloed van materie en energie. Alle materie volgt de kromming van de ruimtetijd. Een kracht wordt uitgeoefend wanneer iets de materie ervan weerhoudt deze kromming te volgen. Wat wij ervaren als de zwaartekracht, is een gevolg van het feit dat het aardoppervlak ons weerhoudt van het volgen van de kromming van de ruimtetijd. Toen je van de duikplan afstapte, was je even in vrije val en kon je wel ongehinderd de kromming van de ruimte volgen. Zodra je het water inplonst voel je weer dat je gewicht hebt, omdat het water je vrije pad door de ruimte blokkeert.

Einstein's algemene relativiteitstheorie is uitzonderlijk elegant (alles wat het voorspelt, is op te schrijven in één korte formule) en vooral in 1915, toen het voor het eerst werd voorgelegd aan de Pruisische Academie van Wetenschappen, was het vrij onorthodox. Uiteraard, zoals het werkt in de wetenschap, moest ook deze elegante theorie getest worden.

De eerste tests van de algemene relativiteitstheorie

De algemene relativiteitstheorie beschrijft een heelal waarbij alles in vrije val – enkel onder invloed van zwaartekracht – de kromming van de ruimtetijd volgt. Om een idee te geven hoe dit ongeveer werkt maak ik graag gebruik van een bekend voorbeeld: een trampoline met een zware bal in het midden. Je zult zien dat het doek van de trampoline rond de bal buigt. Je kan het oppervlak van de trampoline zien als de ruimtetijd die kromt onder de invloed van de zon, die we in dit voorbeeld hebben vervangen voor een bal. Rol nu een knikker richting de put zonder onze zon te raken. Het pad wat de knikker aflegt is een kromme lijn. Als je de knikker precies op de juiste manier laat rollen kan deze zelfs een rondje om de zon maken.

De knikker kan gezien worden als de aarde die rond de zon beweegt. In tegenstelling tot de knikker zal de aarde natuurlijk niet zomaar richting de zon storten. Bij de knikker gebeurt dit wel, door wrijving. Door de wrijving mindert de knikker namelijk vaart, en zonder genoeg vaart rolt deze de put in. De aarde volgt de kromming van de ruimtetijd rond de zon, en omdat de snelheid precies goed is, en er vrijwel geen vrijwing is, blijft deze rondjes om de zon draaien.

Voor licht gelden precies dezelfde regels als voor alle andere materie. Ook licht volgt dus de kromming van de ruimtetijd. Met dit in gedachten komen we even terug op het voorbeeld van de trampoline. Zolang we de knikker genoeg vaart geven, zal deze niet in de put rollen: zijn pad zal na het passeren van de zon hooguit een beetje zijn afgebogen. Nu we weten dat licht ook de kromming volgt, hebben we een idee wat er met licht gebeurt wanneer het de zon passeert: licht is erg snel en zal dus hooguit een beetje van koers wijzigen.

Dit effect, het feit dat materie het pad van licht kan buigen, is de eerste voorspelling die de ART deed die getest kon worden. Astronomen over de hele wereld zette expedities op om als eerste Einstein's bijzondere voorspelling te bewijzen of weerleggen. Ze keken naar de schijnbare verandering van de positie van sterren, doordat het sterlicht afgebogen werd door de zon. Dit effect is zeer klein. De beste kans om dit te kunnen meten is door naar verre sterren te kijken die net achter de zon vallen. In Fig. 6.14 is geschetst hoe de schijnbare verandering van de positie van deze sterren tot stand komt.

Het uitvoeren van deze observaties was geen simpele taak, maar het idee erachter is eenvoudig toe te lichten met wat schetsjes zoals in Fig. 6.14. Wanneer het licht van een verre ster langs de zon scheert kromt het pad een klein beetje waardoor de positie van de ster lijkt te verschuiven. In het linkse vak van Fig. 6.14 zien we de waargenomen posities van sterren wanneer de zon niet in de buurt is. Het vak in het midden laat zien hoe het licht afkomstig van een verre ster afbuigt onder de invloed van de zon. Het rechter-vak laat zien hoe deze afwijking zich uit in de schijnbaar veranderde posities van dezelfde sterren. Het effect is zeer klein en kon alleen gemeten worden voor sterren die dichtbij de rand van de zon stonden. Uiteraard zijn deze sterren onder normale omstandigheden niet te zien, zo dichtbij de felle zon. Eddington was de eerste die het bewijs leverde door sterposities te observeren tijdens een volledige zonsverduistering.

En zo doorstond de ART zijn eerste test. De afwijking van sterposities die de ART voorspelt, zijn na Eddington's expeditie nog vele malen succesvol gemeten met

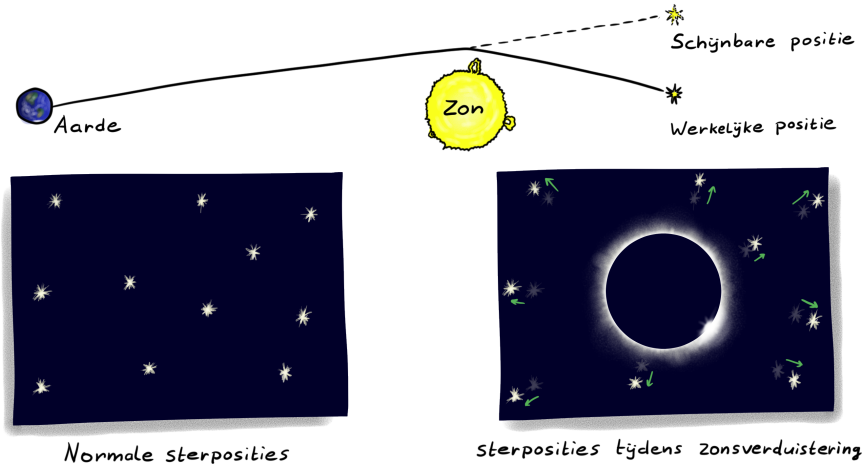


Figure 6.14: Een illustratie van hoe de buiging van licht rond een zwaar object geobserveerd kon worden. De bovenste illustratie laat zien hoe het pad van een lichtstraal afkomstig van een verre ster afgebogen wordt. In het linker-vak zien we een groep sterren aan de nachthemel. Het rechter-vak laat zien hoe de positie van deze sterren schijnbaar verandert wanneer we ze zien tijdens een volledige zonsverduistering.

steeds nauwkeurigere resultaten.

De ART doorstaat al deze, en andersoortige testen moeiteloos, zelfs de meest nauwkeurige experimenten laten geen spoor zien van enige afwijking van de theorie. We hebben echter slechts het topje van de ijsberg gezien, zowel in deze tekst als in de werkelijke expirimenten. Voordat we het kunnen hebben over de rest van de ijsberg zullen we eerst even een klein uitstapje moeten maken naar het leven van een ster.

Een beetje ster-evolutie

We weten nu dat ruimtetijd kromt onder invloed van materie. Hoe meer materie in een stuk volume, hoe sterker de kromming. Dit houden we even in ons achterhoofd. Denk nu weer even terug aan de trampoline. Wanneer we een knikker op een bepaalde afstand langs de grote bal rollen, is het de snelheid die bepaalt of de knikker er langs rolt, om de bal heen draait, of naar het midden toe stort. Laten we nu even afstand nemen van het simplistische voorbeeld van de trampoline en een grote massa beschouwen die een kromming in de ruimte veroorzaakt. Een zwaar object veroorzaakt in feite een gravitationele put. We kunnen in gedachten vanalles op allerlei manieren richting die put schieten. Fig. 6.15 schetst wat situaties.

Wanneer er meer massa is, zal de gravitationele put steiler en dieper worden. In dat geval is er dus meer snelheid nodig om een deeltje op dezelfde afstand langs het object te schieten. We weten inmiddels ook dat licht de kromming van de ruimtetijd volgt. Dat betekent dat wanneer de massa van het centrale object maar groot genoeg

is, zelfs licht, het snelste fenomeen wat wij kennen, naar het object zal vallen. Een dergelijk object wordt een zwart gat genoemd.

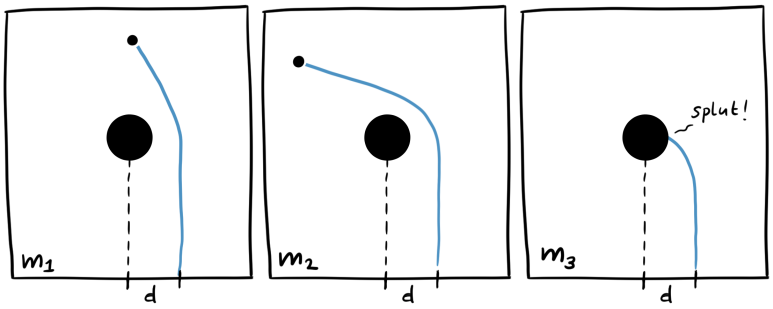


Figure 6.15: Een demonstratie van het effect van massa op een naderend object. Het naderend object beweegt zich voort met dezelfde beginsnelheid in elk plaatje en passeert op dezelfde afstand (d). De centrale massa is in ieder plaatje anders: deze wordt steeds zwaarder van links naar rechts. Op een gegeven moment heeft het naderend object niet genoeg snelheid meer om aan de centrale massa te ontsnappen en zal het er naartoe storten.

Zwarte gaten zijn best vreemd. De massa van zwarte gaten gaat van een paar keer dat van de zon, tot wel miljarden keer zo veel. Aan het oppervlak van een zwart gat is de ontsnappingssnelheid gelijk aan de lichtsnelheid. Dit betekent dat er geen licht kan ontsnappen uit een zwart gat en daarom is deze dus zwart. Om een zwart gat te kunnen vormen, moet er genoeg materie samengedrukt worden in een klein genoeg volume. Dit is waar wij wat begrip van sterevolutie nodig hebben. Sterevolutie vertelt ons dat wanneer een ster zwaar genoeg is, deze uiteindelijk inelkaar zal storten om een zwart gat te vormen.

Onze zon schijnt door het fuseren van waterstof-atomen. De zwaartekracht zorgt ervoor dat de waterstofatomen in het centrum van de zon heel hard tegenelkaar botsen; zo hard dat ze fuseren. Bij de fusie van waterstof komt energie vrij. De energie die naar buiten straalt door fusie in het centrum van de zon is sterk genoeg om de druk van de zwaartekracht, die naar binnen is gericht, te compenseren. In de zon en elke andere ster heerst een evenwicht tussen inwaartse druk als gevolg van zwaartekracht en uitwaartse druk verzoorzaakt door fusie. In Fig. 6.16 is een illustratie van deze balans te zien.

Het fuseren van lichte elementen zoals waterstof produceert zwaardere elementen zoals helium. Deze zwaardere elementen blijven ook tegenelkaar op botsen en produceren nog zwaardere elementen, zoals koolstof en zuurstof. Deze keten aan steeds zwaardere elementen gaat door tot ijzer. Het fuseren van twee ijzer-atomen levert niet langer energie op, maar kost juist energie. Het fusieproces binnen een ster is dus een keer uitgewerkt omdat er geen energie meer vrij komt zodra het centrum van de ster vooral uit ijzer bestaat. Een ster raakt dus door zijn brandstof heen[3]. Op dat

[3]Geen zorgen, we hebben nog een ruime 4 miljard jaar voordat onze zon geen brandstof meer heeft.

moment is er geen uitwaartse druk meer en neemt de zwaartekracht het over: de ster zal instorten.

Er zijn verschillende manieren waarop een ster kan instorten, afhankelijk van de stermassa. Een ster zoals onze zon zal niet helemaal instorten, maar een compacte bal vormen bestaande uit koolstof en zuurstof, een witte dwerg genoemd. Zwaardere sterren, vanaf ongeveer twee keer zo zwaar als de zon, zullen instorten tot een nog compacter object: een neutronenster. Neutronensterren zijn bolletjes met een diameter zo groot als Terschelling, die bestaan uit anderhalf keer zoveel massa als de zon! Een druppel van deze vorm van materie weegt rond een miljard kilogram. Wat gebeurd er met sterren die nog meer massa hebben?

Als er genoeg massa in een klein stukje volume wordt gedrukt, zal de dichtheid zo groot worden dat geen enkele vorm van materie de druk van de zwaartekracht tegen kan gaan die door deze massa veroorzaakt wordt. Sterren die het leven beginnen met een massa van vijf keer of meer dat van de zon zullen uiteindelijk instorten tot een zwart gat. Op dat moment kan licht niet langer van het oppervlak ontsnappen. Het is interessant om te realiseren dat voor een dergelijk object de exacte vorm van de materie verborgen is voor ons: alles wat buiten het zwart gat gebeurt wordt beschreven door pure algemene relativiteitstheorie. Voor de observeerder had het net zo goed een lege oneindig diepe put in de ruimte kunnen zijn. Dit maakt zwarte gaten een ideaal laboratorium om de ART te bestuderen gezien de complexe invloeden van materie geen rol meer spelen.

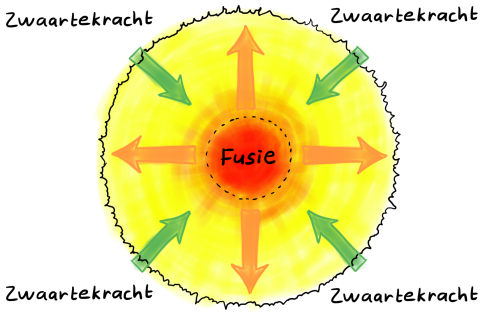


Figure 6.16: De twee belangrijkste krachten die spelen binnen een ster: fusie produceert uitwaartse druk terwijl zwaartekracht in tegenovergestelde richting drukt. De twee krachten houden de ster in balans zolang er brandstof is voor het fusieproces.

Ook al zijn zwarte gaten vreemd, er zijn sterke aanwijzingen voor het werkelijk bestaan van deze objecten. Bij observaties zien we gas draaien rond zware donkere objecten, opgesloten in kleine volumes. Ook zien we in het centrum van ons sterrenstelsel sterren bewegen rond een donkere leegte waarin zich de massa van miljoenen zonnen moet bevinden.

De reden dat we alleen nog aanwijzingen hebben voor het bestaan van zwarte gaten, en geen direct bewijs, is omdat zwarte gaten zelf geen electromagnetische straling kunnen uitzenden. Het blijkt echter dat ons ideaal laboratorium voor de ART niet geheel onzichtbaar is: de ART voorspelt een vorm van straling die zelfs uit

de diepste zwaartekrachtsputten kan ontsnappen om door ons gezien te worden...

Zwaartekrachtsgolven

De ruimtetijd is dynamisch. Het kan uitgestrekt worden door een stilstaande massa, maar een versnellende massa zal zelfs golven veroorzaken in de ruimtetijd! We zullen nogmaals onze verbeelding doen opleven door een analogie waar we ons wat meer bij voor kunnen stellen: het oppervlak van een vijver. We kunnen het oppervlak van een vijver ook een beetje vervormen door er bijvoorbeeld een klein balletje op te leggen. Dit balletje stelt een massa voor die de ruimte vervormt. Als we het balletje nu heen en weer bewegen veroorzaakt dit golven.

Golven aan een oppervlak ontstaan wanneer het oppervlak tijd nodig heeft om zich aan te passen aan veranderingen. Wanneer we plotseling het klein balletje uit het water halen, zien we een circelvormige golf die zich naar buiten verspreidt. Het putje in het oppervlak zal niet onmiddelijk verdwijnen. Men zegt dat het oppervlak een bepaalde stijfheid heeft: hoe groter de stijfheid in het oppervlak, hoe sneller de golf zich zal voortbewegen en hoe sneller het oppervlak zich aanpast aan veranderingen.

Iets vergelijkbaars gebeurt ook in de structuur van de ruimtetijd. Wanneer een massa versnelt door de ruimte, moet de kromming van de ruimtetijd meebewegen. De kromming op de plek waar de massa zich eerst bevond is nu weg omdat daar geen massa meer is. Zoals op het oppervlak van de vijver heeft ook de ruimtetijd een bepaalde stijfheid en kan dus niet onmiddelijk veranderen. Deze verandering kost tijd en net als in de vijver zal ook in de ruimtetijd een versnellende massa golven veroorzaken. De stijfheid van de ruimte blijkt gigantisch te zijn. Er is veel massa en energie nodig om degelijke golven te veroorzaken. De golven bewegen zich voort met de snelheid van het licht.

Een zeer sterke aanwijzing voor het bestaan van zwaartekrachtsgolven is al in de zeventiger jaren gevonden door Russel Alan Hulse en Joseph Hooton Taylor. Zij hebben jaren lang twee neutronensterren geobserveerd die rond elkaar draaien. Ze zagen dat de afstand tussen de twee objecten steeds kleiner werd, precies volgens wat de ART voorspelt. De reden hiervoor, volgens de ART, is dat energie uit het systeem wordt weggedragen door de uitgezonden zwaartekrachtsgolven.

Eerder vertelde ik dat een knikker op een trampoline geen perfecte omwenteling rond de centrale massa kon maken vanwege wrijving met het oppervlak. De twee neutronensterren in het Hulse-Taylor-systeem bewegen door de ruimte. Zelfs in de lege ruimte ervaart een systeem zoals dit iets wat op wrijving lijkt. Hierdoor zullen deze objecten uiteindelijk op elkaar botsen. In het zon-aarde-systeem is dit wrijvingseffect miniem, omdat het zo traag beweegt. Voor het Hulse-Taylor systeem is dit een ander verhaal: deze objecten draaien elke acht uur omelkaar heen! Maar zelfs in dit systeem zijn de zwaartekrachtsgolven lang niet sterk genoeg om op aarde gemeten te kunnen worden.

Het dynamische sterkeveldenregime van de ART

Het lot van het Hulse-Taylor-systeem – en elk ander binair neutronenstersysteem – is een onvoorkombare botsing tussen de twee objecten. Wat gebeurd er tot het zover is?

Zwaartekrachtsgolven dragen energie weg uit het systeem, waardoor de twee objecten dichter omelkaar gaan draaien. Dit zorgt ervoor dat ze sneller om elkaar heen draaien. Nu de objecten nog sneller omelkaar heen draaien, worden er sterkere zwaartekrachtsgolven geproduceerd en dus verliest het systeem nog sneller energie. Het gevolg is dat de objecten wederom dichter en sneller om elkaar heen gaan draaien. Dit produceert weer sterkere golven, enzovoorts. Uiteindelijk klappen de twee objecten met enorme snelheid op elkaar, met een enorme brul aan zwaartekrachtsgolven. Voor het Hulse-Taylor-systeem zal dit nog ongeveer driehonderdmiljoen jaar duren. Helaas voor ons zullen de uitgezonden golven slechts in de laatste minuut sterk genoeg zijn om op aarde te kunnen detecteren (al kunnen wij over driehonderdmiljoen jaar misschien ter plaatse van de show genieten).

We hopen binnekort zwaartekrachtsgolven van dit soort systemen te zien, maar voor de rest van deze samenvatting zullen we ons bezighouden met een nog extremer systeem: binaire zwarte gaten (BZG), waarin twee zwarte gaten om elkaar heen draaien.

Er zijn verschillende theorieën die wijzen op het mogelijke bestaan van BZG-systemen, maar inschatten hoeveel dit er precies zijn (en of ze inderdaad bestaan) is erg moeilijk omdat ze simpelweg niet direct zichtbaar zijn. Het was dan ook een fantastische verrassing om zwaartekrachtsgolven afkomsting van een BZG te detecteren op 14 september 2015.

Beide zwarte gaten in dit systeem hadden een massa van ongeveer dertig maal zo groot als de massa van de zon. We hebben met deze detectie gezien hoe de twee zwarte gaten tien keer omelkaar heen draaiden om vervolgens op elkaar te storten en te eindigen als een enkel zwart gat. De klap was gigantisch en dit was dan ook de meest krachtige gebeurtenis die de mensheid ooit heeft waargenomen. De ruimtetijd rond dit systeem werd met groot geweld gebogen, uitgerekt en gedraaid tijdens de laatste tien omwentelingen, in slechts een fractie van een seconde. De laatste momenten van een systeem zoals deze onthullen de aard van de ART onder de meest extreme omstandigheden.

Zoals eerder genoemd zijn zwarte gaten het ideale laboratorium om de ART te kunnen testen en een samensmelting van een binair zwart gat laat alles zien wat de ART te bieden heeft. Nog nooit eerder hadden wij de mogelijkheid dit te zien.

De eerste tests van het dynamische sterkeveldenregime van de ART

Alle tests die de ART tot nog toe had doorstaan, hadden betrekking op ofwel zwakke zwaartekrachtsvelden (subtiele krommingen in de ruimtetijd) of systemen die relatief traag veranderen (weinig dynamiek).

Deeltjes die dicht rond een zwart gat bewegen bevinden zich in een sterk zwaartekrachtsveld, maar ze zijn prima te beschrijven met slechts een benadering van de ART. De complexe aspecten van de ART zijn hier te subtiel om gezien te worden. Dit systeem is statisch; de massa die de kromming veroorzaakt, hangt daar maar een beetje. Het Hulse-Taylor-systeem daarentegen is niet statisch en geeft dan ook een iets dieper inzicht in de ART, maar zelfs hier zijn slechts kleine correcties ten opzichte van Newtoniaanse mechanica te meten. De effecten als gevolg van de ART zijn in dit soort sytemen nog steeds subtiel en we kunnen ze dan ook niet gebruiken om te testen of alle dynamische aspecten die de ART voorspelt correct zijn.

Voor die speciale ontdekking op 15 september 2015 hebben wij nog niet eerder de mogelijkheid gehad om het dynamische sterkeveldenregime te observeren. Maar er zit nog iets diepers achter ons enthousiasme om juist het meest extreme regime van de ART te willen onderzoeken: er is iets vreemds aan de hand met ons huidige begrip van de het heelal. De ART is in staat het allegrootste in ons heelal tot ongekende nauwkeurigheid te voorspellen, terwijl de quantummechanica dat uitzonderlijk goed doet voor het allerkleinste. Wat er nu zo vreemd is, is dat deze twee theorieën niet goed samenwerken. De ART werkt niet goed op het extreem kleine, terwijl de quantummechanica niets te vertellen heeft over het gedrag van het allergrootste in ons heelal. Iets mist dus aan een van de theorieën en als er iets mist aan de ART, dan zal het antwoord verstopt zijn in het tot nu toe onverkende dynamische sterkeveldenregime.

Het verkennen en testen van dit regime is de hoofdlijn van deze scriptie. In de rest van dit hoofdstuk zal ik toelichten hoe wij dit hebben gedaan met behulp van de eerste detecties van zwaartekrachtsgolven.

De vorm van een zwaartekrachtsgolfsignaal

Ik zal de manier waarop we zwaartekrachtsgolven detecteren overslaan omdat dat niet het onderwerp van deze scriptie is. Het gaat er hier om wat voor fysica we uit de gedetecteerde signalen kunnen halen. Wat wel belangrijk is om te begrijpen, is dat de vorm van het signaal zoals het uit de detector komt nooit schoon is. Het detecteren van een zwaartekrachtsgolf gebeurt doormiddel van het meten van zeer kleine lengteverschillen in twee kilometerslange detectorarmen. Helaas is het niet enkel de vervorming van de ruimtetijd die lengteverschillen in de armen veroorzaken. Allerlei bronnen van trillingen zijn verantwoordelijk voor het vervormen van het signaal, waardoor de gemeten data rommelig zijn. Het signaal wat door alles behalve een zwaartekrachtsgolf veroorzaakt wordt, noemen wij ruis. In Fig. 6.17 is een illustratie te zien van hoe de data afkomstig van een detector eruit kan zien. Het blauwe signaal is ruis en het rode signaal is de zwaartekrachtsgolf die voorbij komt.

We zijn in staat zeer kleine signaaltjes uit de ruis te graven. Wanneer er een signaal gevonden is in de ruis, kunnen wij ons werk beginnen op het stukje data wat het signaal bevat. Het signaal zal sterk vervormen door de ruis en om het te kunnen vinden en er analyses op uit te voeren, moeten we een goed idee hebben van hoe een dergelijk signaal eruit zou moeten zien.

De ART voorspelt precies hoe een zwaartekrachtsgolfsignaal eruit ziet, afhankelijk van de bron die de golf veroorzaakt heeft. Dit noemen we de golfvorm. In Fig. 6.18 is

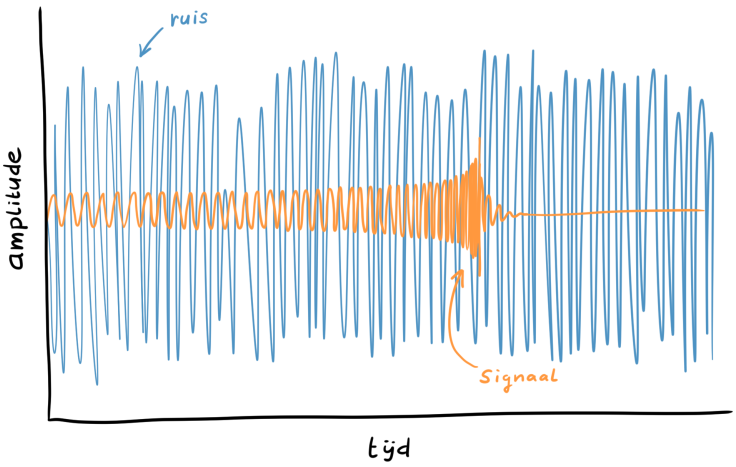


Figure 6.17: Een klein stukje data zoals het eruit zou kunnen zien met een zwaartekrachtsgolfsignaal. De blauwe rommelige lijnen is ruis die altijd aanwezig is in de detector. Het kleine rode golfje is een zwaartekrachtsgolfsignaal verborgen in de ruis.

een signaal afkomstig van een BZG geillustreerd. De twee zwarte gaten beginnen ver uit elkaar. Ze draaien steeds dichter omelkaar heen waardoor snelheid, en daarmee frequentie van de golf, toeneemt. Dit proces gaat steeds sneller. De frequentie neemt sneller en sneller toe terwijl de zwarte gaten steeds dichter bij elkaar komen. Ook de sterkte van het signaal neemt steeds sneller toe. Uiteindelijk vallen de twee zwarte gaten naar elkaar toe en vormen een enkel zwart gat. Dit zwarte gat zal dan nog even galmen (net zoals een bel galmt wanneer je er op slaat met een hamer), totdat de rust in dat stuk van de ruimte is wedergekeerd.

Het signaal varieert van bron to bron. In het geval van twee zwarte gaten hangt de vorm van het signaal af van de twee massas (m_1 en m_2) en de manier waarop ze om hun eigen as draaien, wat beschreven wordt door zes spin-parameters (een parameter voor elke richting in de driedimensionele ruimte per zwart gat). Zodra een van deze acht waardes verandert, zal de golfvorm ook veranderen.

Het meten van systeemparameters

Wanneer wij een signaal analyseren om te achterhalen wat de parameters (de massa's en de spin) van de bron waren, passen we de parameters van ons model aan tot deze overeenkomt met het signaal in de data. Dit proces noemen wij parameterschatting. De ruis in de data zal het signaal vervormen, dus we kunnen nooit exact weten wat de waardes van de parameters echt waren, maar we kunnen wel een reeks waardes vinden die het beste passen bij het gemeten signaal. Deze reeks van mogelijke waardes wordt een 'posterior' of waarschijnlijkheidsverdeling genoemd. In Fig. 6.19 is geschetst hoe

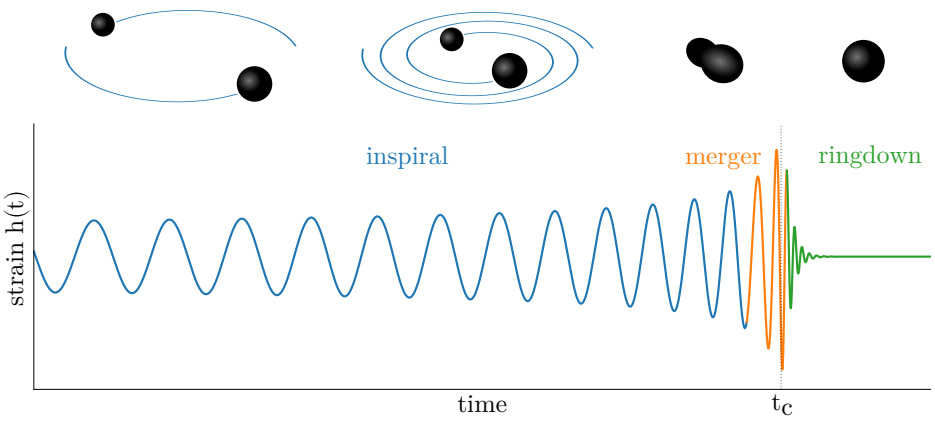


Figure 6.18: De vorm van een zwaartekrachtsgolf afkomstig van twee objecten die steeds dichter om elkaar heen draaien en uiteindelijk op elkaar botsen. Er zijn ruwweg drie fases die gekleurd zijn met blauw, oranje en groen. Er bestaan hier geen Nederlandse termen voor. In het engels worden ze inspiral, merger (samensmelting) en ringdown (het galmen van het overgebleven zwarte gat) genoemd. Het tijdstip t_c is ongeveer waar de twee objecten samen komen tot een enkel object.

een posterior eruit kan zien. De breedte van deze verdeling geeft aan hoe zeker wij kunnen zijn van de best passende waarde voor de parameter. Hoe breder de verdeling, hoe groter het korreltje zout is wat we bij de meest waarschijnlijke waarde in de piek moeten nemen.

Voor elke parameter hebben we een soort verstelknop. We draaien als het ware deze knoppen heen en weer en noteren elke waarde die we proberen, samen met hoe goed de huidige golfvorm past op het werkelijke signaal. Gelukkig hoeven we dit niet met de hand te doen en gebruiken we in werkelijkheid slimme en efficiente algoritmes om in korte tijd enorm veel parameterwaardes te proberen.

Hoe we de tests van de ART hebben uitgevoerd

De golfvorm die we steeds aanpassen, wordt berekend door gebruik te maken van de ART. Als de ART inderdaad de correcte onderliggende theorie is, dan zouden enkel de massas en spinwaardes een effect mogen hebben op de vorm van het signaal. Klopt de ART echter niet met wat er in de natuur gebeurt, dan zouden we de theorie zelf moeten aanpassen – en dus niet alleen de parameters van de bron – om de golfvorm overeen te laten komen met de detectie.

Er zijn veel theorieën die de ART proberen uit te breiden of te veranderen. We zouden in principe de parameterschatting kunnen uitvoeren voor elk van deze theorieën en kijken welke het best overeenkomt met de detectie. Wij doen dit echter anders. We gaan er bijvoorbeeld niet van uit dat we nu al weten op welke punten de ART wel of niet faalt. Om de theorie zo algemeen mogelijk te testen, introduceren we nog een reeks verstelknoppen waarvan elk overeenkomt met een bepaald aspect van

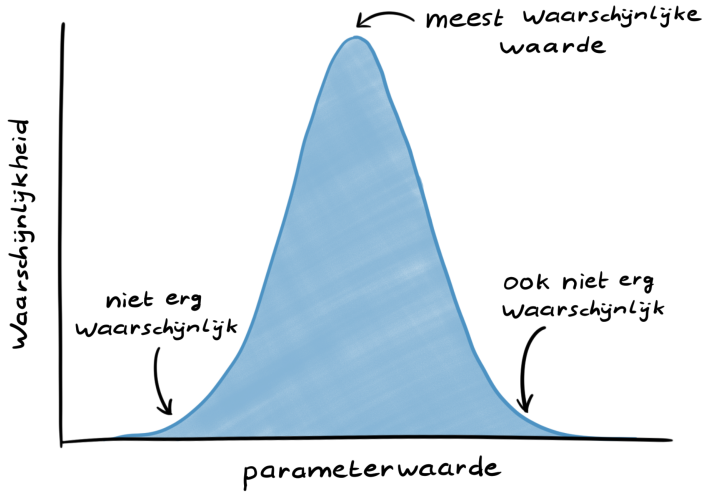


Figure 6.19: De waarschijnlijkheidsverdeling van mogelijke waardes voor een bepaalde parameter na het uitvoeren van parameterschatting. Deze verdeling, die ook wel 'posterior' wordt genoemd, is het hoogst bij de meest waarschijnlijke waarde. De breedte geeft aan hoe onzeker wij zijn over de meest waarschijnlijke waarde. Waardes in de twee staarten van de verdeling komen vrijwel zeker niet overeen met de werkelijke waarde van de parameter.

de ART.

Dus, als de ART niet klopt, dan zouden we deze extra verstelknoppen ook moeten verzetten om onze golfvorm overeen te laten komen met de detectie. De verstelknoppen, ofwel testparameters geven we interessante labels: $\delta\hat{\varphi}_1$, $\delta\hat{\varphi}_2$, enzovoorts. De testparameters zijn speciaal: hun waardes moeten allemaal gelijk aan nul zijn als de ART inderdaad correct is.

We voeren nu nogmaals de parameterschatting uit, maar dit keer nemen we ook de testparameters mee in het proces. Net als met de andere parameters krijgen we ook voor de testparameters posteriors (waarschijnlijkheidsverdelingen).

De posteriors die wij vinden voor de testparameters zijn een van de belangrijkste resultaten van het werk in deze scriptie, omdat hun implicaties groot zijn. Net als bij de posteriors van de bronparameters vertelt de breedte iets over hoe zeker wij zijn van onze metingen. Er zijn nu twee dingen waar we op moeten letten: 1) de posteriors moeten overlappen met nul wil de ART een aannemelijke theorie zijn voor het beschrijven van de detectie en 2) de breedte van de posteriors vertelt ons hoeveel ruimte er nog is voor eventuele veranderingen aan de theorie. We kunnen bijvoorbeeld zeggen dat als de posterior voor $\delta\hat{\varphi}_3$ erg smal is en de piek dichtbij nul ligt, wij zeker kunnen zijn dat het aspect van de ART wat overeenkomt met $\delta\hat{\varphi}_3$ zeer nauwkeurig is.

De testparameters zijn slim gekozen zodat ze bijna allemaal een bepaald fysisch proces binnen de ART representeren. Beschouw als voorbeeld $\delta\hat{\varphi}_3$. Deze parameter

representeert interacties tussen de spins (hoe een object om zijn as draait) van de twee objecten in een binair systeem. De enige manier om afwijkingen te kunnen zien in dit fysisch process, is doormiddel van het observeren van zwaartekrachtsgolven uit een systeem wat zich in het dynamisch sterkeveldenregime bevindt. Alleen in dat regime worden de observeerbare effecten sterk genoeg.

De breedte van de posterior die wij zien bij het meten van de testparameter betekent dat we in staat zijn om een begrenzing te bepalen op mogelijke afwijkingen van wat de ART voorspelt. Dit wil zeggen dat wanneer er een alternatieve theorie voor zwaartekracht bedacht wordt, deze moet passen binnen de begrenzingen die wij hebben bepaald.

Als we nu al de posteriors voor de testparameters op hun kant draaien en naast elkaar weergeven, dan krijgen we Fig. 6.20. Dit is het werkelijke gepubliceerde resultaat van de allereerste tests van het dynamisch sterkeveldenregime van de ART! We kunnen uit deze figuren concluderen dat de ART nog steeds sterk in zijn schoenen staat en overeen-komt met wat wij in de natuur zien gebeuren.

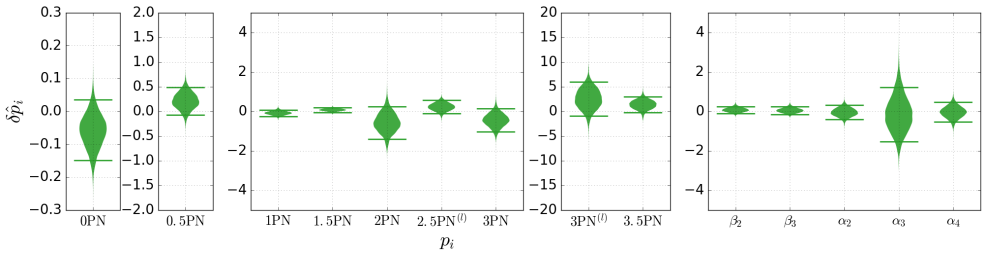


Figure 6.20: Dit zijn de posteriors (90 graden gekanteld en gespiegeld) die wij hebben gevonden met de parameterschatting van de testparameters. De waardes op de verticale as geven fracties aan: 0,1 betekent een 10% verschil met de ART voor de parameter in kwestie en 0,0 betekent geen verschil. De resultaten die ik hier laat zien zijn de gecombineerde resultaten van de twee eerste detecties. Elke posterior geeft aan hoeveel ruimte er nog is voor afwijkingen van de ART. De parameter die de label '1.5PN' heeft, komt bijvoorbeeld overeen met spin-interacties tussen de twee zwarte gaten. We laten hiermee zien dat dit aspect niet meer dan 3% kan afwijken van de ART. Merk op dat *alle* posteriors overlappen met 0,0 wat wil zeggen dat onze metingen consistent zijn met de ART!

In Fig. 6.20 zien we dat we al redelijk sterke begrenzingen hebben kunnen opleggen aan een aantal van de testparameters, maar er is nog steeds veel ruimte voor eventuele afwijkingen van de ART. Sommige posteriors zijn nog steeds vrij breed, maar deze zullen smaller worden naarmate meer informatie van detecties wordt toegevoegd. Wie weet zit binnen de huidige beweegruimte wel het geheim van de stap tussen de quantumwereld en de ART.

Alles wat sneller maken

In de voorgaande sectie heb ik uitgelegd dat we erg veel verschillende zwaartekrachtsgolfsignalen moeten uitrekenen om parameterschattingen te kunnen doen en dus om de posteriors op te kunnen bouwen. Dit vereist uiteraard een hoop rekenkracht. In sommige gevallen duurde het een maand van continue berekeningen op een supercomputer (vergelijkbaar met een stuk of duizend standaard-laptops). Dit is een grote drempel die het testen van de ART moeilijk maakt.

In mijn scriptie stel ik een methode voor die het uitrekenen van een zwaartekrachtsgolfsignaal en het uitrekenen hoe goed deze overeenkomt met het echte signaal, significant sneller maakt. Ook al is de kernmethode van dit concept niet nieuw binnen de wereld van zwaartekrachtsgolven, is het een uitdaging om deze methode toe te passen wanneer je extra testparameters aan het model toevoegt.

Het process waarmee we de berekeningen veel sneller kunnen maken, is in feite gewoon door lui te zijn: we rekenen alleen uit wat essentieel is. Stel je even voor dat de golfvorm die we willen berekenen enkel afhangt van de massas van de twee objecten die de golf veroorzaken. Als we de massas willen vinden die het best overeenkomen met het signaal in de detector, dan lijkt het erop dat we een golfvorm moeten uitrekenen voor elke denkbare combinatie van mogelijke massas. Misschien is de beste massa 10, maar het kan ook zijn dat het 10,1 is, of zelfs 10,0001. Je kan je waarschijnlijk voorstellen dat als we de golfvorm uitrekenen voor een massa van 10,0001, we waarschijnlijk niet nog eens een golfvorm hoeven uit te rekenen voor 10,0002. Op een gegeven moment zijn deze golfvormen niet meer te onderscheiden.

Het is dus wel duidelijk dat we niet voor elke mogelijke combinatie van massas een golfvorm hoeven uit te rekenen. Dat is maar goed ook, want dat zijn er namelijk oneindig veel. Het blijkt dat wanneer we de golfvormen hebben uitgerekend voor een paar slim gekozen massas, we deze kunnen gebruiken als basis om daarmee zonder enige moeite alle andere golfvormen te kunnen uitrekenen. Het is een beetje vergelijkbaar met een recept uit een kookboek: een paar ingrediënten kunnen met verschillende hoeveelheden gecombineerd worden om verschillende gerechten te maken. Deze ingrediënten vormen de basis voor al die recepten.

Wanneer we dit idee vertalen naar zwaartekrachtsgolven kunnen we bijvoorbeeld een set van combinaties van massas maken en vervolgens de golfvorm voor elk van deze combinaties uitrekenen. Dit zijn dan de ingediënten. We noemen ze e_i, waarbij i alles kan zijn tussen 0 en hoeveel ingrediënten we hebben. We kunnen nu elke andere golfvorm maken door een snufje e_1, een theelepeltje e_2, een teentje e_3, etcetera toe te voegen.

Het vinden van de optimale ingrediënten, of beter gezegd, de basisfuncties voor zwaartekrachtsgolven, is een uitdaging. Het wordt nog lastiger wanneer we meer en meer parameters toevoegen aan ons model. Voor het uitvoeren van onze ART-tests hebben we extra testparameters nodig naast de al grote hoeveelheid standaardparameters. In deze scriptie beschrijf ik de methode waarmee we deze uitdagingen te lijf kunnen gaan, en hoe we de beste e_i kunnen vinden. Met deze methode heb ik laten zien dat het uitvoeren van een test van ART wel tot 300 maal zo snel kan!

Om een idee te geven van hoe we dan die ingredienten e_i vinden, nemen wij

wederom aan dat de golfvorm enkel afhangt van de twee masssas. We beginnen door een hele hoop combinaties van massas te noteren. Als we tien waardes hebben voor de ene massa en tien voor de ander, hebben we in totaal honderd mogelijke combinaties. We rekenen nu de golfvorm uit voor elk van deze combinaties, pakken er een en noemen deze e_1. We vergelijken e_1 vervolgens met elke andere combinatie in het rooster. De golfvorm die het grootste verschil maakt met e_1 noemen we e_2 en deze onthouden we. Deze twee golfvormen vormen nu onze basis. Als we weer even terugdenken aan de receptenanalogie, kunnen we inzien dan we met slechts twee ingredienten waarschijnlijk niet alle recepten kunnen maken die we willen. We gaan dus door met het proces van het vinden van e_i totdat we vinden dat we met genoeg nauwkeurigheid de gewenste recepten kunnen reproduceren.

Dit is hoe we een basis bouwen voor onze golfvormen. Bij elke stap in het proces vergelijken we onze huidige basis met al de honderd andere golfvormen en bewaren we die golfvorm die het meest heeft toe te voegen. We gaan door met toevoegen van e_i tot de basis in staat is met de gewenste nauwkeurigheid alle golfvormen te produceren die we willen.

Zoals gezegd is deze uitdaging groter naarmate we meer parameters toevoegen aan het model. In het voorbeeld hierboven hadden we twee parameters, die elk tien verschillende waardes hadden. Dat vormt een rooster van $10 \times 10 = 100$ punten. Als we nu alle parameters willen meenemen (twee massas, zes spin-parameters en misschien 1 testparameter) en weer voor elke parameter slechts tien waardes gebruken, wordt het rooster gigantisch, namelijk 10^9, ofwel een tiencijferig getal. Elk van die combinaties komt overeen met een golfvorm die ofwel opgeslagen moet worden in het geheugen van een computer, ofwel bij elke stap in het proces opnieuw uitgerekend moet worden. Met het uitrekenen van 10^9 golfvormen ben je al gauw een decennium zoet.

Het opslaan van de golfvormen kost na het uitvoeren van wat slimme trucjes ook al gauw meer dan 200 GB (2 TB als we er slechts 1 parameter aan toevoegen). Door gebruik te maken van een eerder uitgerekende basis van pure ART-golfvormen en een op maat gemaakte selectie van testparameterwaardes, heb ik me uit deze moeilijkheden weten te werken. Ik heb laten zien dat we met deze methode een basis kunnen bouwen voor golfvormen die ook testparameters bevatten.

De titel toegelicht

Ik hoop dat ik met deze samenvatting de lezer heb kunnen overtuigen dat de enorm succesvolle algemene relativiteitstheorie nog steeds getest moet worden en dat het heelal nog altijd meer geheimen voor ons in petto heeft. Wat we hebben gezien van de ART is slechts het topje van de ijsberg. Twee om elkaar heen draaiende zwarte gaten die uiteindelijk in elkaar storten onder uitzending van zwaartekrachtsgolven is het ideale laboratorium voor de ultieme tests van de ART.

Zwarte gaten worden beschreven door pure ART. Voor een observeerder kan er net zo goed geen materie aanwezig zijn, enkel gekromde ruimtetijd: Zwarte gaten zijn een leegte (void) voor ons. We hebben deze leegte “gesubstantiëerd”[4] en daarmee ook

[4]Het engelse woord Substantiate, betekent ondersteuning bieden van of het bewijzen van.

de ART, door de eerste tests van het dynamische sterkeveldenregime van de ART uit te voeren en te concluderen dat de ART nog altijd sterk in zijn schoenen staat.

Acknowledgements

Time to lose all formality and get right into the ramble that is the Acknowledgements. I'm not much for this Hollywood-style-Oscars-thank-you-ramble, but it's tradition, so what can I do. I'll just start with a disclaimer: This text will contain tackiness.

My four and half years here at Nikhef have been an amazing and privileged experience. I have been very lucky to have had the opportunity to work in the field of gravitational waves right around the major discovery that everyone in the field had to work and wait for so long. I would like to thank my advisors Jo van den Brand and Chris Van Den Broeck for making me part of the group at Nikhef. Chris, I could not have wished for a better supervisor, you have been a great help and example throughout my time here at Nikhef.

The gravitational waves group at Nikhef has changed quite a bit since I started. Although I have collaborated and had coffee and beers with some people more than others, I can certainly say that the reason I enjoyed these past years so much is thanks to our entire group. Everyone in our group is always ready to chat about anything, share thoughts, or talk nonsense; so, many thanks to all of the gravitational waves group at Nikhef in both past and present!

I shall now throw around a few special thanks as per tradition. It all began with a large Scotsman; John, it was a pleasure working with you, you've been a great source of knowledge and fun discussions. Of course I thank Michalis, the only fellow PhD student in the data analysis "group" after all those post-docs left us behind. Then of course you had to live it up in Cambridge and I had to fend for myself here. In all seriousness though, you've been a fun colleague and friend.Thankfully, even after so many people left the group I was never completely alone. Of course I still had Reinier, Kazuhiro and Joris to keep me company, thanks guys for many wacky conversations at lunch, coffee and drinks! To Reinier a special thanks, not only for always being ready to help me out in a tight spot concerning pesky compiler linking issues, but also for being a good friend and colleague right from the start. After a while where I was the only CBC data analyst came the age of Laura, and with it a surge of new colleagues in both data analysis and instrumentation. Maria and Boris, any group you'll join will be the better for it! Archisman and Ka Wa, I'm glad you guys joined our group well before I left! And of course, thanks Laura, for deciding to join our group and making everything better! To all the usual suspects: Let there be plenty of drinks for us in the future!

And Janna, of course I did not forget you, thanks for livening up the office during the year you were here. I hope the exotic town of Birmingham treats you well and I wish you a bright future, whatever you may decide to do after your PhD!

The acknowledgements would not be complete without giving many thanks to the people of the LIGO-Virgo Collaboration. It has been an honor and privilege to be part this collaboration and not just because of the Breakthrough Prize in Fundamental Physics. I would like to highlight a few people in particular: Rory, Vivien, Karl-Johan and Scott for the many useful discussions on ROQs. Many thanks as well to Sarah, Patricia, Chris M, Valeriu, Laura N, Duncan, Tom A and Tom D for being awesome colleagues and friends during my year in Cardiff. Of course also many thanks to Sathya who was a wonderful supervisor there and who inspired me to continue in

the field of gravitational waves. I might never have done a PhD if it weren't for this awesome year in Cardiff!

Before I continue to the very important and rather tacky non-work related acknowledgements, I would like to thank the people from the grid-support department at Nikhef; David, Dennis, Jeffn and Mischa. Thanks a lot for the outstanding and enthusiastic support you provided during my efforts to get my project running on the Dutch e-Infrastructure. To the personeelsvereniging (PV) of Nikhef: So long, and thanks for all the fish!

Aside from my awesome colleagues there are of course my friends and family I would like to thank for support and indulging my (then rather silly) idea of wanting to study astrophysics, even though I was nowhere near the educational level required. I would never have been able to get to this point if it weren't for my family: Mom, Dad, Erik, Suzanne, this pile of pages and the accomplishments it represents are thanks to you as well! (Told you it would get tacky.)

And to my dear friends Ans, Eveline and Wouter; you guys are awesome, you made my ups and downs at Nijmegen many more ups than downs and I am very happy we never lost contact since, I am looking forward to many more dinners for us in the future.

Seriously, thanks.